LECTURE NOTES FOR NT III/IV

HERBERT GANGL

1. Motivation

In September 1994, the British mathematician Andrew Wiles finished his proof of a long-held conjecture which stated that

For $n \geqslant 3$, there are no solutions in positive integers x, y, z of
$$x^n + y^n = z^n\,.$$

Fermat famously had scribbled “I have a truly marvelous proof of this fact but the margin here is too small to contain it” on his copy of Diophantus's Oeuvre “Arithmetica”, and the search for such a proof had challenged number theorists for more than 350 years...

“Fermat's Last Theorem”, as the statement was called, is in a sense an *emblematic problem* for number theory: it is a question about integer solutions of an easily formulated equation but whose proofs are often exceedingly hard. In the quest of finding a solution for it, important structures were found (like ideals, class groups, ...) and amazing connections were uncovered (to elliptic curves, Galois representations, algebraic K-theory, ...).
It should be emphasized that Wiles was building on work of many other mathematicians (Taniyama, Shimura, Weil, Frey, Ribet, Mazur, Langlands, Tunnell, Taylor...).

The proof of FLT is far beyond what we are able to cover in this course. Nevertheless, we will use similar questions which can be treated with considerably easier methods, but which still have a “Diophantus–Fermat-like” flavour.

The main number theorist of ancient Greek times is Diophantus (∼250 A.D.), who studied more generally equations with integer coefficients and found ingenious methods to solve them in integers or also rationals. In honor of this eminent scholar such equations, where one is only interested in rational numbers—or sometimes only integers—as solutions, are called **Diophantine equations**.

For Diophantus, elementary geometry triggered a number of challenging questions, like the following one inspired by Pythagoras's theorem:

Q.1: Are there infinitely many “Pythagorean triples”, i.e. solutions (in positive integers x, y and z) of the equation
$$x^2 + y^2 = z^2\,?$$
Can one list/describe all the solutions?

⟦Note that the square of an odd number is again odd, and since any odd integer $2n+1$ is the difference of two successive squares n^2 and $(n+1)^2$, there are certainly infinitely many Pythagorean triples.⟧

Using a *geometric method* one can parametrise the set of all solutions.

Q.2: Which primes can occur as the hypotenuse of a right-angled triangle with integer sides? (This refines Q.1.) Formally, for which prime p can we write $p^2 = x^2 + y^2$ with x, $y > 0$?

⟦Answer: roughly "half of them": precisely when $p \equiv 1(4)$. ⟧

Q.3: How often does a cube exceed a square by 2? In mathematical notation: what are the solutions (in integers x and y) of

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

⟦There are two rather simple solutions $x = \pm 5$, $y = 3$; and they are in fact the only ones.⟧

This is an equation which can be adequately analysed by a very rich theory, the *arithmetic of elliptic curves* which also plays an important role in Wiles's proof.

One of the first renowned people in "modern" times deserving the name "number theorist" is Pierre de Fermat (1601–1665) who by profession was actually a lawyer in Toulouse. He had obtained one of the six books that Diophantus had left as his legacy, which turned out to be the stimulus for Fermat's ingenuity in inventing new methods (and new interesting, often innocuous-looking, problems) for the solutions of Diophantine equations. Among his findings are the following:

Q.4: Which primes can be expressed as a sum of two (integer) squares? Variations on this question: given an integer N, which primes p can be written as

$$p = x^2 + Ny^2 \,, \qquad x, y \in \mathbb{Z}\,?$$

⟦For $N = 1$, the solutions are $p = 2$ and, again, all primes $p \equiv 1(4)$.
For $N = 2$, one can solve it precisely the primes $p \equiv 1(8)$ and $p \equiv 3(8)$.
For $N = -2$, one can solve it precisely the primes $p \equiv 1(8)$ and $p \equiv 7(8)$. ⟧

Statements like the three ones above led to one of the most celebrated theories of 20th century mathematics, the so-called *class field theory.* The latter establishes e.g. the fact that the factorization of primes in $\mathbb{Z}[i]$ is determined simply by its congruence class modulo 4.

Q.5: Are there finitely many or infinitely many solutions of

$$x^2 - 2y^2 = 1 \ ?$$

Can you describe the set of all solutions?

⟦Write $x_n + y_n\sqrt{2} = (3 + 2\sqrt{2})^n$ for $n \in \mathbb{Z}$, then the pairs $\pm(x_n, y_n)$ describe precisely the—infinitely many—solutions of the above equation.⟧

In the literature, this and similar questions are nowadays referred to as "Pell's equation". It is intimately connected with one of the fundamental objects in algebraic number theory, the *units* in number rings. Furthermore, it is also directly related to continued fractions.

A result which at first glance is very surprising is which integers can be written as a sum of four squares (here terms 0^2 is allowed, e.g. $5 = 2^2 + 1^2 + 0^2 + 0^2$, $1367 = 27^2 + 25^2 + 3^2 + 2^2$, or $1234567891 = 28729^2 + 20229^2 + 3^2 + 0^2$).

Q.6: Show that **all** positive integers are sums of four squares!

⟦We will see a proof shortly.⟧

The first proof is accredited to Lagrange (while Fermat was the first to have claimed the fact, and very likely had a proof).

Even more surprisingly, two centuries after Fermat (who did not pass on a proof of his claim) another renowned number theorist, C.G. Jacobi (1804–1851), in a brilliant piece of work using Fourier analysis of elliptic functions, found an explicit formula expressing the **number of ways** in which an integer can be written as such a sum of four squares.

Fermat did not only look at quadratic equations (although they already provide a wealth of beautiful and intricate structures). For example, another innocuous-looking question about triangles leads naturally to an equation of degree 3:

Q.7: Which integers are *congruent numbers*, i.e. occur as the area of a right-angled triangle with *rational* sides?

⟦For instance, 6 does occur, since $3^2+4^2=5^2$, there is a right-angled triangle with sides of lengths 3, 4 and 5, whose area is 6. Since $\left(\frac{20}{3}\right)^2+\left(\frac{3}{2}\right)^2=\left(\frac{41}{6}\right)^2$, one can conclude the non-obvious fact that 5 is also a congruent number.

Amusingly enough, the number 157 is congruent and, although itself rather small, its least complicated corresponding right triangle has hypotenuse length for which both numerator and denominator have a whopping 45 digits.⟧

In order to tackle problems as the ones above, many ingenious techniques had to be invented. The more elementary ones deal with divisibility questions (often in an ad hoc manner), other more sophisticated approaches use more systematic tools like number rings (like $\mathbb{Z}[\sqrt{2}]$ in Q.5) or even elliptic curves (like the last two questions). Typically one is immediately led to rather profound mathematics.

Acknowledgments. What follows is based in large parts on a course by Steve Wilson (thanks to Ruth Jenni for providing me with the notes).

2. Diophantine equations via divisibility

2.1. Pythagorean triples.

We want to find all triples (a,b,c) of integers which satisfy the "Pythagorean" equation $x^2+y^2=z^2$. Since from each such solution we get (infinitely) many others (ka,kb,kc) by simply multiplying all three by the same number k, we restrict ourselves to the case where they are coprime, i.e. where $\gcd(a,b,c)=1$.

Problem 2.1. *Determine all* primitive Pythagorean triples*, i.e. all triples* (a,b,c)*,* $a,b,c\in\mathbb{N}_{>0}$ *such that* $a^2+b^2=c^2$ *("Pythagorean") and* $\gcd(a,b,c)=1$ *("primitive").*

Solution. We first investigate the parity of a, b and c, working first modulo 2 and then modulo 4.

Observe:

- not all three numbers a, b, c are $\equiv 0(2)$ ⟦otherwise $2\mid\gcd(a,b,c)$ ⟧.
- a, b are not both even ⟦or else c would also be; this we just ruled out⟧.
- a, b are not both odd: consider both sides modulo 4 ⟦consider squares of integers mod 4: m even $\Rightarrow m^2\equiv 0 \pmod 4$; m odd $\Rightarrow m^2\equiv 1 \pmod 4$ ⟧.

  If a and b were odd, then LHS $\equiv 2 \pmod 4$, but RHS $\equiv 0$ or $1 \pmod 4$. This is impossible.
- Therefore precisely one of a and b is odd, and consequently, c must be odd. Swapping roles of a and b, if necessary, we can assume a even, b odd.
- Put $a=2n$; note $a^2=c^2-b^2=(c-b)(c+b)$, and both factors on the right are even (since both b and c are odd).

  Put $c-b=2v$, $c+b=2w$; then we obtain $(2n)^2=2v\cdot 2w$, and thus $n^2=vw$ (*) ⟦n, v and w are all non-zero⟧.
- v and w are coprime ⟦a common factor would divide both $b(=w-v)$ and $c(=w+v)$⟧.
- By unique factorisation in $\mathbb{Z}$, (*) therefore implies $v=r^2$ and $w=s^2$ ⟦a prime factor dividing v, say, does not divide w, due to their being coprime; it also divides the LHS, in fact to an even power, and thus it divides v to that same (even) power⟧.

- So (a, b, c) is necessarily of the form $(2rs, s^2 - r^2, s^2 + r^2)$.
- Conversely, each such triple does satisfy the Pythagorean equation (check!).

In summary, we get as the complete list of primitive Pythagorean triples the following:

$$\{(2rs, s^2 - r^2, s^2 + r^2) \mid r, s \in \mathbb{N}_{>0}\}\,.$$

So by letting r and s run through all positive integers independently, we can create as many Pythagorean triples as we like (they will actually be primitive whenever r and s are coprime)—indeed, we get all those triples in this way. (This is called a *parametrisation* of the solutions.)

Note: This apparently has already been known to the Babylonians (some 3500 years ago), e.g. they listed the example

$$4961^2 + 6480^2 = 8161^2\,.$$

2.2. **How many solutions to $c^2 - b^2 = n$?** We can ask a more refined question: in how many Pythagorean triples does a given a occur (as one of the smaller numbers)? It turns out that in a way it is more convenient to answer a slightly more general question: how often can a number n be represented in the form $c^2 - b^2$ (previously n was a square a^2)?

Interlude. How many (positive) factors does an integer $n(> 0)$ have? Notation: $\sigma_0(n)$ = number of divisors of n. ⟦More generally, in number theory one often considers the function $\sigma_k(n) = \sum_{d|n} d^k$, i.e. the sum of powers d^k where d runs through the divisors of n.⟧ A short table shows:

n	1	2	3	4	5	6	7	8	9	10
$\sigma_0(n)$	1	2	2	3	2	4	2	4	3	4

This suggests the

Claim: $\sigma_0(n)$ is odd precisely when n is a square.

Indeed: as factors come in pairs $(d, n/d)$, it would seem that the number of divisors should always be *even*, except if d and n/d agree (then this divisor $d = n/d$ would only be counted once). But the latter happens precisely when $n = d^2$, i.e., n is a square.

Now we first try to evaluate $\sigma_0(n)$ for the building blocks which in our context are *prime powers*.

Claim: $\sigma_0(p^m) = m + 1$. ⟦Proof: the divisors of p^m are $1, p, p^2, \ldots, p^m$⟧

Mini-exercise: the function $\sigma_0(n)$ is **multiplicative**, i.e., if $\gcd(m, n) = 1$ $(m, n \in \mathbb{N}_{>0})$ then $\sigma_0(mn) = \sigma_0(m)\sigma_0(n)$.

Using the multiplicativity, we get the following result: suppose n has the prime decomposition $n = \prod_i p_i^{m_i}$ (i.e., the p_i are (mutually different) primes), then we get

$$\sigma_0(n) = \prod_i (m_i + 1)\,.$$

Example: Let $n = 55000$. Since $n = 2^3 \cdot 5^4 \cdot 11$, we get $\sigma_0(n) = 4 \cdot 5 \cdot 2 = 40$.

This ends the interlude, and we can now tackle the question stated at the beginning of this subsection..

Problem 2.2. *Let $n > 0$ be an integer.*
How many solutions are there to $x^2 - y^2 = n$, with x and y in $\mathbb{N}_{>0}$?

Solution. As in the previous problem, we first try to find a necessary form for the pairs (x, y).

So suppose (x, y) is a solution. Put $d = x + y$ and $e = x - y$. Then the equation is rewritten as $de = n$. We can deduce parity for d and e: since $d + e = 2x$, we know that $d \equiv e \bmod 2$. Since $d - e = 2y > 0$, we also know that $d > e$.

Thus (x, y) lies in the following set

$$S := \Big\{ \big(\frac{d+e}{2}, \frac{d-e}{2}\big) \text{ such that } de = n,\ d \equiv e \bmod 2 \text{ and } d > e > 0 \Big\} .$$

Again, one checks easily that each element in S indeed provides a solution.

In order to determine the size of S, we distinguish two cases.

I. Case n odd. In this case any divisor of $n = de$ is also odd, so the condition $d \equiv e \pmod 2$ is automatically satisfied. Furthermore, once we know d, the other number e is determined ($e = n/d$). Therefore $|S|$ is the number of divisors of n with $d > n/d$, i.e., such that $d > \sqrt{n}$.

Now to each such $d > \sqrt{n}$ dividing n there is an $e = n/d < \sqrt{n} < d$, so d contributes a member to S. But all factorisations of $n = de$, $d \geqslant e$, entail $d > \sqrt{n} > e$ or $d = \sqrt{n} = e$. The latter occurs precisely if n is a square.

If we denote the number of (positive) divisors of a number n by $\sigma(n)$, we can therefore conclude

$$|S| = \frac{\sigma(n)}{2} ,$$

except when n is a square, in which case it reads

$$|S| = \frac{\sigma(n) - 1}{2} ,$$

II. Case n even. This case can be somehow reduced to the previous case. One of d and e must be even, and due to the condition $d \equiv e \pmod 2$ both have to be. Therefore we can conclude that for $n/2$ odd there are *no solutions*, i.e. $|S| = 0$.

On the other hand, if $4|n$, then we get $d = 2d'$ and $e = 2e'$ with d', e' in $\mathbb{Z}$ and $d'e' = n/4$, and so we can restate the set S for the case n even in terms of d' and e' (the description is slightly simpler as the condition $d \equiv e \pmod 2$ is no longer needed)

$$S = \{(d' + e', d' - e') \text{ such that } d'e' = n/4, \text{ and } d' > e' > 0\} .$$

Proceeding as in Case I, we see that $|S|$ is the number of divisors of $n/4$ which are greater than $n/4$, i.e.

$$|S| = \begin{cases} \frac{\sigma_0(n/4)}{2} & \text{if } n \text{ is not a square,} \\ \frac{\sigma_0(n/4)-1}{2} & \text{if } n \text{ is a square.} \end{cases}$$

2.3. **The four-square theorem.** The following striking statement, together with its proof, should give a first glimpse of the power of ingenious ideas. It is not so difficult to find four squares which add up to 111, say ($111 = 9^2 + 5^2 + 2^2 + 1^2$), but it seems forbidding to achieve such a presentation for a much larger number, say, the prime 1234567891. Fermat had already stated that each natural number can be thus represented, albeit he didn't leave a proof. The first proof came from J.L. Lagrange (1736–1813), and we will follow his argument.

Theorem 2.3. *Any $N \in \mathbb{N}$, there are w, x, y, z in $\mathbb{Z}$ such that*

$$N = w^2 + x^2 + y^2 + z^2 .$$

Proof: **Step 0.** The statement is clear for $N = 2$ since $2 = 1^2 + 1^2 + 0^2 + 0^2$.
Step 1. Reduction to N a prime: we use an identity by l. Euler (1707–1783):

$$\begin{aligned}(a^2+b^2+c^2+d^2)(w^2+x^2+y^2+z^2) \quad = \quad & (aw+bx+cy+dz)^2\\ & +(ax-bw-cz+dy)^2\\ & +(ay+bz-cw-dx)^2\\ & +(az-by+cx-dw)^2\,.\end{aligned}$$

Therefore the product of two four-squares (as on the left) is also a four-square (as on the right). Thus it is enough to show the statement of the theorem for the (multiplicative) building blocks, i.e., for $N = p$ prime.

Step 2. It is rather easy to show that a slightly weaker claim holds: the four-square property holds for a non-zero multiple of the prime p:

$$\exists m > 0 \quad \text{such that} \quad mp = w^2 + x^2 + y^2 + z^2 \quad \text{for some} \quad w, x, y, z \in \mathbb{Z}\,.$$

One actually shows, using the pigeon-hole principle, the following even stronger claim:

Lemma 2.4. *For a prime p, there exists $m < p$ such that mp can be written as a sum of 3 squares; more precisely, $mp = x^2 + y^2 + 1$ for some integers x, y.*
⟦Proof: Exercise; for hints see Problem Sheet 1, Ex. 5.⟧

Step 3. Starting from the claim in Step 2, successively replace m by smaller m', still satisfying the four-square property for $m'p$, until $m' = 1$. Then we are done.

How to replace? Distinguish two cases, according to whether m is even or odd:

I. Case m even. If mp satisfies the four-square property, then so does $\frac{m}{2}p$: More generally, if $2N = w^2 + x^2 + y^2 + z^2$, then there are an even number of odd integers and also an even number of even integers among w, x, y, z. So we can group them in pairs, say $w \equiv x(2)$ and $y \equiv z(2)$. Then $N = \left(\dfrac{w+x}{2}\right) + \left(\dfrac{w-x}{2}\right) + \left(\dfrac{y+z}{2}\right) + \left(\dfrac{y-z}{2}\right)$.

We can assume $p > 2$ (cf. Step 0) and therefore, if m is even, reduce m to $m/2$.

II. Case m odd. By assumption, we have $mp = w^2+x^2+y^2+z^2$ (from Step 2); in fact, we can assume $0 < m < p$ by the lemma above. Now we "switch" the point of view and work modulo m. We choose the unique a, b, c and d which are congruent to w, x, y and z modulo m, respectively, such that $-m/2 < a, b, c, d < m/2$. This immediately implies that

$$a^2 + b^2 + c^2 + d^2 \equiv w^2 + x^2 + y^2 + z^2 \equiv 0 \pmod{m}\,,$$

and in fact that

$$a^2 + b^2 + c^2 + d^2 = km \qquad \text{with} \quad 0 < k < m\,.$$

The latter claim on the size of k follows directly from $a^2 < \left(\frac{m}{2}\right)^2$ (and similarly for b, c, d) so that $a^2 + b^2 + c^2 + d^2 < 4\left(\frac{m}{2}\right)^2 = m^2$ and so $k < m$. Note that $k \neq 0$. ⟦Otherwise $a = b = c = d = 0$ and therefore $w \equiv x \equiv y \equiv z \equiv 0 \pmod{m}$ which implies that m^2 divides $w^2+x^2+y^2+z^2$. But the latter is equal to mp by assumption and so $m \mid p$ which contradicts the outcome of the above lemma ($0 < m < p$).⟧

Finally, all we need is to use Euler's identity again, this time with the specific expressions above. On the left hand side, we get $(a^2 + b^2 + c^2 + d^2)(w^2 + x^2 + y^2 + z^2) = km \cdot mp$, while on the right hand side we have

> the squares of $aw + bx + cy + dz$, $ax - bw - cz + dy$, $ay + bz - cw - dx$ and $az - by + cx - dw$, respectively. But the way we have chosen a, b, c, d implies that all these four expressions are divisible by m. Therefore we can conclude that
>
> $$kp = W^2 + X^2 + Y^2 + Z^2 \,,$$
>
> where W, X, Y and Z are these expressions divided by m, e.g. $W = (aw + bx + cy + dz)/m$, $X = (ax - bw - cz + dy)/m$, etc. which are all *integers* by the above.

This finishes the reduction step for m odd, and therefore also the proof of the theorem.

The above proof does not provide any specific decomposition, but one can give a "constructive" proof, e.g., check at http://www.alpertron.com.ar/4SQUARES.HTM, where one can find an applet (http://www.alpertron.com.ar/FSQUARES.HTM) by Dario Alpern which gives in our case above

$$1234567891 = 28729^2 + 20229^2 + 3^2 + 0^2 \,.$$

2.4. **The descent method.** Many Diophantine equations have either *no* solution or *infinitely many* solutions. Fermat invented a technique which can deal with either situation! This technique is called the *descent (method)*. The idea, roughly, is to devise a mechanism which produces from a given "old" solution a "new" (different) one.

More precisely, the new solution should be in some sense "smaller" than the old one (typically one takes as measure the smallest—in absolute value—member in a given solution). Note that a variant of this has already been used in the proof of the 4-square theorem (when passing from a solution for mp to a solution for $m'p$, $0 < m' < m$). Surprisingly, the descent also works when there is *no* solution.

A good example for the method is Fermat's last theorem (FLT) for the exponent 4.

Proposition 2.5. *The equation*

$$x^4 + y^4 = z^4$$

has no (non-trivial) solution in integers.

For the proof, we will use the "descent technique", but also our knowledge of the shape of Pythagorean triples. Again, we will actually show a slightly *stronger* statement:

Claim 2.6. *The equation $x^4 + y^4 = z^2$ has no (non-trivial) solution in integers.*

Proof: Assume we had a *primitive* solution (x, y, z) of this equation (i.e., where $\gcd(x, y, z) = 1$), then, writing it as $(x^2)^2 + (y^2)^2 = z^2$, this is a Pythagorean triple, so necessarily of the form (up to possibly swapping the roles of x and y)

$$x^2 = 2rs \,, \qquad y^2 = s^2 - r^2 \,, \qquad z = s^2 + r^2$$

for some $r, s \in \mathbb{N}$, $s > r$. Note that $\gcd(r, s) = 1$ ⟦otherwise $\gcd(x^2, y^2, z) \neq 1$, but then also $\gcd(x, y, z) \neq 1$, contrary to our assumption⟧.

We can rewrite the equation as

$$x^4 = (z - y^2)(z + y^2) \,.$$

As before, we would like to conclude that each of the factors on the right is itself a fourth power. (This is not quite true, but it is not far from being correct.) So suppose p prime divides both factors, then $p|(\text{sum}=)2z$ and $p|(\text{diff }=)2y^2$, so $p|2$ ⟦as $(z, y) = 1$ implies also $(z, y^2) = 1$⟧. Therefore $(z - y^2, z + y^2) = 2$ ⟦check that no higher power of 2 can divide the gcd⟧.

Although we thus cannot conclude that both $z-y^2$ and $z+y^2$ are fourth powers, we get at least that

- either $\quad z-y^2 = 2a^4$, a odd, $\quad z+y^2 = 2^3b^4$
- or $\quad z-y^2 = 2^3a^4$, $\quad z+y^2 = 2b^4$, b odd.

But the first alternative would imply $2y^2 = 2^3b^4 - 2a^4$, and so $y^2 = 4b^4 - a^4$, which is impossible as we see upon reducing both sides modulo 4 ⟦LHS$\equiv 1 \pmod 4$, while RHS$\equiv 0-1=-1 \pmod 4$⟧.

Therefore we can only have the second alternative, from which we deduce

$$y^2 = b^4 - 4a^4\,, \qquad z = b^4 + 4a^4\,.$$

Note that the latter equation implies $0<b<z$, while the former gives

$$4a^4 = (b^2-y)(b^2+y)\,.$$

Similar to our reasoning above, the gcd of the two factors on the RHS is 2 ⟦check this!⟧, so we have

$$b^2 - y = 2c^4\,, \qquad b^2 + y = 2d^4\,,$$

and by eliminating y from them (add them up and then divide both sides by 2) we get

$$b^2 = c^4 + d^4\,,$$

which constitutes a *new* solution ⟦recall $0<b<z$⟧.
Conclusion: From each solution we can construct a new, in fact "smaller" one (as $b<z$), which is also non-trivial (as $0<b$).

Now in order to finish the proof, suppose we took the solution of $x^4+y^4=z^2$ with the smallest possible z. Then by the above we could fabricate an even smaller one. Contradiction.

Therefore we have shown: there cannot be a (non-trivial) solution of $x^4+y^4=z^2$ ⟦we could always reduce it to an even smaller one, and after a finite number of steps it would have to be reduced to the *smallest* one—which we just showed cannot exist⟧. □

From this Claim we can immediately deduce the above Proposition, i.e., the case $n=4$ of FLT. ⟦If we cannot find solutions to $x^4+y^4=z^2$, then we have an even harder time finding a solution with the further constraint that z be a square.⟧

2.5. **Rings larger than $\mathbb{Z}$ and (the lack of) uniqueness of factorisation.**

Our final motivational example gives a short indication of the Pell equation (Q.5 above). The problem is to find solutions to

$$x^2 - 10y^2 = 1 \tag{1}$$

in (positive) integers x, y. With trial and error we can find $x=19$, $y=6$. Is this the only solution—and if not, how can we find more?

An elegant way to deal with this question uses several notions from ANTII: the idea is to pass to a slightly larger number system, in order to be able to factorise the left hand side of (1): we consider (as in ANTII) the following **ring**:

$$\mathbb{Z}[\sqrt{10}] = \{a + b\sqrt{10} \mid a, b \in \mathbb{Z}\}$$

⟦Recall that a ring $(R,+,\cdot)$ is a non-empty set R with two binary operations $+$ and $\cdot$ (i.e., it is closed under addition $+$ and multiplication $\cdot$), and is an abelian group wrt. $+$, is associative under $\cdot$, and the two operations are compatible as formulated in the distributive law.⟧
(this is a subring of $\mathbb{R}$), for which we can write the LHS as

$$(x + y\sqrt{10})(x - y\sqrt{10})\,.$$

From ANTII we may remember that this is precisely the *norm function*

$$N : \mathbb{Z}[\sqrt{10}] \longrightarrow \mathbb{Z}\,, \qquad N(a + b\sqrt{10}) = a^2 - b^2 \cdot 10\,.$$

So we can rephrase (1) as saying that $x + y\sqrt{10}$ is a *unit* in $\mathbb{Z}[\sqrt{10}]$.

⟦Recall that a **unit** u in a ring R satisfies $uv = 1$ for some v in R.⟧

The norm is *multiplicative*, so we can derive from each solution (x, y) of (1) infinitely many others:

$$N\big((x + y\sqrt{10})^r\big) = \big(N(x + y\sqrt{10})\big)^r = 1\,,$$

since $(x + y\sqrt{10})^r \in \mathbb{Z}[\sqrt{10}]$ ⟦use closedness under multiplication⟧. All these solutions turn out to be different.

Important properties that we have used in our deduction:

- we enlarged the ring of integers (e.g., to $\mathbb{Z}[\sqrt{D}]$ for some D),
- we (implicitly) have used the uniqueness fo factorisation in $\mathbb{Z}$, e.g.:

$$\left\{ \begin{array}{c} x^2 = vw \\ (v, w) = 1 \end{array} \right\} \Rightarrow v = \pm\Box,\ w = \pm\Box\,. \tag{2}$$

Crucial fact: The latter uniqueness is no longer guaranteed in $\mathbb{Z}[\sqrt{D}]$!

Example: 1. In the ring $\mathbb{Z}[\sqrt{10}]$ above, we have

$$(\sqrt{10} + 1)(\sqrt{10} - 1) = 9 = 3^2\,. \tag{3}$$

But one can check that all the factors on the left and on the right of this equation are irreducible in $\mathbb{Z}[\sqrt{10}]$.

⟦Recall that an element a in a ring R is **irreducible** if for any decomposition $a = bc$ with b, c in R one has that b or c must be a unit.⟧

In particular, neither $1 + \sqrt{10}$ nor $1 - \sqrt{10}$ is a square in $\mathbb{Z}[\sqrt{10}]$, so we cannot conclude as in (2). (Also, the gcd might not exist in such larger rings.) In summary, we have encountered the new phenomenon of an *ambiguity of decomposition* of a number into irreducibles.

2. This phenomenon sincerely limits our capability to solve Diophantine equations. For example, we could rather easily solve "half" of Fermat's Last Theorem if we always had uniqueness of factorisation in the "cyclotomic" number rings $\mathbb{Z}[\zeta_p]$ given as follows: let p be an odd prime and ζ_p a primitive pth root of unity (e.g., $\zeta_p = e^{2\pi i/p}$), then

$$\mathbb{Z}[\zeta_p] := \{a_0 + a_1\zeta_p + \cdots + a_{p-2}\zeta^{p-2} \mid a_0, \ldots, a_{p-2} \in \mathbb{Z}\}\,.$$

We could factor

$$z^p = x^p + y^p = (x + y)(x + \zeta_p y)(x + \zeta_p^2 y) \cdots (x + \zeta_p^{p-1} y)$$

and then we could conclude that each factor on the right is itself a pth power (times a unit) which would readily lead to a contradiction.

2.6. **A way out.** The big question thus is: how to overcome this ambiguity in the decomposition? The solution was suggested by E.E. Kummer (1810–1893) who postulated "ideal elements" into which numbers in such a larger ring then would decompose. We illustrate with our previous example $\mathbb{Z}[\sqrt{10}]$: suppose there were "ideal elements" π_1, π_2 with the following properties

$$\begin{cases} 3 = \pi_1 \cdot \pi_2\,, \\ \sqrt{10} + 1 = \pi_1^2\,, \\ \sqrt{10} - 1 = \pi_2^2\,, \end{cases}$$

then (3) would become

$$\pi_1^2 \cdot \pi_2^2 = (\pi_1\pi_2)^2\,,$$

which looks very good already. We would still need certain important properties of these ideal elements: they should respect the divisibility properties that we want to use $\big($e.g. $(\pi \mid \alpha$ and $\pi \mid \beta) \Rightarrow \pi \mid (\alpha \pm \beta)$ $\big)$. Furthermore, we would need to add and multiply them. Kummer showed that this can be done consistently.

But where can we find these ideal numbers? The complex numbers will not be of much help. ⟦This is not quite true, one can in fact view the ideal numbers as being represented by certain algebraic numbers (keyword "Hilbert class field") which can be embedded into the complex numbers. But this would take us too far afield.⟧ Instead, R. Dedekind (1831–1916) had a very nice point of view: one can characterise an ideal number π by the "shadow" that it throws in the underlying ring of integers R in the following sense: the shadow of π is the set of all integers in R which are *divisible by* π. From this idea is derived the notion of an *ideal* (=the above shadow) in a ring, which replaces Kummer's notion of an ideal element.

This concludes our motivation for the study of such (number) rings and ideals.

3. Recap of Rings and Ideals

We collect a number of properties of rings and ideals from ANTII, occasionally recalling definitions.

General assumption: A **ring** in this course is always understood to be **commutative with identity** (unless otherwise stated).

Definition 3.1. *An* **integral domain** *is a ring R (i.e., commutative with identity by our general assumption) without zero divisors, i.e.*

$$a, b \in R - \{0\} \;\Rightarrow\; a \cdot b \in R - \{0\}\,.$$

Note. The units of R form an (abelian) group, denoted by R^*. We can think of them as the "divisors of 1".

Examples:

1. $R = \mathbb{Z}[\sqrt{-5}]$ is a subring of $\mathbb{C}$, in fact a integral domain. Its group of units is given by $R^* = \{\pm 1\}$.
2. $R = \mathbb{Z}[i]$ has units $R^* = \{\pm 1, \pm i\}$.
3. For $R = \mathbb{Z}[\sqrt{10}]$ we have seen that $\{(19 + 6\sqrt{10})^r \mid r \in \mathbb{Z}\} \subset R^*$. This is actually not the full story: it will turn out (later in the course) that
$$R^* = \{\pm(3 \pm \sqrt{10})^r \mid r \in \mathbb{Z}\}\,.$$

Definition 3.2. *Two elements a, b in a ring R are called* **associate** *(to each other), denoted*

$$a \sim b$$

if $a = ub$ for some unit $u \in R^*$.

Examples:

1. In $\mathbb{Z}[i]$, we have
$$2 + i \sim -1 + 2i \sim -2 - i \sim 1 - 2i\,.$$
(More generally $a + bi \sim -b + ai \sim \dots$)
2. In the integral domain $\mathbb{Q}[X]$ (polynomials in one variable with coefficients in $\mathbb{Q}$), we have $f(x) \sim af(x)$ for any $a \in \mathbb{Q}^*(= \mathbb{Q} - \{0\})$.

Definition 3.3. *An element a in the ring R* **divides** $b \in R$*—or “b* **is divisible by** $a \in R$*”—if $b = a \cdot c$ for some $c \in R$. If, furthermore, $a \not\sim b$ (i.e., the c above $\notin R^*$), then a is called a* **proper divisor** *of b.*

Lemma 3.4. *Let $R = \mathbb{Z}[\sqrt{-d}] \subset \mathbb{C}$ where $d \in \mathbb{Z}_{>0}$, and let $\alpha,\ \beta \in R^*$. Then*

1. *$\alpha\bar{\alpha} \in \mathbb{Z}_{>0}$ (here $\bar{\alpha}$ is the complex conjugate of α). Note that $\alpha\bar{\alpha} = N(\alpha)$ in our previous notation.*
2. *If $\alpha \mid \beta$ in R, then $\alpha\bar{\alpha} \mid \beta\bar{\beta}$ in $\mathbb{Z}$.*
3. *Let $\alpha \mid \beta$. Then α is a* proper *divisor of β if and only if $\alpha\bar{\alpha} < \beta\bar{\beta}$.*

Lemma 3.5. *Let a, b be elements in a ring R. Then we have*

1. *$a \mid b$ and $b \mid a \quad \Rightarrow \quad a \sim b$.*
2. *$a \sim 1 \quad \Leftrightarrow \quad a$ is a unit in R.*

Definition 3.6. *An element $r \in R \setminus R^*$ is* **irreducible** *if*

$$r = a \cdot b, \quad \textit{with } a, b \in R \quad \Rightarrow \quad a \in R^* \textit{ or } b \in R^*\,.$$

In other words: any proper divisor of an irreducible element is a unit.

The above definition of irreducible is what we typically use to characterise *prime numbers*. Instead, the algebraic definition of being *prime* is the following:

Definition 3.7. *An element $r \in R \setminus R^*$ is* **prime** *if $r \mid ab$ for some $a,\ b \in R$ implies that $r \mid a$ or $r \mid b$.*

For $\mathbb{Z}$ both concepts (prime and irreducible) turn out to be the same.

Examples:

1. Prime numbers in $\mathbb{Z}$ are irreducible.
2. In $\mathbb{Q}[X]$, the irreducible polynomials are indeed irreducible in the above sense.
3. $\delta = 1 - 3\sqrt{-6}$ in $\mathbb{Z}[\sqrt{-6}]$ is irreducible.

Proof of (3).

- δ is not a unit ⟦we know that the units in $\mathbb{Z}[\sqrt{-d}]$, $d > 1$, are only ± 1: their norm has to be 1, i.e., $a^2 + b^2 d = 1$, and this is only possible for $b = 0$, whence $a = \pm 1$.⟧
- Suppose α is a proper divisor of δ. Need to show: α is a unit. By the above lemma we know $\alpha\bar{\alpha} \mid \delta\bar{\delta} (= 55)$ and so $\alpha\bar{\alpha} < \delta\bar{\delta}$. Therefore $\alpha \in \{1, 5, 11\}$.

  But $\alpha\bar{\alpha} = 5$ entails $a^2 + b^2 \cdot 6 = 5$, whence $b = 0$ and $a^2 = 5$ which is not possible. Similarly $\alpha\bar{\alpha} = 11$ would give either $b = 0$ and $a^2 = 11$, or else $b = \pm 1$ and $a^2 = 5$; both cases are not possible.

  Therefore $\alpha\bar{\alpha} = 1$, i.e., α is a unit.

Problem: Factorise $\beta = 16 + 7\sqrt{-6}$ into irreducibles in $R = \mathbb{Z}[\sqrt{-6}]$.

Solution: Suppose $\alpha | \beta$, then also $N(\alpha) | N(\beta) = 550$. Now we only need to check divisors of 550 up to $\sqrt{550} < 24$, i.e. $1, 2, 5, 10, 11, 22$.

Putting $\alpha = a + b\sqrt{-6} = a^2 + 6b^2$ can not become 2 or 5. ⟦b would have to be $0 \ldots$⟧ On the other hand, $N(\alpha) = 10$ is possible: $b = \pm 1$, $a = \pm 2$.

So we check whether we can divide β by any of these four numbers—which, up to associates, are only two different ones, e.g., $2 \pm \sqrt{-6}$. Division gives

$$\frac{16 + 7\sqrt{-6}}{2 \pm \sqrt{-6}} = \frac{16 + 7\sqrt{-6}}{2 \pm \sqrt{-6}} \frac{2 \mp \sqrt{-6}}{2 \mp \sqrt{-6}} = \frac{32 \pm 42 + (14 \mp 16)\sqrt{-6}}{10}\,.$$

This shows that the "upper" sign gives a number which is *not* in R, while the lower sign gives $-1+3\sqrt{-6}$, and this number we happen to have just recognized as irreducible (see above). Thus we get

$$\beta = (2-\sqrt{-6})(-1+\sqrt{-6})\,,$$

and both factors are irreducible (any proper divisor of $2-\sqrt{-6}$ would have norm 2 or 5, but we just saw that there are no such...).

Two central notions in an integral domain which are particular interesting for us are the notions of *prime* and *irreducible*. The former implies the latter, but in general not vice versa:

Proposition 3.8. *Let $\pi\in R$, where R is an integral domain. Then if π is prime (in R), then it is also irreducible (in R).*

Proof. Write $\pi = ab$. We want to show: a or b is a unit.

Since in particular $\pi\,|\,ab$, we have (use that π is prime) $\pi\,|\,a$ or $\pi\,|\,b$.
Up to swapping a and b, we can suppose $\pi\,|\,a$, i.e., $a=\pi\rho$ for some $\rho\in R$. Then

$$\pi = ab = (\pi\rho)b = \pi(\rho b)$$

and hence $\rho b = 1$, i.e., b is a unit.
Conclusion: $\pi = ab \ \Rightarrow\ b$ is a unit or a is a unit (keep above swapping in mind). □

Many of our proofs of statements about, say, Diophantine equations so far have invoked the (implicit) use of unique factorisation into irreducibles, but we have seen that for more general rings we cannot expect this property to hold. Therefore we distinguish this class:

Definition 3.9. *An integral domain R is called a* **unique factorisation domain (=UFD)** *if every non-zero element factors into a product of irreducible elements and the factorisation is unique, up to replacing each irreducible element by an associate one, and up to reordering the factors. In less verbose terms:*

for any $x\in R$, if $x = up_1\cdots p_r = vq_1\cdots q_s$ for $u,v\in R^*$ with p_i, q_j irreducible in R, then $r=s$ and, after possible reordering of the q_j, we have $p_j\sim q_j$ $(1\leqslant j\leqslant r)$.

For these especially nice rings we have a converse of the above proposition:

Proposition 3.10. *In a UFD, any irreducible element is also prime.*

Proof. Let π be irreducible in the UFD R (in particular, π is not a unit).

Suppose $\pi\,|\,ab$ for some $a,b\in R$. Then we need to show that $\pi\,|\,a$ or $\pi\,|\,b$.

Start by decomposing both a and b into irreducibles p_i and units u_a and u_b, respectively:

$$a = u_a p_1\cdots p_\ell\,,\qquad b = u_b p_{\ell+1}\cdots p_{\ell+r}\,,\qquad \text{and so}\qquad ab = u_a u_b p_1\cdots p_{\ell+r}\,.$$

By assumption, the decomposition of ab is *unique*, up to replacing each p_i by an associate and up to reordering the p_i.

Now $\pi\,|\,ab$ implies $ab=\pi\rho$, where $\rho = u_\rho q_1\cdots q_s$ is some decomposition into irreducibles. Since the factorisation of ab is unique (in the above sense), π must be associate to one of the p_i as well $[\![$compare the two decompositions $u_\rho q_1\cdots q_s\cdot\pi = u_a u_b p_1\cdots p_{\ell+r}]\!]$. If $1\leqslant i\leqslant \ell$, then $\pi\,|\,a$, otherwise $\pi\,|\,b$. □

Examples:

(1) The following are UFDs: $\mathbb{Z}$, $\mathbb{Z}[i]$ and $\mathbb{Z}[\zeta_p]$ with p prime $\leqslant 19$.
(2) The following are no UFDs: $\mathbb{Z}[\sqrt{-6}]$, in fact most rings of the form $\mathbb{Z}[\sqrt{-d}]$, $d>0$ squarefree, are not UFDs; nor are $\mathbb{Z}[\zeta_p]$ with p prime $\geqslant 23$.

This motivates the quest for criteria to

- to find UFDs, or at least,
- in non-UFDs, to "measure" the ambiguity in how many wways we can decompose a number ⟦this will be the number of *ideal classes* below⟧.

3.1. **Passing from one ring to another.** We have used before, that we can transfer a problem about the integers (an *infinite* ring), e.g. solving $x^2 - 4y^2 = 3$ in integers, to a—hopefully easier—problem about $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ (a *finite* ring); e.g. we can take $m = 4$ and see immediately that the resulting reduced equation $x^2 \equiv 3 \pmod 4$ has no equations in $\mathbb{Z}_m$.

In the process we need to keep the relevant structures, which leads to the notion of a *homo*(=same)*morphism*(=structure):

Definition 3.11. *Let A and B be rings. A homomorphism of rings $\varphi : A \to B$ is a map respecting both ring operations, i.e.,*

$$\begin{aligned} \varphi(a +_A b) &= \varphi(a) +_B \varphi(b)\,, \\ \varphi(a *_A b) &= \varphi(a) *_B \varphi(b)\,. \end{aligned}$$

In the following we will drop the subscripts indicating in which ring we are working.

Examples:

(1) For any $m \in \mathbb{N}$, we have the reduction homomorphism $\varphi : \mathbb{Z} \longrightarrow \mathbb{Z}_m$, where $\varphi(a) = \bar{a} = a + m\mathbb{Z} = \{a + mn \mid n \in \mathbb{Z}\}$.

(2) For any $a \in \mathbb{C}$ there is the specialisation homomorphism $\varphi : \mathbb{Z}[X] \to \mathbb{Z}[a]$, where $\varphi\big(f(X)\big) = f(a)$.

Note that both homomorphisms are surjective. What are their kernels? Recall:

Definition 3.12. *The* **kernel** *of a ring homomorphism $\varphi : A \to B$, denoted by* $\ker(\varphi)$*, is the set $\varphi^{-1}(0_B)$(= $\{a \in A \mid \varphi(a) = 0_B\}$).*

Note further that $\ker(\varphi)$ is always a subring (but not necessarily with identity!) of A. It is in fact an ideal (see below).

Examples: For the previous example, we have

(1) in the first case

$$\ker(\varphi) = \{a \in \mathbb{Z} \mid \bar{a} = \bar{0} \text{ in } \mathbb{Z}/m\mathbb{Z}\} = \{a \in \mathbb{Z} \mid a \in m\mathbb{Z}\} = m\mathbb{Z}\,,$$

(2) in the second case

$$\ker(\varphi) = (X^2 + 1)\mathbb{Z}[X]\,. \qquad \text{(Exercise)}$$

This gives us yet another motivation to introduce the following

Definition 3.13. *An ideal I in the ring R is a* sub*group of $(R, +)$ which is closed under multiplication by elements in R, i.e.,*

$$\forall a \in I\ \forall r \in R\ :\quad ar \in I\,,$$
$$\textit{i.e.}\quad I \cdot R \subset I\,.$$

⟦You can think of the ideal as a "black hole" swallowing everything which comes "near" it. . . ⟧

We can see the connection of ideals to divisibility questions:

(1) The subgroup property: if $b \in R$ divides a and a' in R, then b divides $a - a'$ as well.

(2) Furthermore, if $b \in R$ divides a, then $b \in R$ divides ar for any $r \in R$.

Examples: For the previous example, we have

(1) For $m \in \mathbb{Z}$, we have the ideal $(m)_{\mathbb{Z}} = \{rm \mid r \in \mathbb{Z}\}$

(2) For $a, b \in \mathbb{Z}$, the set $I = (a + bi)_{\mathbb{Z}[i]} \subset \mathbb{Z}[i]$ forms an ideal.

In either case, the ideals given are the only ones.

We can compute with ideals (just as we would expect to compute with "ideal elements/numbers"):

Lemma 3.14. *If I, J are ideals in R, then so are $I+J$, $I\cdot J$ and $I\cap J$.*

Using the ideals in the previous example, we can get a feel for the corresponding operations:

- $(m)_{\mathbb{Z}}+(n)_{\mathbb{Z}}$ corresponds to taking the multiples of m and the multiples of n together; if we allow to add them, we get the $\gcd(m,n)$ and all its multiples, i.e., $\big(\gcd(m,n)\big)_{\mathbb{Z}}$.
- $(m)_{\mathbb{Z}}(n)_{\mathbb{Z}}$ corresponds to taking among the numbers which are divisible by m those numbers which are further divisible by n, i.e., the multiples of mn, or as an ideal $(mn)_{\mathbb{Z}}$.
- $(m)_{\mathbb{Z}}\cap(n)_{\mathbb{Z}}$ corresponds to taking numbers which are at the same time multiples of m and n, i.e., the multiples of the $\operatorname{lcm}(m,n)$.

We still should recall how we are allowed to compute with sets: for any subgroups A and B of $(R,+)$ we define

$$\begin{aligned} A+B &:= \{\,a+b \mid a\in A, b\in B\,\}\,,\\ A\cdot B &:= \{\sum_{\text{finite}} a_ib_i \mid a_i\in A, b_i\in B\} = \langle ab \mid a\in A, b\in B\rangle_{\text{gp}}\,. \end{aligned}$$

Lemma 3.15. (i) $IJ\subset I, \quad I+J\supset I\,,$

(ii) $I\cdot J\subset I\cap J\subset \begin{Bmatrix} I\\ J\end{Bmatrix}\subset I+J\,.$

3.2. **Principal and non-principal ideals.** The simplest ideals in R are given as "all the multiples of a given $a\in R$":

Lemma-Definition 3.16. *For $a\in R$, the set $\{ar\mid r\in R\}$ is an ideal. It is called the* **principal ideal** *generated by a. We write it as $aR=(a)_R=(a)$ (the latter notation, albeit sloppy, is the standard one, while in the book of Stewart–Tall, it is denoted $\langle a\rangle$).*

We collect a few simple immediate consequences of the definitions.

Lemma 3.17. *Let $I\subset R$ be an ideal, and let a, $b\in R$.*

(i) *For any $a\in R$, we have $(a)_R\subset I$.*
(ii) $a\mid b \quad\Leftrightarrow\quad (a)_R\supset(b)_R \quad\Leftrightarrow\quad b\in(a)_R\,;$
(iii) $a\sim b \quad\Leftrightarrow\quad (a)_R=(b)_R\,;$
(iv) $(a)_R\cdot(b)_R=(ab)_R\,;$
(v) $a\in R^* \quad\Rightarrow\quad (a)_R=R\,.$

Notation. For $a,b\in R$, we write

$$(a,b)_R=(a)_R+(b)_R=\{ar+bs\mid r,s\in R\}\,,$$

and more generally

$$(a_1,\dots,a_n)_R=\{\sum_{i=1}^n a_ir_i\mid r_i\in R\}\,,$$

the **ideal generated by** $\{a_1,\dots,a_n\}$.

Proposition 3.18. *Let a, b, c, $d\in R$, and let $I\subset R$ be an ideal. Then*

(i) $(a)_RI=aI\big(:=\{ar\mid r\in I\}\big)\,;$
(ii) $(a,b)_R\cdot(c)_R=(ac,bc)_R;$

(iii) $(a,b)_R \cdot (c,d)_R = (ac, bc, ad, bd)_R$ *and so forth for more generators:*

$$(a_1, \dots, a_m)_R \cdot (b_1, \dots, b_n)_R = (\dots, a_i b_j, \dots)_R \,.$$

We just indicate the proof of (ii), leaving the rest as a simple exercise:

$$(a,b)_R \cdot (c)_R = \big((a)_R + (b)_R\big) \cdot (c)_R = (a)_R(c)_R + (b)_R(c)_R = (ac)_R + (bc)_R \,.$$

For (iii), we need to apply the distributive law several times.

Example (of a non-principal ideal): take $R = \mathbb{Z}[\sqrt{-6}]$.

Claim: $I = (2, \sqrt{-6})$ is not principal.

Proof. Suppose I were principal, then for some $\alpha \in R$ (we can put $\alpha = a + b\sqrt{-6}$ for some $a, b \in \mathbb{Z}$) we have

$$I = (\alpha)_R = (a + b\sqrt{-6})_R \,.$$

Then $\alpha \mid 2$ and $\alpha \mid \sqrt{-6}$ ⟦as $\alpha = (2, \sqrt{-6})$ contains both (2) and $(\sqrt{-6})$⟧. Applying the norm map $N : a + b\sqrt{-6} \mapsto a^2 + 6b^2$ yet again gives $N(\alpha) \mid N(2) = 4$ and $N(\alpha) \mid N(\sqrt{-6}) = 6$, from which we deduce $N(\alpha) \mid 2$, i.e. $a^2 + 6b^2 = 1$ or $= 2$; but the latter is obviously not possible. Therefore we can colclude that $b = 0$ and $a = \pm 1$, i.e. $\alpha = \pm 1$, a unit.

But then we know that $I = (\pm 1)_R = R$ ⟦Lemma 3.17(v)⟧, so in particular $1 \in I$, and we should be able to write

$$1 = 2\beta + \sqrt{-6}\,, \qquad \text{for some } \beta, \gamma \in R\,.$$

Putting $\beta = r + s\sqrt{-6}$, $\gamma = t + u\sqrt{-6}$, then we find $1 = 2r - 6u + (2s+t)\sqrt{-6}$, and taking the real part on both sides of the latter equation gives $1 = 2r - 6u$ which obviously cannot hold.

Conclusion: our supposition (that I is principal) cannot hold. Therefore we have found that I is *not* principal. □

Although in general we cannot take the gcd of two numbers in a ring R (with identity denoted by $\mathbb{1}_R$), we still have it for the numbers $m \cdot \mathbb{1}_R$ which correspond to the integers $m \in \mathbb{Z}$:

Lemma 3.19. *Let R be an integral domain. If $m, n \in \mathbb{Z} \setminus \{0\}$ with $d = gcd(m,n)$, then*

$$(m \cdot \mathbb{1}_R, n \cdot \mathbb{1}_R)_R = (d \cdot \mathbb{1}_R)_R \,.$$

Proof. Since $d \mid m$ and $d \mid n$, we have $(d \cdot \mathbb{1}_R)_R \supset (m \cdot \mathbb{1}_R)_R$ and $(d \cdot \mathbb{1}_R)_R \supset (n \cdot \mathbb{1}_R)_R$, from which we deduce that the LHS equals $(m \cdot \mathbb{1}_R)_R + (n \cdot \mathbb{1}_R)_R \subset (d \cdot \mathbb{1}_R)_R$, the latter just being the RHS.

Moreover, since $d = am + bn$ for some $a, b \in \mathbb{Z}$, we have

$$d \cdot \mathbb{1}_R = a(m \cdot \mathbb{1}_R) + b(n \cdot \mathbb{1}_R) \in (m \cdot \mathbb{1}_R, n \cdot \mathbb{1}_R)_R$$

and so the RHS is contained in the LHS as well. □

Now we can "remedy" the non-uniqueness of factorisation, if only on the "level of ideals":

Example: In $R = \mathbb{Z}[\sqrt{-6}]$, we have

$$(1 + 3\sqrt{-6})(1 - 3\sqrt{-6}) = 5 \cdot 11 \qquad \text{as numbers in } R\,.$$

In terms of ideals this gives

$$(1 + 3\sqrt{-6})_R(1 - 3\sqrt{-6})_R = (5)_R \cdot (11)_R \qquad \text{as ideals in } R\,. \tag{4}$$

Now define two ideals

$$\mathfrak{p}_5 = (5, 1 + 3\sqrt{-6})_R\,, \qquad \mathfrak{p}_5' = (5, 1 - 3\sqrt{-6})_R\,,$$

and similarly

$$\mathfrak{p}_{11} = (11, 1+3\sqrt{-6})_R\,, \qquad \mathfrak{p}'_{11} = (11, 1-3\sqrt{-6})_R\,.$$

Then we have $\mathfrak{p}_5 \cdot \mathfrak{p}'_5 = (5)_R$ and $\mathfrak{p}_{11} \cdot \mathfrak{p}'_{11} = (11)_R$:

$$\begin{aligned}\mathfrak{p}_5 \cdot \mathfrak{p}'_5 &= (5, 1+3\sqrt{-6})_R \cdot (5, 1-3\sqrt{-6})_R \\ &= (25, 5\cdot(1-3\sqrt{-6}), (1+3\sqrt{-6})\cdot 5, 55)_R \\ &= (25, 55, 5\cdot(1-3\sqrt{-6}), 5\cdot(1+3\sqrt{-6}))_R \\ &= (25, 5, 5\cdot(1-3\sqrt{-6}), 5\cdot(1+3\sqrt{-6}))_R \\ &= (5)_R\,,\end{aligned}$$

the latter identity holds because all four generators are multiples of the second one, 5, so can be discarded.

A similar fact holds for $\mathfrak{p}_{11} \cdot \mathfrak{p}'_{11}$.

Now another possible product of the four ideals under consideration is

$$\begin{aligned}\mathfrak{p}_5 \cdot \mathfrak{p}_{11} &= (5, 1+3\sqrt{-6})_R \cdot (11, 1+3\sqrt{-6})_R \\ &= (55, 5\cdot(1+3\sqrt{-6}), (1+3\sqrt{-6})\cdot 11, (1+3\sqrt{-6})^2)_R \\ &= (55, 5\cdot(1+3\sqrt{-6}), 1+3\sqrt{-6}, (1+3\sqrt{-6})^2)_R \\ &= (1+3\sqrt{-6})_R\,,\end{aligned}$$

since all four generators are divisible by the third one, $1+3\sqrt{-6}$.

In a similar way, we can find that $\mathfrak{p}'_5 \cdot \mathfrak{p}'_{11} = (1-3\sqrt{-6})_R$.

Finally, (4) becomes

$$(\mathfrak{p}_5 \cdot \mathfrak{p}_{11})_R \cdot (\mathfrak{p}'_5 \cdot \mathfrak{p}'_{11})_R = (\mathfrak{p}_5 \cdot \mathfrak{p}'_5)_R \cdot (\mathfrak{p}_{11} \cdot \mathfrak{p}'_{11})_R\,,$$

which indicates that the original ambiguity of the decomposition is now resolved.

It turns out that the above ideals $\mathfrak{p}_i$ and $\mathfrak{p}'_i$ ($i \in \{5, 11\}$) can be viewed as "building blocks" among the ideals in $\mathbb{Z}[\sqrt{-6}]$, in a similar fashion as the prime numbers are building blocks for $\mathbb{Z}$. In particular, we will be able to deduce that if one of them divides one side of some equation, then it also has to divide the other side. So the following notion should be not particularly surprising.

Definition 3.20. *An ideal $\mathfrak{p} \subsetneq R$ is called* **prime** *if it satisfies the condition*

$$\forall a, b \in R \text{ with } a\cdot b \in \mathfrak{p} \text{ we have } a \in \mathfrak{p} \text{ or } b \in \mathfrak{p}\,.$$

Note that, just as $1 \in \mathbb{Z}$ is *not* a prime, we do not consider $(1)_R$ (which is equal to R itself) as a prime ideal ($(1)_R$ would "destroy" unique factorisation). On the other hand, $(0)_R$ is considered to be a prime ideal.

Proposition 3.21. *Let I, J and $\mathfrak{p}$ be non-zero ideals in R, let $\mathfrak{p}$ be prime. Then*

$$\mathfrak{p} \supset IJ \quad \Leftrightarrow \quad \mathfrak{p} \supset I \ \text{ or } \ \mathfrak{p} \supset J\,.$$

Proof. "$\Rightarrow$" is obvious, as e.g. $I \supset IJ$.

"$\Leftarrow$": Suppose $\mathfrak{p} \supset IJ$, but $\mathfrak{p} \not\supset I$. Then $\exists a \in I \setminus \mathfrak{p}$. Now for any $b \in J$ we have $a \cdot b \in I \cdot J \subset \mathfrak{p}$, so $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. But $a \notin \mathfrak{p}$ ⟦by the choice of a⟧, so $b \in \mathfrak{p}$.

Conclusion: $J \subset \mathfrak{p}$. □

Note: Let us define *divisibility of ideals* in the obvious manner, i.e., $I \mid J$ (for two ideals I and J in R) if there is an ideal K such that $J = I \cdot K$. Then it is clear that $I \mid J$ implies $I \supset J$ ⟦since $J = IK \subset IR \subset I$⟧. The converse holds only for special rings—e.g., for so-called "Dedekind rings", to be introduced later—in which case the proposition says: $\mathfrak{p} \mid IJ \ \Rightarrow \ \mathfrak{p} \mid I$ or $\mathfrak{p} \mid J$. In other words: prime ideals then "behave" analogously to prime elements. Good news: most of the rings in the course will indeed turn out to be "Dedekind rings".

Definition 3.22. *An ideal $\mathfrak{m} \subsetneq R$ is called* **maximal** *if there is no ideal properly containing it except R itself, i.e., for any ideal I in R, we have $I \supsetneq \mathfrak{m} \Rightarrow I = R$.*

Recall that, for a ring R and an ideal I in R, the set of cosets $r + I$, $r \in R$, forms a ring, the **quotient ring of R with respect to I**, which is denoted R/I. ⟦This is compatible to our previous notation: $r + I = \{r\} + I = \{r + i \mid i \in I\}$. Furthermore, we have an addition of cosets: $(r + I) + (s + I) = (r + s) + I$, and a multiplication of cosets: $(r + I)(s + I) = (rs) + I$.⟧ Note that $a + I = I \Leftrightarrow a \in I$.

Now there is a very useful characterisation of prime and maximal ideals, respectively, in terms of the corresponding quotient rings.

Theorem 3.23. *Let R be an integral domain.*

(1) *An ideal $\mathfrak{p} \subset R$ is prime $\Leftrightarrow$ $R/\mathfrak{p}$ is an integral domain.*
(2) *An ideal $\mathfrak{m} \subset R$ is maximal $\Leftrightarrow$ $R/\mathfrak{m}$ is a field.*

Proof

(1) Let a, $b \in R$. They correspond to cosets $a + \mathfrak{p}$, $b + \mathfrak{p}$ in $R/\mathfrak{p}$.

The *prime condition* $ab \in \mathfrak{p} \Rightarrow a \in \mathfrak{p}$ or $b \in \mathfrak{p}$ translates into the *integral domain condition* "no zero divisors"

$$a \cdot b \in \mathfrak{p} = \bar{0} \text{ in } R/\mathfrak{p} \Rightarrow a + \mathfrak{p} = \bar{0} \text{ or } b + \mathfrak{p} = \bar{0} \text{ in } R/\mathfrak{p}\,.$$

Note that, moreover, $\mathbb{1}_R \in R$ maps to an identity $\mathbb{1}_{R/\mathfrak{p}} (= \mathbb{1}_R + \mathfrak{p})$ in $R/\mathfrak{p}$.

(2) "$\Rightarrow$": Suppose $\mathfrak{m}$ is maximal. **Need to show:** any class $a + \mathfrak{m}$, $a \notin \mathfrak{m}$, has an inverse. ⟦Here $a + \mathfrak{m} = \{a\} + \mathfrak{m} = \{a + m \mid m \in \mathfrak{m}\}$ is the coset notation, not to be confused with the ideal addition.⟧

Since $(a)_R + \mathfrak{m} \supsetneq \mathfrak{m}$, it must be equal to R ⟦by the maximality of $\mathfrak{m}$⟧. In particular, we have $\mathbb{1}_R \in (a)_R + \mathfrak{m}$, i.e., $1_R = ba + cm$ for some $b, c \in R$. For the corresponding cosets with respect to $\mathfrak{m}$, we get

$$1_R + \mathfrak{m} = ba + cm + \mathfrak{m} = ba + \mathfrak{m} = (b + \mathfrak{m})(a + \mathfrak{m})\,.$$

Conclusion: for $a \notin \mathfrak{m}$, we have found an inverse $b + \mathfrak{m}$ in $\mathbb{R}/\mathfrak{p}$.

"$\Leftarrow$": Suppose $R/\mathfrak{m}$ is a field. Take an ideal $\mathfrak{n}$ such that $\mathfrak{m} \subsetneq \mathfrak{n} \subset R$. **Need to show:** $\mathfrak{n} = R$. ⟦Then we can conclude that $\mathfrak{m}$ has to be maximal.⟧

Choose $a \in \mathfrak{n} \setminus \mathfrak{m}$ ⟦this is possible, as our assumption on $\mathfrak{n}$ implies $\mathfrak{n} \setminus \mathfrak{m} \neq \emptyset$⟧. Then $a + \mathfrak{m} \neq \mathfrak{m}$, so it must have an inverse, say $b + \mathfrak{m}$. ⟦Note that necessarily $b + \mathfrak{m} \neq \mathfrak{m}$, i.e., $b \notin \mathfrak{m}$.⟧ Thus $ab + \mathfrak{m} = \mathbb{1}_R + \mathfrak{m}$ and in particular $\mathbb{1}_R \in (a)_R + \mathfrak{m} \subset \mathfrak{n}$, which implies that $\mathfrak{n} = R$. □

Corollary 3.24. *Every maximal ideal is also a prime ideal.*

3.3. **Principal ideal domains and Euclidean domains.** We have seen above that it is preferable to work in a unique factorisation domain. But it is not clear how to make sure that a given ring is indeed a UFD. If we could actually argue with ideals as we are used to do for the integers, say, then we should be in a good position to prove a statement like unique factorisation. A "nice" ring R in this respect would be one in which any ideal came from a *single* element in R.

Definition 3.25. *An integral domain R is called a* **principal ideal domain (PID)** *if all its ideals are principal ideals (i.e., can be written with a single generator).*

Examples:

1) In $\mathbb{Z}$, every ideal has the form $(m)_{\mathbb{Z}}$, for some $m \in \mathbb{Z}$. Thus $\mathbb{Z}$ is a PID.

2) In $\mathbb{Q}[X]$, every ideal has the form $(f(X))_{\mathbb{Q}[X]}$, for some polynomial $f(X) \in \mathbb{Q}[X]$, and so $\mathbb{Q}[X]$ is a PID.
3) The rings $\mathbb{Z}[\sqrt{-5}]$ and $\mathbb{Z}[\sqrt{-6}]$ are *not* PIDs (see our examples above).
4) The ring $\mathbb{Z}[X]$ is *not* a PID: e.g., the ideal $(2, X)_{\mathbb{Z}[X]}$ cannot be written with a single generator.

Theorem 3.26. *Every PID is a UFD.*

An important step in the proof of the theorem is the following

Proposition 3.27. *In a PID R, every irreducible element is prime.*

Proof. Let π be irreducible in R, and suppose that $\pi \mid \alpha\beta$ for some $\alpha, \beta \in R$.

We have to show: $\quad \pi \mid \alpha \quad$ or $\quad \pi \mid \beta$.

Consider the ideal generated by π and α, denote it by $I = (\pi, \alpha)_R$. Since R is a PID, there is a $\gamma \in R$ such that $I = (\gamma)_R$, in particular $\gamma \mid \pi$ (and $\gamma \mid \alpha$).

But π is irreducible, so either I) $\gamma \sim \pi$ or II) $\gamma \sim 1$ ⟦i.e., γ is a unit⟧.

Case I) implies $\pi \mid \alpha$ ⟦as $\gamma \mid \alpha$⟧, while Case II) implies $1 = \lambda\pi + \mu\alpha$, and multiplying both sides by β gives

$$\beta = \beta\lambda\pi + \mu\alpha\beta\,.$$

Now since π divides the RHS, we have that $\pi \mid$ LHS as well i.e., $\pi \mid \beta$.

Conclusion: in either case the claim is shown. □

The rest of the proof of the theorem involves claims like

Proposition 3.28. *In a PID R, each element can be factored into (a finite number of) irreducibles.*

The proof of the latter is somewhat more involved, one typically introduces the notion of a **Noetherian ring**: a ring in which every ideal is finitely generated. The rings that we consider in the course will typically be of that type. (An example of a non-Noetherian ring is the polynomial ring over $\mathbb{Q}$ in *infinitely many* generators $\mathbb{Q}[X_1, X_2, X_3, \dots]$.) One shows that the above condition (that every ideal is finitely generated) can be equivalently stated as saying that each ascending chain of ideals $I_1 \subseteq I_2 \subseteq \cdots \subseteq I_n \subseteq \dots$ becomes stationary, i.e. $I_m = I_{m+1}$ for all large enough $m \in \mathbb{N}$. Yet another equivalent condition is that every (non-empty) set of ideals has a *maximal element*, i.e., an element which is not properly contained in any other element of that set. (Cf., e.g., Proposition 4.5 in Stewart-Tall.) The above proposition then is a corollary of the fact that the corresponding statement indeed holds for *any Noetherian ring* (cf. Theorem 4.6 in Stewart–Tall). ⟦Note that a PID is (rather obviously) a Noetherian ring.⟧

Finally one shows that, granted one can factor into irreducibles, a ring is a UFD if (and only if) every irreducible element is prime (cf., e.g., Theorem 4.13 in Stewart-Tall.) Putting this together with the two propositions above then provides a proof of the Theorem.

What have we won so far? Instead of checking whether an integral domain is a UFD, we are now left with the task of checking whether it is a PID. Now if we had a way to always replace, in an ideal $I = (a_1, \dots, a_n)_R$, two generators by a single one, then we would succeed—since after a finite number of steps we are left with a single generator only, i.e., I would indeed turn out to be a principal ideal.

Recall how this is achieved for $\mathbb{Z}$: $(m, n)_{\mathbb{Z}} = (d)_{\mathbb{Z}}$, where $d = \gcd(m, n)$; and the gcd can be obtained by the Euclidean algorithm, the basis of which is division with remainder.

Examples:

1) In $\mathbb{Z}$, divide a by b: we can find q and r such that $a = q \cdot b + r$ and with the crucial condition on r being $0 \leqslant r < b$.
2) In $\mathbb{Q}[X]$, divide similarly two polynomials, say, $a(X)$ by $b(X)$. This time there is no "smaller" relation among the elements in $\mathbb{Q}[X]$, but still we can introduce some notion of size: the degree of the polynomial. Then there are $q(X)$ and $r(X)$ such that $a(X) = q(X)b(X) + r(X)$ and with the crucial condition on $r(X)$ being: either $r = 0$ or $\deg(r(X)) < \deg(b(X))$.

This suggests the following: whenever we have a "good" way to measure the size of elements in R, there is a chance that a gcd can be taken ⟦and then R has a chance to be a PID, and in particular a UFD⟧. Some consistencies should be kept in mind, though: the size should be measured by, say, numbers in $\mathbb{N} \cup \{0\}$ (it is not enough to take $\mathbb{Z}$, otherwise there may not be a stopping criterion); furthermore, the size should somehow be compatible with divisibilities (if $a \mid b$ then size$(a) \leqslant$ size(b)).

Definition 3.29. *Let R be an integral domain. A* **Euclidean function (or norm)** *for R is a function $\varphi : R \setminus \{0\} \to \mathbb{N}$ such that*

(i) *for $a, b \in R \setminus \{0\}$, one has* $\quad a \mid b \;\Rightarrow\; \varphi(a) \leqslant \varphi(b)$;
(ii) $\forall a, b \in R \setminus \{0\}\ \exists\, q, r \in R: \quad a = b \cdot q + r$ *with either* $r = 0$ *or* $\varphi(r) < \varphi(b)$.

Examples:

1) For $\mathbb{Z}$, consider $\varphi : \mathbb{Z} \setminus \{0\} \to \mathbb{N}$ given by $a \mapsto |a|$ (and extend by $0 \mapsto 0$).
2) For $\mathbb{Q}[X]$, consider $\varphi : \mathbb{Q}[X] \setminus \{0\} \to \mathbb{N}$ given by $a(X) \mapsto \deg(a(X))$ (and we can extend it by putting $\varphi(0) = -\infty$).
3) For $\mathbb{Z}[i]$, consider $\varphi : \mathbb{Z}[i] \setminus \{0\} \to \mathbb{N}$ given by $a + bi \mapsto N(a + bi) = a^2 + b^2$.

Definition 3.30. *An integral domain for which a Euclidean function exists is called a* **Euclidean domain***.*

Geometric idea to prove 3) above, i.e., that $\mathbb{Z}[i]$ is Euclidean: consider the elements in $\mathbb{Z}[i] \subset \mathbb{C}$ as lattice points ((a, b) with $a,\ b \in \mathbb{Z}$) in the plane (where a complex number $x + iy$ is identified as usual with the point $(x, y) \in \mathbb{R}^2$). To visualise the division with remainder for two elements α, β in $\mathbb{Z}[i]$, take the point in the plane corresponding to their quotient α/β (which certainly lies in $\mathbb{Q}[i] \subset \mathbb{C}$) and choose a nearest lattice point (s, t) to approximate it (this need not be unique!). Then the corresponding point $\gamma = s + it$ satisfies

$$\left|\frac{\alpha}{\beta} - \gamma\right| \leqslant \frac{1}{2}\sqrt{2} < 1\,,$$

and putting $r := \alpha - \beta\gamma$, we get $|r| = |\alpha - \beta\gamma| < |\beta|$.

Theorem 3.31. *A Euclidean domain R is also a PID.*

Proof. Let I be an ideal in the Eculidean domain R, and let φ be a Euclidean function for R.

To show: I is principal.

We can assume that $I \neq (0)_R$ ⟦$I = (0)_R$ *is* principal⟧and so we can choose an $x \neq 0$ in I.

Main point: We can choose x such that $\varphi(x)$ is *minimal.*

Now take any $y \in I$ and show that it is a multiple of x: division with remainder of y by x gives $y = qx + r$ for some $q, r \in R$, with $r = 0$ or $\varphi(r) < \varphi(x)$.

Both y and x are in I, so r, as a linear combination of the two, must also be. Due to the minimality of $\varphi(x)$ we have in fact $r = 0$, whence $y = qx$, a multiple of x.

Conclusion: since any $y \in I$ is a multiple of $x \in I$, it follows that I is principal (with generator x). □

It is clear now how to define a gcd for elements in a *Euclidean domain* R: as the last “divisee” in the Euclidean algorithm which results from a Euclidean function on R.

Lemma 3.32. *Let α, beta, gamma $\in R$, an integral domain. Then*

$$\operatorname{gcds}(\alpha,\beta)=\operatorname{gcds}(\alpha,\beta-\gamma\alpha)\,,$$

wher “gcds” detoes the set of all possible gcd’s.

⟦Proof: Common divisors on the left are also common divisors on the right and vice versa.⟧

Remark 3.33.

1) *There are comparatively few Euclidean domains known; e.g. one knows around two dozens among $\mathbb{Z}[\sqrt{m}]$ or, if $m\equiv 1(4)$, among the $\mathbb{Z}\big[\frac{1+\sqrt{m}}{2}\big]$.*
2) *One can weaken the condition on the Eulidean function somewhat, and still deduce that the corresponding ring is a UFD. With that generalization, we may produce a few more examples.*

The remark makes it clear that this approach (i.e., trying to find UFDs by establishing a Euclidean function on them) is not really the way to go if we want to develop a general theory. Instead, we will find unique factorization into prime *ideals*, in particular for so-called “number rings” (like $\mathbb{Z}[\sqrt{m}]$ or $\mathbb{Z}[\zeta_n]$, to be defined more precisely below) which naturally lie inside “number fields” (like $\mathbb{Z}$ inside $\mathbb{Q}$, or $\mathbb{Z}[i]$ inside $\mathbb{Q}[i]$).

3.4. **Number fields.** We have already encountered fields like $\mathbb{Q}$ or $\mathbb{Q}(\sqrt{m})$. They can be viewed as subfields of $\mathbb{C}$. ⟦Not all fields are subfields of $\mathbb{C}$: for example, the finite fields $\mathbb{Z}/p^r\mathbb{Z}$ (p prime, $r\geqslant 1$) cannot be embedded into $\mathbb{C}$—where “embedded” means via a homomorphism, not just as a set; other example: $\mathbb{C}(X)$, the field of rational functions in one variable X.⟧

There is an obvious (ring) homomorphism $\mathbb{Q}\to\mathbb{Q}(\sqrt{m})$, sending $q\in\mathbb{Q}$ to $q+0\cdot\sqrt{m}$. Thus we can view $\mathbb{Q}$ as a subfield of $\mathbb{Q}(\sqrt{m})$ or, conversely, $\mathbb{Q}(\sqrt{m})$ as an “overfield” or as a “field extension” of $\mathbb{Q}$. More generally:

Definition 3.34. *Let K and L be fields. If K is contained in L, then K is a* subfield *of L; conversely, L is a* field extension *of F.*

Here “contained” means “contained as a subring” (i.e. 0 and 1 agree, and F is closed under $+$ and $\cdot$.)

Remark 3.35. *If L is a field extension of F, then L is in particular a* vector space *over F ⟦recall: F-vector space = abelian group with scalar multiplication by elements of F⟧.*

Example: $\mathbb{Q}(\sqrt{-2})=\{a+b\sqrt{-2}\mid a,b,\in\mathbb{Q}\}$ is isomorphic, as a vector space only, to $\{(a,b)\mid,b\in\mathbb{Q}\}\simeq\mathbb{Q}\oplus\mathbb{Q}$, a 2-dimensional vector space over $\mathbb{Q}$.

We have the following correspondence:

$$\begin{array}{rcl} & \text{addition:} & \\ a_1+b_1\sqrt{-2} & \leftrightarrow & (a_1,b_1)\,, \\ +_R\ (a_2+b_2\sqrt{-2}) & \leftrightarrow & \oplus\ (a_2,b_2)\,, \\ =(a_1+a_2)+(b_1+b_2)\sqrt{-2} & \leftrightarrow & =\ (a_1+a_2,b_1+b_2)\,, \\ \text{scalar} & \text{multiplication:} & \\ r(a_1+b_1\sqrt{-2}),\ r\in\mathbb{Q} & \leftrightarrow & r(a_1,b_1)\,, \\ =ra_1+rb_1\sqrt{-2}, & \leftrightarrow & (ra_1,rb_1)\,. \end{array}$$

Think of 1 and $\sqrt{-2}$ as basis vectors in $\mathbb{Q}(\sqrt{-2})$ corresponding to $(1,0)$ and $(0,1)$ in $\mathbb{Q}+\mathbb{Q}$, respectively.

Definition 3.36. *Let L be a field extension of F. Then the* **degree** $[L:F]$ *of L over F is given by the dimension* $\dim_L(F)$ *of L as a vector space over F.*

Example:

1) $[\mathbb{C}:\mathbb{R}]=2$, with standard basis $\{1,i\}$;
2) Similarly, for m a non-square in $\mathbb{Z}$, we have
$$[\mathbb{Q}(\sqrt{m}):\mathbb{Q}]=2\,,$$
with basis, e.g., $\{1,\sqrt{m}\}$.
3) $$\begin{aligned}\mathbb{Q}(\sqrt[3]{2}) &= \{a+b\sqrt[3]{2}+\left(\sqrt[3]{2}\right)^2 \mid a,b,c\in\mathbb{Q}\}\\ &\simeq \mathbb{Q}\oplus\mathbb{Q}\oplus\mathbb{Q}=\mathbb{Q}^3\,,\end{aligned}$$
a 3-dimensional vector space over $\mathbb{Q}$ (the sign $\simeq$ here denotes isomorphism of vector spaces). Here $\sqrt[3]{2}$ is a root of the (by Eisenstein irreducible) polynomial x^3-2. ⟦An elementary way to see that 1, $\sqrt[3]{2}$ and $(\sqrt[3]{2})^2$ are linearly independent: suppose they were linearly dependent, i.e., for some a, b, c in $\mathbb{Z}$ with gcd 1 we have $a+b\sqrt[3]{2}=c(\sqrt[3]{2})^2$. Taking cubes on both sides gives $a^3+2b^3+6abc=4c^3$, and now considering successively mod 2, mod 4 and mod 8 we can conclude that $2\mid a$, $2\mid b$ and $2\mid c$, respectively, contradicting the gcd 1 condition on a, b and c.⟧

Definition 3.37. *Let L be a field extension of F. An element $\alpha\in L$ is* **algebraic** *over F if it satisfies $f(\alpha)=0$ for some polynomial $f(X)\in F[X]$. If all elements of L are algebraic over F, then L is called an* algebraic extension *of F (or simply "is algebraic over F")*

Examples:

1) $\mathbb{C}$ is algebraic over $\mathbb{R}$ with standard basis $\{1,i\}$, but it is not algebraic over $\mathbb{Q}$ ⟦e.g., the famous number $\pi=\sqrt{6\sum_{n=1}^{\infty}n^{-2}}=3.1415\ldots$ is not⟧.
2) $\mathbb{Q}(\sqrt[m]{n})$ is algebraic over $\mathbb{Q}$, for any $m\geqslant 2$.
3) $\mathbb{Q}(\sqrt[2]{5})(X)$ is algebraic over $\mathbb{Q}(X)$.

Proposition 3.38. *If $[L:F]=d<\infty$, then L is algebraic over F.*

Proof. Take any $\alpha\in L$ and form the set $\{1,\alpha,\alpha^2,\ldots,,\alpha^d\}$ of cardinality $d+1$, the elements of which lie in L. They are linearly dependent (since $\dim_F(L)=d$), i.e. for some $r_i\in F$ one has $\sum_{i=0}^{d}r_i\alpha^i=0$, i.e., α is root of $f(X)=\sum_{i=0}^{d}r_iX^i$; in particular, α is algebraic over F. □

Definition 3.39. *A number $\alpha\in\mathbb{C}$ which is algebraic over $\mathbb{Q}$ is called an* **algebraic number**. *A field with $\mathbb{Q}\subset F\subset\mathbb{C}$ and $[F:\mathbb{Q}]<\infty$ is called an* **(algebraic) number field**.

Examples:

- $\sqrt[17]{13}-\sqrt{3\sqrt{-5}+\frac{1}{\sqrt[3]{-7^5}}}$ is algebraic.
- One can show: e (Euler's number) and π are *not* algebraic (instead they are called "transcendental").
- $\mathbb{Q}(\sqrt[n]{m})$, $n\geqslant 2$, $m\in\mathbb{Z}$, defines a number field.

- In fact, any number field is isomorphic to a quotient ring
$$\mathbb{Q}[X]/\big(f(X)\big)_{\mathbb{Q}[X]}$$
for some irreducible polynomial $f(X)$. ⟦Since $f(X)$ is irreducible in the Euclidean domain $\mathbb{Q}[X]$, it follows that $\big(f(X)\big)$ is a maximal ideal (cf. Problem Sheet 4, 4(i)); therefore the above quotient ring is indeed a field.⟧

Definition 3.40. *Let α be algebraic over a field F. The* **minimum polynomial of** *α is the monic polynomial of smallest degree in $\mathbb{Q}[X] \setminus \{0\}$ such that $f(\alpha) = 0$.*

⟦This is unique, and in fact irreducible.⟧

Examples:

- The minimum polynomial of $i = \sqrt{-1}$ and $\sqrt{3}$ over $\mathbb{Q}$ are given by $X^2 + 1$ and $X^2 - 3$, respectively.
- The minimum polynomial of $\sqrt[n]{m}$ over $\mathbb{Q}$ is *not always* given by $X^n - m$: $\sqrt[2]{25}$ is either 5 or -5, so its minimum polynomial equals $X - 5$ or $X + 5$.
- The minimum polynomial of $\alpha = 3 + i$ over $\mathbb{Q}$ is given by $X^2 - 6 + 10$, since α satisfies $(\alpha - 3)^2 = i^2 = -1$ (and it obviously cannot have a linear (i.e. degree 1) minimum polynomial over $\mathbb{Q}$).
- What is the minimum polynomial of $\alpha = \sqrt{3} + i$ over $\mathbb{Q}$? We square both sides of the equation $\alpha - i = \sqrt{3}$, thus getting rid of at least one square root: $(\alpha - i)^2 = 3$, and the resulting identity $\alpha^2 - 2 = 2\alpha i$ (we again try to separate one of the square roots from the rest) gets squared a second time, yielding that α is a root of the polynomial $X^4 - 4X^2 + 16$. Note that α^2 satisfies the quadratic equation $X^2 - 4X + 16$, so we can first solve for α^2 and then take the square root, which gives the degree $2 \cdot 2 = 4$ for α. This is an instance of the following

Theorem 3.41. *(The Tower Theorem) Let $L \supset K \supset F$ be algebraic field extensions. Then*
$$[L : F] = [L : K] \cdot [K : F]\,.$$
More precisely, if $\{\alpha_1, \ldots, \alpha_r\}$ is a basis for K over F and $\{\beta_1, \ldots, \beta_s\}$ is a basis for L over K, then $\mathcal{B} := \{\alpha_j\beta_k \mid 1 \leqslant j \leqslant r,\ 1 \leqslant k \leqslant s\}$ is a basis for L over F.

Proof. Let $\gamma \in L$, then $\gamma = \sum_{k=1}^{s} \lambda_k\beta_k$ for some $\lambda_k \in K$, and each λ_k can be written as $\lambda_k = \sum_{j=1}^{r} \mu_{jk}\alpha_j$ for some $\mu_{jk} \in F$, whence $\gamma = \sum_k \sum_j \mu_{jk}\alpha_j\beta_k$. Thus $\mathcal{B}$ *spans* L over F.

We still need to show the linear independence of the vectors $\alpha_j\beta_k$, in order to establish the basis property of $\mathcal{B}$: so suppose $\sum_j \sum_k \mu_{jk}\alpha_j\beta_k$. Regrouping terms gives
$$\sum_k \Big(\sum_j \mu_{jk}\alpha_j\Big)\beta_k = 0\,,$$
but the β_k are a basis of K over F, thus necessarily $\mu_{jk}\alpha_j = 0$ for all $k = 1, \ldots, s$. Now use that the α_j in turn form a basis of L over K, so that necessarily all $\mu_{jk} = 0$.

This establishes the linear independence of $\mathcal{B}$. □

Example: Let $L = \mathbb{Q}\big(\sqrt{2}, \sqrt[3]{5}\big) \supset K = \mathbb{Q}(\sqrt{2}) \supset F = \mathbb{Q}$. (By $\sqrt[3]{5}$ we understand the *real* root of the (Eisenstein-)irreducible polynomial $X^3 - 5$.)

We first note that $\alpha = \sqrt[3]{5} \notin K$ ⟦α has degree 3, while any element $a + b\sqrt{2} \in K$ $(a, b \in \mathbb{Q})$ has degree $\leqslant 2$⟧. The other (non-real) roots of $X^3 - 5$ are also not in $\mathbb{Q}(\sqrt{2})$, from which we deduce that α has the same minimum polynomial *over K*.

But $L = \mathbb{Q}\big(\sqrt{2}, \sqrt[3]{5}\big) = \Big(\mathbb{Q}(\sqrt{2})\Big)(\sqrt[3]{5}) = K(\sqrt[3]{5})$ and so $[L : K] = 3$. Furthermore, we have of course $[K : \mathbb{Q}] = 2$, and so the Tower Theorem gives $[L : \mathbb{Q}] = 6$, a basis of $L/\mathbb{Q}$ can e.g. be given by $\{1, \sqrt{2}, \sqrt[3]{5}, \sqrt[3]{5}\sqrt{2}, (\sqrt[3]{5})^2, (\sqrt[3]{5})^2\sqrt{2}\}$.

Can we perhaps generate L by a single element? A typical candidate is $\sqrt{2}+\sqrt[3]{5}$ (or also the product of the two generators, as a member of the audience suggested in the lecture) ⟦squaring still leaves us with a cube root, while taking cubes still leaves us with a square root, the smallest conceivable power which would make both terms rational thus being 6⟧. Indeed, we have more generally the

Theorem 3.42. *(Simple Extension Theorem) Every algebraic number field K (i.e. $[K:\mathbb{Q}]<\infty$) has the form $K=\mathbb{Q}(\theta)$ for some $\theta\in K$.*

⟦Idea of proof: reduce the number of generators successively, a typical reduction step being—with α and β generating algebraic elements over $\mathbb{Q}$—the following: $\mathbb{Q}(\alpha)(\beta)=\mathbb{Q}(\alpha,\beta)\overset{!}{=}\mathbb{Q}(\alpha+\lambda\beta)$ for some $\lambda\in\mathbb{Q}$, in fact, most λ do the trick, but one needs to perform this carefully (see, e.g., Theorem 2.2 in Stewart–Tall).⟧

We still need to justify the notation $\mathbb{Q}(\alpha)$ (which indicates a quotient field) for the ring $\mathbb{Q}[\alpha]$, if α is an algebraic number.

⟦Aside: recall that one can obtain $\mathbb{Q}$ as a **quotient field** $\mathbb{Q}=\mathrm{frac}(\mathbb{Z})$ of the ring of integers. One introduces pairs (a,b) which correspond to rational numbers $\frac{a}{b}$, defines a multiplication on those pairs which exactly mirrors the one for rational numbers (simply put $(a,b)*(a',b'):=(aa',bb')$). Inversion corresponds to swapping the two members of such a pair, addition is defined as $(a,b)+(a',b')=(ab'+a'b,bb')$, and finally one identifies two such pairs if the corresponding rational expressions represent the same fraction $\big($i.e., $(a,b)\sim(a',b')$ if there are $c,d\in\mathbb{Z}$ such that $(ac,bc)=(a'd,b'd)\big)$. Analogously we can form the fraction field $\mathrm{frac}(R)$ of any integral domain R.⟧

We already know that for $\deg(\alpha)=2$ the ring $\mathbb{Q}[\alpha]$ agrees with its quotient field $\mathbb{Q}(\alpha)$: by using the corresponding *norm map* we can invert each element in a quadratic field (if $\beta=a+b\sqrt{D}$, then $1/\beta=(a-b\sqrt{D})/N(\beta)=a/N(\beta)-b/N(\beta)\sqrt{D}\in\mathbb{Q}[\alpha]$). For number fields of higher degree this is less obvious, but should become clear in the following.

We can think of a "hierarchy of structures" for a number field K; we illustrate this first in the case $[K:\mathbb{Q}]=2$.

as a $\mathbb{Q}$-vector space	⇐	as a ring	⇐	as a field
$\mathbb{Q}+\mathbb{Q}\cdot\sqrt{D}$ addition and multipl. by scalars (elts. in $\mathbb{Q}$)		$\mathbb{Q}[\sqrt{D}]$ addition and multipl. by elts. in K		$\mathbb{Q}(\sqrt{D})$ addition and multipl. in K + inverses

How would the ring multiplication in $\mathbb{Q}[\sqrt{D}]$ look like on the underlying vector space? We compute it on the obvious (ordered) basis $\{\beta_1:=1,\beta_2:=\sqrt{D}\}$:

$$\begin{aligned}\beta_1 &: a+b\sqrt{D} &\mapsto\quad& a+b\sqrt{D}=a\cdot\beta_1+b\cdot\beta_2\,,\\ \beta_2 &: a+b\sqrt{D} &\mapsto\quad& a\sqrt{D}+bD=bD\cdot\beta_1+a\cdot\beta_2\,,\end{aligned}$$

and we obtain, after identifying $\beta_1\widehat{=}(1,0)$ and $\beta_2\widehat{=}(0,1)$ in $\mathbb{Q}+\mathbb{Q}$ that

$$\begin{pmatrix}1\\0\end{pmatrix}\mapsto\begin{pmatrix}a&*\\b&*\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix},\qquad\begin{pmatrix}0\\1\end{pmatrix}\mapsto\begin{pmatrix}*&bD\\ *&a\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix},$$

which together produces the matrix

$$A=\begin{pmatrix}a&bD\\b&a\end{pmatrix}.$$

Therefore we can view $\alpha = a + b\sqrt{D} \in \mathbb{Q}[\sqrt{D}]$ as producing a linear map with the above matrix A. ⟦Furthermore, we know that $\alpha \neq 0$ has an inverse in $\mathbb{Q}[\sqrt{D}]$, and we can check that the group $\mathbb{Q}[\sqrt{D}]^*$ (=the units) *acts* on the vector space $\mathbb{Q} + \mathbb{Q} \cdot \sqrt{D}$ in the sense of representation theory.⟧

Example: Consider a field $K = \mathbb{Q}[\theta]$ of degree 3 over Q, where $\theta^3 - \theta + 2 = 0$ (i.e., the minimum polynomial $p_\theta(X) = X^3 - X + 2$, which obviously is irreducible). In this case we get

$$\begin{aligned}
\beta_1 &: a + b\theta + c\theta^2 &\mapsto\ & a \cdot \beta_1 + b \cdot \beta_2 + c \cdot \beta_3 \,, \\
\beta_2 &: a + b\theta + c\theta^2 &\mapsto\ & a\theta + b\theta^2 + \underbrace{c\theta^3}_{=c\theta - 2c} = -2c \cdot \beta_1 + (a + c) \cdot \beta_2 + b \cdot \beta_3 \,, \\
\beta_3 &: a + b\theta + c\theta^2 &\mapsto\ & a\theta^2 + \underbrace{b\theta^3}_{=b\theta - 2b} + \underbrace{c\theta^4}_{=c\theta^2 - 2c\theta} = -2b \cdot \beta_1 + (b - 2c) \cdot \beta_2 + (a + c) \cdot \beta_3 \,.
\end{aligned}$$

The corresponding matrix therefore has the form

$$A = \begin{pmatrix} a & -2c & -2b \\ b & a + c & b - 2c \\ c & b & a + c \end{pmatrix} \,.$$

Any $\alpha = a + b\theta + c\theta^2$ thus defines the multiplication-by-α map

$$\begin{aligned}
\hat{\alpha} : \mathbb{Q}[\theta] &\to \mathbb{Q}[\theta] \,, \\
\lambda &\mapsto \alpha \cdot \lambda \,,
\end{aligned}$$

which is *linear* (i.e., $\hat{\alpha}(r\lambda) = r\hat{\alpha}(\lambda)$ if $r \in \mathbb{Q}$ and $\hat{\alpha}(\lambda + \mu) = \hat{\alpha}(\lambda) + \hat{\alpha}(\mu)$ for $\lambda, \mu \in \mathbb{Q}[\theta]$), and this in turn gives a map of vector spaces of $\mathbb{Q} + \mathbb{Q}\theta + \mathbb{Q}\theta^2$ to itself which on our standard basis $\{1, \theta, \theta^2\}$ is given by the above matrix.

Recall from linear algebra that the matrix associated to a linear map of vector spaces depends on the choice of a basis, but we can derive from it basis invariant information: its determinant and its trace, or better even its characteristic polynomial.

Definition 3.43. *Let K be a number field. The (**absolute**) **norm** and **trace** of $\alpha \in K$ **from** K **to** $\mathbb{Q}$ are defined as*

$$\begin{aligned}
\mathrm{N}_K(\alpha) &= \det(A) \,, \\
\mathrm{Tr}_K(\alpha) &= \mathrm{trace}(A) \,,
\end{aligned}$$

where A denotes the matrix representing the $\mathbb{Q}$-linear map $\hat{\alpha}$ associated to α.

Note: Both $\mathrm{N}_K(\alpha)$ and $\mathrm{Tr}_K(\alpha)$ lie in $\mathbb{Q}$ ⟦since the entries in the corresponding matrix A do⟧.

Examples:

1) Let $\alpha := a + b\sqrt{D} \in K = \mathbb{Q}[\sqrt{D}]$, then $\mathrm{N}_K(\alpha) = \det \begin{pmatrix} a & bD \\ b & a \end{pmatrix} = a^2 - b^2 D$, which fittingly coincides with our old norm map (for fields of degree 2 over $\mathbb{Q}$).
2) Let $K = \mathbb{Q}[\theta]$, where $\theta^3 = \theta - 2$ (as in one of the examples above). Then

$$\begin{aligned}
\mathrm{N}_K(\alpha) &= a^3 - 2b^3 + 4c^3 + 2a^2 c + ac^2 - ab^2 + 2bc^2 + 6abc \\
\text{and} \quad & \mathrm{Tr}_K(\alpha) = 3a + 2c \,.
\end{aligned}$$

Proposition 3.44. *Let K be a number field, Then*

(i) *for $\alpha \in K$, we have: $\mathrm{N}_K(\alpha) = 0 \iff \alpha = 0$;*

(ii) multiplicativity *of the norm (certainly the most important property of the "old" norm that we have used so far):*

$$\forall\alpha,\beta\in K: \qquad \mathrm{N}_K(\alpha\beta)=\mathrm{N}_K(\alpha)\mathrm{N}_K(\beta)\,;$$

(iii) $\mathbb{Q}$-linearity *of the trace:*

$$\forall\alpha,\beta\in K\,,\forall\lambda,\mu\in\mathbb{Q}: \qquad \mathrm{Tr}_K(\lambda\alpha+\mu\beta)=\lambda\mathrm{Tr}_K(\alpha)+\mu\mathrm{Tr}_K(\beta)\,,$$

i.e., $\mathrm{Tr}_K: K\to\mathbb{Q}$ *is a* $\mathbb{Q}$*-linear map;*

(iv) *for* $\alpha\in\mathbb{Q}$*, we have*

$$\mathrm{N}_K(\alpha)=\alpha^{[K:\mathbb{Q}]}\,,\qquad \mathrm{Tr}_K(\alpha)=[K:\mathbb{Q}]\alpha\,.$$

Proof. (i) The statement is easy to see on the level of *rings*, i.e., by considering the multiplication-by-α map $\hat\alpha:\mathbb{Q}[\theta]\to\mathbb{Q}[\theta]$, $\lambda\mapsto\alpha\lambda$ (instead of α itself). This map is bijective if and only $\alpha\neq 0$. ⟦Note that in $\mathbb{Q}[\theta]$ there are no zero divisors.⟧

(ii) Follows from the corresponding properties for the determinant:

$$\mathrm{N}_K(\alpha\beta)=\det(\widehat{\alpha\beta})=\det(\hat\alpha\hat\beta)=\det(\hat\alpha)\det(\hat\beta)=\mathrm{N}_K(\alpha)\mathrm{N}_K(\beta)\,.$$

(iii) Obvious since trace(A) equals the sum of all the *diagonal* elements of A.

(iv) The corresponding matrix is simply the diagonal matrix $\alpha\cdot Id$.

The above brings us closer to seeing $\mathbb{Q}[\theta]=\mathbb{Q}(\theta)$ for α algebraic. ⟦Note that for X an indeterminate the corresponding equality does not hold, as the field of fractions $\mathbb{Q}(X)$ is considerably larger than the polynomial ring $\mathbb{Q}[X]$—since X does not satisfy any algebraic relation over $\mathbb{Q}$, we can consider it as being "transcendental".⟧

Any $\alpha\in K=\mathbb{Q}[\sqrt{D}]$ divides its own norm $\mathrm{N}_K(\alpha)=(a+b\sqrt{D})(a-b\sqrt{D})\in\mathbb{Q}$, since we just multiply by its "conjugate" $a-b\sqrt{D}$ ⟦for $D<0$, this coincides with the "complex conjugate" for the complex numbers⟧.

In general, consider $\mathbb{Q}[\theta]$, where the minimum polynomial $p_\theta(X)$ of θ is of degree n, say; then we will see that any $\alpha=a_0+a_1\theta+\cdots+\alpha_{n-1}\theta^{n-1}$ divides its own norm $\mathrm{N}_{\mathbb{Q}[\theta]}(\alpha)=\alpha\cdot\beta\in\mathbb{Q}$, for some $\beta\in\mathbb{Q}[\theta]$, which then allows us to invert, since $1/\alpha=\beta/\mathrm{N}_{\mathbb{Q}[\theta]}(\alpha)\in\mathbb{Q}[\theta]$. In order to figure out what that β looks like (in terms of α), it is useful to consider the minimum polynomial again.

Proposition 3.45. *The minimum polynomial* $p_\alpha(X)\in\mathbb{Q}[X]$ *of an algebraic number* α *has no repeated roots.*

Proof. Note first that $\gcd(p_\alpha(X),p'_\alpha(X))=1$. ⟦We have $p'_\alpha(X)\neq 0$ and $\deg(p'_\alpha(X))<\deg(p_\alpha(X))$; so a common factor must be different from $p_\alpha(X)$ itself and cannot have positive degree, otherwise $p_\alpha(X)$ would be reducible.⟧

Therefore we can write

$$q(X)p_\alpha(X)+r(X)p'_\alpha(X)=1\,, \tag{5}$$

with some $q(X)$, $r(X)\in\mathbb{Q}[X]$. A repeated root ρ of $p_\alpha(X)$ would also be a root of $p'_\alpha(X)$ ⟦since then $p_\alpha(X)=(X-\rho)^2\cdot s(X)$ for some $s(X)\in\mathbb{C}[X]$, and so $p'_\alpha(X)=2(X-\rho)\,s(X)+(X-\rho)^2\,s'(X)$⟧. Plugging in β into (5) would give $0=1$, a contradiction.

Conclusion: $p_\alpha(X)$ cannot have a repeated root. □

Definition 3.46. *For an algebraic number* α*, the roots in* $\mathbb{C}$ *of its minimum polynomial* $p_\alpha(X)$ *(over* $\mathbb{Q}$*) are called the* **conjugates** *of* α *(over* $\mathbb{Q}$*).* ⟦*We can replace here* $\mathbb{Q}$ *by any more general fields, in particular by a number field,* K *and some algebraic number* α *over* K*.*⟧

Depending on the shape of $p_\alpha(X)$, there may be hidden symmetries among the roots—they were discovered by Galois and Abel when trying to solve the general quintic equation. Nowadays those symmetries are usually made apparent using "field homomorphisms", studied in detail in *Galois theory.*

Lemma 3.47. *A homomorphism $\varphi : K \to L$ of fields K and L is always injective.*

⟦The kernel of φ (which is in particular a *ring* homomorphism) is an ideal in K; but there are only two such ideals: $(0)_K$ and $(1)_K$, as any non-zero element in K is a unit. Now $(1)_K$ cannot be the kernel, since $\mathrm{im}(\varphi)$ is necessarily a field and thus must have at least two elements (the neutral elements for addition and multiplication, which have to be different).⟧

Proposition 3.48. *For all the conjugates α_i, $i = 1, \dots, n$, of an algebraic integer α of degree n, one has*

$$\mathbb{Q}[\alpha_i] \simeq \mathbb{Q}[\alpha]\,.$$

Idea of proof: One has $p_\alpha(X) = p_{\alpha_i}(X)$ $\forall i$, now $\mathbb{Q}[\alpha] \simeq \frac{\mathbb{Q}[X]}{(p_\alpha(X))}$ by the first isomorphism theorem for rings. . .

In the proposition, we should think of the quotient ring $\frac{\mathbb{Q}[X]}{(p_\alpha(X))}$ as being an "abstract" polynomial ring. Now we can try to view it more "concretely" by mapping (embedding) it into $\mathbb{C}$:

$$\sigma_i : \quad \frac{\mathbb{Q}[X]}{\big(p_\alpha(X)\big)} \longrightarrow \mathbb{C} \qquad (i = 1, \dots, n)$$
$$g(X) \mapsto g(\alpha_i)$$

in n *different* ways.

Examples: 1. For $n = 2$, consider $\mathbb{Q}[\lambda] := \frac{\mathbb{Q}[X]}{(X^2+1)}$, with the two embeddings

$$\sigma_1 : \quad g(\lambda) \mapsto g(i)\,,$$
$$\sigma_2 : \quad g(\lambda) \mapsto g(-i)$$

for any polynomial $g(\lambda)$.

2. For $n = 3$, consider $\mathbb{Q}[\lambda] := \frac{\mathbb{Q}[X]}{(X^3-5)}$, with the embeddings

$$\sigma_i : g(\lambda) \mapsto g(\alpha_i)$$

with $\alpha_1 = \sqrt[3]{5}$, $\alpha_2 = \sqrt[3]{5} \cdot \omega$, $\alpha_3 = \sqrt[3]{5} \cdot \omega^2$, where $\omega = \frac{-1+\sqrt{-3}}{2}$. Note that α_2 and α_3 are in $\mathbb{C} \setminus \mathbb{R}$.

Thus we obtain 3 different field homomorphisms $\mathbb{Q}[\lambda] \to \mathbb{C}$, and also *among* the $\mathbb{Q}[\alpha_i]$: $\mathbb{Q}[\alpha_i] \simeq \mathbb{Q}[\alpha_j]$ $1 \leqslant i, j \leqslant 3$.

Better even: consider $L = \mathbb{Q}[\alpha_i, \omega]$ (here we can take *any* of the three indices $i = 1, 2, 3$), which can be also written as $L = \mathbb{Q}[\alpha_1, \alpha_2, \alpha_3, \omega]$ or also as $L = \mathbb{Q}[\alpha_1, \alpha_2, \alpha_3]$. This is a field of degree 6 over $\mathbb{Q}$, and, e.g., the map sending $g(\alpha_1, \alpha_2, \alpha_3, \omega)$ to $g(\alpha_2, \alpha_3, \alpha_1, \omega)$ ⟦cyclic shift of the elements α_i⟧is an isomorphism of the field with itself.

Definition 3.49. *An isomorphism φ of a field L with itself is a* **(field) automorphism**. *If φ leaves a subfield K fixed pairwise, then φ is a K-automorphism.*

The key point of the above discussion in our context is the following: for an algebraic number α as above, the conjugates are precisely the roots of $p_\alpha(X) \in \mathbb{Q}[X]$, so over $\mathbb{C}$ we have

$$p_\alpha(X) = \prod_{i=1}^n (X - \alpha_i) = X^n - \Big(\underbrace{\sum_{i=1}^n \alpha_i}_{=\mathrm{Tr}_{\mathbb{Q}[\alpha]}(\alpha_j)} \Big) X^{n-1} \pm \ldots + (-1)^n \underbrace{\prod_{i=1}^n \alpha_i}_{=\mathrm{N}_{\mathbb{Q}[\alpha]}(\alpha_j)}\,, \qquad (\text{any } j)$$

which is invariant under permutations of the α_i, and since the coefficients are in $\mathbb{Q}$, we get

$$\frac{1}{\alpha_i} = \frac{\prod_{j \neq i} \alpha_j}{\prod_{\text{all } j} \alpha_j} \quad \in \mathbb{Q}[\alpha]\,,$$

since the denominator, being a norm, lies in $\mathbb{Q}$. Therefore we get

Corollary 3.50. *For any algebraic number α, we have*

$$\mathbb{Q}[\alpha] = \mathbb{Q}(\alpha)\,.$$

This implies that the algebraic numbers form a field, which we will denote by $\overline{\mathbb{C}}$.

3.5. **Algebraic integers.** An algebraic number is a root of a polynomial in $\mathbb{Q}[X]$, in fact, in $\mathbb{Z}[X]$. After clearing denominators, we see that for $\frac{m}{n} \in \mathbb{Q}$ $(m \in \mathbb{Z}, n \in \mathbb{N})$ we can take the polynomial $x - \frac{m}{n} \in \mathbb{Q}[X]$ or $nx - m \in \mathbb{Z}[X]$, and for $m \in \mathbb{Z}$ we can simply take $x - m \in \mathbb{Z}[X]$. The integers are thus characterized as satisfying a *monic* (linear) polynomial $\in \mathbb{Z}[X]$. In general, one defines

Definition 3.51. *An* **algebraic integer** *is the root of a* monic *polynomial in* $\mathbb{Z}[X]$.

Examples: 1. $\sqrt[m]{D}$ $(m \in \mathbb{N},\ D \in \mathbb{Z})$ is a root of $x^m - D$ and thus is an algebraic integer (note that we do not require the monic polynomial to be irreducible).

2. A surprise, maybe: $\frac{1+\sqrt{-3}}{2}$ is a root of $X^6 - 1$ (or also of the irreducible polynomial $X^2 - X + 1$), so is—despite appearances—an algebraic *integer*. More generally, for $m \equiv 1 \pmod 4$, we have that $\alpha = \frac{1+\sqrt{m}}{2}$ is an algebraic integer. Note that $\alpha^2 = \frac{m+1}{2} + \frac{\sqrt{m}}{2}$ has only denominator 2, since m is odd, and $\alpha^2 - \alpha = \frac{m-1}{4}$ lies in $\mathbb{Z}$ by our assumption on m. Thus α is a root of $X^2 - X - \frac{m-1}{4} \in \mathbb{Z}[X]$.

Our next aim is to see that sums and products of algebraic integers are again algebraic integers, i.e., the algebraic integers form a *ring*. This is not obvious (try to check directly, say, that $\sqrt[3]{5} + \frac{1+\sqrt{17}}{2} - 3i$ is an algebraic number...).

The idea is the following: in the above example, $\alpha = \frac{1+\sqrt{m}}{2}$ was found to be "okay" since α^2 still had *bounded denominator* ($\leqslant 2$). For instance, $\beta = \frac{\sqrt{m}}{2}$ would not work: β^2 has "worse" denominator, and in general β^n has denominator 2^n. Thus the denominators of these powers are unbounded as n grows, so the set of all powers of β cannot be captured by linear combinations of a *finite* set of numbers.

This idea is made more precise in the following

Theorem 3.52. *Let α be an algebraic number with minimum polynomial $p_\alpha(X) \in \mathbb{Q}[X]$. Then the following are equivalent (="TFAE")*

(i) *α is an algebraic integer,*
(ii) *$p_\alpha(X)$ is in $\mathbb{Z}[X]$,*
(iii) *$\mathbb{Z}[\alpha]$ is a finitely generated abelian group $[\![$whence $\exists n \in \mathbb{N}$ such that $\mathbb{Z}[\alpha] = \mathbb{Z} + \mathbb{Z}\alpha + \mathbb{Z}\alpha^2 + \cdots + \mathbb{Z}\alpha^{n-1}\,]\!]$,*
(iv) *there is a finitely generated abelian subgroup $G \subset \mathbb{Q}[\alpha]$, $G \neq 0$, such that*

$$\alpha G \subseteq G\,.$$

Proof. (i)⇒(ii): Let $f(X)$ be a monic polynomial in $\mathbb{Z}[X]$ of smallest degree such that $f(\alpha) = 0$ $[\![$this exists by definition of an algebraic integer$]\!]$.

Then $f(X)$ is irreducible in $\mathbb{Z}[X]$ $[\![$otherwise we can find a decomposition $f(X) = q(X) \cdot r(X)$ in $\mathbb{Z}[X]$ with $\deg(q(X)), \deg(r(X)) < \deg(f(X))$, and since $f(\alpha) = 0$ in the integral domain $\mathbb{Q}[\alpha]$, it follows that $q(\alpha)$ or $r(\alpha)$, contradicting the minimality of $\deg f(X)\,]\!]$. By the Gauss lemma, $f(X)$ is irreducible in $\mathbb{Q}[X]$ as well, which is a Euclidean domain.

Note that $f(X)$ lies in the *ideal*

$$I := \{g(X) \in \mathbb{Q}[X] \mid g(\alpha) = 0\}\,.$$

⟦Check that this is indeed an ideal!⟧ Now in a Euclidean domain any ideal is principal and generated by an element of *smallest* (non-zero) Euclidean norm ⟦we've seen this argument before⟧, which here is the degree.

Certainly $p_\alpha(X) \in I$ and it *is* of smallest degree, i.e. generates I, and so $f(X)$ must be a multiple of $p_\alpha(X)$. But both are irreducible and monic, so must coincide.

(ii)⇒(iii): Let $p_\alpha(X)$ be of degree n, i.e., $= X^n + a_{n-1}X^{n-1} + \cdots + a_0$, $a_i \in \mathbb{Z}$. Then

$$\alpha^n = -a_{n-1}\alpha^{n-1} - \cdots - a_0 \in \langle 1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\rangle_{\rm gp} \big(= \mathbb{Z} + \mathbb{Z}\alpha + \mathbb{Z}\alpha^2 + \cdots + \mathbb{Z}\alpha^{n-1}\big)\,.$$

Inductively, let $m > n$, and assume we know $\alpha^k \in \langle 1, \alpha, \ldots, \alpha^{n-1}\rangle_{\rm gp}$ for $k = 0, 1, \ldots, m-1$, then

$$\alpha^m = \alpha^{m-n} \cdot \alpha^n \in \langle \alpha^{m-n}\ \alpha^{m-n+1}, \ldots, \alpha^{m-1}\rangle_{\rm gp} \subseteq \langle 1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\rangle_{\rm gp}\,.$$

Thus *any* power of α lies in the finitely generated abelian group $\langle 1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\rangle_{\rm gp}$.

(iii)⇒(iv): Take $G = \mathbb{Z}[\alpha]$, then
$\alpha G = \alpha\mathbb{Z}[\alpha] = \langle \alpha, \alpha^2, \ldots, \alpha^n\rangle_{\rm gp} \subseteq \langle 1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\rangle_{\rm gp} = \mathbb{Z}[\alpha] = G$.

(iv)⇒(i): Let $G \subseteq \mathbb{Q}[\alpha]$ be a finitely generated abelian subgroup, generated, say, by $\gamma_1, \ldots, \gamma_r$, i.e., $G = \mathbb{Z}\gamma_1 + \cdots + \mathbb{Z}\gamma_r$ ⟦over $\mathbb{Z}$!⟧. By assumption on G, we can express

$$\alpha\gamma_i = \sum_{j=1}^r \mu_{ij}\gamma_j\,, \qquad i = 1, \ldots, r \text{ with } \mu_{ij} \in \mathbb{Z}\,.$$

We can combine this and state it in terms of matrices as

$$\alpha \begin{pmatrix} \gamma_1 \\ \vdots \\ g_r \end{pmatrix} = \underbrace{\begin{pmatrix} \mu_{11} & \cdots & \mu_{1r} \\ & \vdots & \\ \mu_{r1} & \cdots & \mu_{rr} \end{pmatrix}}_{=:M} \begin{pmatrix} \gamma_1 \\ \vdots \\ g_r \end{pmatrix}.$$

In other words, α is an eigenvalue (to the eigenvector $(\gamma_1, \ldots, \gamma_r)^t$), in particular α is a root of the characteristic polynomial of M, given by $\det(Id \cdot X - M)$ which is monic with coefficients in $\mathbb{Z}$. □

Corollary 3.53. *The algebraic integers form a ring, sometimes denoted $\overline{\mathbb{Z}}$ (in analogy with $\overline{\mathbb{Q}}$, the field of algebraic numbers).*

Definition 3.54. *For a number field K, denote*

$$\mathcal{O}_K := \{\alpha \in K \mid \alpha \textit{ is an algebraic integer}\}\,,$$

the **ring of integers in** K *or* **number ring of** K.

Note that $\mathcal{O}_K$ is indeed a ring ⟦it is equal to the intersections of the two rings $\overline{\mathbb{Z}}$ and K⟧.

As expected, the algebraic integers among the rational numbers are precisely the integers:

Proposition 3.55. $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$.

Remark 3.56. *(cf. Problem Sheet 6)*

(1) We can write

$$K = \{\frac{\alpha}{\beta} \mid \alpha, \beta \in \mathcal{O}_K\}\,,$$

in fact somewhat better

$$K = \{\frac{\alpha}{n} \mid \alpha \in \mathcal{O}_K,\, n \in \mathbb{N}\}\,.$$

(2) Let S be a subring of a field K and suppose there are $\alpha, \beta \in S \setminus \{0\}$ such that

(i) $\frac{\alpha}{\beta} \notin S$, *yet*

(ii) $\frac{\alpha}{\beta}$ *is a root of a monic polynomial in* $S[X]$.

Then S cannot be a UFD.

4. Quadratic fields and their rings of integers

Definition 4.1. *Let $d \in \mathbb{Z} \setminus \{0\}$. We call d* **squarefree** *if there is no integer $m > 1$ such that $m^2 \mid d$.*

Note: If $n \in \mathbb{Z} \setminus \{0\}$ and s is the largest integer such that $s^2 \mid n$, then $\frac{n}{s^2}$, the **squarefree part of** n is indeed squarefree. ⟦Check!⟧

Theorem 4.2. *Let K be an extension of $\mathbb{Q}$ of degree 2. Then $K = \mathbb{Q}(\sqrt{d})$ for some squarefree $d \in \mathbb{Z} \setminus \{0, 1\}$.*

Definition 4.3. *A field as in the theorem is called a* **quadratic field**. *More precisely, it is called* $\left\{\begin{array}{l}\textit{real quadratic}\\ \textit{imaginary quadratic}\end{array}\right\}$ *if* $\left\{\begin{array}{l} d > 0 \\ d < 0\end{array}\right\}$.

Proof. Choose an $\alpha \in K \setminus \mathbb{Q}$ ⟦this exists since $K = \mathbb{Q}$ would be an extension of $\mathbb{Q}$ of degree 1⟧. As a vector space, K is 2–dimensional, so $1, \alpha, \alpha^2$ are linearly independent over $\mathbb{Q}$, i.e.,

$$R\alpha^2 + S\alpha + T = 0 \qquad \text{for some } R, S, T \in \mathbb{Q}, \quad R \neq 0\,.$$

Solving the quadratic equation, we get $\alpha = A \pm \sqrt{D}$, for some $A, D \in \mathbb{Q}$, $D \neq 0$. Now "pull out" the squarefree integer part of $D = \frac{B}{C}$, where $B, C \in \mathbb{Z}$, so

$$\sqrt{D} = \sqrt{\frac{BC}{C^2}} = \sqrt{\frac{n^2 d}{C^2}} = \pm\frac{n}{c}\sqrt{d}\,,$$

where d is the squarefree part of BC. Solving $\alpha = A \pm \frac{n}{c}\sqrt{d}$ for $\sqrt{d}$ gives $\sqrt{d} = \mp(\alpha - A)\frac{C}{n} \in K$, i.e., $\mathbb{Q}(\sqrt{d}) \subseteq K$.
But both fields also have the same dimension (= 2) over $\mathbb{Q}$, so must coincide. □

In the following we want to determine its ring of integers.

Lemma 4.4. *Let $K = \mathbb{Q}(\sqrt{d})$, with $d \equiv 1(4)$ squarefree ($\neq 1$). Then*

(i) $\mathbb{Z}[\frac{1+\sqrt{d}}{2}] \subseteq \mathcal{O}_K$;

(ii) $\mathbb{Z}[\dfrac{1+\sqrt{d}}{2}] = \{\dfrac{r + s\sqrt{d}}{2} \mid r, s \in \mathbb{Z},\ r \equiv s \pmod 2\}$.

Proof. (i) has been checked before.

(ii) Put $\theta = \frac{1+\sqrt{d}}{2}$. If $\beta \in$LHS, then it can be written as $x + y\theta$ for some $x, y \in \mathbb{Z}$, i.e., as $\frac{2x+y+y\sqrt{d}}{2}$ and indeed $2x + y \equiv y \pmod 2$, as required ⟦i.e., $\beta \in$RHS⟧.

Conversely, if $\beta \in$RHS, then $\beta = \frac{r-s}{2} + s\big(\frac{1+\sqrt{d}}{2}\big) \in \mathbb{Z} + \mathbb{Z}\theta =$LHS. □

Theorem 4.5. *Let $K = \mathbb{Q}(\sqrt{d})$, d squarefree ($\neq 0, 1$). Then*

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}[\sqrt{d}] & \textit{if } d \equiv 2, 3 \pmod 4\,, \\ \mathbb{Z}[\frac{1+\sqrt{d}}{2}] & \textit{if } d \equiv 1 \pmod 4\,. \end{cases}$$

Proof. RHS $\subset \mathcal{O}_K$ is clear from our previous considerations.

Conversely, put $\alpha = \frac{a+b\sqrt{d}}{c}$, with $a, b, c \in \mathbb{Z}$, $\gcd(a,b,c) = 1$. Then

$$p_\alpha(X) = \Big(X - \frac{a+b\sqrt{d}}{c}\Big)\Big(X - \frac{a-b\sqrt{d}}{c}\Big) = X^2 - 2\,\frac{a}{c}\,X + \frac{a^2-b^2d}{c^2}\,.$$

Now a and c are coprime [[if a prime p divides $gcd(a,c)$, then p^2 divides b^2d, but d is squarefree, so necessarily $p \mid b$, a contradiction to the assumption $\gcd(a,b,c) = 1$]].

- The case $c = 1$ is okay, as then $\alpha = a + b\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$.
- The case $c = 2$ implies a, b odd, and furthermore $\frac{a^2-b^2d}{4} \in \mathbb{Z}$, i.e., $a^2 - bd^2 \equiv 0 \pmod 4$ with $a^2 \equiv b^2 \equiv 1 \pmod 4$. This entails $d \equiv 1 \pmod 4$.

Conversely, $d \equiv 1 \pmod 4$ gives for a, b odd that $\frac{a+b\sqrt{d}}{2}$ is an algebraic integer.

Conclusion: if $d \not\equiv 1 \pmod 4$, then $c = 1$ and $\mathcal{O}_K \subseteq \mathbb{Z}[\sqrt{d}]$, while if $d \equiv 1 \pmod 4$, then either $c = 1$ or $c = 2$ *and* a, b odd, so in this case we get $\mathcal{O}_K \subseteq \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$.

Hence equality must hold in both cases.

Proposition 4.6. *$\mathcal{O}_d := \mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ is a factorization domain (not necessarily a* unique *factorization domain, though), i.e., each element can be decomposed into finitely many irreducibles.*

[[Idea of proof: Use $\psi(\alpha) := |N_{\mathbb{Q}(\sqrt{d})}|$ satisfies the condition of Problem Sheet 4, Q2, hence by Q2c) is a factorization domain.]]

Lemma 4.7. *Let $\alpha \in \mathcal{O}_K$ be prime. Then*

1) *$\alpha \mid p$ for some prime p in $\mathbb{Z}$, and then p factorises in three possible ways into irreducibles:*
   (i) *p is prime also in $\mathcal{O}_K$, so $p \sim \alpha$; p is then called* **inert***;*
   (ii) *$p = \pm\alpha\overline{\alpha}$ and $\alpha \not\sim \overline{\alpha}$; p is then called* **split***;*
   (iii) *$p = \pm\alpha\overline{\alpha}$ and $\alpha \sim \overline{\alpha}$; p is then called* **ramified***.*
   Note that $\overline{\alpha}$ is also prime in $\mathcal{O}_d$.
2) *If $\mathcal{O}_d$ is a UFD, then any prime $p \in \mathbb{Z}$ has a* prime *factorization of one of the above types. Moreover,*

$$p \quad \textit{is not inert} \quad \Leftrightarrow \begin{cases} p = \pm(a^2 - b^2d) & \textit{if } d \equiv 2,3 \pmod 4\,, \\ 4p = \pm(a^2 - b^2d) & \textit{if } d \equiv 1 \pmod 4 \end{cases}$$

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

Proof. 1) $\alpha \mid N_K(\alpha) = \pm(\text{product of primes})$. Hence α divides (at least) one of these primes [[being prime itself]]; denote one of those by p. Then $N_K(\alpha) \mid N_K(p) = p^2$.

Thus either $N_K(\alpha) = \pm p^2$ and so necessarily $\alpha \sim p$ [[a *proper* factor of p would have a norm which is a proper factor of the integer $N_K(p)$]] or $N_K(\alpha) = \pm p$.

2) In a UFD, the above factorization into irreducibles is also a factorization into primes [[since then "irreducible $\Leftrightarrow$ prime"]].

Moreover, p is not inert $\Leftrightarrow p = \pm\alpha\overline{\alpha}$ and α is of the form $\alpha = a + b\sqrt{d}$ (if $d \equiv 2,3 \pmod 4$) or $\alpha = \frac{a+b\sqrt{d}}{2}$ (if $d \equiv 1 \pmod 4$), for some a, b in $\mathbb{Z}$.

Examples:

1) $d = -1$: $\mathcal{O}_d = \mathbb{Z}[\sqrt{3}]$. We have
   - $2 = (1+i)(1-i)$ and we have $1+i = i(1-i) \sim 1-i$, whence $2 \sim (1-i)^2$ is ramified in $\mathbb{Z}[i]$;
   - $3 \neq a^2 + b^2$ for a, $b \in \mathbb{Z}$, thus 3 is *inert* in $\mathbb{Z}[i]$;

- $5 = 1^2 + 2^2 = (1+2i)(1-2i)$ and $1+2i \not\sim 1-2i$ ⟦the units in $\mathbb{Z}[i]$ are $\{\pm 1, \pm i\}$⟧, thus 5 *splits* in $\mathbb{Z}[i]$.

More generally, we have seen that all primes $p \equiv 1 \pmod 4$ can be written as a sum of two integer squares, and thus are split in $\mathbb{Z}[i]$ ⟦$p = a^2 + b^2 = (a+ib)(a-ib)$⟧, and all primes $\equiv 3 \pmod 4$ cannot be written as such a sum, hence are inert ⟦reduce modulo 4⟧.

2) $d = 3$: $\mathcal{O}_3 = \mathbb{Z}[\sqrt{3}]$.

- $3 = (\sqrt{3})^2$ is ramified;
- $2 = (\sqrt{3}+1)(\sqrt{3}-1)$ is not inert. But $\frac{\sqrt{3}+1}{\sqrt{3}-1} = 2 + \sqrt{3} \in \mathcal{O}_3$, and similarly for the reciprocal expression, so $\sqrt{3}+1 \sim \sqrt{3}-1$. Hence 2 is ramified in $\mathcal{O}_3$.
- Is $5 = a^2 - 3b^2$ possible with $a, b \in \mathbb{Z}$? If so, then $5 \nmid b$ ⟦otherwise $5 \mid a$ and in fact 5^2 would divide the RHS, but not the LHS, a contradiction⟧.

So choose $c \pmod 5$ such that $bc \equiv 1 \pmod 5$. Since $a^2 \equiv 3b^2 \pmod 5$, we get $(ac)^2 \equiv 3(bc)^2 \equiv 3 \pmod 5$, a contradiction.

Hence 5 is inert in $\mathcal{O}_3$.

In general, it turns out that precisely the primes $\equiv \pm 1 \pmod{12}$ are split, and the primes $\equiv \pm 5 \pmod{12}$ are inert in $\mathcal{O}_3$.

Examples:

(i) How many solutions in integers a, b are there to

$$a^2 + 2b^2 = M\,, \qquad \text{where } M = 2^9 \cdot 11^5 \cdot 13^2 \cdot 19\,?$$

Recognize the left hand side as the "norm form" on the UFD $\mathcal{O}_2 = \mathbb{Z}[\sqrt{-2}]$: $\alpha = a + b\sqrt{-2}$ has norm $\mathrm{N}(\alpha) = a^2 + 2b^2$.

So try to find α such that $\alpha\bar{\alpha} = M$.

Possible prime factors for α must also occur in M, where M is viewed as a number in $\mathcal{O}_2$. Hence we check the prime factorizations of 2, 11, 13 and 19 in $\mathcal{O}_2$:

(a) • $2 = -(\sqrt{-2})^2$ is ramified;

(b) • $11 = (3+\sqrt{-2})(3-\sqrt{-2})$ is split (the two factors are not associate since the only units in $\mathcal{O}_2$ are ± 1);

(c) • $13 = 13$ is prime in $\mathcal{O}_2$;

(d) • $19 = (1+3\sqrt{-2})(1-3\sqrt{-2})$ is also split.

Altogether: every prime in $\mathcal{O}_2$ dividing α is associated to $\sqrt{-2}$, $3 \pm \sqrt{-2}$, 13 or $1 \pm 3\sqrt{-2}$, and α has the prime power decomposition

$$\alpha = \text{unit} \times (\sqrt{-2})^r (3+\sqrt{-2})^s (3-\sqrt{-2})^t 13^u (1+3\sqrt{-2})^v (1-3\sqrt{-2})^w\,. \qquad (6)$$

This decomposition is unique, as $\mathcal{O}_2$ is a UFD. The factor "unit" here represents ± 1. ⟦Note that for other number rings there may be more choices, e.g. for $\mathcal{O}_{-1}$ it would represent the four units i^n, $n = 0, \ldots, 3$.⟧

Now $\mathrm{N}(\alpha) = M$ precisely if

$$2^r \cdot 11^{s+t} \cdot 13^{2u} \cdot 19^{v+w} = 2^9 \cdot 11^5 \cdot 13^2 \cdot 19\,,$$

i.e., precisely if $r = 9$, $s + t = 5$, $u = 1$ and $v + w = 1$ ($r, s, t, u, w \geqslant 0$). Hence we get $1 \cdot 6 \cdot 1 \cdot 2 \cdot 2 = 24$ possibilities, where the last $\cdot 2$ comes from the number of units in $\mathcal{O}_2$.

(ii) How many of these solutions are in positive integers?

To each solution (a, b) there correspond four solutions $(\pm a, \pm b)$ in (i), where all four are different since $a = 0$ and $b = 0$ cannot occur for $a^2 + 2b^2 = M$ with M as above. Hence the solutions come in packets of four, and we get $24/4 = 6$ solutions in *positive* integers.

(iii) Note that there would be *no* solutions for $M = \cdots 13^{\text{odd}} \cdot \ldots$, since then u above would have had to be a half-integer…

We have seen above that the decomposition behaviour of a prime p in a quadratic field $\mathbb{Q}(\sqrt{d})$ depends on whether d is a square modulo p or not, and more precisely the case when d is a square mod p is further subdivided into d being 0 modulo p or not. It is convenient to introduce the following concept:

Definition 4.8. *The* **Legendre symbol** $\left(\frac{n}{p}\right)$ *of an integer n with respect to a prime p is defined as*

$$\left(\frac{n}{p}\right) = \begin{cases} 1 & \text{if } n \pmod p \text{ is a square, } p \nmid n, \\ 0 & \text{if } p|n, \\ -1 & \text{if } n \pmod p \text{ is not a square.} \end{cases}$$

An important property of the Legendre symbol is its multiplicativity:

$$\left(\frac{m}{p}\right)\left(\frac{n}{p}\right) = \left(\frac{m}{p}\right), \qquad m, n \in \mathbb{Z}\,.$$

⟦Note that $(\mathbb{Z}/p\mathbb{Z})^*$ consists of $\frac{p-1}{2}$ squares mod p and $\frac{p-1}{2}$ squares mod p, and "square·square = non-square·non-square = square" .⟧

Using this notion, we can rewrite our criterion to distinguish the three possible cases how a prime in $\mathbb{Z}$ decomposes in a quadratic field.

Theorem 4.9. *Suppose $\mathcal{O}_d$ is a UFD and p an* odd *prime integer. Then*

(i) *if* $\left(\frac{d}{p}\right) = -1$*, then p is also prime in $\mathcal{O}_d$, and we call p* **inert in** $\mathcal{O}_d$*;*

(ii) *if* $\left(\frac{d}{p}\right) = 1$*, then $p = \pm\alpha_p\overline{\alpha_p}$, $\alpha_p \not\sim \overline{\alpha_p}$, is a prime decomposition of p, and p* **splits** *in $\mathcal{O}_d$;*

(iii) *if* $\left(\frac{d}{p}\right) = 0$*, then $p = \pm\alpha_p\overline{\alpha_p}$, $\alpha_p \sim \overline{\alpha_p}$ is a prime decomposition of p, and p* **is ramified** *in $\mathcal{O}_d$.*

Proof. Claim: If p is not prime in $\mathcal{O}_d$, then d is a square mod p.

Proof of claim: Since $\mathcal{O}_d$ is a UFD, p is divisible by a prime $\alpha_p = \frac{1}{2}(r + s\sqrt{d})$ with r, s in $\mathbb{Z}$. Then, by the lemma,

$$p = \pm\alpha_p\overline{\alpha_p} = \pm\frac{1}{4}(r^2 - ds^2)\,, \qquad \text{i.e., } 4p = \pm(r^2 - ds^2)\,. \tag{$*$}$$

But $p \nmid s$ ⟦otherwise $p|r^2$ hence $p|r$ hence $p^2|(r^2 - ds^2) = \pm 4p$, p odd, contradiction⟧, hence has an inverse t mod p. Now (*) implies $r^2 \equiv ds^2 \pmod p$, hence $d = (rt)^2 \pmod p$, which proves the claim.

The logical negation of the claim gives

(i) If $\left(\frac{d}{p}\right) = -1$, i.e., d is *not* a square mod p, then p must be prime.

Converse claim: if d is a square mod p, then p is not prime in $\mathcal{O}_d$.

Proof of "converse claim": Suppose $d \equiv x^2 \pmod p$, for some $x \in \mathbb{Z}$. Then $p|(d - x^2) = (\sqrt{d} - x)(\sqrt{d} + x)$, but $\frac{\sqrt{d}\pm x}{p} \notin \mathcal{O}_d$ (as $p \neq 2$), so p is not prime in $\mathcal{O}_d$. This proves the "converse claim".

Hence for $\left(\frac{d}{p}\right) \neq -1$, (i.e., for d a square mod p) we have by the lemma $p = \pm\alpha_p\overline{\alpha_p}$.

(ii) Note that $\alpha + p_p \sim \overline{\alpha_p}$ implies $d \equiv 0 \pmod p$. ⟦Since then $p|\alpha_p\overline{\alpha_p}$, $p|\alpha_p^2$, $p|\overline{\alpha_p}^2$, and hence $p|(\alpha_p - \overline{\alpha_p})^2 = \alpha_p - 2\alpha_p\overline{\alpha_p} + \overline{\alpha_p}^2$, hence $p|d$ by the above, as we had seen that $p \nmid s$.⟧

Negation again gives $\left(\frac{d}{p}\right) = 1$, whence $\alpha_p \not\sim \overline{\alpha_p}$.

(iii) Conversely, if $d \equiv 0(p)$ ⟦i.e., $\left(\frac{d}{p}\right) = 0$, then $p|d|ds^2 = (\alpha_p - \overline{\alpha_p})^2$, and so $\alpha_p|p|(\alpha_p - \overline{\alpha_p})^2$, hence since α_p is prime also $\alpha_p|(\alpha_p - \overline{\alpha_p})$ and then also $\alpha_p|\overline{\alpha_p}$. Similarly $\alpha_p|\overline{\alpha_p}$, so $\alpha_p \sim \overline{\alpha_p}$. □

What happens to the even prime?

Theorem 4.10. *Suppose $\mathcal{O}_d$ is a UFD. Then*

(i) *if $d \equiv 5 \pmod 8$, then 2 is prime in $\mathcal{O}_d$ (and 2 is* **inert***);*

(ii) *if $d \equiv 1 \pmod 8$, then $2 = \pm\alpha_2\overline{\alpha_2}$, $\alpha_2 \not\sim \overline{\alpha_2}$, is a prime decomposition in $\mathcal{O}_d$ (and 2 is* **split***);*

(iii) *if $d \equiv 2, 3 \pmod 4$, then $2 = \pm\alpha_2\overline{\alpha_2}$, $\alpha_2 \sim \overline{\alpha_2}$, is a prime decomposition in $\mathcal{O}_d$ (and 2 is* **ramified***).*

Proof. Claim: If 2 is not prime in $\mathcal{O}_d$. then $d \equiv 1 \pmod 8$ or $d \equiv 2, 3(4)$.

⟦Proof of Claim: Since $\mathcal{O}_d$ is a UFD, 2 is divisible by a prime $\alpha_2 = \frac{1}{2}(r + s\sqrt{d})$, say, with r, $s \in \mathbb{Z}$. Then, by the lemma,

$$2 = \pm\alpha_2\overline{\alpha}_2 = \pm\frac{1}{4}(r^2 - s^2 d)\,, \quad \text{i.e.} \quad r^2 - s^2 d = \pm 8\,.$$

Case $r \equiv s \equiv 1(2)$ then implies $r^2 \equiv s^2 \equiv 1(8)$, and so $1 - d \equiv 0(8)$. Case $r \equiv s \equiv 0(2)$ implies $a = \frac{r}{2}$, $b = \frac{s}{2} \in \mathbb{Z}$ and

$$a^2 - db^2 = \pm 2\,,$$

which cannot hold for $d \equiv 1(4)$. ⟧

Therefore we get (i) by "negation":

(i) if $d \equiv 5(8)$ then 2 must be prime in $\mathcal{O}_d$.

Now for the other two cases

(ii) Suppose $d \equiv 1(8)$, then $2|\frac{d-1}{4} = \left(\frac{1-\sqrt{d}}{2}\right) \cdot \left(\frac{1+\sqrt{d}}{2}\right)$, but 2 does not divide any of the factors ⟦$\frac{1\pm\sqrt{d}}{4} \notin \mathcal{O}_d$⟧, hence 2 is not prime and so

$$2 = \pm\alpha_2\overline{\alpha}_2\,, \text{ and again } r^2 - s^2 d = \pm 8 \quad \text{for} \quad \alpha_2 = \frac{r + s\sqrt{d}}{2}\,.$$

From the proof of the Claim above, we must have $r \equiv s \equiv 1(2)$, as $d \equiv 1(8)$, hence in particular $d \equiv 1(4)$. Therefore $\alpha_2 \not\sim \overline{\alpha}_2$ ⟦otherwise $2|(\alpha_2 - \overline{\alpha}_2)^2 = s^2 d$ and $2|d$, a contradiction⟧.

(iii) Suppose $d \equiv 2$ or $3(4)$. Then $\mathcal{O}_d = \mathbb{Z}[\sqrt{d}]$.

2 is *not* prime, since $2|d(d-1) = (d - \sqrt{d})(d + \sqrt{d})$ and $\frac{d\pm\sqrt{d}}{2} \notin \mathbb{Z}[\sqrt{d}]$, hence $2 = \pm\alpha_2\overline{\alpha}_2$, where $\alpha_2 = a + b\sqrt{d}$ (a, $b \in \mathbb{Z}$).

But then $\alpha_2|2$ and, since $(\alpha_2 - \overline{\alpha}_2)^2 = 4b^2 d$, also $2|(\alpha_2 - \overline{\alpha}_2)^2$. Putting this together gives

$$\alpha_2|(\alpha_2 - \overline{\alpha}_2)^2\,,$$

but α_2 is prime, sp we also get

$$\alpha_2|(\alpha_2 - \overline{\alpha}_2)\,.$$

Hence $\alpha_2|\overline{\alpha}_2$ and similarly $\overline{\alpha}_2|\alpha_2$, so we get $\alpha_2 \sim \overline{\alpha}_2$.

Conclusion: for $d \equiv 2, 3(4)$ we have $2 = \text{unit} \cdot \alpha_2^2$. □

We can rephrase the above in terms of factorisations of ideals as follows: if $\mathcal{O}_d$ is a UFD, we get

$$\begin{cases} \left(\frac{d}{p}\right) = -1 & \Rightarrow (p) \text{ is prime} \\ \left(\frac{d}{p}\right) = 1 & \Rightarrow (p) = (\alpha_p)(\overline{\alpha}_p) \\ \left(\frac{d}{p}\right) = 0 & \Rightarrow (p) = (\alpha_p)^2. \end{cases}$$

We will see later, that we get a similar statement for *any* $\mathcal{O}_d$, except the fact that the prime ideals into which (p) factors, need not be principal: i.e., one has $(p) = \wp_1 \cdot \overline{\wp}_1$ (with two prime ideals $\wp_i$).

Again, we get a glimpse of how ideals make up for the lack of unique factorization.

4.1. **Quadratic residues*. Note:** $\left(\frac{a}{p}\right)$ only depends on the *residue class* (or coset) mod p, i.e.,

$$\left(\frac{a}{p}\right) = \left(\frac{a'}{p}\right) \quad \text{if} \quad a \equiv a'(p)\,.$$

Recall: For an odd prime p, the group $(\mathbb{Z}/p\mathbb{Z})^*$ is cyclic of order $p-1$ (in particular even). Let g be a generator of this cyclic group, then $g^2, g^4, \dots, g^{p-1} = 1$ are the non-zero squares in $\mathbb{Z}/p\mathbb{Z}$, while g^1, g^3, …, g^{p-2} (all the odd powers) are the non-squares in $\mathbb{Z}/p\mathbb{Z}$.

Example: Consider the case $p = 7$: the group $(\mathbb{Z}/7\mathbb{Z})^*$ has generator $g = \overline{3}$, since $\overline{3}^2 = \overline{2}$, $\overline{3}^3 = \overline{6}$, $\overline{3}^4 = \overline{4}$, $\overline{3}^5 = \overline{5}$, $\overline{3}^6 = \overline{1}$. (We could also take $g = \overline{5}(= \overline{3}^{-1})$.) The squares mod 7 therefore are $\overline{2}$, $\overline{4}$, $\overline{1}$, the non-squares are $\overline{3}$, $\overline{6}$, $\overline{5}$.

Proposition 4.11. *(Euler's criterion) For an odd prime p and a not divisible by p, we have*

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}\,.$$

Here is a second way (given by Gauss) to describe $\left(\frac{a}{p}\right)$: as a preparatory step, represent the cosets (mod p) as numbers in the interval $\left[-\frac{p-1}{2}, \frac{p-1}{2}\right]$, and denote "negatives" and "positives" by

$$N = \Big\{-\frac{p-1}{2}, -\frac{p-3}{2}, \dots, -1\Big\} \quad \text{and} \quad P = \Big\{1, 2, \dots, \frac{p-1}{2}\Big\}\,,$$

respectively. Then we have

Proposition 4.12. *(Gauss's criterion) For an odd prime p and a not divisible by p, we have*

$$\left(\frac{a}{p}\right) = (-1)^m\,, \qquad \text{where } m = \#(\overline{a} \cdot P \cap N)\,.$$

Example: Is 5 a quadratic residue mod 17? We have $N = \{-8, -7, \dots, -1\}$, $P = \{1, 2, \dots, 8\}$, and for $5P = \{5, 10, 15, 3, 8, 13, 1, 6\}$ the numbers 10, 15 and 13 are not in P, so $m = 3$ in this case, giving

$$\left(\frac{5}{17}\right) = (-1)^3 = -1\,.$$

We will only need to find a good expression for the *parity of m* in the above Gauss criterion.

Proposition 4.13. *Let p be an odd prime and a an odd number not divisible by p. Denoting by m the exponent in the above Gauss criterion, we have*

$$m \equiv \sum_{k=1}^{\frac{p-1}{2}} \lfloor \frac{ka}{p} \rfloor \pmod{2}\,.$$

Proof. Multiplying $\lfloor \frac{ka}{p} \rfloor + \left\{\frac{ka}{p}\right\} = \frac{ka}{p}$ by p gives

$$ka = p\lfloor \frac{ka}{p} \rfloor + \underbrace{p\left\{\frac{ka}{p}\right\}}_{=:R_k}\,.$$

Note that $0 < R_k < p$.

Now
$$\sum_{k=1}^{\frac{p-1}{2}} ka - \sum_{k=1}^{\frac{p-1}{2}} p\lfloor\frac{ka}{p}\rfloor \ = \ \sum_{k=1}^{\frac{p-1}{2}} R_k = \sum_{j=1}^{m'} \underbrace{\pi_j}_{\in P} + \sum_{k=1}^{m} \underbrace{\nu_k}_{\in N},$$
where $m + m' = \frac{p-1}{2}$. On the other hand,
$$\sum_{k=1}^{\frac{p-1}{2}} k \ = \ \sum_{j=1}^{m'} \underbrace{\pi_j}_{\in[1,\frac{p-1}{2}]} + \sum_{k=1}^{m} \underbrace{p-\nu_k}_{\in[\frac{p-1}{2},p-1]} \ = \ mp + \sum \pi_j - \sum \nu_k \,.$$
If we add the above two equations mod 2, we are left with
$$mp \equiv p \sum_{k=1}^{\frac{p-1}{2}} \lfloor\frac{ka}{p}\rfloor \,,$$
and we can cancel the odd factor p on both sides. □

Theorem 4.14. *(Quadratic Reciprocity Law) Let p and q be odd primes, $p \neq q$. Then*
$$\Big(\frac{p}{q}\Big)\Big(\frac{q}{p}\Big) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}} \,.$$

Proof. Using the Gauss criterion, we write
$$\Big(\frac{q}{p}\Big) \ = \ (-1)^m \,, \quad \text{where } \ m = \sum_{k=1}^{\frac{p-1}{2}} \lfloor\frac{kq}{p}\rfloor \,,$$
$$\Big(\frac{p}{q}\Big) \ = \ (-1)^n \,, \quad \text{where } \ n = \sum_{\ell=1}^{\frac{q-1}{2}} \lfloor\frac{kp}{q}\rfloor \,.$$
Now the crucial idea is to interpret each term of the sums as points of a lattice along a certain interval:

Altogether we get
$$\Big(\frac{p}{q}\Big)\Big(\frac{q}{p}\Big) = (-1)^{m+n} = (-1)^{\#\text{lattice points in rectangle}} = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}} \,. \qquad \square$$

Epiphany term

We have seen that it can be very hard to find solutions (in $\mathbb{Z}$ or in $\mathbb{Q}$) to diophantine equations. When we were able to solve them, it typically involved intricate divisibility properties, and in fact the interrelationship of such divisibilities. As a prominent example, Fermat's method of infinite descent comes to mind.

By extending $\mathbb{Z}$ to somewhat larger rings (i.e., number rings), we obtain a bit more "wiggle room" for refined divisibility arguments, e.g., for proving impossibility (in case there is no solution), for counting numbers of solutions (in case there are finitely many), and sometimes even parametrizing the solutions (in case there are infinitely many) and finding structure (e.g. of a group) among them.

We encountered obstacles in those larger rings: we often run into non-UFDs whose building blocks (=irreducibles) need no longer be prime. As a remedy, we saw "ideal numbers" appear, whose properties then were captured by the notion of an ideal; in the context of ideals, the building blocks (=the prime ideals) will indeed have the property of being *prime*, and the factorization into these will turn out to be essentially unique.

So far, we have made the passage to *quadratic* extensions $\mathbb{Z} \to \mathcal{O}_d$ $(= \mathbb{Z}[\sqrt{d}]$ or, if $d \equiv 1(4)$, $= \mathbb{Z}[\frac{1+\sqrt{d}}{2}])$, and under which a prime ideal $(p)_\mathbb{Z} = p\mathbb{Z}$ goes into $(p)_{\mathcal{O}_d}$ and factor in $\mathcal{O}_d$ in three possible ways: either it stays prime or it ramifies into the square $\wp^2$ of a prime ideal $\wp$ in $\mathcal{O}_d$ or it splits into a prime ideal $\wp$ and its "conjugate" $\overline{\wp}$.

Although this indicates that we have made progress in understanding how to work in $\mathcal{O}_d$, we still haven't yet established the "full arithmetic" in those rings: ideals "ignore" units, e.g. $(ux)_R = (x)_R$ for $u \in R^*$ in a ring R. Hence we need to treat them separately.

⟦Note that once prime ideals + units are understood, we are closer to this "full arithmetic", but we will still be missing an important point: a measure for the ambiguity in a non-UFD, which is reflected by a group that is concocted from ideals (or more precisely *classes of* ideals, modulo principal ideals).⟧

Our next goal is therefore to understand the units in $\mathcal{O}_d$.

5. Units in Quadratic fields

The general assumption for this section is the following: unless mentioned otherwise, let $d \in \mathbb{Z}\backslash\{0\}$, d not a square, $K = \mathbb{Q}(\sqrt{d})$. We will consider either $S = \mathbb{Z}[\sqrt{d}]$ (for any such d) or possibly $S = \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ (only in the case $d \equiv 1(4)$).

Note that we do not suppose d to be squarefree!

We recapitulate our state of knowledge about the units in S, first in the imaginary quadratic case.

Theorem 5.1. (i) $S^* = \{\alpha \in S \mid \mathrm{N}(\alpha) = \pm 1\}$.

(ii) *(a) For* $d < -1$ *get*

$$\mathbb{Z}[\sqrt{d}]^* = \{\pm 1\}\,.$$

(b) $\mathbb{Z}[\sqrt{-1}]^* = \{\pm 1, \pm i\}$.

(iii) *(a) For* $d \equiv 1 \pmod 4$, $d < -3$, *get*

$$\mathbb{Z}[\frac{1+\sqrt{d}}{2}]^* = \{\pm 1\}\,.$$

(b) $\mathbb{Z}[\frac{1+\sqrt{-3}}{2}]^* = \{\pm 1, \pm\omega, \pm\omega^2\}$, $\omega = \frac{1+\sqrt{-3}}{2}$.

Proof. Items (i), (ii) have been dealt with earlier.

(iii) If $\alpha \in \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$, then $\alpha = \frac{r+s\sqrt{d}}{2}$ with $r \equiv s \pmod 2$.

(a) We have $d \equiv 1 \pmod 4$, $d \leqslant -7$.
Furthermore, $\alpha \in \mathcal{O}_d^* \Leftrightarrow \alpha\overline{\alpha} = +1$, i.e. $r^2 + s^2|d| = 4$.
But $|d| \geqslant 7$ then implies $s = 0$, hence $r = \pm 2$, so $\alpha = \pm 1$.

(b) $\alpha \in \mathcal{O}_{-3}^* \Leftrightarrow r^2 + s^2 \cdot 3 = 4$, hence ($s = 0$ and $r = \pm 2$), i.e. $\alpha = \pm 1$, or else ($s = \pm 1$ and $r = \pm 1$), i.e. $\alpha = \frac{\pm 1 \pm \sqrt{-3}}{2}$. □

Notation: If $d > 1$ and $\alpha = a + b\sqrt{d}$, put $\widetilde{\alpha} := a - b\sqrt{d}$.
Also note that we write $\sqrt{d}$ for the *positive* root of $x^2 - d$ (this agrees with the usual conventions in analysis, say) and often think of it as embedded in $\mathbb{R}$. With this identification we can (and will) use the ordering in $\mathbb{R}$.
But note that *algebraically* we cannot favour any of the two roots (cf. Galois theory).

Main Theorem 5.1. *(the real quadratic case) Let $d > 1$. Then*

(i) *S has a* least *unit $u > 1$.*

(ii) *$S^* = \{\pm u^r \mid r \in \mathbb{Z}\} = \langle u, -1 \rangle$.*

Examples:

(i) $d = 3$: $u = 2 + \sqrt{3}$.

(ii) $d = 94$: $u = 2143295 + 221064\sqrt{94}$ (it is indeed the smallest unit > 1 in this case!).

Definition 5.2. *A unit u as in the main theorem is called the* **fundamental unit of** *S. If furthermore $S = \mathcal{O}_d$, then is it also called the* **fundamental unit of** *the field $\mathbb{Q}(\sqrt{d})$.*

Strategy of proof: units in S give better "approximations" to $\sqrt{d}$ than the average element in S; we will find a unit > 1 using a set of "positive elements with small conjugates".

Preconsideration: Given $n \in \mathbb{Z}_{>0}$, denote by m the nearest integer to $n\sqrt{d}$, such that $|m - n\sqrt{d}| < \frac{1}{2}$. Then

$$|\sqrt{d} - \frac{m}{n}| < \frac{1}{2n}, \tag{7}$$

so $\frac{m}{n}$ is the *best* approximation with denominator n.

But now take a *unit* $\alpha = a + b\sqrt{d} \in S^*$ with $a, b > 0$. ⟦One of the four units $\{\pm\alpha, \pm\widetilde{\alpha}\}$ has both coefficients > 0.⟧

Then

$$|b\sqrt{d} - a| = |\widetilde{\alpha}| = \frac{1}{|\alpha|} = \frac{1}{\alpha} < \frac{1}{b\sqrt{d}}. \qquad \text{as } \alpha = a + b\sqrt{d} > b\sqrt{d})$$

Hence

$$|\sqrt{d} - \frac{a}{b}| < \frac{1}{b^2\sqrt{d}}.$$

This is a far better (quadratic rather than linear) approximation than (7).

Now define the set of "positive elements in S with small conjugates" as

$$A = \{\alpha = a + b\sqrt{d} \mid a, b \in \mathbb{Z}_{>0} \text{ and } |\widetilde{\alpha}| < \frac{1}{b}\}.$$

⟦Note that $\approx \frac{1}{4}$ of all units lie in here.⟧

Lemma 5.3. $|A| = \infty$.

Proof. Suppose $|A|$ were finite, then we could choose $n \in \mathbb{Z}_{>0}$ such that

$$\frac{1}{n} < |\widetilde{\alpha}| \quad \forall \alpha \in A. \tag{8}$$

We prepare for applying the pigeonhole principle.

- Consider the $n+1$ multiples $r\sqrt{d}$ ($r = 0, \dots, n$) and take their fractional parts $\lambda_r := r\sqrt{d} - \lfloor r\sqrt{d} \rfloor \in [0,1)$.
- Divide $[0,1)$ into n subintervals $[\frac{i}{n}, \frac{i+1}{n})$ of length $\frac{1}{n}$.

By the pigeonhole principle, there are two of the λ_r, say λ_s and λ_t ($s < t$), in one subinterval, i.e.

$$\left| s\sqrt{d} - \lfloor s\sqrt{d} \rfloor - t\sqrt{d} + \lfloor t\sqrt{d} \rfloor \right| = |\lambda_s - \lambda_t| < \frac{1}{n}\,.$$

Put $a := \lfloor t\sqrt{d} \rfloor - \lfloor s\sqrt{d} \rfloor$ and $b := t - s$, so that $|a - b\sqrt{d}| < \frac{1}{n}$.

Furthermore, $a > 0$, $b > 0$ $[\![t > s$ and $\sqrt{d} \geqslant 1]\!]$ and also $b \leqslant n$ $[\![s, t \in \{0, \dots, n\}]\!]$. From this we deduce that $\alpha := a + b\sqrt{d}$ lies in A, since

$$|\widetilde{\alpha}| = |a - b\sqrt{d}| < \frac{1}{n} \leqslant \frac{1}{b}\,.$$

But this contradicts our assumption (8). □

We *cannot* claim that *all* elements in A are units, but at least we can bound their norm:

Lemma 5.4. *If $\alpha \in A$, then* $|\mathrm{N}(\alpha)| < 1 + 2\sqrt{d}$.

$[\![$Pf: $\alpha = a + b\sqrt{d}$ implies $\widetilde{\alpha} = a - b\sqrt{d}$ hence $\widetilde{\alpha} = (\alpha - b\sqrt{d}) - b\sqrt{d} = \alpha - 2b\sqrt{d}$ and, since $\alpha \in A$, also $|\widetilde{\alpha}| < \frac{1}{b}$. Hence $|\mathrm{N}(\alpha)| = |\alpha\widetilde{\alpha}| = \alpha \cdot |\widetilde{\alpha}| < (2b\sqrt{d} + \frac{1}{b})\frac{1}{b} \leqslant 2\sqrt{d} + 1\,.]\!]$

The idea is now to use that there must be two elements of the same norm in A, hence whose quotient is of norm ± 1. But we still need to ensure that this quotient will be an algebraic *integer* rather than just an algebraic number. For this we break up the set A into finitely many appropriately chosen subsets and form that quotient in a given such subset.

Lemma 5.5. *There are two elements $\alpha = a + b\sqrt{d}$, $\alpha' = a' + b'\sqrt{d}$ in A with $\alpha > \alpha'$ and* $|\mathrm{N}(\alpha)| = |\mathrm{N}(\alpha')| =: n$ and *such that*

$$a \equiv a' \pmod{n}\,, \qquad b \equiv b' \pmod{n}\,.$$

Proof. As foreshadowed in the above remark, we partition A into classes (r, s, $n \in \mathbb{Z}$)

$$A_{n,r,s} := \{a \in A \mid \ |\mathrm{N}(\alpha)| = n,\ a \equiv r(n),\ b \equiv s(n)\}.$$

By the previous lemma, there are only finitely many non-empty such classes, as these sets are empty except possibly for $1 \leqslant n \leqslant 1 + 2\sqrt{d}$ and $0 \leqslant r, s < n$.

By the pigeonhole principle, we obtain that at least one of the $A_{n,r,s}$ has at least two (in fact infinitely many) different elements α, α' of A. □

From this lemma we can concoct a unit by dividing two such elements.

Theorem 5.6. *There is a unit in $\mathbb{Z}[d]^*$ such that $u > 1$.*

Proof. We take $\alpha = a + b\sqrt{d}$, $\alpha' = a' + b'\sqrt{d}$ as in Lemma 5.5, with $\alpha > \alpha'$, say. Then we put $u := \frac{\alpha}{\alpha'} \in \mathbb{Q}(\sqrt{d})$.
Clearly $u > 1$ by our assumption $\alpha > \alpha'$.
Furthermore, $u \in \mathbb{Z}[\sqrt{d}]$: here we use the congruences $a \equiv a'(n)$ and $b \equiv b'(n)$, which guarantee that $\gamma := \frac{1}{n}(\alpha - \alpha') = \frac{a-a'}{n} + \frac{b-b'}{n}\sqrt{d}$ lies in $\mathbb{Z}[\sqrt{d}]$.
Hence the proof is complete after realising that

$$u = \frac{\alpha}{\alpha'} = \frac{\alpha' + n\gamma}{\alpha'} = 1 + \frac{n}{\alpha'}\gamma = 1 + (\pm\widetilde{\alpha}')\gamma \in \mathbb{Z}[\sqrt{d}]\,,$$

where the last equality stems from $n = \mathrm{N}(\alpha') = \pm\alpha'\widetilde{\alpha}'$. □

Before proving the main theorem, we give a convenient way to rephrase the "positivity condition" $a > 0$, $b > 0$ in the definition of A.

Lemma 5.7. *Let $\alpha = a + b\sqrt{d} \in \mathbb{Q}(\sqrt{d})$. Then*

$$\alpha > \sqrt{|\mathrm{N}(\alpha)|} \;\Leftrightarrow\; a > 0,\; b > 0\,.$$

Proof. Note that $a = \dfrac{\alpha + \alpha'}{2}$, $b = \dfrac{\alpha - \alpha'}{2\sqrt{d}}$.
" $\Rightarrow$ ": Suppose that $\alpha > \sqrt{|\mathrm{N}(\alpha)|}$, so in particular $\alpha > 0$.
Then $\alpha^2 = |\mathrm{N}(\alpha)| = |\alpha\alpha'| = \alpha|\alpha'| \;\Rightarrow\; \alpha > |\widetilde{\alpha}'| = \pm\widetilde{\alpha}$, hence $\alpha \pm \alpha' > 0$ and so $a > 0, b > 0$.
" $\Leftarrow$ ": Suppose that $a > 0$, $b > 0$. Then $\alpha = a + b\sqrt{d} > |a - b\sqrt{d}| = |\widetilde{\alpha}|$ and so $\alpha^2 > \alpha|\widetilde{\alpha}| = |\mathrm{N}(\alpha)|$. □

We are now ready to prove our Main Theorem 5.1.

Proof. (i) From the above, we get a unit $v > 1$ in S.
Now form

$$U_v = \{\alpha \in S^* | 1 < \alpha \leqslant v\}\,.$$

Clearly $U_v \neq \emptyset$, as $v \in U_v$.
Moreover, any $\alpha \in U_v$ satisfies $\alpha > \sqrt{|\mathrm{N}(\alpha)|}(= 1)$. But then $\alpha = \frac{a+b\sqrt{d}}{2}$ (note that S here can stand for $\mathbb{Z}[\sqrt{d}]$ and $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$) satisfies $a > 0$, $b > 0$ by the above lemma.
Furthermore, we know from $\alpha \leqslant v$ and $a, b > 0$ that $\frac{a}{2}$, $\frac{b}{2} < v$.
Hence $\#U_v \leqslant (2v)^2 < \infty$.
Therefore there exists a least element u in (the finite set) U_v, and hence also a least element > 1 in S^*.
Conclusion: this latter element is the fundamental unit in S.

(ii) Clearly $S^* \supset \{\pm u^m \mid m \in \mathbb{Z}\}$, since $u \in S^*$ and the norm is multiplicative. Now we show the other inclusion by reducing any unit x in S to one of the above form. First we can assume, up to replacing x by its negative, that $x > 0$. Next there is a (unique!) $r \in \mathbb{Z}$ such that $u^r \leqslant x < u^{r+1}$. (Explicitly, we can write $r = \lfloor \frac{\log x}{\log u} \rfloor$.)
Therefore we can write $1 \leqslant xu^{-r} < u$ and the unit xu^{-r} must be $= 1$, since u is the fundamental unit, i.e. $x = u^r$.
Conclusion: $S^* = \{\pm u^m \mid m \in \mathbb{Z}\}$. □

Examples: We will verify below the following examples:

(1) For $d = 2$, a rather obvious unit is $1 + \sqrt{2}$ (its norm is -1). Indeed, it turns out to be the *fundamental* unit in $\mathbb{Z}[\sqrt{2}]$, hence
$$\mathbb{Z}[\sqrt{2}]^* = \{\pm(1 + \sqrt{2})^m \mid m \in \mathbb{Z}\}\,.$$
(2) For $d = 5$, a unit (of infinite order) is $u_5 = 2 + \sqrt{5}$, which is a fundamental unit in $\mathbb{Z}[\sqrt{5}]$, but *not* a fundamental unit in $\mathcal{O}_5 = \mathbb{Z}[\frac{1+\sqrt{5}}{2}]$; for the latter one, we have
$$\mathbb{Z}[\frac{1+\sqrt{5}}{2}]^* = \{\pm(\frac{1+\sqrt{5}}{2})^m \mid m \in \mathbb{Z}\}\,,$$
and $u_5 = \left(\frac{1+\sqrt{5}}{2}\right)^3$.

These two examples arise very easily, once we have established the following

Theorem 5.8. *Let $d > 1$, d not a square.*

(1) *If $S = \mathbb{Z}[\sqrt{d}]$ and $a > 0$, $b > 0$ be a solution of*
$$a^2 - db^2 = \pm 1$$
with b least possible. Then $a + b\sqrt{d}$ is a fundamental unit of S.

(2) *If* $S = \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$*, and in particular* $d \equiv 1 \pmod 4$*, then we have the following cases:*

(a) *For* $S = \mathbb{Z}[\frac{1+\sqrt{5}}{2}]$*, the fundamental unit is* $\frac{1+\sqrt{5}}{2}$*.*

(b) *For* $S = \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$*, with* $d > 5$*, the fundamental unit is* $\frac{s+t\sqrt{d}}{2}$ *where* $s^2 - t^2d = \pm 4$ *with* $s, t > 0$ *and* t *least possible.*

Proof. We only prove part (ii), as part (i) is rather similar (and easier).
(a) Let $d = 5$ and $u = \frac{1+\sqrt{5}}{2}$, which is a unit such that $u > 1$.
By our previous lemma $[\![\alpha = a + b\sqrt{5} > \sqrt{|\mathrm{N}(\alpha)|} \Leftrightarrow a, b > 0]\!]$ we have, for any unit $w = \frac{s+t\sqrt{5}}{2}$ with $w > 1$ that $s, t > 0$.
But then also $s + \sqrt{5} > 1 + \sqrt{5}$ hence $w > u$.
We conclude that u is the least unit > 1, i.e., u is the fundamental unit of S.
(b) Let $d \neq 5$ and $\frac{m+n\sqrt{d}}{2} =: v$, the fundamental unit in S. By definition $v > 1$ and hence (again by the previous lemma) $m, n > 0$.
We now compare this to the unit as in the statement, i.e. $w := \frac{s+t\sqrt{d}}{2}$ with $s, t > 0$ and t least possible.
• First we need to verify that $w \in S^*$ $[\![$it is in S since $s^2 - dt^2 = \pm 4$ implies $s \equiv t(2)$, and the equality moreover implies that w is a unit$]\!]$.
• Furthermore, $w > 1$ $[\![$again, we can invoke the lemma$]\!]$.
• Clearly $m^2 - n^2d = \pm 4$ (as v is a unit), so by our choice of w we have $n \geqslant t$.
By assumption v is the fundamental unit, and so $w \geqslant v$, more precisely $w = v^r$ for some $r > 0$. To show: $r = 1$.
We now use positivity of each term in the following (binomial) expansion:

$$\frac{s + t\sqrt{d}}{2} = \Big(\frac{m + n\sqrt{d}}{2}\Big)^r = \frac{m^r + \binom{r}{1} m^{r-1} n\sqrt{d} + \dots}{2^r}$$

and compare the coefficients of $\sqrt{d}$ on both sides to get

$$\frac{t}{2} = \frac{rm^{r-1}n + \dots}{2^r} \geqslant \frac{rm^{r-1}n}{2^r} \qquad \Rightarrow \qquad 2^{r-1}t \geqslant rm^{r-1}n \geqslant rm^{r-1}t\,,$$

and so $r = 1$ (in which case we are done) or $m = 1$, implying $\pm 4 = m^2 - n^2d = 1 - n^2d$ which is only possible (still assuming d, n positive) for $n = 1$ and $d = 5$, contradicting our choice of d.
Conclusion: $r = 1$, from which we deduce $w = v$. □

Examples: Now the above examples are easily verified:

(1) For $d = 2$, the smallest possible $s, t > 0$ (i.e. $s = t = 1$) already give a unit which by the Theorem must be a fundamental unit in $\mathbb{Z}[\sqrt{2}]$.

(2) For $d = 5$, the solution $a = 2$, $b = 1$ of $a^2 - 5b^2 = -1$ has the smallest possible b and hence gives a fundamental unit for $\mathbb{Z}[\sqrt{5}]$.
The case $\mathbb{Z}[\frac{1+\sqrt{5}}{2}]$ is treated in the Theorem. Note that *both* $u = \frac{1+\sqrt{5}}{2}$ and $u^2 = \frac{3+\sqrt{5}}{2}$ have the smallest possible least coefficient for $\sqrt{5}$ which is why we had to differentiate between the cases in the proof.

(3) For $d = 11$ we find the following table:
for successive b we solve for $a^2 - 11b^2 = \pm 1$ and obtain

b	1	2	3
$11b^2 - 1$	10	43	98
$11b^2 + 1$	12	45	100

and the latter entry 100 is indeed a square (note that not both $11b^2 \pm 1$ can be squares), so the smallest b to give a solution is $b = 3$, accompanied by $a = \sqrt{100} = 10$.
Conclusion: the fundamental unit in $\mathbb{Z}[\sqrt{11}]$ is $10 + 3\sqrt{11}$.

We can now apply our new insight to solve—in fact completely—many more Diophantine equations than before, most prominently

Pell's equation (for $d > 1$ not a square): $\quad x^2 - y^2 d = \pm 1$.

Examples.

(1) For $d = 2$ we consider $S = \mathbb{Z}[\sqrt{2}]$ with fraction field $\mathbb{Q}(\sqrt{2})$, and with fundamental unit $u = 1 + \sqrt{2}$, of norm -1.
A solution of the equation

$$x^2 - 2y^2 = 1$$

corresponds to $N(x + y\sqrt{2}) = +1$, and hence to all *even* powers of u, and we can conclude that the possibilities are precisely given by the norms of $\pm u^{2n}$, for $n \in \mathbb{Z}$.

Moreover, we can reconstruct from u the coefficients x and y, since we have

$$\begin{aligned} x + y\sqrt{2} &= \pm u^{2n} , \\ x - y\sqrt{2} &= \pm \widetilde{u}^{2n} , \end{aligned}$$

from which we get x and y from u^{2n} and its conjugate via

$$x = \pm\Big(\frac{u^{2n} + \widetilde{u}^{2n}}{2}\Big) , \qquad y = \pm\Big(\frac{u^{2n} - \widetilde{u}^{2n}}{2\sqrt{2}}\Big) ,$$

so we find, using $u^2 = 3 + 2\sqrt{2}$, that

$$x = \pm\frac{(3 + 2\sqrt{2})^n + (3 - 2\sqrt{2})^n}{2} , \qquad y = \pm\frac{(3 + 2\sqrt{2})^n - (3 - 2\sqrt{2})^n}{2\sqrt{2}} .$$

(2) In a similar way, since the fundamental unit $2 + \sqrt{5}$ in $\mathbb{Z}[\sqrt{5}]$ has norm -1 we can "parametrise" the solutions to Pell's equation for $d = 5$ by invoking $u^2 = 9 + 4\sqrt{5}$ as

$$x = \pm\frac{(9 + 4\sqrt{5})^n + (9 - 4\sqrt{5})^n}{2} , \qquad y = \pm\frac{(9 + 4\sqrt{5})^n - (9 - 4\sqrt{5})^n}{2\sqrt{5}} .$$

(3) A slightly more subtle case arises when d is not squarefree.
For $d = 75$, say, the quotient field of $S = \mathbb{Z}[\sqrt{75}]$ is $\mathbb{Q}(\sqrt{75}) = \mathbb{Q}(\sqrt{3})$, but $S \subsetneq \mathbb{Z}[\sqrt{3}] = \mathcal{O}_3$.
The fundamental unit in S is of course also a unit in $\mathbb{Z}[\sqrt{3}]$ and must be a power of the fundamental unit $u = 2 + \sqrt{3}$ of the latter ring (both are positive).
In fact, the third power of u is $v := u^3 = 26 + 15\sqrt{3} = 26 + 3\sqrt{75} \in S$.
Hence the solutions of $x^2 - 75y^2 = 1$ are given by

$$x = \pm\frac{v^n + \widetilde{v}^n}{2} , \qquad y = \pm\frac{v^n - \widetilde{v}^n}{2\sqrt{75}} .$$

We can in fact combine the method with a previous one to treat even more equations.

Examples:

(1) Find the solutions $(x, y) \in \mathbb{Z}^2$ to
  (i) $x^2 - 14y^2 = 5$,
  (ii) $x^2 - 14y^2 = -5$.

In order to treat those cases, we will need to invoke prime factorisation for the right hand side. So we first need to know that $\mathbb{Z}[\sqrt{14}]$ is a *UFD*–which is indeed the case, so let us assume it for now.
Then we determine the fundamental unit which is $u = 15 + 4\sqrt{14}$, of norm $+1$.

(i) the prime factorisation of 5 in $\mathbb{Z}[\sqrt{14}]$ is given by $5 = -\beta\widetilde{\beta}$, where $\beta = 3 + \sqrt{14}$, so any α with $\mathrm{N}(\alpha) = +5$ (those correspond bijectively to the solutions of (i)) is associate to either β or $\widetilde{\beta}$ (here we use unique factorisation), i.e. $\alpha = \pm u^r\beta$ or $\alpha = \pm u^r\widetilde{\beta}$.
But since *all* units have positive norm and β has a negative norm, there cannot be any such α (of norm 5).
Conclusion: (i) has no solution (in integers).

(ii) On the other hand, we can indeed solve $\mathrm{N}(\alpha) = -5$, e.g. with $\alpha = \beta$ as above. Moreover, since the norm of all units are $+1$, we get $\mathrm{N}(\pm u^r\beta) = -5$ for any $r \in \mathbb{Z}$; similarly for $\widetilde{\beta}$. So the general solution of (ii) is given by using a similar "trick" as above to express the coefficients in terms of $u^m\beta$ and its conjugate via

$$x = \pm\frac{u^m\beta + \widetilde{u}^m\widetilde{\beta}}{2}, \qquad y = \pm\frac{u^m\beta - \widetilde{u}^m\widetilde{\beta}}{2\sqrt{14}}, \qquad m \in \mathbb{Z},$$

so e.g. $x = \pm\frac{1}{2}\big((15+4\sqrt{14})^m(3+\sqrt{14}) + (15-4\sqrt{14})^m(3-\sqrt{14})\big)$, and a similar expression for y.

(2) Find all integer solutions of

$$x^2 - 126y^2 = -5\,.$$

Now $\mathbb{Z}[\sqrt{126}]$ is not a UFD, but the slightly larger ring $\mathbb{Z}[\sqrt{14}]$ is, as we have used above: the non-squarefree number 126 satisfies $126 = 3^2 \cdot 14$.
So we rewrite the equation as

$$x^2 - 14(3y)^2 = -5\,, \tag{9}$$

and we can reduce the problem to the previous one (i.e. to solutions (a, b) of $a^2 - 14b^2 = -5$), with the extra condition that $3 \mid b$.
We can rephrase the latter: any such solution (a, b) corresponds to an $\alpha = a + b\sqrt{14}$ such that $\alpha \equiv a \pmod{3\mathbb{Z}[\sqrt{14}]}$.
So we work "modulo 3", keeping in mind that this means we can add any $3x' + 3y'\sqrt{14}$ with x', $y' \in \mathbb{Z}$.
In particular, we get, with $u = 15 + 4\sqrt{14}$, as determined above,

$$\begin{aligned} u^{\pm 1} &\equiv 15 \pm 4\sqrt{14} \equiv \pm\sqrt{14} \pmod 3\,, \\ u^{\pm m} &\equiv (\pm\sqrt{14})^m \pmod 3\,. \end{aligned}$$

Moreover, we have

$$\beta = 3 + \sqrt{14} \equiv \sqrt{14} \pmod 3\,.$$

The upshot now is that we get a solution (a, b) of (9) precisely if $\alpha = a + b\sqrt{14}$ is congruent to an *integer* modulo $3\mathbb{Z}[\sqrt{14}]$. Using the above, we find

$$\alpha = \pm u^m\beta \equiv \sqrt{14}^{\,m+1} \pmod 3\,,$$

which is an integer exactly if m is odd.
Conclusion: the set of solutions of (9) is given by

$$x = \pm\frac{u^{2k-1}\beta + \widetilde{u}^{2k-1}\widetilde{\beta}}{2}, \qquad y = \pm\frac{u^{2k-1}\beta - \widetilde{u}^{2k-1}\widetilde{\beta}}{3\cdot 2\sqrt{14}}\,.$$

As an example, take $k = 1$ and compute
$x = \frac{1}{2}\big((15+4\sqrt{14})\cdot(3+\sqrt{14})+(15-4\sqrt{14})\cdot(3-\sqrt{14})\big) = 101$,
$y = \frac{1}{6\sqrt{14}}\big((15+4\sqrt{14})\cdot(3+\sqrt{14})-(15-4\sqrt{14})\cdot(3-\sqrt{14})\big) = 9$,
for which we verify

$$101^2 - 126\cdot 9^2 = -5\,.$$

Similarly, $k = 2$ gives
$x = \frac{1}{2}\big((15+4\sqrt{14})^3\cdot(3+\sqrt{14})+(15-4\sqrt{14})^3\cdot(3-\sqrt{14})\big) = 90709$,
$y = \frac{1}{6\sqrt{14}}\big((15+4\sqrt{14})^3\cdot(3+\sqrt{14})-(15-4\sqrt{14})^3\cdot(3-\sqrt{14})\big) = 8081$,
and indeed

$$90709^2 - 126\cdot 8081^2 = -5\,.$$

In contrast to the above examples, we get only finitely many examples for equations of the form

$$x^2 + dy^2 = M\,, \qquad d > 0 \ \text{ a non-square}\,,$$

which can be viewed as a "norm equation" in the imaginary quadratic field $\mathbb{Q}(\sqrt{d})$ where we only have *finitely many* units. (Another way to quickly see finiteness is by realising that both terms on the left are positive, so both a and b are bounded by $\sqrt{|M|}$.)

In such a situation, we can actually often give the number of solutions (and also parametrise them), as in the Example following Lemma 4.7 treated above in the text (but not yet in the course).

Examples:

(i) How many solutions in integers a, b are there to

$$a^2 + 2b^2 = M\,, \qquad \text{where } M = 2^9\cdot 11^5\cdot 13^2\cdot 19\,? \tag{10}$$

Recognize the left hand side as the "norm form" on the UFD $\mathcal{O}_2 = \mathbb{Z}[\sqrt{-2}]$: $\alpha = a + b\sqrt{-2}$ has norm $\mathrm{N}(\alpha) = a^2 + 2b^2$.

So try to find α such that $\alpha\widetilde{\alpha} = M$.

Possible prime factors for α must also occur in M, where M is viewed as a number in $\mathcal{O}_2$. Hence we check the prime factorizations of 2, 11, 13 and 19 in $\mathcal{O}_2$:

- $2 = -(\sqrt{-2})^2$ is ramified;
- $11 = (3+\sqrt{-2})(3-\sqrt{-2})$ is split (the two factors are not associate since the only units in $\mathcal{O}_2$ are ±1);
- $13 = 13$ is prime in $\mathcal{O}_2$;
- $19 = (1+3\sqrt{-2})(1-3\sqrt{-2})$ is also split.

Altogether: every prime in $\mathcal{O}_2$ dividing α is associated to $\sqrt{-2}$, $3\pm\sqrt{-2}$, 13 or $1\pm3\sqrt{-2}$, and so α has the prime power decomposition

$$\alpha = \text{unit}\times(\sqrt{-2})^r(3+\sqrt{-2})^s(3-\sqrt{-2})^t 13^u(1+3\sqrt{-2})^v(1-3\sqrt{-2})^w\,. \tag{11}$$

This decomposition is unique, as $\mathcal{O}_2$ is a UFD. The factor "unit" here represents ±1. ⟦Note that for other number rings there may be more choices, e.g. for $\mathcal{O}_{-1}$ "unit" would represent the four units i^n, $n = 0,\ldots,3$, and for real-quadratic number rings it would represent infinitely many.⟧

Now $\mathrm{N}(\alpha) = M$ precisely if

$$2^r\cdot 11^{s+t}\cdot 13^{2u}\cdot 19^{v+w} = 2^9\cdot 11^5\cdot 13^2\cdot 19\,,$$

i.e., precisely if $r = 9$, $s+t = 5$, $u = 1$ and $v+w = 1$ ($r,s,t,u,w \geqslant 0$). Hence we get $1\cdot6\cdot1\cdot2\cdot2 = 24$ possibilities, where the last $\cdot2$ comes from the number of units in $\mathcal{O}_2$.

(ii) How many of these solutions are in positive integers?

To each solution (a, b) there correspond four solutions $(\pm a, \pm b)$ in (i), which are different since $a = 0$ or $b = 0$ cannot occur for $a^2 + 2b^2 = M$ with M as above (some exponents in (10) are odd). Hence the solutions come in packets of four, and we get $24/4 = 6$ solutions in *positive* integers.

(iii) Note that there would be *no* solutions for $M = \cdots \cdot 13^{\text{odd}} \cdot \ldots$, since then u above would have to be a half-integer...

Challenge: Show that there are infinitely many integer solutions (x, y) to

$$2x(x+1) = y(y+1)\,.$$

6. Fractional ideals

As foreshadowed in the motivation, we want to understand the "ambiguities" of decomposition into irreducibles in number fields. One part of this is to understand the units which we have studied up to now—at least for quadratic fields. A second kind of ambiguity can be captured using ideals for which we will encounter a unique factorisation *into prime ideals.*

In order to formulate the latter we enlarge the set of non-zero ideals in a number ring to a set with group structure (where the group operation is given via multiplication of ideals). We already know how to multiply ideals, but we don't know yet how to take inverses—this is our next goal.

For $K = \mathbb{Q}(\sqrt{d})$ and a rational prime p, we recall the following situation: in about "half" the cases the prime splits, i.e. the principal ideal $(p)_{\mathcal{O}_d})$ decomposes into a product of two prime ideals, $(p)_{\mathcal{O}_d} = \mathfrak{p}\widetilde{\mathfrak{p}}$, where $\mathfrak{p} \not\sim \widetilde{\mathfrak{p}}$.

Now suppose we define

$$\mathfrak{q} := \frac{1}{p}\widetilde{\mathfrak{p}} = \frac{1}{p}\{\alpha \mid \alpha \in \widetilde{\mathfrak{p}}\} = \{\frac{\alpha}{p} \mid \alpha \in \widetilde{\mathfrak{p}}\}\,,$$

then ideal multiplication would give

$$\begin{aligned}\mathfrak{p}\mathfrak{q} &= \{\sum_{\text{finite}} a_i b_i \mid a_i \in \mathfrak{p},\ b_i \in \mathfrak{q}\}\\ &= \{\frac{1}{p}\sum_{\text{finite}} a_i b_i \mid a_i \in \mathfrak{p},\ b_i \in \widetilde{\mathfrak{p}}\}\,.\end{aligned}$$

In particular, we would have $1 \in \mathfrak{p}\mathfrak{q}$ $[\![$as $p \in \mathfrak{p}\widetilde{\mathfrak{p}} = (p)]\!]$, and so $\mathcal{O}_d = (1)_{\mathcal{O}_d} \subset \mathfrak{p}\mathfrak{q}$.

What is more, since $\mathfrak{p}\widetilde{\mathfrak{p}} = (p)$ consists of all the multiples of p (in $\mathcal{O}_d$), we find $\mathfrak{p}\mathfrak{q} \subset \mathcal{O}_d = (1)_{\mathcal{O}_d}$.

Combining the two statements above, we get $\mathfrak{p}\mathfrak{q} = (1)_{\mathcal{O}_d}$.

Therefore $\mathfrak{q}$ plays the role of an inverse of $\mathfrak{p}$ with respect to ideal multiplication, and this motivates to define inverses of ideals as follows.

General assumption for this chapter: Let K be a number field, with ring of integers $R = \mathcal{O}_K$.

Definition 6.1. (1) *A* **fractional ideal** *of R is a subset of K (!) of the following form:*

$$\lambda I = \{\lambda\alpha \mid \alpha \in I\} \subset K\,,$$

where I is a non-zero ideal in $\mathcal{O}_K$ and $\lambda \in K^*$.

(2) *A fractional ideal $\mathfrak{a}$ is called* **invertible** *if there is a fractional ideal $\mathfrak{b}$ of R such that*

$$\mathfrak{a} \cdot \mathfrak{b} = (1)_R\,.$$

Notation: We denote $\mathcal{J}(R) := \{\text{fractional ideals of } R\}$.

So our goal can be rephrased as showing that each fractional ideal is invertible, which in turn shows that $\mathcal{J}(R)$ is a group—in fact an abelian one.

To invoke some kind of analogue, recall that the positive integers do not form a group under multiplication (only a "semi-group", also called a "monoid"), but that the positive *rational* numbers do, and we can embed the former into the latter (by allowing denominators).

Definition 6.2. *A fractional ideal of R of the form λR, where $\lambda \in K^*$, is called a* principal fractional ideal*, denoted*

$$(\lambda)_R \subset K\,.$$

Notation: We denote $\mathcal{P}(R) := \{\text{principal fractional ideals of } R\}$.

This group $\mathcal{P}(R)$ turns out to be a sub*group* of $\mathcal{J}(R)$; note that the two sets coincide for a PID (by definition).

Now fractional ideals behave pretty much like the "usual" (also called *integral*) ideals in R.

Proposition 6.3. *Let $\mathfrak{a}$, $\mathfrak{b} \in \mathcal{J}(R)$ and α, β, γ, $\delta \in K^*$. Then*

(1) *If $\mathfrak{a} \subset R$, then it is in fact an ideal in R (i.e. an integral ideal).*
(2) $\alpha\mathfrak{a} = (\alpha)_R\mathfrak{a}$.
(3) $\mathfrak{a} + \mathfrak{b}$, $\mathfrak{a} \cap \mathfrak{b}$ *and* $\mathfrak{a} \cdot \mathfrak{b}$ *are fractional ideals.*
(4) *Associativity and distributivity still hold, e.g.* $\mathfrak{a}(\mathfrak{b} + \mathfrak{b}') = \mathfrak{a}\mathfrak{b} + \mathfrak{a}\mathfrak{b}'$.
(5) *In terms of generators of ideals, we have*
$(\alpha)_R \cdot (\beta)_R = (\alpha\beta)_R$,
$(\alpha, \beta)_R \cdot (\gamma, \delta)_R = (\alpha\gamma, \alpha\delta, \beta\gamma, \beta\delta)_R\,.$

Proof. The proof is essentially a verification—we treat only parts (i) and (iii), the remaining parts use similar ideas.
(i) For $\mathfrak{a}$ we can find a $\lambda \in K^*$ such that $\mathfrak{a} = \lambda I$, and so for $a_i \in \mathfrak{a}$ $(i = 1, 2)$ we find b_i such that $a_i = \lambda b_i$. But then also

$$a_1 + a_2 = \lambda(b_1 + b_2) \in \lambda I = \mathfrak{a}\,, \qquad r a_i = r\lambda b_i \in \lambda I = \mathfrak{a}\,.$$

(iii) We can assume $\mathfrak{a}$ and $\mathfrak{b}$ non-zero, and then write $\mathfrak{a} = \lambda I$, $\mathfrak{b} = \mu I'$ for some ideals I, $J \subset R$ and some numbers λ, $\mu \in K^*$.
Then we can find ℓ, $m \in \mathbb{N}$ such that $\ell\lambda \in R$, $m\mu \in R$.
Putting $n = \operatorname{lcm}(\ell, m)$, we find

$$n\mathfrak{a} = \underbrace{\frac{n}{\ell}}_{\in\mathbb{Z}} \cdot \underbrace{\ell\lambda}_{\in R}\underbrace{I}_{\subset R} \subset R\,,$$

and similarly $n\mathfrak{b} \subset R$, and so by (i) both are ideals in R. Thus

$$\mathfrak{a} + \mathfrak{b} = \frac{1}{n}(n\mathfrak{a} + n\mathfrak{b})\,, \qquad \mathfrak{a} \cdot \mathfrak{b} = \frac{1}{n^2}(n\mathfrak{a} \cdot n\mathfrak{b})\,,$$

and the ideals $(n\mathfrak{a} + n\mathfrak{b})$ and $(n\mathfrak{a} \cdot n\mathfrak{b})$ are indeed ideals in R (similar for $\cap$). □

It will be very useful to attach a numerical invariant for any such ideal, its "ideal norm", which e.g. enters in the definition of the inverse of an ideal. We will be mainly interested in the case for *quadratic* fields, in which case the definition can—and will—be given in a rather explicit, albeit somewhat ad hoc, form.

6.1. The ideal norm in quadratic fields.

The ideal norm will help us in defining an appropriate inverse of an ideal (inside the fractional ideals).

Assumption: For the rest of this subsection we will focus on quadratic fields only.

In any quadratic field $\mathbb{Q}(\sqrt{d})$, the conjugation map $\alpha \mapsto \widetilde{\alpha}$ is determined by sending $\sqrt{d} \mapsto -\sqrt{d}$ which determines a field automorphism. We can extend it to ideals.

Lemma 6.4. *For any subset $I \subset K$, put $\widetilde{I} := \{\widetilde{\alpha} \mid \alpha \in I\} \subset K$. Then*

1. $\widetilde{R} = R$.
2. $I \in \mathcal{J}(R) \Rightarrow \widetilde{I} \in \mathcal{J}(R)$.
3. $\widetilde{(\alpha, \beta)_R} = (\widetilde{\alpha}, \widetilde{\beta})_R$.
4. *For subgroups I, $J < K$, have $\widetilde{I \cdot J} = \widetilde{I} \cdot \widetilde{J}$.*

Definition 6.5. *For a non-zero integral ideal $I \subset \mathcal{O}_d$, we define the* **ideal norm** *$N(I)$ of I as the smallest positive integer of $I \cdot \widetilde{I}$.*

Note that any non-zero integral ideal indeed contains a non-zero element and hence also an integer $\mathrm{N}(\alpha)$, and in particular a positive one $|\mathrm{N}(\alpha)|$.

Remark 6.6. *A more general definition of the ideal norm of $(0) \neq I \subset S$ in the case of a number ring S is to take the quotient ring S/I which is* finite, *and to define the ideal norm as the order of this quotient.*

Lemma 6.7. *The least positive integer in $(m)_R$, for $m \in \mathbb{Z}_{>0}$, is m itself.*

⟦Suppose $n \in \mathbb{Z}_{>0}$ lies in $(m)_R$. Then m divides n in R, i.e. $n = \lambda m$, $\lambda \in R$; but $\lambda = \frac{m}{n}$ is also in $\mathbb{Q}$, hence in $\mathbb{Q} \cap R = \mathbb{Z}$.⟧

We now observe that any ideal in R can be generated by at most two elements (this is a fact that holds in more generality for any number ring).

Lemma 6.8. *An (integral) ideal I in R can be generated by two elements in R. More precisely, there exist γ, $\delta \in R$ such that*

$$I = \underbrace{\langle \gamma, \delta \rangle_{\text{gp}}}_{=\mathbb{Z}\gamma + \mathbb{Z}\delta} = \underbrace{(\gamma, \delta)}_{=R\gamma + R\delta} .$$

Proof. We know that $R = \mathbb{Z}[\theta]$ where $\theta = \sqrt{d}$ or $= \frac{1+\sqrt{d}}{2}$.
Therefore, as an abelian group, we get $(R, +) = (\mathbb{Z} + \mathbb{Z}\theta, +)$, which is torsion-free.
Hence any ideal I, viewed as an additive subgroup, is torsion-free of order $\leqslant 2$.
⟦Cf. proof of fundamental theorem of finitely generated abelian groups: can intersect $I \cap \mathbb{Z}$ which is a subgroup of $\mathbb{Z}$, hence of the form $m\mathbb{Z}$ for some $m \geqslant 0$, and choose an element $v = a + b\theta$ in I such that b is least positive (if this doesn't exist, then $I = m\mathbb{Z}$, a torsion-free group, and we are done); then I is isomorphic to $m\mathbb{Z} \oplus \langle v \rangle \cong \mathbb{Z} \oplus \mathbb{Z}$, using the usual criterion for direct products of groups, say.⟧
In particular, we can write $I = \langle \gamma, \delta \rangle_{\text{gp}}$ $(= \mathbb{Z}\gamma + \mathbb{Z}\delta)$ for some γ, $\delta \in R$ (not necessarily different). And so clearly $I = \mathbb{Z}\gamma + \mathbb{Z}\delta \subset R\gamma + R\delta$.
On the other hand, since γ, $\delta \in I$, we also get the other inclusion $I \supset R\gamma + R\delta$. □

An important very useful tool is the following "Hurwitz lemma".

Theorem 6.9. *For an ideal $I \subset R$, choose α, $\beta \in R$ such that $I = (\alpha, \beta)_R$. Then*

1. $N(I) = \gcd(\alpha\widetilde{\alpha}, \beta\widetilde{\beta}, \widetilde{\alpha}\beta + \alpha\widetilde{\beta})$.
2. $I \cdot \widetilde{I} = \big(N(I)\big)_R$.

Proof. If we put $m := \gcd(\alpha\widetilde{\alpha}, \beta\widetilde{\beta}, \widetilde{\alpha}\beta + \alpha\widetilde{\beta})$, then we first realise that $m \mid \widetilde{\alpha}\beta$: we get the integer polynomial

$$\big(x - \frac{\widetilde{\alpha}\beta}{m}\big)\big(x - \frac{\alpha\widetilde{\beta}}{m}\big) = x^2 - x\underbrace{\left(\frac{\widetilde{\alpha}\beta + \alpha\widetilde{\beta}}{m}\right)}_{\in\mathbb{Z}} + \underbrace{\frac{\alpha\widetilde{\alpha}}{m}}_{\in\mathbb{Z}}\underbrace{\frac{\beta\widetilde{\beta}}{m}}_{\in\mathbb{Z}} \in \mathbb{Z}[x]$$

and since it is monic, we find that $\gamma := \frac{\widetilde{\alpha}\beta}{m} \in K$ is an algebraic *integer*, hence in R. This allows to successively reduce the number of generators in $I \cdot \widetilde{I}$:

$$\begin{aligned} I \cdot \widetilde{I} &= (\alpha, \beta)_R(\widetilde{\alpha}, \widetilde{\beta})_R = (\alpha\widetilde{\alpha}, \alpha\widetilde{\beta}, \beta\widetilde{\alpha}, \beta\widetilde{\beta})_R = (\underbrace{\alpha\widetilde{\alpha}, \beta\widetilde{\beta}, \widetilde{\alpha}\beta + \alpha\widetilde{\beta}}_{\gcd = m}, \widetilde{\alpha}\beta)_R \\ &= (m, \widetilde{\alpha}\beta)_R = (m)_R\,. \end{aligned}$$

By Lemma 6.7, m is the smallest positive integer in $(m)_R$, so $m = N(I)$. □

Part (ii) of the Hurwitz lemma suggests to define the inverse of any non-zero ideal I in R as

$$I^{-1} := \frac{1}{N(I)}\widetilde{I}\,.$$

Examples: For $K = \mathbb{Q}(\sqrt{-11})$, we consider the ideal $I = (5 + 7\sqrt{-11}, 13 - 10\sqrt{-11})_R$, where $R = \mathcal{O}_{-11} = \mathbb{Z}\big[\frac{1+\sqrt{-11}}{2}\big]$.
Its norm is obtained as follows:

$$\begin{aligned} N(I) &= \gcd(\alpha\widetilde{\alpha}, \alpha\widetilde{\beta}, \widetilde{\alpha}\beta + \alpha\widetilde{\beta}) \\ &= \gcd(25 + 11 \cdot 49, 169 + 11 \cdot 100, -705 + 14\sqrt{-11} - 705 - 14\sqrt{-11}) \\ &= \gcd(564, 1269, -1410) \\ &= 141\,. \end{aligned}$$

Hence $N(I) = 141$ and $(N(I)) = (141) = I \cdot \widetilde{I}$, so $(1) = I \cdot \frac{1}{141}\widetilde{I}$, and the inverse of I is $I^{-1} = \frac{1}{141}\widetilde{I}$.

Corollary 6.10. *Any non-zero integral ideal $I \subset R$ is invertible. More precisely, its inverse is given by the fractional ideal*

$$I^{-1} = \frac{1}{N(I)} \cdot \widetilde{I}\,.$$

⟦We verify that $I \cdot \widetilde{I}^{-1} = \frac{1}{N(I)} \cdot I \cdot \widetilde{I} = \frac{1}{N(I)}\big(N(I)\big) = (1)$. ⟧

Corollary 6.11. *$\mathcal{J}(R)$ is a group.*

Proof. Take any fractional ideal $\mathfrak{a}$, necessarily of the form λI for some integral ideal I and $\lambda \in K^*$.
Define $\mathfrak{b} = \frac{1}{\lambda} \cdot \frac{1}{N(I)} \cdot \widetilde{I}$. Then we claim that $\mathfrak{b}$ is an inverse to $\mathfrak{a}$:

$$\mathfrak{a} \cdot \mathfrak{b} = \lambda I \cdot \frac{1}{\lambda}\frac{1}{N(I)}I \cdot \widetilde{I} = \frac{1}{N(I)}I \cdot \widetilde{I} = \frac{1}{N(I)}\big(N(I)\big) = (1)\,. \quad \square$$

We see furthermore that $\mathcal{P}(R)$ is a *subgroup* of $\mathcal{J}(R)$, i.e., it is also closed under taking inverses: for $\gamma \in K^*$ we have

$$(\gamma) \cdot (\gamma^{-1}) = (1) = R\,,$$

hence the principal ideal (γ^{-1}) is the inverse of (γ).

Example We give an example of the inverse of a non-principal ideal.
The ring $R = \mathbb{Z}[\sqrt{-6}]$ is not a principal ideal domain (PID), and the ideals

$$\mathfrak{p}_5 = (5, 1 + 3\sqrt{-6})_R\,, \qquad \widetilde{\mathfrak{p}}_5 = (5, 1 - 3\sqrt{-6})_R$$

are both non-principal.
We have seen earlier that $\mathfrak{p}_5\widetilde{\mathfrak{p}}_5 = (5)_R$, hence $\mathfrak{q} := \frac{1}{5}\widetilde{\mathfrak{p}}_5 = R + \frac{1-3\sqrt{-6}}{5}R$ can be viewed as an inverse of $\mathfrak{p}_5$.

We will often make use of the very convenient properties of the ideal norm that it is *multiplicative* and that it *respects inclusion*:

Corollary 6.12. *Let I, $J \subset R = \mathcal{O}_d$ be (integral) ideals. Then we have*

(i) $N(I \cdot J) = N(I)N(J)$.
(ii) *If* $I \supset J$ *then* $N(I) \mid N(J)$.
(iii) *If* $I \supset J$ *and* $N(I) = N(J)$*, then* $I = J$.

Proof. (i) We have the following equalities:

$$\begin{aligned} \big(N(I\cdot J)\big) &= IJ \cdot \widetilde{IJ} = IJ \cdot \widetilde{I}\widetilde{J} = I\widetilde{I} \cdot J\widetilde{J} \\ &= \big(N(I)\big)\big(N(J)\big) = \big(N(I)N(J)\big)\,. \end{aligned}$$

Now the least positive integer of the principal ideal on the left is $N(IJ)$, while the one on the right is $N(I)N(J)$, which proves (i).
(ii) From $I \supset J$, we also get $\widetilde{I} \supset \widetilde{J}$, and hence $\underbrace{I\widetilde{I}}_{(N(I))} \supset \underbrace{J\widetilde{J}}_{(N(J))}$, and so $N(J)$ lies in $\big(N(I)\big)$ which implies that $N(I) \mid N(J)$.
(iii) Put $m = N(I) = N(J)$, then $\frac{1}{m}I\widetilde{I} = (1)$ and so

$$J = \frac{1}{m}I\widetilde{I}J \supset \frac{1}{m}I\widetilde{J}J = I\,. \qquad \square$$

Example. Consider for $R = \mathbb{Z}[\sqrt{14}]$ the ideals $I_1 = (5)$, $I_2 = (-2 + \sqrt{14})_R$ and $I_3 = (5, -2 + \sqrt{14})$. They have norms 5^2 (smallest positive element in $(5 \cdot \widetilde{5}) = (25)$, 10 (smallest positive element in $\big((-2 + \sqrt{14}) \cdot (-2 + \sqrt{14})\big) = (-10)$ and 5 (use Hurwitz lemma), respectively.
Note that divisibilities are respected: I_1 and I_2 clearly are contained in I_3, and $N(I_3)$ divides the norm of the other two ideals.

Remark 6.13. *The ideal norm of a* principal *ideal is closely related to the norm of its generator. For $\alpha \in R \setminus \{0\}$ we get*

$$\begin{aligned} N\big((\alpha)_R\big) &= N\big((\alpha, 0)_R\big) = \gcd(\alpha\widetilde{\alpha}, \alpha \cdot 0, 0 \cdot \alpha, 0 \cdot 0) \\ &= |\alpha\widetilde{\alpha}| = |\mathrm{N}(\alpha)|\,. \end{aligned}$$

Example. In the above example, we had $I_2 = (-2 + \sqrt{14})_R$ with $\mathrm{N}(-2 + \sqrt{14}) = -10$ and the ideal norm $N\big((2 + \sqrt{10})_R\big)$ is 10.

We can extend the definition of ideal norm also to fractional ideals. By extending the multiplicativity property of the norm, we must have that the inverse I^{-1} of the integral ideal I has the inverse norm of I, a number in $\mathbb{Q}^*$.

Definition 6.14. *The norm of a fractional ideal λI with $I \subset R$ and $\lambda \in K^*$ is given by*

$$N(\lambda I) = |\mathrm{N}(\lambda)| \cdot N(I)\,.$$

$[\![$ *This is indeed well-defined due to multiplicativity of the norm.* $]\!]$

6.2. Uniqueness of factorisation into prime ideals for a general number field.

We indicate now a proof of the uniqueness of factorisation of an ideal into prime ideals in general, postponing a proof of the following three properties (which we know to be true for quadratic fields).

Assumptions (i)–(iii):

(i) maximal ideals are invertible;

(ii) there is a multiplicative map $N : \{\text{non-zero ideals}\} \to \mathbb{Z}_{>0}$ which respects strict inclusions;

(iii) every ideal is finitely generated as an abelian group.

Note that we can immediately get from (i) that any *product* of maximal ideals is invertible.

Lemma 6.15. *Let I, J be ideals such that I is invertible and $I \supset J$. Then*

(1) *$I^{-1}J$ is an (integral) ideal in R, and hence $I \mid J$.*

(2) *$I^{-1}J \supset J$, furthermore $I^{-1}J \neq J \;\Leftrightarrow\; I \neq R$.*

Proof. Put $I' := I^{-1}J$.
(1) From $I \supset J$ we get $I^{-1}J \subset I^{-1}I = R$, and so we have for the integral(!) ideal I' that $I \mid I \cdot I' = I(I^{-1}J) = J$.
(2) From $R \supset I$ we get $I' \supset I' \cdot I = J$.
Moreover, suppose equality, i.e., $(I' =)I^{-1}J = J$, then we have $\alpha J \subset J$ for any $\alpha \in I^{-1}$. But we have seen that any integral ideal in R is finitely generated, in particular for J, and then $\alpha J \subset J$ implies that α is an algebraic *integer* (cf. characterisation of an algebraic integer last term, Theorem 3.52, (iv)).
Hence we can indeed conclude that $\alpha \in R$ for any $\alpha \in I^{-1}$, i.e. $I^{-1} \subset R$ and $R = I \cdot I^{-1} \subset I \cdot R \subset I$. Hence $R = I$, which proves "⇒".
The other direction is easy (check it!). □

This allows us now to decompose all proper ideals (i.e., not zero, not the full ideal) in terms of the *maximal* ideals in R.

Theorem 6.16. *Any non-trivial ideal $J \subset R$ is a product of $\geqslant 1$ maximal ideals.*

Proof. Suppose we had a counterexample J (i.e. an ideal which is *not* a product of maximal ideals). Then we choose one with *minimal ideal norm* $N(J)$ ⟦this is possible since we know that all non-trivial ideals, i.e. ideals different from $(0)_R$ and $(1)_R = R$ itself, have *positive* norm⟧.
Note that J cannot itself be a maximal ideal.
Hence we can sandwich an ideal I between J and R, i.e. $J \subsetneq I \subsetneq R$, and by the above Corollary 6.12 we get $N(I) < N(J)$.
Moreover, we again invoke Corollary 6.12 to get $N(I^{-1}J) < N(J)$, since $I^{-1} \supset R$ implies $I^{-1}J \supset J$.
But by the minimality of (the ideal norm of) J we get that both I and $I^{-1}J$ can be written as a product of maximal ideals, and so can $J = I \cdot I^{-1}J$, a contradiction. □

Corollary 6.17. (1) *All non-zero ideals in R are invertible.*

(2) *$\mathcal{J}(R)$ is a group.*

As a way to remember how inclusion of ideals is related to divisibility, think what Julius Caesar allegedly said w.r.t. the countries conquered by the Roman Empire: "**To divide is to contain**".
For I, J ideals in R, we have indeed $I \mid J \;\Leftrightarrow\; I \supset J \ldots$

We now show that for any number ring the notions of irreducible, prime and maximal ideal are essentially interchangeable, the only exception being (0) which is prime but not maximal.

Proposition 6.18. *Let $\mathfrak{p}$ be a proper ideal in the number ring R (i.e. $\mathfrak{p} \neq (0)$, $\mathfrak{p} \neq R$). The following are equivalent:*

(i) $\mathfrak{p}$ *is irreducible;*
(ii) $\mathfrak{p}$ *is maximal ideal;*
(iii) $\mathfrak{p}$ *is prime ideal;*
(iv) $\mathfrak{p}$ *is a* prime object*, i.e.* $\mathfrak{p} \mid IJ \;\Rightarrow\; \mathfrak{p} \mid I$ *or* $\mathfrak{p} \mid J$*.*

Proof. (i) ⇒ (ii): A non-maximal ideal, by the theorem above, is a product of $\geqslant 2$ maximal ideals, hence cannot be irreducible.
(ii) ⇒ (iii): Clear. ⟦Pass to the quotient $R/\mathfrak{p}$, and use that a field is also an integral domain.⟧
(iii) ⇒ (iv): The defining property of a prime *ideal* (i.e. $\mathfrak{p} \supset IJ \;\Rightarrow\; \mathfrak{p} \supset I$ or $\mathfrak{p} \supset J$) translates, using "Caesar's maxim", into the defining property of a prime *object* (i.e. $\mathfrak{p} \mid IJ \;\Rightarrow\; \mathfrak{p} \mid I$ or $\mathfrak{p} \mid J$).
(iv) ⇒ (i): We show the contrapositive: suppose $\mathfrak{p}$ is not irreducible, hence $\mathfrak{p} = \mathfrak{a} \cdot \mathfrak{b}$ for some proper ideals $\mathfrak{a}$, $\mathfrak{b}$, hence in particular $\mathfrak{p} \mid \mathfrak{a} \cdot \mathfrak{b}$ and so, by the "prime object property", also $\mathfrak{p} \mid \mathfrak{a}$ or $\mathfrak{p} \mid \mathfrak{b}$; we can assume the first one, and then $\mathfrak{a} = \mathfrak{p} \cdot \mathfrak{q}$ for some $\mathfrak{q} \subset R$, hence $\mathfrak{p} = \mathfrak{a} \cdot \mathfrak{b} = \mathfrak{p} \cdot \mathfrak{q}\mathfrak{b}$ which necessitates $\mathfrak{q}\mathfrak{b} = R$.
But since both are integral ideals in R, the latter equality can only hold when $\mathfrak{q} = \mathfrak{b} = R$.
Conclusion: $\mathfrak{p}$ is irreducible. □

Corollary 6.19. *Any ideal in R has a unique factorisation into prime ideals (up to ordering).*

Proof. (i) Existence: clear by combining the previous theorem with the proposition.
(ii) Uniqueness: if an ideal I has two factorisations into prime ideals

$$I = \mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{q}_1 \cdots \mathfrak{q}_s; \tag{12}$$

then we need to show that $r = s$ and $\mathfrak{p}_i = \mathfrak{q}_i$ (*), up to possible renumbering.
Suppose that there is a counterexample to this uniqueness; then there is one such with smallest r, $r \leqslant s$ and $\mathfrak{p}_j$, $\mathfrak{q}_j$ violating (*).
The idea is now to construct a *smaller* counterexample.
We can assume $r > 0$ ⟦otherwise the left hand side equals R, hence also the right hand side, and $s = 0$⟧.
As $\mathfrak{p}_r$ is prime, it divides one of the factors on the right, say $\mathfrak{q}_s$. But $\mathfrak{p}_r \mid \mathfrak{q}_s$ means that both ideals have to agree (they are both maximal). Hence if we now multiply both sides by $\mathfrak{q}_s^{-1}$ (which exists by assumption), we obtain a counterexample with fewer factors violating (*).
Conclusion: there can be no counterexample to uniqueness. □

6.3. **Ideal factorisations in quadratic fields.** We now return to the case of quadratic fields to see very explicit decompositions into prime ideals.
Our **general assumption** is now $K = \mathbb{Q}(\sqrt{d})$, with d **squarefree**(!) different from 0 and 1, and $R = \mathcal{O}_d$.
Seen earlier: how prime ideals $(p)_\mathbb{Z}$ get decomposed after passage $\mathbb{Z} \to R$, where $(p)_\mathbb{Z} \mapsto (p)_R$, at least if R is a UFD (which is actually rather rare).

Clearly any ideal I of prime norm p is maximal ⟦an ideal J between I and R would have to have a norm which divides p but can neither be 1, otherwise J would equal R, nor p, otherwise J would have to be equal to I, as it contains it⟧.

We had earlier treated the example $\mathfrak{p}_5 = (5, 1 + 3\sqrt{-6})$ in $\mathbb{Z}[\sqrt{-6}]$.
Its norm is 5 and hence is maximal.

Theorem 6.20. *Let $\mathfrak{p}$ be a maximal ideal, then*

(i) *There is a* unique *prime $p \in \mathbb{Z}$ such that $\mathfrak{p} \mid (p)_R$.*
(ii) *Either $N(\mathfrak{p}) = p^2$ and $\mathfrak{p} = (p)_R$*
or $N(\mathfrak{p}) = p$ and $\mathfrak{p}\widetilde{\mathfrak{p}} = (p)_R$.

Proof. By the Hurwitz lemma, we know that $N(\mathfrak{p}) \in \mathfrak{p}$; factoring $N(\mathfrak{p})$ in $\mathbb{Z}$ gives that $\mathfrak{p}$ divides one of its prime factors by the prime property; denote this by p, then $\mathfrak{p} \supset (p)_R$. $\llbracket$Clearly there are not two primes in the maximal ideal $\mathfrak{p}$, otherwise it would contain 1 as well.$\rrbracket$
(ii) From (i) we deduce

$$N(\mathfrak{p}) \mid N((p)_R) = |\mathrm{N}_K(p)| = p^2 \,.$$

Hence we have two possibilities: 1) $N(\mathfrak{p}) = p^2$ and hence $\mathfrak{p} = (p)_R$ $\llbracket$inclusion of ideals of the same norm$\rrbracket$
or 2) $N(\mathfrak{p}) = p$ and then $\mathfrak{p}\widetilde{\mathfrak{p}} = (N(\mathfrak{p})) = (p)_R$. □

Our next task is to find *explicit* generators for each such maximal ideal, depending on d. We will distinguish the cases of p odd and p even.

Proposition 6.21. *For an odd prime p, suppose $\left(\dfrac{d}{p}\right) \neq -1$, i.e. $d \equiv x^2 \pmod{p}$ for some $x \in \mathbb{Z}$.*
Then we have $(p)_p = \mathfrak{p}_p\widetilde{\mathfrak{p}}_p$, where $\mathfrak{p}_p = (p, x - \sqrt{d})$.
Furthermore, $\mathfrak{p}_p = \widetilde{\mathfrak{p}}_p \;\Leftrightarrow\; p \mid d$.

Proof. 1) Case $p \mid d$, then also $p \mid x$ and

$$\mathfrak{p}_p = (p, x - \sqrt{d}) = (p, \sqrt{d}) = \widetilde{\mathfrak{p}}_p \,.$$

2) Case $p \nmid d$, then $p \nmid x$ and

$$N(\mathfrak{p}_p) = \gcd(p^2, x^2 - d, 2xp) = p \,.$$

In this case, we cannot have $\mathfrak{p}_p = \widetilde{\mathfrak{p}}_p$ $\llbracket$otherwise $2x = (x + \sqrt{d}) + (x - \sqrt{d}) \in \mathfrak{p}_p\rrbracket$and $1 = \gcd(2x, p) \in \mathfrak{p}_p$ which contradicts the assumption that $\mathfrak{p}$ is maximal (and hence not $= R$). □

The story for *even* primes:

Theorem 6.22. *For $p = 2$, we get*

(i) *If $d \equiv 2 \pmod{4}$, then put $\mathfrak{p}_2 = (2, \sqrt{d})$, giving $(2)_R = \mathfrak{p}_2^2$.*
(ii) *If $d \equiv 3 \pmod{4}$, then put $\mathfrak{p}_2 = (2, 1 + \sqrt{d})$, giving $(2)_R = \mathfrak{p}_2^2$.*
(iii) *If $d \equiv 1$ (mod)8, then put $\mathfrak{p}_2 = (2, \frac{1+\sqrt{d}}{2})$, giving $(2)_R = \mathfrak{p}_2\widetilde{\mathfrak{p}}_2 \neq \mathfrak{p}_2^2$.*
(iv) *If $d \equiv 5 \pmod{8}$, then $(2)_R$ is prime.*

Proof. (i) $N(\mathfrak{p}_2) = \gcd(4, -d, 2\widetilde{\sqrt{d}} + \widetilde{2}\sqrt{d}) = 2$; $\widetilde{\mathfrak{p}}_2 = \mathfrak{p}_2$.
(ii) $N(\mathfrak{p}_2) = \gcd(4, 1 - d, 2) = 2$; hence $\widetilde{\mathfrak{p}}_2 = (2, 1 - \sqrt{d}) = (2, -1 - \sqrt{d}) = \mathfrak{p}_2$.
(iii) $N(\mathfrak{p}_2) = \gcd(4, \frac{1-d}{4}, 2) = 2$; this time $\widetilde{\mathfrak{p}}_2 \neq \mathfrak{p}_2$ $\llbracket$otherwise $1 = \frac{1+\sqrt{-d}}{2} + \frac{1-\sqrt{d}}{2} \in \mathfrak{p}_2\rrbracket$.
(iv) If $(2)_R$ is non-prime, hence non-maximal $\llbracket$previous proposition$\rrbracket$, then $\mathfrak{p}_2 \supsetneq (2)_R$ with 2 generators, i.e. $\mathfrak{p}_2 = (\alpha, \beta)$, not both of which are divisible by 2.
Can assume $2 \nmid \alpha = \frac{r+s\sqrt{d}}{2}$, say, with $r, s \in \mathbb{Z}$.
By the Hurwitz lemma, $2 \mid \gcd(\alpha\widetilde{\alpha}, \alpha\widetilde{\beta}, \widetilde{\alpha}\beta, \beta\widetilde{\beta})$, hence in particular $2 \mid \alpha\widetilde{\alpha} =$

$\frac{r^2-ds^2}{4}$, i.e., $r^2 \equiv ds^2 \pmod 8$. Since d is odd, we must have $r \equiv s \pmod 2$.
We now derive a contradiction for the cases r even and r odd separately:
if r is odd, then we have $\text{odd}^2 \equiv d \cdot \text{odd}^2 \pmod 8$, which contradicts $\text{odd}^2 \equiv 1 \pmod 8$.
If r is even, then $(r')^2 \equiv (s')^2 \pmod 2$ with $r' = r/2$, $s' = s/2$, and so $r' \equiv s' \pmod 2$.
Hence $2 \mid 2 \cdot \frac{r'+s'\sqrt{d}}{2} = r' + s'\sqrt{d} = \alpha$. Contradiction. □

The story for *odd* primes:

Theorem 6.23. *For p an odd prime, we get*

(i) *If $d \equiv 0 \pmod p$, then put $\mathfrak{p}_p = (p, \sqrt{d})$, giving $(p)_R = \mathfrak{p}_p^2$.*
(ii) *If $d \equiv m^2 \not\equiv 0 \pmod p$, then put $\mathfrak{p}_p = (p, m - \sqrt{d})$, giving $(p)_R = \mathfrak{p}_p\widetilde{\mathfrak{p}}_p \neq \mathfrak{p}_p^2$.*
(iii) *If $d \not\equiv \square \pmod p$, then $(p)_R$ is prime ideal.*

Proof. (i) and (ii) from Proposition 6.21
(iii) Similar to (iv) in the Theorem above, $(p)_R$ non-prime must be contained in a maximal ideal $\mathfrak{p}_p$ with two generators α, β, not both of which are divisible by p, except this time from $p \nmid \alpha = \frac{r+s\sqrt{d}}{2}$ we get $r^2 \equiv s^2 d \pmod{4p}$.
Now s must be invertible modulo p ⟦otherwise $p \mid s$, whence $p \mid r$ and $p \mid \alpha$⟧.
Hence $d \equiv (rs^{-1})^2$, a square. □

Typical problems associated to this: to show that certain rings have non-principal ideals of prescribed norm.

Examples.
Problem 1: Show that $\mathbb{Z}[\sqrt{-26}]$ has non-principal ideals of norm 30.
Solution: Consider $K = \mathbb{Q}(\sqrt{-26})$, then $\mathcal{O}_K = \mathbb{Z}[\sqrt{-26}]$. Hence the ideal factorization theorems apply.
Claim (i): $\mathcal{O}_K$ has 4 ideals of norm $30 = 2 \cdot 3 \cdot 5$.
We check the decomposition/factorization of (2), (3), and (5).
$p = 2$: $-26 \equiv 2 \pmod 4$, so $(2)_R = \mathfrak{p}_2^2$, where $\mathfrak{p}_2 = (2, \sqrt{-26})_R$.
$p = 3$: $-26 \equiv 1^2 \pmod 3$, so $(3)_R = \mathfrak{p}_3\widetilde{\mathfrak{p}}_3$, where $\mathfrak{p}_3 = (3, 1 - \sqrt{-26})_R$.
$p = 5$: $-26 \equiv 2^2 \pmod 5$, so $(5)_R = \mathfrak{p}_5\widetilde{\mathfrak{p}}_5$, where $\mathfrak{p}_5 = (5, 2 - \sqrt{-26})_R$.
Note $N(\mathfrak{p}_p) = p$ in each case $p = 2, 3, 5$.
Suppose I is an ideal of norm 30, then

$$I \mid (30)_R = (2)_R(3)_R(5)_R = \mathfrak{p}_2^2\mathfrak{p}_3\widetilde{\mathfrak{p}}_3\mathfrak{p}_5\widetilde{\mathfrak{p}}_5 \,.$$

Now use unique factorization into prime ideals to get

$$I = \mathfrak{p}_2^r\mathfrak{p}_3^s\widetilde{\mathfrak{p}}_3^t\mathfrak{p}_5^u\widetilde{\mathfrak{p}}_5^v \,, \qquad 0 \leqslant r \leqslant 2 \,, \qquad 0 \leqslant s, t, u, v \leqslant 1 \,.$$

Taking norms, we find

$$2 \cdot 3 \cdot 5 = N(I) = 2^r 3^{s+t} 5^{u+v} \,.$$

Comparing exponents on both sides gives

$$r = 1, \; s + t = 1, \; u + v = 1, \qquad s, t, u, v \geqslant 0$$

and hence $1 \cdot 2 \cdot 2 = 4$ different ideals of norm 30. Explicitly, we find

$$\mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_5, \quad \mathfrak{p}_2\widetilde{\mathfrak{p}}_3\mathfrak{p}_5, \quad \mathfrak{p}_2\mathfrak{p}_3\widetilde{\mathfrak{p}}_5, \quad \mathfrak{p}_2\widetilde{\mathfrak{p}}_3\widetilde{\mathfrak{p}}_5 \,.$$

Claim (ii): $\mathbb{Z}[\sqrt{-26}]$ has only 2 principal ideals of norm 30.
Suppose the ideal I of norm 30 is principal, i.e. $I = (\gamma)$ for some $\gamma = a + b\sqrt{-26} \in R$, $a, b \in \mathbb{Z}$.
Then $30 = N(I) = \mathrm{N}_K(\gamma) = a^2 + 26b^2$, whence $a = \pm 2$, $b = \pm 1$. But these only give two different ideals ⟦γ and $-\gamma$ give the same ideal⟧.

Problem 2: Show that $\mathfrak{p} = (10, 5 + \sqrt{10})_R$ is *not* a principal ideal in $R = \mathbb{Z}[\sqrt{10}]$.
Solution: Consider $K = \mathbb{Q}(\sqrt{10})$, where $\mathcal{O}_K = \mathbb{Z}[\sqrt{10}] = R$.
Check that its norm satisfies $N(\mathfrak{p}) = \gcd(10^2, 25 - 10, 2 \cdot 10 \cdot 5) = 5$. ⟦Hurwitz lemma, as usual.⟧
Suppose $\mathfrak{p}$ is principal, i.e. $\mathfrak{p} = (\gamma)_R$ for some $\gamma = a + b\sqrt{10}$, $a, b \in \mathbb{Z}$. Then $|\mathrm{N}_K(\gamma)| = N\big((\gamma)_R\big) = 5$ and $a^2 - 10b^2 = 5$ (*).
Consider this modulo 8 (using that the squares mod 8 are 0, 1 and 4).
Obviously, a is odd, hence $a^2 \equiv 1 \pmod 8$, and so (*) gives

$$1 - 2b^2 \equiv 5 \pmod 8\,, \qquad \text{hence} \qquad -2b^2 \equiv 4 \pmod 8\,,$$

which is not possible.
Conclusion: $\mathfrak{p}$ cannot be principal.

Problem 3: Factorise the principal ideal $I = 57(8 - 11\sqrt{-10})_R$ of $R = \mathbb{Z}[\sqrt{-10}] = \mathcal{O}_{-10}$, and determine which of its prime factors are principal.
Solution: Put $\alpha = 8 - 11\sqrt{-10}$. Its norm in $K = \mathbb{Q}(\sqrt{-10})$ is given as

$$\mathrm{N}(\alpha) = 64 + 1210 = 2 \cdot 7^2 \cdot 13\,.$$

Decomposing the relevant principal ideals $(p)_R$ $(p = 2, 7, 13)$ gives
for $(2)_R$: $-10 \equiv 2 \pmod 4$, hence $(2) = \mathfrak{p}_2^2$, where $\mathfrak{p}_2 = (2, \sqrt{-10})$ which is of norm 2 and the only prime ideal above 2;
for $(7)_R$: $-10 \equiv 2^2 \pmod 7$, hence $(7) = \mathfrak{p}_7\widetilde{\mathfrak{p}}_7$, where $\mathfrak{p}_7 = (7, 2 - \sqrt{-10})$ which is of norm 7;
for $(13)_R$: $-10 \equiv 4^2$ (mod 1)3, hence $(13) = \mathfrak{p}_{13}\widetilde{\mathfrak{p}}_{13}$, where $\mathfrak{p}_{13} = (13, 4 - \sqrt{-10})$ which is of norm 13.

First we get a bound on the exponent of prime factors in $(\alpha)_R$:

$$(\alpha)_R \mid (\alpha\widetilde{\alpha})_R = \big(N(\alpha)\big)_R = (2)_R(7)_R^2(13)_R = \mathfrak{p}_2^2\mathfrak{p}_7^2\widetilde{\mathfrak{p}}_7^2\mathfrak{p}_{13}\widetilde{\mathfrak{p}}_{13}\,. \qquad (**)$$

We get more precise information from taking norms:

$$N\big((\alpha)_R\big) = \mathrm{N}_K(\alpha) = 2 \cdot 7^2 \cdot 13\,.$$

This can only be achieved for a factorisation in $(**)$ if

$$(\alpha)_R = \mathfrak{p}_2\mathfrak{p}_7^a\widetilde{\mathfrak{p}}_7^{2-a}\mathfrak{p}_{13}^b\mathfrak{p}_{13}^{1-b}$$

for some $0 \leqslant a \leqslant 2$ and $0 \leqslant b \leqslant 1$.

In order to check which of those prime ideals can occur in a factorisation of (α), we check for each one whether α actually *lies in* that ideal; so e.g. we find that $\alpha \in \mathfrak{p}_7$, which is generated by 7 and $2 - \sqrt{-10}$:

$$\alpha = 8 - 11\sqrt{-10} = 11(2 - \sqrt{-10}) - 14 \in \mathfrak{p}_7\,;$$

and so $\mathfrak{p}_7 \supset (\alpha)_R$, i.e. $\mathfrak{p}_7 \mid (\alpha)_R$ ⟦think Caesar⟧and so $a = \{1, 2\}$.
But we cannot have $\widetilde{\mathfrak{p}}_7$, as otherwise $(7)_R = \mathfrak{p}_7\widetilde{\mathfrak{p}}_7 \mid (\alpha)_R$ and so $7 \mid \alpha$, which is obviously not the case. We conclude that $a = 2$.
In a similar way, we find that $\alpha \in \widetilde{\mathfrak{p}}_{13}$, giving $b = 0$.
Conclusion: $(\alpha)_R = \mathfrak{p}_2\mathfrak{p}_7^2\widetilde{\mathfrak{p}}_{13}$.

Now in order to factorise $I = (57\alpha)_R$, we still need to decompose $57 = 3 \cdot 19$, and in a similar vein we find

$$-10 \equiv 2 \not\equiv \square \pmod 3 \Rightarrow 3 \text{ is inert, i.e. } (3)_R \text{ is a prime ideal}\,;$$

$$-10 \equiv 3^2 \pmod{19} \Rightarrow 19 \text{ is split, i.e. } (19)_R = \mathfrak{p}_{19}\widetilde{\mathfrak{p}}_{19}\,,$$

where $\mathfrak{p}_{19} = (19, 3 - \sqrt{-10})_R$. Altogether we find

$$I = (3)_R\mathfrak{p}_2\mathfrak{p}_7^2\widetilde{\mathfrak{p}}_{13}\mathfrak{p}_{19}\widetilde{\mathfrak{p}}_{19}\,.$$

6.4. **The ideal class group.** Let $K = \mathbb{Q}(\theta)$, $|K:\mathbb{Q}| = n$, $R = \mathcal{O}_K$ with our running assumptions $[\![$(1) a maximal ideal is invertible, (2) $\exists$ a norm function for ideals, (3) ideals are finitely generated as abelian groups.$]\!]$
Seen, as a consequence: $\mathcal{J}(R)$ is an abelian group.

Now for principal ideals one can often use very similar arguments as for elements, so in this sense we will consider those as "understood", giving rise to the subgroup $\mathcal{P}(R)$.

Definition: Two ideals $I, J \subset R$ are considered *equivalent*, denoted $I \sim J$, if $\exists \lambda \in K^*$ such that $I = (\lambda)_R J$.

Note: (1) In particular, all principal ideals are considered equivalent; in fact they are equivalent to R itself:

$$(\lambda)_R = \lambda(1)_R\,, \qquad \text{i.e. } (\lambda)_R \sim (1)_R = R\,.$$

(2) In other words, we have

$$I \sim J \Leftrightarrow I \in J\cdot\mathcal{P}(R) \Leftrightarrow I\cdot\mathcal{P}(R) = J\cdot\mathcal{P}(R)\,.$$

In particular $I \sim R \Leftrightarrow I \in \mathcal{P}(R)\,.$

Definition 6.24. (1) *The quotient group*

$$\mathcal{C}\ell(R) = \mathcal{J}(R)/\mathcal{P}(R)$$

is called the **(ideal) class group** *of R (or also of K, where $R = \mathcal{O}_K$).*

(2) *The order of $\mathcal{C}\ell(R)$ is denoted by*

$$h = h_R = h_K$$

(in particular have another variant $h_{\sqrt{d}}$ for $K = \mathbb{Q}(\sqrt{d})$, called the **class number** *of R (or also of K).*

(3) *For an ideal $I \in \mathcal{J}(R)$, denote by*

$$[I] := I\cdot\mathcal{P}(R)$$

its **class** *in $\mathcal{C}\ell(R)$.*

We have the following simple consequences of the definition:

Proposition 6.25. *Let I, $J \in \mathcal{J}(R)$. Then denote the identity in the class group by $e = e_{\mathcal{C}\ell(R)}$. We have*

(1)
$$[I] = e \Leftrightarrow I \in \mathcal{P}(R) \Leftrightarrow I \text{ is principal.}$$

(2)
$$\begin{aligned}[I] = [J] \Leftrightarrow I \sim J \quad &\Leftrightarrow \quad I = (\lambda)_R J \text{ for some } \lambda \in K^* \\ &\Leftrightarrow \quad (\alpha)_R I = (\beta)_R J \text{ for some } \alpha, \beta \in R\setminus\{0\}.\end{aligned}$$

(3)
$$\begin{aligned}[I]\cdot[J] &= I\cdot\mathcal{P}(R)\cdot J\cdot\mathcal{P}(R) \\ &= I\cdot J\cdot\mathcal{P}(R) = [I\cdot J]\end{aligned}$$

(4)
$$[I]^{-1} = [I^{-1}]\,.$$

(5)
$$[I]^m = e \;\Leftrightarrow\; I^m \text{ is principal.}$$

In order to get an *upper* bound on the class number, it will be helpful to find relations among ideal classes, e.g. by factorising *principal* ideals:

$$(\alpha) = \mathfrak{p}\mathfrak{q} \quad \Rightarrow \quad [\mathfrak{p}] = [\mathfrak{q}]^{-1}\,, \qquad \text{as } e = [(\alpha)] = [\mathfrak{p}\mathfrak{q}] = [\mathfrak{p}][\mathfrak{q}]\,.$$

Proposition: If $K = \mathbb{Q}(\sqrt{d})$, then we get

(i) $[\widetilde{I}] = [I]^{-1}$,
(ii) $I\widetilde{J}$ principal $\Leftrightarrow$ $[I] = [J]$.

Problem: For $R = \mathcal{O}_K = \mathbb{Z}[\sqrt{-29}]$ (i.e. $K = \mathbb{Q}(\sqrt{-29})$),
(a) show that R contains ideals $\mathfrak{p}_2$, $\mathfrak{p}_3$, $\mathfrak{p}_5$ of norm 2, 3, and 5, respectively, and that $[\mathfrak{p}_2]$, $[\mathfrak{p}_3]$ and $[\mathfrak{p}_5]$ have order 2, 6 and 3, respectively, in $\mathcal{C}\ell(R)$.
(b) Show that R has an ideal of norm 11 whose order in $\mathcal{C}\ell(R)$ is of order 6.

Solution: We use a previous theorem to decompose primes in R:
$-29 \equiv 3 \pmod 4$, hence $(2) = \mathfrak{p}_2^2$, where $\mathfrak{p}_2 = (2, \sqrt{-29})$;
$-29 \equiv 1^2 \pmod 4$, hence $(3) = \mathfrak{p}_3\widetilde{\mathfrak{p}}_3$, where $\mathfrak{p}_3 = (3, 1 - \sqrt{-29})$;
$-29 \equiv 1^2 \pmod 4$, hence $(5) = \mathfrak{p}_5\widetilde{\mathfrak{p}}_5$, where $\mathfrak{p}_5 = (5, 1 - \sqrt{-29})$.

Question: are there any *principal* ideals of norm 2, 3 or 5?
Answer: No. ⟦$a^2 + 29b^2$ is never 2, 3 or 5 (for $a, b \in \mathbb{Z}$).⟧
Hence we get $[\mathfrak{p}_2] \neq e$, $[\mathfrak{p}_3] \neq e$, $[\mathfrak{p}_5] \neq e$, and the same is then of course true for $[\widetilde{\mathfrak{p}}_3]$ and $[\widetilde{\mathfrak{p}}_5]$.

But we can say more: the precise orders can be obtained as follows:
- $\mathfrak{p}_2$: as $[\mathfrak{p}_2]^2 = [\mathfrak{p}_2^2] = [(2)] = e$, it is clear that the class $[\mathfrak{p}_2]$ has precisely order 2.
- $\mathfrak{p}_5$: Instead of tediously multiplying out $\mathfrak{p}_5^j$ and checking that $\mathfrak{p}_5^2$ is non-principal, while $\mathfrak{p}_5^3$ is principal, we can use relations among the ideal classes to determine their respective orders.

Main idea: decompose *principal* ideals whose norm involve only the primes we consider (here 2, 3, 5).
For example, we find that $\beta = 3 + 2\sqrt{-19}$ has norm 5^3, hence $(\beta) = \mathfrak{p}_5^a\widetilde{\mathfrak{p}}_5^{3-a}$ for some $0 \leqslant a \leqslant 3$. In the class group, this implies

$$e = [(\beta)] = [\mathfrak{p}_5]^a[\widetilde{\mathfrak{p}}_5]^{3-a}\,.$$

But we have $\beta \in \mathfrak{p}_5$ ⟦as $\beta = 5 - 2(1 - \sqrt{-29})$⟧and $\beta \notin \widetilde{\mathfrak{p}}_5$ ⟦otherwise also $\widetilde{\beta} \in \mathfrak{p}_5$ and $1 = \beta + \widetilde{\beta} - 5 \in \mathfrak{p}_5$⟧. Hence $(\beta)_R = \mathfrak{p}_5^3$ and so $[e] = [\mathfrak{p}_5]^3$, so that the order of $[\mathfrak{p}_5]$ *divides* 3; as we have checked that $\mathfrak{p}_5$ is not principal, we can conclude that $[\mathfrak{p}_5]$ has precise order 3.
- $\mathfrak{p}_3$: Consider $\gamma = 1 + \sqrt{-29}$ of norm $30 = 2 \cdot 3 \cdot 5$, to get

$$(\gamma) = \mathfrak{p}_2\mathfrak{p}_3^a\widetilde{\mathfrak{p}}_3^{1-a}\mathfrak{p}_5^b\widetilde{\mathfrak{p}}_5^{1-b}\,.$$

But we easily see that $\gamma \in \widetilde{\mathfrak{p}}_3$ and $\gamma \in \widetilde{\mathfrak{p}}_5$, so we must have

$$(\gamma) = \mathfrak{p}_2\widetilde{\mathfrak{p}}_3\widetilde{\mathfrak{p}}_5\,,$$

giving the class group relation $e = [\mathfrak{p}_2][\widetilde{\mathfrak{p}}_3][\widetilde{\mathfrak{p}}_5]$, hence $[\mathfrak{p}_3] = [\mathfrak{p}_2][\widetilde{\mathfrak{p}}_5]$, and by elementary properties in groups we get that the order of $[\mathfrak{p}_3]$ equals $\operatorname{lcm}(2, 3) = 6$.

(b) is now easy, provided we can relate some prime $\mathfrak{p}_{11}$ of norm 11 to some prime of norm 3, say $\mathfrak{p}_3$, in the class group, e.g. if $\mathfrak{p}_{11} \cdot \mathfrak{p}_3$ or $\mathfrak{p}_{11} \cdot \widetilde{\mathfrak{p}}_3$ is principal.
Its norm would be 33, and a principal ideal of this norm would have a generator $\alpha \in \mathcal{O}_K$ such that $\mathrm{N}_K(\alpha) = 33$; and indeed, $\alpha = 2 + \sqrt{-29}$ ($\in \mathfrak{p}_3$) does it. □

6.5. **Finiteness of the class group.** One of our goals is to show that h_K is finite for any number field K.

The main idea, due to Minkowski, is to list prime ideals by norm size, say, and then to show that primes of norm beyond a certain bound (the **Minkowski bound**) do not contribute anything new to the class group (i.e. there is always a relation expressing their classes in terms of a given finite set of classes), so one reduces everything to a finite set of prime ideals.

An important ingredient is the insight that the ideal norm of any ideal I and the "best possible" norm $\mathrm{N}_K(\alpha)$ of an element $\alpha \in I$ are not too far apart (we will make this statement more precise below).

In order to formulate the Minkowski bound, we will need to introduce the notion of a discriminant of a number field; we will first give a quick definition which works for most number fields that we are working with, so that we can immediately give examples, and we will give a more general definition later.

Definition 6.26. (i) *A monic polynomial $f(x) \in \mathbb{C}[x]$ of degree $n > 0$ with roots $\theta_1, \ldots, \theta_n$ has a factorization $f(x) = \prod_{j=1}^n (x - \theta_j)$, and we define the* **discriminant** *of $f(x)$ as*

$$\operatorname{discr}(f) = \prod_{1 \leqslant i < j \leqslant n} (\theta_i - \theta_j)^2 .$$

In particular, the discriminant "detects" multiple roots, i.e.

$$\operatorname{discr}(f) = 0 \;\Leftrightarrow\; f \textit{ has a multiple root}.$$

(ii) *The* **discriminant Δ_K of a number field** $K = \mathbb{Q}(\theta)$ **with** $\mathcal{O}_K = \mathbb{Z}[\theta]$ *for some $\theta \in K$ can be given as the discriminant of the minimal polynomial $p_\theta(x)$ of θ over $\mathbb{Q}$.*

Caveat: The above definition (ii) is *not always applicable*: there is not always a $\theta \in K$ such that $\mathcal{O}_K = \mathbb{Z}[\theta]$!

Example: $K = \mathbb{Q}(\theta)$ with θ a root of $x^3 - 10$ has no such description.

On the positive side, though, for quadratic fields (ii) always applies, and we get:

$$\Delta_{\mathbb{Q}(\sqrt{d})} = \begin{cases} 4d & \text{if } d \equiv 2,3 \pmod 4, \\ d & \text{if } d \equiv 1 \pmod 4. \end{cases}$$

For $d \equiv 2,3 \pmod 4$ we get the minimum polynomial

$$p_{\sqrt{d}}(x) = (x - \sqrt{d})(x + \sqrt{d}),$$

so the discriminant of $\mathbb{Q}(\sqrt{d})$ is $\operatorname{discr}(p_{\sqrt{d}}(x)) = (\sqrt{d} - (-\sqrt{d}))^2 = 4d$.

For $d \equiv 1 \pmod 4$, we find the minimal polynomial of $\theta = \frac{1+\sqrt{d}}{2}$ as

$$p_\theta(x) = (x - \frac{1+\sqrt{d}}{2})(x + \frac{1+\sqrt{d}}{2}),$$

which implies that the discriminant equals d.

Now we can formulate the Minkowski bound:

Definition 6.27. *Let $K = \mathbb{Q}(\theta)$ be a number field of degree $n = |K : \mathbb{Q}| = s + 2t$ where $s = \#\{$real roots of $p_\theta(x)\}$ and $t = \#\{$pairs of complex conjugate roots of $p_\theta(x)\}$. The* **Minkowski bound** *for K is the number*

$$B_K := \left(\frac{4}{\pi}\right)^t \frac{n!}{n^n} \sqrt{|\Delta_K|}\,.$$

We have the following extremely useful result by Minkowski stating roughly that all the ideals in R can be uniformly approximated by their respective elements:

Theorem 6.28. *For any number field K and any integral ideal $J \subset \mathcal{O}_K$, $J \neq (0)$, there is a non-zero $\alpha \in J$ such that*

$$\frac{|\mathrm{N}_K(\alpha)|}{N(J)} \leqslant B_K\,.$$

Corollary 6.29. *For each ideal class $\mathfrak{c}$ in $\mathcal{C}\ell(\mathcal{O}_K)$ we can find a representative I which is an* integral *ideal in R and satisfies*

$$N(I) \leqslant B_K\,.$$

Proof. We can choose, up to multiplying by a principal ideal, an *integral* ideal J in the *inverse* ideal class $\mathfrak{c}^{-1}$, i.e. $\mathfrak{c}^{-1} = [J]$.
By the theorem above, we can choose $\alpha \in J \setminus \{0\}$ such that $|\mathrm{N}_K(\alpha)| \leqslant N(J) \cdot B_K$.
Clearly, $J \supset (\alpha)$, hence $(\alpha) = J \cdot I$ for some integral ideal $I \subset R$.
This ideal I satisfies the requirements in the Corollary:
(i) $I \in \mathfrak{c}$; ⟦we have $[I] = [\alpha J^{-1}] = [J^{-1}] = [J]^{-1} = (\mathfrak{c}^{-1})^{-1} = \mathfrak{c}\,$.⟧
(ii) $N(I) \leqslant B_k$
⟦$N(I)N(J) = N((\alpha)_R) = \mathrm{N}_K(\alpha) \leqslant B_K \cdot N(J)$, from which we conclude $N(I) \leqslant B_K$ as $N(J) \neq 0$.⟧ □

Example: The class number of $R = \mathbb{Z}\left[\frac{1+\sqrt{19}}{2}\right]$ is 1.
In particular, we can conclude that R is a PID.
⟦Note that one can show (e.g. Stewart–Tall, Thm 4.18) that R is not a Euclidean ring.⟧

The Minkowski bound gives, using that $t = 1$, $n = 2$ and $\Delta_{\mathbb{Q}(\sqrt{-19})}$. Hence

$$B_{\mathbb{Q}(\sqrt{-19})} = \left(\frac{4}{\pi}\right)\frac{2!}{2^2}\sqrt{|-19} = \frac{2}{\pi}\sqrt{19} < 3\,.$$

Therefore we can conclude that all ideal classes have a representative of norm $\leqslant 2$.
Moreover, there are no ideals of norm 2 in this number ring: as $-19 \equiv 5 \pmod 8$, we know that 2 is inert in $\mathcal{O}_{-19}$, and so its norm is equal to 4.
So we can strengthen our result by stating that all ideal classes have a representative of norm 1; but an ideal of norm 1 is the full ring $\mathcal{O}_{-19}$, and its ideal class is the identity in $\mathcal{C}\ell(\mathcal{O}_{-19})$.
Conclusion: the ideal class group has only a single element, hence $h_{\sqrt{-19}} = 1$.

Note: For any number ring there is only a finite number of ideals of norm below a given bound.
⟦It suffices to show that there are only finitely many ideals of a given norm m, say. If I is an ideal with $N(I) = m$, then $I \mid \big(N(I)\big)$ and we can decompose into prime ideals as $(M) = \mathfrak{p}_1 \dots \mathfrak{p}_r$; hence $I = \prod_{j=1}^{r} \mathfrak{p}_j^{\varepsilon_j}$ with $\varepsilon_j \in \{0, 1\}$, of which we get $\leqslant 2^r$ possibilities. □

Corollary 6.30. *$\mathcal{C}\ell(\mathcal{O}_K)$ is finite for any number field K ⟦with our running assumptions⟧.*

Proof. We only need to combine the Minkowski bound with the proposition. □

What is more: we can use the Minkowski bound and a similar reasoning as before to get the *fulll structure* of the class group.

Example: Consider $R = \mathbb{Z}[\sqrt{-14}]$, the ring of integers in $K = \mathbb{Q}(\sqrt{-14})$; find the structure of $\mathcal{C}\ell(R)$.

Since $n = 2$, $t = 1$, and $\Delta_K = 4 \cdot -14 = -56$.
This produces the Minkowski bound

$$B_K = \left(\frac{4}{\pi}\right)^1 \frac{2!}{2^2}\sqrt{56} = \frac{2}{\pi}2\sqrt{14} = \frac{4}{\pi}\sqrt{14} < 5\,.$$

So we only need to check all ideals of norm $\leqslant 4$:
norm 2: $\mathfrak{p}_2 = (2, \sqrt{-14})_R \qquad (-14 \equiv 2 \pmod 4)$, ramified;
norm 3: $\mathfrak{p}_3 = (3, 1 - \sqrt{-14})_R \qquad (-14 \equiv 1^2 \pmod 3)$, split;
norm 4: $\mathfrak{p}_2^2 = (2)_R$.

Hence the class group, as a *set*, is $\{e, [\mathfrak{p}_2], [\mathfrak{p}_3], [\widetilde{\mathfrak{p}}_3]\}$, since we can drop $[\mathfrak{p}_2^2] = [(2)] = e$.
This already give us a good upper bound on the class number: $h_{\sqrt{-14}} \leqslant 4$.
In a second step, we try to eliminate principal ideals: but none of the ideals $\mathfrak{p}_2$, $\mathfrak{p}_3$ and $\widetilde{\mathfrak{p}}_3$ is principal ⟦we cannot solve $a^2 + 14b^2 = 2 or 3$ in integers⟧.
Moreover, we cannot solve $a^2 + 14b^2 = 6$ either, and hence neither one of the two ideals $\mathfrak{p}_2\mathfrak{p}_3$ and $\mathfrak{p}_2\widetilde{\mathfrak{p}}_3$ of norm 6 can be principal.
This implies that $[\mathfrak{p}_2]^{-1} \neq [\mathfrak{p}_3]$ and we get a lower bound on h: indeed $h \geqslant 3$.

Similarly $[\mathfrak{p}_2]^{-1} \neq [\widetilde{\mathfrak{p}}_3]$.

Finally we check that the square of $\mathfrak{p}_3$ is not principal: its norm is 9, but the only solution in integers of $a^2 + 14b^2 = 9$ is via $a = \pm 3$, $b = 0$, so as principal ideals we would need to have $\mathfrak{p}_3^2 = (3)$ which is not true as $\mathfrak{p}_3 \neq \widetilde{\mathfrak{p}}_3$.
Hence we find $[\mathfrak{p}_3] \neq [\widetilde{\mathfrak{p}}_3]$ ⟦multiply both sides by $[\mathfrak{p}_3]$⟧and hence $h \geqslant 4$.
Conclusion: $h = 4$ and $\mathcal{C}\ell(R) = \{e, [\mathfrak{p}_2], [\mathfrak{p}_3], [\widetilde{\mathfrak{p}}_3]\}$ is cyclic with generator $[\mathfrak{p}_3]$, say.