Chapter 1

The Fundamental Theorem of Arithmetic

1.1 Prime numbers

If $a, b \in \mathbb{Z}$ we say that a divides b (or is a divisor of b) and we write $a \mid b$, if

$$b = ac$$

for some $c \in \mathbb{Z}$.

Thus $-2 \mid 0$ but $0 \nmid 2$.

Definition 1.1 *The number $p \in \mathbb{N}$ is said to be* prime *if p has just 2 divisors in $\mathbb{N}$, namely 1 and itself.*

Note that our definition excludes 0 (which has an infinity of divisors in $\mathbb{N}$) and 1 (which has just one).

Writing out the prime numbers in increasing order, we obtain the *sequence of primes*

$$2, 3, 5, 7, 11, 13, 17, 19, \ldots$$

which has fascinated mathematicians since the ancient Greeks, and which is the main object of our study.

Definition 1.2 *We denote the nth prime by p_n.*

Thus $p_5 = 11,\ p_{100} = 541$.

It is convenient to introduce a kind of inverse function to p_n.

Definition 1.3 *If $x \in \mathbb{R}$ we denote by $\pi(x)$ the number of primes $\leq x$:*

$$\pi(x) = \|\{p \leq x : p \text{ prime}\}\|.$$

Thus

$$\pi(1.3) = 0, \ \pi(3.7) = 2.$$

Evidently $\pi(x)$ is monotone increasing, but discontinuous with jumps at each prime $x = p$.

Theorem 1.1 ***(Euclid's First Theorem)*** *The number of primes is infinite.*

Proof ► Suppose there were only a finite number of primes, say

$$p_1, p_2, \ldots, p_n.$$

Let

$$N = p_1 p_2 \cdots p_n + 1.$$

Evidently none of the primes $p_1, \ldots, p_n$ divides N.

Lemma 1.1 *Every natural number $n > 1$ has at least one prime divisor.*

Proof of Lemma ▷ The smallest divisor $d > 1$ of n must be prime. For otherwise d would have a divisor e with $1 < e < d$; and e would be a divisor of n smaller than d. ◁

By the lemma, N has a prime factor p, which differs from $p_1, \ldots, p_n$. ◄

Our argument not only shows that there are an infinity of primes; it shows that

$$p_n < 2^{2^n};$$

a very feeble bound, but our own. To see this, we argue by induction. Our proof shows that

$$p_{n+1} \leq p_1 p_2 \cdots p_n + 1.$$

But now, by our inductive hypothesis,

$$p_1 < 2^{2^1}, \ p_2 < 2^{2^2}, \ \ldots, \ p_n < 2^{2^n}.$$

It follows that

$$p_{n+1} \leq 2^{2^1 + 2^2 + \cdots + 2^n}$$

But

$$2^1 + 2^2 + \cdots + 2^n = 2^{n+1} - 1 < 2^{n+1}.$$

Hence

$$p_{n+1} < 2^{2^{n+1}}.$$

It follows by induction that

$$p_n < 2^{2^n},$$

for all $n \geq 1$, the result being trivial for $n = 1$.

This is not a very strong result, as we said. It shows, for example, that the 5th prime, in fact 11, is
$$< 2^{2^5} = 2^{32} = 4294967296.$$

In general, any bound for p_n gives a bound for $\pi(x)$ in the opposite direction, and vice versa; for
$$p_n \le x \Longleftrightarrow \pi(x) \ge n.$$

In the present case, for example, we deduce that

$$\pi(2^{2^y}) \ge [y] > y - 1$$

and so, setting $x = 2^{2^y}$,

$$\pi(x) \ge \log_2 \log_2 x - 1 > \log \log x - 1.$$

for $x > 1$. (We follow the usual convention that if no base is given then $\log x$ denotes the logarithm of x to base e.)

The *Prime Number Theorem* (which we shall make no attempt to prove) asserts that
$$p_n \sim n \log n,$$
or, equivalently,
$$\pi(x) \sim \frac{x}{\log x}.$$

This states, roughly speaking, that the probability of n being prime is about $1/\log n$. Note that this includes even numbers; the probability of an *odd* number n being prime is about $2/\log n$. Thus roughly 1 in 6 odd numbers around 10^6 are prime; while roughly 1 in 12 around 10^{12} are prime.

(The Prime Number Theorem is the central result of *analytic number theory* since its proof involves complex function theory. Our concerns, by contrast, lie within *algebraic number theory*.)

There are several alternative proofs of Euclid's Theorem. We shall give one below. But first we must establish the Fundamental Theorem of Arithmetic (the Unique Factorisation Theorem) which gives prime numbers their central rôle in number theory; and for that we need Euclid's Algorithm.

1.2 Euclid's Algorithm

Proposition 1.1 *Suppose* $m, n \in \mathbb{N}$, $m \ne 0$. *Then there exist unique* $q.r \in \mathbb{N}$ *such that*
$$n = qm + r, \quad 0 \le r < m.$$

Proof ► For uniqueness, suppose

$$n = qm + r = q'm + r',$$

where $r < r'$, say. Then

$$(q' - q)m = r' - r.$$

The number of the right is $< m$, while the number on the left has absolute value $\geq m$, unless $q' = q$, and so also $r' = r$.

We prove existence by induction on n. The result is trivial if $n < m$, with $q = 0,\ r = n$. Suppose $n \geq m$. By our inductive hypothesis, since $n - m < n$,

$$n - m = q'm + r,$$

where $0 \leq r < m$. But then

$$n = qm + r,$$

with $q = q' + 1$. ◄

Remark: One might ask why we feel the need to justify division with remainder (as above), while accepting, for example, proof by induction. This is not an easy question to answer.

Kronecker said, "*God gave the integers. The rest is Man's.*" Virtually all number theorists agree with Kronecker in practice, even if they do not accept his theology. In other words, they believe that the integers exist, and have certain obvious properties.

Certainly, if pressed, one might go back to Peano's Axioms, which are a standard formalisation of the natural numbers. (These axioms include, incidentally, proof by induction.) Certainly any properties of the integers that we assume could easily be derived from Peano's Axioms.

However, as I heard an eminent mathematician (Louis Mordell) once say, "If you deduced from Peano's Axioms that $1+1 = 3$, which would you consider most likely, that Peano's Axioms were wrong, or that you were mistaken in believing that $1 + 1 = 2$?"

Proposition 1.2 *Suppose $m, n \in \mathbb{N}$. Then there exists a unique number $d \in \mathbb{N}$ such that*

$$d \mid m,\ d \mid n,$$

and furthermore, if $e \in \mathbb{N}$ then

$$e \mid m,\ e \mid n \Longrightarrow e \mid d.$$

Definition 1.4 *We call this number d the* greatest common divisor *of m and n, and we write*

$$d = \gcd(m, n).$$

Proof ► Euclid's Algorithm is a simple technique for determining the greatest common divisor $\gcd(m, n)$ of two natural numbers $m, n \in \mathbb{N}$. It proves incidentally — as the Proposition asserts — that any two numbers *do* indeed have a greatest common divisor (or highest common factor).

First we divide the larger, say n, by the smaller. Let the quotient be q_1 and let the remainder (all we are really interested in) be r_1:

$$n = mq_1 + r_1.$$

Now divide m by r_1 (which must be less than m):

$$m = r_1 q_2 + r_2.$$

We continue in this way until the remainder becomes 0:

$$\begin{aligned} n &= mq_1 + r_1, \\ m &= r_1 q_2 + r_2, \\ r_1 &= r_2 q_3 + r_3, \\ &\cdots \\ r_{t-1} &= r_{t-2} q_{t-1} + r_t, \\ r_t &= r_{t-1} q_t. \end{aligned}$$

The remainder must vanish after at most m steps, for each remainder is strictly smaller than the previous one:

$$m > r_1 > r_2 > \cdots$$

Now we claim that the last non-zero remainder, $d = r_t$ say, has the required property:

$$d = \gcd(m, n) = r_t.$$

In the first place, working up from the bottom,

$$\begin{aligned} d = r_t &\mid r_{t-1}, \\ d \mid r_t \text{ and } d \mid r_{t-1} &\Longrightarrow d \mid r_{t-2}, \\ d \mid r_{t-1} \text{ and } d \mid r_{t-2} &\Longrightarrow d \mid r_{t-3}, \\ &\cdots \\ d \mid r_3 \text{ and } d \mid r_2 &\Longrightarrow d \mid r_1, \\ d \mid r_2 \text{ and } d \mid r_1 &\Longrightarrow d \mid m, \\ d \mid r_1 \text{ and } d \mid m &\Longrightarrow d \mid n. \end{aligned}$$

Thus

$$d \mid m, n;$$

so d is certainly *a* divisor of m and n.

On the other hand, suppose e is a divisor of m and n:

$$e \mid m, n.$$

Then, working *downwards*, we find successively that

$$\begin{aligned} e \mid m \text{ and } e \mid n &\Longrightarrow e \mid r_1, \\ e \mid r_1 \text{ and } e \mid m &\Longrightarrow e \mid r_2, \\ e \mid r_2 \text{ and } e \mid r_1 &\Longrightarrow e \mid r_3, \\ &\dots \\ e \mid r_{t-2} \text{ and } e \mid r_{t-1} &\Longrightarrow e \mid r_t. \end{aligned}$$

Thus

$$e \mid r_t = d.$$

We conclude that our last non-zero remainder r_t is number we are looking for:

$$\gcd(m, n) = r_t.$$

◄

It is easy to overlook the power and subtlety of the Euclidean Algorithm. The algorithm also gives us the following result.

Theorem 1.2 *Suppose* $m, n \in \mathbb{N}$. *Let*

$$\gcd(m, n) = d.$$

Then there exist integers $x, y \in \mathbb{Z}$ *such that*

$$mx + ny = d.$$

Proof ► The Proposition asserts that d can be expressed as a linear combination (with integer coefficients) of m and n. We shall prove the result by working backwards from the end of the algorithm, showing successively that d is a linear combination of r_s and r_{s+1}, and so, since r_{s+1} is a linear combination of r_{s-1} and r_s, d is also a linear combination of r_{s-1} and r_s.

To start with,

$$d = r_t.$$

From the previous line in the Algorithm,

$$r_{t-2} = q_t r_{t-1} + r_t.$$

Thus

$$d = r_t = r_{t-2} - q_t r_{t-1}.$$

But now, from the previous line,

$$r_{t-3} = q_{t-1} r_{t-2} + r_{t-1}.$$

Thus

$$r_{t-1} = rt - 3 - q_{t-1} r_{t-2}.$$

Hence

$$\begin{aligned} d &= r_{t-2} - q_t rt - 1 \\ &= r_{t-2} - q_t(r_{t-3} - q_{t-1} r_{t-2}) \\ &= -q_t r_{t-3} + (1 + q_t q_{t-1}) r_{t-2}. \end{aligned}$$

Continuing in this way, suppose we have shown that

$$d = a_s r_s + b_s r_{s+1}.$$

Since

$$r_{s-1} = q_{s+1} r_s + r_{s+1},$$

it follows that

$$\begin{aligned} d &= a_s r_s + b_s(r_{s-1} - q_{s+1} r_s) \\ &= b_s r_{s-1} + (a_s - b_s q_{s+1}) r_s. \end{aligned}$$

Thus

$$d = a_{s-1} r_{s-1} + b_{s-1} r_s,$$

with

$$a_{s-1} = b_s,\ b_{s-1} = a_s - b_s q_{s+1}.$$

Finally, at the top of the algorithm,

$$\begin{aligned} d &= a_0 r_0 + b_0 r_1 \\ &= a_0 r_0 + b_0(m - q_1 r_0) \\ &= b_0 m + (a_0 - b_0 q_1) r_0 \\ &= b_0 m + (a_0 - b_0 q_1)(n - q_0 m) \\ &= (b_0 - a_0 q_0 + b_0 q_0 q_1) m + (a_0 - b_0 q_0) n, \end{aligned}$$

which is of the required form. ◀

Example: Suppose $m = 39,\ n = 99$. Following Euclid's Algorithm,

$$\begin{aligned} 99 &= 2 \cdot 39 + 21, \\ 39 &= 1 \cdot 21 + 18, \\ 21 &= 1 \cdot 18 + 3, \\ 18 &= 6 \cdot 3. \end{aligned}$$

Thus

$$\gcd(39, 99) = 3.$$

Also

$$\begin{aligned} 3 &= 21 - 18 \\ &= 21 - (39 - 21) \\ &= -39 + 2 \cdot 21 \\ &= -39 + 2(99 - 2 \cdot 39) \\ &= 2 \cdot 99 - 5 \cdot 39. \end{aligned}$$

Thus the *Diophantine equation*

$$99x + 39y = 3$$

has the solution

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

(By a Diophantine equation we simply mean a polynomial equation to which we are seeking integer solutions.)

This solution is not unique; we could, for example, add 39 to x and subtract 99 from y. We can find the general solution by subtracting the particular solution we have just found to give a *homogeneous* linear equation. Thus if $x', y' \in \mathbb{Z}$ also satisfies the equation then $X = x' - x,\ Y = y' - y$ satisfies the homogeneous equation

$$99X + 39Y = 0,$$

ie

$$33X + 13Y = 0,$$

the general solution to which is

$$X = 13t,\ Y = -33t$$

for $t \in \mathbb{Z}$. The general solution to this diophantine equation is therefore

$$x = 2 + 13t,\ y = -5 - 33t \qquad (t \in \mathbb{Z}).$$

It is clear that the Euclidean Algorithm gives a complete solution to the general linear diophantine equation

$$ax + by = c.$$

This equation has no solution unless

$$\gcd(a, b) \mid c,$$

in which case it has an infinity of solutions. For if (x, y) is a solution to the equation

$$ax + by = d,$$

and $c = dc'$ then $(c'x, c'y)$ satisfies

$$ax + by = c,$$

and we can find the general solution as before.

Corollary 1.1 *Suppose $m, n \in \mathbb{Z}$. Then the equation*

$$mx + ny = 1$$

has a solution $x, y \in \mathbb{Z}$ if and only if $\gcd(m, n) = 1$.

It is worth noting that we can improve the efficiency of Euclid's Algorithm by allowing negative remainders. For then we can divide with remainder $\leq m/2$ in absolute value, ie

$$n = qm + r,$$

with $-m/2 \leq r < m/2$. The Algorithm proceeds as before; but now we have

$$m \geq |r_0/2| \geq |r_1/2^2| \geq \ldots,$$

so the Algorithm concludes after at most $\log_2 m$ steps.

This shows that the algorithm is *in class P*, ie it can be completed in polynomial (in fact linear) time in terms of the lengths of the input numbers m, n — the *length* of n, ie the number of bits required to express n in binary form, being

$$[\log_2 n] + 1.$$

Algorithms in class P (or *polynomial time* algorithms) are considered *easy* or *tractable*, while problems which cannot be solved in polynomial time are considered *hard* or *intractable*. RSA encryption — the standard techniqhe for encrypting confidential information — rests on the belief — and it should be emphasized that this is a belief and not a proof — that factorisation of a large number is intractable.

Example: Taking $m = 39,\ n = 99$, as before, the Algorithm now goes

$$\begin{aligned} 99 &= 3 \cdot 39 - 18, \\ 39 &= 2 \cdot 18 + 3, \\ 18 &= 6 \cdot 3, \end{aligned}$$

giving (of course)

$$\gcd(39, 99) = 3,$$

as before.

1.3 Ideals

We used the Euclidean Algorithm above to show that if $\gcd(a, b) = 1$ then there we can find $u, v \in \mathbb{Z}$ such that

$$au + bv = 1.$$

There is a much quicker way of proving that such u, v exist, without explicitly computing them.

Recall that an *ideal* in a commutative ring A is a non-empty subset $\mathfrak{a} \subset A$ such that

1. $a, b \in \mathfrak{a} \Longrightarrow a + b \in \mathfrak{a}$;
2. $a \in \mathfrak{a},\ c \in A \Longrightarrow ac \in \mathfrak{a}$.

As an example, the multiples of an element $a \in A$ form an ideal

$$\langle a \rangle = \{ac : c \in A\}.$$

Such an ideal is said to be *principal*.

Proposition 1.3 *Every ideal $\mathfrak{a} \subset \mathbb{Z}$ is principal.*

Proof ▶ If $\mathfrak{a} = 0$ (by convention we denote the ideal $\{0\}$ by 0) the result is trivial: $\mathfrak{a} = \langle 0 \rangle$. We may suppose therefor that $\mathfrak{a} \neq 0$.

Then $\mathfrak{a}$ must contain integers $n > 0$ (since $-n \in \mathfrak{a} \Longrightarrow n \in \mathfrak{a}$). Let d be the least such integer. Then

$$\mathfrak{a} = \langle d \rangle.$$

For suppose $a \in \mathfrak{a}$. Dividing a by d,

$$a = qd + r,$$

where

$$0 \leq r < d.$$

But

$$r = a + (-q)d \in \mathfrak{a}.$$

Hence $r = 0$; for otherwise r would contradict the minimality of d. Thus

$$a = qd,$$

ie every element $a \in \mathfrak{a}$ is a multiple of d. ◀

Now suppose $a, b \in \mathbb{Z}$. Consider the set of integers

$$I = \{au + bv : u, v \in \mathbb{Z}\}.$$

It is readily verified that I is an ideal.

According to the Proposition above, this ideal is principal, say

$$I = \langle d \rangle.$$

But now

$$a \in I \Longrightarrow d \mid a, \quad b \in I \Longrightarrow d \mid b.$$

On the other hand,

$$\begin{aligned} e \mid a,\ e \mid b &\Longrightarrow e \mid au + bv \\ &\Longrightarrow e \mid d. \end{aligned}$$

It follows that
$$d = \gcd(a, b);$$
and we have shown that the diophantine equation
$$au + bv = d$$
always has a solution.

In particular, if $\gcd(a, b) = 1$ we can $u, v \in \mathbb{Z}$ such that
$$au + bv = 1.$$

This proof is much shorter than the one using the Euclidean Algorithm; but it suffers from the disadvantage that it provides no way of computing
$$d = \gcd(a, b),$$
and no way of solving the equation
$$au + bv = d.$$
In effect, we have taken d as the least of an infinite set of positive integers, using the fact that the natural numbers $\mathbb{N}$ are *well-ordered*, ie every subset $S \subset \mathbb{N}$ has a least element.

1.4 The Fundamental Theorem of Arithmetic

Proposition 1.4 *(Euclid's Lemma) Suppose $p \in \mathbb{N}$ is a prime number; and suppose $a, b \in \mathbb{Z}$. Then*
$$p \mid ab \Longrightarrow p \mid a \text{ or } p \mid b.$$

Proof ► Suppose $p \mid ab,\ p \nmid a$. We must show that $p \mid b$. Evidently
$$\gcd(p, a) = 1.$$
Hence, by Corollary 1.1, there exist $x, y \in \mathbb{Z}$ such that
$$px + ay = 1.$$
Multiplying this equation by b,
$$pxb + aby = b.$$
But $p \mid pxb$ and $p \mid aby$ (since $p \mid ab$). Hence
$$p \mid b.$$

◄

Theorem 1.3 *Suppose $n \in \mathbb{N},\ n > 0$. Then n is expressible as a product of prime numbers,*

$$n = p_1 p_2 \cdots p_r,$$

and this expression is unique up to order.

Remark: We follow the convention that an empty product has value 1, just as an empty sum has value 0. Thus the theorem holds for $n = 1$ as the product of *no* primes.

Proof ► We prove existence by induction on n, the result begin trivial (by the remark above) when $n = 1$. We know that n has at least one prime factor p, by Lemma 1.1, say

$$n = pm.$$

Since $m = n/p < n$, we may apply our inductive hypothesis to m,

$$m = q_1 q_2 \cdots q_s.$$

Hence

$$n = pq_1 q_2 \cdots q_s.$$

Now suppose

$$n = p_1 p_2 \cdots p_r = m = q_1 q_2 \cdots q_s.$$

Since $p_1 \mid n$, it follows by repeated application of Euclid's Lemma that

$$p_1 \mid q_j$$

for some j. But then it follows from the definition of a prime number that

$$p_1 = q_j.$$

Again, we argue by induction on n. Since

$$n/p_1 = p_2 \cdots p_r = q_1 \cdots \hat{q_j} \cdots q_s$$

(where the 'hat' indicates that the factor is omitted), and since $n/p_1 < n$, we deduce that the factors $p_2, \ldots, p_r$ are the same as $q_1, \ldots, \hat{q_j}, \ldots, q_s$, in some order. Hence $r = s$, and the primes $p_1, \cdots, p_r$ and $q_1, \ldots, q_s$ are the same in some order. ◄

We can base another proof of Euclid's Theorem (that there exist an infinity of primes) on the fact that if there were only a finite number of primes there would not be enough products to "go round".

Thus suppose there were just m primes

$$p_1, \ldots, p_m.$$

Let $N \in \mathbb{N}$. By the Fundamental Theorem, each $n \le N$ would be expressible in the form

$$n = p_1^{e_1} \cdots p_m^{e_m}.$$

(Actually, we are only using the existence part of the Fundamental Theorem; we do not need the uniqueness part.)

For each i $(1 \le i \le m)$,

$$\begin{aligned} p_i^{e_i} \mid n &\Longrightarrow p_i^{e_i} \le n \\ &\Longrightarrow p_i^{e_i} \le N \\ &\Longrightarrow 2^{e_i} \le N \\ &\Longrightarrow e_i \le \log_2 N. \end{aligned}$$

Thus there are at most $\log_2 N + 1$ choices for each exponent e_i, and so the number of numbers $n \le N$ expressible in this form is

$$\le (\log_2 N + 1)^m.$$

So our hypothesis implies that

$$(\log_2 N + 1)^m \ge N$$

for all N.

But in fact, to the contrary,

$$X > (\log_2 X + 1)^m = \left(\frac{\log X}{\log 2} + 1\right)^m$$

for all sufficiently large X. To see this, set $X = e^x$. We have to show that

$$e^x > \left(\frac{x}{\log 2} + 1\right)^m.$$

Since

$$\frac{x}{\log 2} + 1 < 2x$$

if $x \ge 3$, it is sufficient to show that

$$e^x > (2x)^m$$

for sufficiently large x. But

$$e^x > \frac{x^{m+1}}{(m+1)!}$$

if $x > 0$, since the expression on the right is one of the terms in the power-series expansion of e^x. Thus the inequality holds if

$$\frac{x^{m+1}}{(m+1)!} > (2x)^m,$$

ie if

$$x > 2^m(m+1)!.$$

We have shown therefore that m primes are insufficient to express all $n \le N$ if

$$N \ge e^{2^m(m+1)!}.$$

Thus our hypothesis is untenable; and Euclid's theorem is proved.

Our proof gives the bound

$$p_n \le e^{2^m(m+1)!}.$$

which is even worse than the bound we derived from Euclid's proof. (For it is easy to see by induction that

$$(m+1)! > e^m$$

for $m \ge 2$. Thus our bound is worse than e^{e^n}, compared with 2^{2^n} by Euclid's method.)

We can improve the bound considerably by taking out the square factor in n. Thus each number $n \in \mathbb{N}$ $(n > 0)$ is uniquely expressible in the form

$$n = d^2 p_1 \dots p_r,$$

where the primes $p_1, \dots, p_r$ are distinct. In particular, if there are only m primes then each n is expressible in the form

$$n = d^2 p_1^{e_1} \cdots p_m^{e_m},$$

where now each exponent e_i is either 0 or 1.

Consider the numbers $n \le N$. Since

$$d \le \sqrt{n} \le \sqrt{N},$$

the number of numbers of the above form is

$$\le \sqrt{N} 2^m.$$

Thus we shall reach a contradiction when

$$\sqrt{N} 2^m \ge N,$$

ie

$$N \le 2^{2m}.$$

This gives us the bound

$$p_n \le 2^{2n},$$

better than 2^{2^n}, but still a long way from the truth.

1.5 The Fundamental Theorem, recast

We suppose throughout this section that A is an integral domain. (Recall that an integral domain is a commutative ring with 1 having no zero divisors, ie if $a, b \in A$ then

$$ab = 0 \Longrightarrow a = 0 \text{ or } b = 0.)$$

We want to examine whether or not the Fundamental Theorem holds in A — we shall find that it holds in some commutative rings and not in others. But to make sense of the question we need to re-cast our definition of a prime.

Looking back at $\mathbb{Z}$, we see that we could have defined primality in two ways (excluding $p = 1$ in both cases):

1. p is prime if it has no proper factors, ie

$$p = ab \Longrightarrow a = 1 \text{ or } b = 1.$$

2. p is prime if

$$p \mid ab \Longrightarrow p \mid a \text{ or } p \mid b.$$

The two definitions are of course equivalent in the ring Z. However, in a general ring the second definition is stronger: that is, an element satisfying it must satisfy the first definition, but the converse is not necessarily true. We shall take the second definition as our starting-point.

But first we must deal with one other point. In defining primality in Z we actually restricted ourselves to the semi-ring $\mathbb{N}$, defined by the *order* in $\mathbb{Z}$:

$$\mathbb{N} = \{n \in \mathbb{Z} : n \geq 0\}.$$

However, a general ring A has no natural order, and no such semi-ring, so we must consider all elements $a \in A$.

In the case of Z this would mean considering $-p$ as a prime on the same footing as p. But now, for the Fundamental Theorem to make sense, we would have to regard the primes $\pm p$ as essentially the same.

The solution in the general ring is that to regard two primes as *equivalent* if each is a multiple of the other, the two multiples necessarily being *units*.

Definition 1.5 *An element $\epsilon \in A$ is said to be a* unit *if it is invertible, ie if there is an element $\eta \in A$ such that*

$$\epsilon\eta = 1.$$

We denote the set of units in A by $A^\times$.

For example,

$$\mathbb{Z}^\times = \{\pm 1\}.$$

Proposition 1.5 *The units in A form a multiplicative group $A^\times$.*

Proof ► This is immediate. Multiplication is associative, from the definition of a ring; and $\eta = \epsilon^{-1}$ is a unit, since it has inverse ϵ. ◄

Now we can define primality.

Definition 1.6 *Suppose $a \in A$ is not a unit, and $a \neq 0$. Then*

1. *a is said to be* irreducible *if*

$$a = bc \Longrightarrow b \text{ or } c \text{ is a unit}.$$

2. *a is said to be* prime *if*

$$a \mid bc \Longrightarrow a \mid b \text{ or } p \mid b.$$

Proposition 1.6 *If $a \in A$ is prime then it is irreducible.*

Proof ► Suppose

$$a = bc.$$

Then

$$a \mid b \text{ or } a \mid c.$$

We may suppose without loss of generality that $a \mid b$. Then

$$a \mid b,\ b \mid a \Longrightarrow a = b\epsilon,$$

where ϵ is a unit; and

$$a = bc = b\epsilon \Longrightarrow c = \epsilon.$$

◄

Definition 1.7 *The elements $a, b \in A$ are said to be* equivalent*, written*

$$a \sim b,$$

if

$$b = \epsilon a$$

for some unit ϵ.

In effect, the group of units $A^\times$ acts on A and two elements are equivalent if each is a transform of the other under this action.

Now we can re-state the Fundamental Theorem in terms which make sense in any integral domain.

Definition 1.8 *The integral domain A is said to be a* unique factorisation domain *if each non-unit $a \in A,\ a \neq 0$ is expressible in the form*

$$a = p_1 \cdots p_r,$$

where $p_1, \ldots, p_r$ are prime, and if this expression is unique up to order and equivalence of primes.

In other words, if

$$a = q_1 \cdots q_s$$

is another expression of the same form, then $r = s$ and we can find a permutation π of $\{1, 2, \ldots, r\}$ and units $\epsilon_1, \epsilon_2, \ldots, \epsilon_r$ such that

$$q_i = \epsilon_i p_{\pi(i)}$$

for $i = 1, 2, \ldots, r$.

Thus a unique factorisation domain (UFD) is an integral domain in which the Fundamental Theorem of Arithmetic is valid.

1.6 Principal ideals domains

Definition 1.9 *The integral domain A is said to be a* principal ideal domain *if every ideal $\mathfrak{a} \in A$ is* principal, *ie*

$$\mathfrak{a} = \langle a \rangle = \{ac : c \in A\}$$

for some $a \in A$.

Example: By Proposition 1.3, $\mathbb{Z}$ is a principal ideal domain.

Our proof of the Fundamental Theorem can be divided into two steps — this is clearer in the alternative version outlined in Section 1.3 — first we showed that that $\mathbb{Z}$ is a principal ideal domain, and then we deduced from this that $\mathbb{Z}$ is a unique factorisation domain.

As our next result shows this argument is generally available; it is the technique we shall apply to show that the Fundamental Theorem holds in a variety of integral domains.

Proposition 1.7 *A principal ideal domain is a unique factorisation domain.*

Proof ► Suppose A is a principal ideal domain.

Lemma 1.2 *A non-unit $a \in A,\ a \neq 0$ is prime if and only if it is irreducible, ie*

$$a = bc \implies a \text{ is a unit or } b \text{ is a unit}.$$

Proof of Lemma ▷ By Proposition 1.6, a prime is always irreducible.

Conversely, if

$$a = p_1 \cdots p_r$$

is irreducible then evidently $r = 1$, and a is prime. ◁

Now suppose a is neither a unit nor 0; and suppose that a is *not* expressible as a product of primes. Then a is reducible, by the Lemma above: say

$$a = a_1 b_1,$$

where a_1, b_1 are non-units. One at least of a_1, b_1 is not expressible as a product of primes; we may assume without loss of generality that this is true of a_1.

It follows by the same argument that

$$a_1 = a_2 b_2,$$

where a_2, b_2 are non-units, and a_2 is not expressible as a product of primes.

Continuing in this way,

$$a = a_1 b_1,\ a_1 = a_2 b_2,\ a_2 = a_3 b_3, \ldots .$$

Now consider the ideal

$$\mathfrak{a} = \langle a_1, a_2, a_3, \ldots \rangle.$$

By hypothesis this ideal is principal, say

$$\mathfrak{a} = \langle d \rangle.$$

Since $d \in \mathfrak{a}$,

$$d \in \langle a_1, \ldots, a_r \rangle = \langle a_r \rangle$$

for some r. But then

$$a_{r+1} \in \langle d \rangle = \langle a_r \rangle.$$

Thus

$$a_r \mid a_{r+1},\ a_{r+1} \mid a_r \implies a_r = a_{r+1}\epsilon \implies b_{r+1} = \epsilon,$$

where ϵ is a unit, contrary to construction.

Thus the assumption that a is not expressible as a product of primes is untenable;

$$a = p_1 \cdots p_r.$$

To prove uniqueness, we argue by induction on r, where r the smallest number such that a is expressible as a product of r primes.

Suppose

$$a = p_1 \cdots p_r = q_1 \cdots q_s.$$

Then

$$p_1 \mid q_1 \cdots q_s \implies p_1 \mid q_j$$

for some j. Since q_j is irreducible, by Proposition 1.6, it follows that

$$q_j = p_1 \epsilon,$$

where ϵ is a unit.

We may suppose, after re-ordering the q's that $j = 1$. Thus

$$p_1 \sim q_1.$$

If $r = 1$ then

$$a = p_1 = \epsilon p_1 q_2 \cdots q_s \Longrightarrow 1 = \epsilon q_2 \cdots q_s.$$

If $s > 1$ this implies that $q_2, \ldots, q_s$ are all units, which is absurd. Hence $s = 1$, and we are done.

If $r > 1$ then

$$q_1 = \epsilon p_1 \Longrightarrow p_2 p_3 \cdots p_r = (\epsilon q_2) q_3 \cdots q_s$$

(absorbing the unit ϵ into q_2). The result now follows by our inductive hypothesis. ◄

1.7 Polynomial rings

If A is a commutative ring (with 1) then we denote by $A[x]$ the ring of polynomials

$$p(x) = a_n x^n + \cdots + a_0 \quad (a_0, \ldots, a_n \in A).$$

Note that these polynomials should be regarded as formal expressions rather than maps $p : A \to A$; for if A is finite two different polynomials may well define the same map.

We identify $ainA$ with the *constant* polynomial $f(x) = a$. Thus

$$A \subset A[x].$$

Proposition 1.8 *If A is an integral domain then so is $A[x]$.*

Proof ► Suppose

$$f(x) = a_m x^m + \cdots + a_0, \quad g(x) = b_n x^n + \cdots + b_0,$$

where $a_m \neq 0,\ b_n \neq 0$. Then

$$f(x)g(x) = (a_m b_n) x^{m+n} + \cdots + a_0 b_0;$$

and the leading coefficient $a_m b_n \neq 0$. ◄

Proposition 1.9 *The units in $A[x]$ are just the units of A:*

$$(A[x])^\times = A^\times.$$

Proof ► It is clear that $a \in A$ is a unit (ie invertible) in $A[x]$ if and only if it is a unit in A.

On the other hand, no non-constant polynomial $F(x) \in A[x]$ can be invertible, since

$$\deg F(x)G(x) \geq \deg F(x)$$

if $G(x) \neq 0$. ◄

If A is a *field* then we can divide one polynomial by another, obtaining a remainder with lower *degree* than the divisor. Thus degree plays the rôle in $k[x]$ played by size in $\mathbb{Z}$.

Proposition 1.10 *Suppose k is a field; and suppose $f(x), g(x) \in k[x]$, with $g(x) \neq 0$. Then there exist unique polynomials $q(x), r(x) \in k[x]$ such that*

$$f(x) = g(x)q(x) + r(x),$$

where

$$\deg r(x) < \deg g(x).$$

Proof ► We prove the existence of $q(x), r(x)$ by induction on $\deg f(x)$.

Suppose

$$f(x) = a_m x^m + \cdots + a_0, \quad g(x) = b_n x^n + \cdots + b_0,$$

where $a_m \neq 0,\ b_n \neq 0$.

If $m < n$ then we can take $q(x) = 0,\ r(x) = f(x)$. We may suppose therefore that $m \geq n$. In that case, let

$$f_1(x) = f(x) - (a_m/b_n)x^{m-n}.$$

Then

$$\deg f_1(x) < \deg f(x).$$

Hence, by the inductive hypothesis,

$$f_1(x) = g(x)q_1(x) + r(x),$$

where

$$\deg r(x) < \deg g(x);$$

and then

$$f(x) = g(x)q(x) + r(x),$$

with

$$q(x) = (a_m/b_n)x^{m-n} + q_1(x).$$

For uniqueness, suppose

$$f(x) = g(x)q_1(x) + r_1(x) = g(x)q_2(x) + r_2(x).$$

On subtraction,

$$g(x)q(x) = r(x),$$

where

$$q(x) = q_2(x) - q_1(x), \quad r(x) = r_1(x) - r_2(x).$$

But now, if $q(x) \neq 0$,

$$\deg(g(x)q(x)) \geq \deg g(x), \quad \deg r(x) < \deg g(x).$$

This is a contradiction. Hence

$$q(x) = 0,$$

ie

$$q_1(x) = q_2(), \quad r_1(x) = r_2().$$

◄

Proposition 1.11 *If k is a field then $k[x]$ is a principal ideal domain.*

Proof ► As with $\mathbb{Z}$ we can prove this result in two ways: constructively, using the Euclidean Algorithm; or non-constructively, using ideals. This time we take the second approach.

Suppose

$$\mathfrak{a} \subset k[x]$$

is an ideal. If $\mathfrak{a} = 0$ the result is trivial; so we may assume that $\mathfrak{a} \neq 0$.

Let

$$d(x) \in \mathfrak{a}$$

be a polynomial in $\mathfrak{a}$ of minimal degree. Then

$$\mathfrak{a} = \langle d(x) \rangle.$$

For suppose $f(x) \in \mathfrak{a}$. Divide $f(x)$ by $d(x)$:

$$f(x) = d(x)q(x) + r(x),$$

where $\deg r(x) < \deg d(x)$. Then

$$r(x) = f(x) - d(x)q(x) \in \mathfrak{a}$$

since $f(x), d(x) \in \mathfrak{a}$. Hence, by the minimality of $\deg d(x)$,

$$r(x) = 0,$$

ie

$$f(x) = d(x)q(x).$$

◄

By Proposition 1.7 this gives the result we really want.

Corollary 1.2 *If k is a field then $k[x]$ is a unique factorisation domain.*

Every non-zero polynomial $f(x) \in k[x]$ is equivalent to a unique monic polynomial, namely that obtained by dividing by its leading term. Thus each prime, or irreducible, polynomial $p(x) \in k[x]$ has a unique monic representative; and we can restate the above Corollary in a simpler form.

Corollary 1.3 *Each monic polynomial*

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$$

can be uniquely expressed (up to order) as a product of irreducible monic polynomials:

$$f(x) = p_1(x) \cdots p_r(x).$$

1.8 Postscript

We end this Chapter with a result that we don't really need, but which we have come so close to it would be a pity to omit.

Suppose A is an integral domain. Let K be the *field of fractions* of A. (Recall that K consists of the formal expressions

$$\frac{a}{b},$$

with $a, b \in A,\ b \neq 0$; where we set

$$\frac{a}{b} = \frac{c}{d} \quad \text{if} \quad ad = bc.$$

The map

$$a \mapsto \frac{a}{1} : A \to K$$

is injective, allowing us to identify A with a subring of K.)

The canonical injection

$$A \subset K$$

evidently extends to an injection

$$A[x] \subset K[x].$$

Thus we can regard $f(x) \in A[x]$ as a polynomial over K.

Proposition 1.12 *If A is a unique factorisation domain then so is $A[x]$.*

Proof ► First we must determine the primes in $A[x]$.

Lemma 1.3 *The element $p \in A$ is prime in $A[x]$ if and only if it is prime in A.*

Proof of Lemma ▷ It is evident that

$$p \text{ prime in } A[x] \Longrightarrow p \text{ prime in } A.$$

Conversely, suppose p is prime in A; We must show that if $F(x), G(x) \in A[x]$ then

$$p \mid F(x)G(x) \Longrightarrow p \mid F(x) \text{ or } p \mid G(x).$$

In other words,

$$p \nmid F(x),\ p \nmid G(x) \Longrightarrow p \nmid F(x)G(x).$$

Suppose

$$F(x) = a_m x^m + \cdots + a_0, \quad G(x) = b_n x^n + \cdots + b_0;$$

and suppose

$$p \nmid F(x), \quad p \nmid G(x).$$

Let a_r, b_s be the highest coefficients of $f(x), g(x)$ *not* divisible by p. Then the coefficient of x^{r+s} in $f(x)g(x)$ is

$$a_0 b_{r+s} + a_1 b_{r+s-1} + \cdots + a_r b_s + \cdots + a_{r+s} b_0 \equiv a_r b_s \bmod p,$$

since all the terms except $a_r b_s$ are divisible by p. Hence

$$p \mid a_r b_s \Longrightarrow p \bmod a_r \text{ or } p \bmod b_s,$$

contrary to hypothesis. In other words,

$$p \nmid F(x)G(x).$$

◁

Lemma 1.4 *Suppose $f(x) \in K[x]$. Then $f(x)$ is expressible in the form*

$$f(x) = \alpha F(x),$$

where $\alpha \in K$ and

$$F(x) = a_n x^n + \cdots + a_0 \in A[x]$$

with

$$\gcd(a_0, \ldots, a_n) = 1;$$

and the expression is unique up to multiplication by a unit, ie if

$$f(x) = \alpha F(x) = \beta G(x),$$

where $G(x)$ has the same property then

$$G(x) = \epsilon F(x), \quad \alpha = \epsilon\beta$$

for some unit $\epsilon \in A$.

Proof of Lemma ▷ Suppose

$$f(x) = \alpha_n x^n + \cdots + \alpha_0.$$

Let

$$\alpha_i = \frac{a_i}{b_i},$$

where $a_i, b_i \in A$; and let

$$b = \prod b_i.$$

Then

$$bf(x) = b_n x^n + \cdots + b_0 \in A[x].$$

Now let

$$d = \gcd(b_0, \ldots, b_n).$$

Then

$$f(x) = (b/d)(c_n x^n + \cdots + c_0)$$

is of the required form, since

$$\gcd(c_0, \ldots, c_n) = 1.$$

To prove uniqueness, suppose

$$f(x) = \alpha F(x) = \beta G(x).$$

Then

$$G(x) = \gamma F(x),$$

where $\gamma = \alpha/\beta$.

In a unique factorisation domain A we can express any $\gamma \in K$ in the form

$$\gamma = \frac{a}{b},$$

with $\gcd(a, b) = 1$, since we can divide a and b by any common factor.

Thus

$$aF(x) = bG(x).$$

Let p be a prime factor of b. Then

$$p \mid aF(x) \Longrightarrow p \mid F(x),$$

contrary to our hypothesis on the coefficients of $F(x)$. Thus b has no prime factors, ie b is a unit; and similarly a is a unit, and so γ is a unit. ◁

Lemma 1.5 *A non-constant polynomial*

$$F(x) = a_n x^n + \cdots + a_0 \in A[x]$$

is prime in $A[x]$ if and only if

1. *$F(x)$ is prime (ie irreducible) in $K(x)$; and*
2. $\gcd(a_0, \ldots, a_n) = 1$.

Proof of Lemma $\triangleright$ Suppose $F(x)$ is prime in $A[x]$. Then certainly

$$\gcd(a_0, \ldots, a_n) = 1,$$

otherwise $F(x)$ would be reducible.

Suppose $F(x)$ factors in $K[x]$; say

$$F(x) = g(x)h(x).$$

By Proposition 1.4,

$$g(x) = \alpha G(x), \quad h(x) = \beta H(x),$$

where $G(x), H(x)$ have no factors in A. Thus

$$F(x) = \gamma G(x)H(x),$$

where $\gamma \in K$. Let $\gamma = a/b$, where $a, b \in A$ and $\gcd(a, b) = 1$. Then

$$bF(x) = aG(x)H(x).$$

Suppose p is a prime factor of b. Then

$$p \mid G(x) \quad \text{or} \quad p \mid H(x),$$

neither of which is tenable. Hence b has no prime factors, ie b is a unit. But now

$$F(x) = ab^{-1}G(x)H(x);$$

and so $F(x)$ factors in $A[x]$.

Conversely, suppose $F(x)$ has the two given properties. We have to show that $F(x)$ is prime in $A[x]$.

Suppose

$$F(x) \mid G(x)H(x)$$

in $A[x]$.

If $F(x)$ is constant then

$$F(x) = a \sim 1$$

by the second property, so

$$F(x) \mid G(x) \quad \text{and} \quad F(x) \mid H(x).$$

We may suppose therefore that $\deg F(x) \geq 1$. Since $K[x]$ is a unique factorisation domain (Corollary to Proposition 1.11),

$$F(x) \mid G(x) \quad \text{or} \quad F(x) \mid H(x)$$

in $K[x]$. We may suppose without loss of generality that

$$F(x) \mid G(x)$$

in $K[x]$, say

$$G(x) = F(x)h(x),$$

where $h(x) \in K[x]$.

By Lemma 1.4 we can express $h(x)$ in the form

$$h(x) = \alpha H(x),$$

where the coefficients of $H(x)$ are factor-free. Writing

$$\alpha = \frac{a}{b},$$

with $\gcd(a, b) = 1$, we have

$$bG(x) = aF(x)H(x).$$

Suppose p is a prime factor of b. Then

$$p \mid a \quad \text{or} \quad p \mid F(x) \quad \text{or} \quad p \mid H(x),$$

none of which is tenable. Hence b has no prime factors, ie b is a unit. Thus

$$F(x) \mid G(x)$$

in $A[x]$. $\triangleleft$

Now suppose

$$F(x) = a_n x^n + \cdots a_0 \in A[x]$$

is not a unit in $A[x]$.

If $F(x)$ is constant, say $F(x) = a$, then the factorisation of a into primes in A is a factorisation into primes in $A[x]$, by Lemma 1.3. Thus we may assume that $\deg F(x) \geq 1$.

Since $K[x]$ is a unique factorisation domain (Corollary to Proposition 1.11), $F(x)$ can be factorised in $K[x]$:

$$F(x) = a_n p_1(x) \cdots p_s(x),$$

where $p_1(x), \ldots, p_s(x)$ are irreducible monic polynomials in $K[x]$. By Lemmas 1.4 and 1.5 each $p_i(x)$ is expressible in the form

$$p_i(x) = \alpha_i P_i(x),$$

where $P_i(x)$ is prime in $A[x]$.

Thus

$$F(x) = \alpha P_1(x) \cdots P_r(x),$$

where

$$\alpha = a_n \alpha_1 \cdots \alpha_r \in K.$$

Let

$$\alpha = \frac{a}{b},$$

where $\gcd(a, b) = 1$. Then

$$bF(x) = aP_1(x) \cdots P_r(x).$$

Let p be a prime factor of b. Then

$$p \mid P_i(x)$$

for some i, contrary to the definition of $P_i(x)$. Hence b has no prime factors, ie b is a unit.

If a is a unit then we can absorb $\epsilon = a/b$ into $P_1(x)$:

$$F(x) = Q(x)P_2(x) \cdots P_r(x),$$

where $Q(x) = (a/b)P_1(x)$.

If a is not a unit then

$$ab^{-1} = p_1 \cdots p_s,$$

where $p_1, \ldots, p_s$ are prime in A (and so in $A[x]$ by Lemma 1.3); and

$$F(x) = p_1 \cdots p_s P_1(x) \cdots P_r(x),$$

as required.

Finally, to prove uniqueness, we may suppose that $\deg F(x) \geq 1$, since the result is immediate if $F(x) = a$ is constant.

Suppose

$$F(x) = p_1 \cdots p_s P_1(x) \cdots P_r(x) = q_1 \cdots q_{s'} Q_1(x) \cdots Q_{r'}(x).$$

Each $P_i(x),\ Q_j(x)$ is prime in $K[x]$ by Lemma 1.5. Since $K[x]$ is a unique factorisation domain (Corollary to Proposition 1.11) it follows that $r = r'$ and that after re-ordering,

$$Q_i(x) = \alpha P_i(x),$$

where $\alpha \in K^\times$. Let

$$\alpha = a/b$$

with $\gcd(a, b) = 1$. Then

$$aP_i(x) = bQ_i(x).$$

If p is a prime factor of b then

$$p \mid bQ_i(x) \Longrightarrow p \mid Q_i(x),$$

contrary to the definition of $Q_i(x)$. Thus b has no prime factors, and is therefore a unit. Similarly a is a unit. Hence

$$Q_i(x) = \epsilon_i P_i(x),$$

where $\epsilon_i \in A$ is a unit.

Setting

$$\epsilon = \prod_i \epsilon_i,$$

we have

$$p_1 \cdots p_s = \epsilon q_1 \cdots q_{s'}.$$

Since A is a unique factorisation domain, $s = s'$ and after re-ordering,

$$q_j = \eta_j p_j,$$

where $\eta_j \in A$ is a unit.

We conclude that the prime factors of $F(x)$ are unique up to order and equivalence (multiplication by units), ie $A[x]$ is a unique factorisation domain. ◄

Example: There is unique factorisation in $\mathbb{Z}[x]$, since $\mathbb{Z}$ is a principal ideal domain by Proposition 1.3 and so a unique factorisation domain by Proposition 1.7.

Note that $\mathbb{Z}[x]$ is *not* a principal ideal domain, since eg the ideal

$$\mathfrak{a} = \langle 2, x \rangle,$$

consisting of all polynomials

$$F(x) = a_n x^n + \cdots + a_0$$

with a_0 even, is not principals:

$$\mathfrak{a} \neq \langle G(x) \rangle.$$

For if it were, its generator $G(x)$ would have to be constant, since $\mathfrak{a}$ contains non-zero constants, and

$$\deg G(x)H(x) \geq \deg G(x)$$

if $H(x) \neq 0$. But if $G(x) = d$ then

$$\mathfrak{a} \cap \mathbb{Z} = \langle 2 \rangle \Longrightarrow d = \pm 2,$$

ie $\mathfrak{a}$ consists of all polynomials with *even* coefficients. Since $x \in \mathfrak{a}$ is not of this form we conclude that $\mathfrak{a}$ is not principal.

Chapter 2

Number fields

2.1 Algebraic numbers

Definition 2.1 *A number $\alpha \in \mathbb{C}$ is said to be* algebraic *if it satisfies a polynomial equation*

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n = 0$$

with rational coefficients $a_i \in \mathbb{Q}$.

For example, $\sqrt{2}$ and $i/2$ are algebraic.

A complex number is said to be *transcendental* if it is not algebraic. Both e and π are transcendental. It is in general extremely difficult to prove a number transcendental, and there are many open problems in this area, eg it is not known if π^e is transcendental.

Proposition 2.1 *The algebraic numbers form a field $\bar{\mathbb{Q}} \subset \mathbb{C}$.*

Proof ► If α satisfies the equation $f(x) = 0$ then $-\alpha$ satisfies $f(-x) = 0$, while $1/\alpha$ satisfies $x^n f(1/x) = 0$ (where n is the degree of $f(x)$). It follows that $-\alpha$ and $1/\alpha$ are both algebraic. Thus it is sufficient to show that if α, β are algebraic then so are $\alpha + \beta, \alpha\beta$.

Suppose α satisfies the equation

$$f(x) \equiv x^m + a_1 x^{m-1} + \cdots + a_m = 0,$$

and β the equation

$$g(x) \equiv x^n + b_1 x^{n-1} + \cdots + b_n = 0.$$

Consider the vector space

$$V = \langle \alpha^i \beta^j : 0 \le i < m,\ 0 \le j < n \rangle$$

over $\mathbb{Q}$ spanned by the mn elements $\alpha^i \beta^j$. Evidently

$$\alpha + \beta, \alpha\beta \in V.$$

But if $\theta \in V$ then the $mn+1$ elements

$$1, \theta, \theta^2, \ldots, \theta^{mn}$$

are necessarily linearly dependent (over $\mathbb{Q}$), since $\dim V \le mn$. In other words θ satisfies a polynomial equation of degree $\le mn$. Thus each element $\theta \in V$ is algebraic. In particular $\alpha + \beta$ and $\alpha\beta$ are algebraic. ◄

2.2 Minimal polynomials and conjugates

Recall that a polynomial $p(x)$ is said to be *monic* if its leading coefficient — the coefficient of the highest power of x — is 1:

$$p(x) = x^n + a_1 x^{n-1} + \cdots + a_n.$$

Proposition 2.2 *Each algebraic number $\alpha \in \bar{\mathbb{Q}}$ satisfies a unique monic polynomial $m(x)$ of minimal degree.*

Proof ► Suppose α satisfies two monic polynomials $m_1(x), m_2(x)$ of minimal degree d. Then α also satisfies the polynomial

$$p(x) = m_1(x) - m_2(x)$$

of degree $< d$; and if $p(x) \neq 0$ then we can make it monic by dividing by its leading coefficient. This would contradict the minimality of $m_1(x)$. Hence

$$m_1(x) = m_2(x).$$

◄

Definition 2.2 *The monic polynomial $m(x)$ satisfied by $\alpha \in \bar{\mathbb{Q}}$ is called the* minimal polynomial *of α. The* degree *of the algebraic number α is the degree of its minimal polynomial $m(x)$.*

Proposition 2.3 *The minimal polynomial $m(x)$ of $\alpha \in \bar{\mathbb{Q}}$ is irreducible.*

Proof ► Suppose to the contrary

$$m(x) = f(x)g(x)$$

where $f(x), g(x)$ are of lower degrees than $m(x)$. But then α must be a root of one of $f(x), g(x)$. ◄

Definition 2.3 *Two algebraic numbers α, β are said to be* conjugate *if they have the same minimal polynomial.*

Proposition 2.4 *An algebraic number of degree d has just d conjugates.*

Proof ► If the minimal poynomial of α is

$$m(x) = x^d + a_1 x^{d-1} + \cdots + a_d,$$

then by definition the conjugates of α are the d roots $\alpha_1 = \alpha, \alpha_2, \ldots, \alpha_d$ of $m(x)$:

$$m(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_d).$$

These conjugates are distinct, since an irreducible polynomial $m(x)$ over $\mathbb{Q}$ is necessarily *separable*, ie it cannot have a repeated root. For if α were a repeated root of $m(x)$, ie

$$(x - \alpha)^2 \mid m(x)$$

then

$$(x - \alpha) \mid m'(x),$$

and so

$$(x - \alpha) \mid d(x) = \gcd(m(x), m'(x)).$$

But

$$d(x) \mid m(x)$$

and

$$1 \leq \deg(d(x)) \leq d - 1,$$

contradicting the irreducibility of $m(x)$. ◄

2.3 Algebraic number fields

Proposition 2.5 *Every subfield $K \subset \mathbb{C}$ contains the rationals $\mathbb{Q}$:*

$$\mathbb{Q} \subset K \subset \mathbb{C}.$$

Proof ► By definition, $1 \in K$. Hence

$$n = 1 + \cdots + 1 \in K$$

for each integer $n > 0$.

By definition, K is an additive subgroup of $\mathbb{C}$. Hence $-1 \in K$; and so

$$-n = (-1)n \in K$$

for each integer $n > 0$. Thus

$$\mathbb{Z} \subset K.$$

Finally, since K is a field, each rational number

$$r = \frac{n}{d} \in K$$

where $n, d \in \mathbb{Z}$ with $d \neq 0$. ◄

We can consider any subfield $K \subset \mathbb{C}$ as a vector space over $\mathbb{Q}$.

Definition 2.4 *An* number field *(or more precisely, an* algebraic number field*) is a subfield $K \subset \mathbb{C}$ which is of finite dimension as a vector space over $\mathbb{Q}$. If*

$$\dim_{\mathbb{Q}} = d$$

then K is said to be a number field of degree d.

Proposition 2.6 *There is a smallest number field K containing the algebraic numbers $\alpha_1, \ldots, \alpha_r$.*

Proof ► Every intersection (finite or infinite) of subfields of $\mathbb{C}$ is a subfield of $\mathbb{C}$; so there is a smallest subfield K containing the given algebraic numbers, namely the intersection of all subfields containing these numbers. We have to show that this field is a number field, ie of finite dimension over $\mathbb{Q}$.

Lemma 2.1 *Suppose $K \subset \mathbb{C}$ is a finite-dimensional vector space over $\mathbb{Q}$. Then K is a number field if and only if it is closed under multiplication.*

Proof of Lemma ▷ If K is a number field then it is certainly closed under multiplication.

Conversely, if this is so then K is closed under addition and multiplication; so we only have to show that it is closed under division by non-zero elements.

Suppose $\alpha \in V,\ \alpha \neq 0$. Consider the map

$$x \mapsto \alpha x : V \to V.$$

This is a linear map over $\mathbb{Q}$; and it is injective since

$$\alpha x = 0 \Longrightarrow x = 0.$$

Since V is finite-dimensional it follows that the map is surjective; in particular,

$$\alpha x = 1$$

for some $x \in V$, ie α is invertible. Hence V is a field. ◁

Now suppose α_i is of degree d_i (ie satisfies a polynomial equation of degree d_i over $\mathbb{Q}$). Consider the vector space (over $\mathbb{Q}$)

$$V = \langle \alpha_1^{i_1} \cdots \alpha_r^{i_r} : 0 \leq i_1 < d_1, \cdots, 0 \leq i_r < d_r \rangle.$$

It is readily verified that

$$\alpha_i V \subset V,$$

and so

$$VV \subset V,$$

ie V is closed under multiplication.

It follows that V is a field; and since any field containing $\alpha_1, \ldots, \alpha_r$ must contain these products, V is the smallest field containing $\alpha_1, \ldots, \alpha_r$. Moreover V is a number field since

$$\dim_{\mathbb{Q}} V \leq d_1 \cdots d_r.$$

◄

Definition 2.5 *We denote the smallest field containing* $\alpha_1, \dots, \alpha_r \in \mathbb{C}$ *by* $\mathbb{Q}(\alpha_1, \dots, \alpha_r)$.

Proposition 2.7 *If* α *is an algebraic number of degree* d *then each element* $\gamma \in \mathbb{Q}(\alpha)$ *is uniquely expressible in the form*

$$a_0 + a_1\alpha + \cdots + a_{d-1}\alpha^{d-1} \quad (a_0, a_1, \dots, a_{d-1} \in \mathbb{Q}).$$

Proof ► It follows as in the proof of Proposition 2.6 that these elements do constitute the field $\mathbb{Q}(\alpha)$. And if two of the elements were equal then α would satisfy an equation of degree $< d$, which could be made monic by dividing by the leading coefficient. ◄

A number field of the form $K = \mathbb{Q}(\alpha)$, ie generated by a single algebraic number α, is said to be *simple*. Our next result shows that, surprisingly, every number field is simple. The proof is more subtle than might appear at first sight.

Proposition 2.8 *Every number field* K *can be generated by a single algebraic number:*

$$K = \mathbb{Q}(\alpha).$$

Proof ► It is evident that

$$K = \mathbb{Q}(\alpha_1, \dots, \alpha_r);$$

for if we successively adjoin algebraic numbers

$$\alpha_{i+1} \in K \setminus \mathbb{Q}(\alpha_1, \dots, \alpha_r)$$

then

$$\dim \mathbb{Q}(\alpha_1) < \dim \mathbb{Q}(\alpha_1, \alpha_2) \dim \mathbb{Q}(\alpha_1, \alpha_2, \alpha_3) <$$

and so K must be attained after at most $\dim_{\mathbb{Q}} K$ adjunctions.

Thus it is suffient to prove the result when $r = 2$, ie to show that, for any two algebraic numbers α, β,

$$\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\gamma).$$

Let $p(x)$ be the minimal polynomial of α, and $q(x)$ the minimal polynomial of β. Suppose $\alpha_1 = \alpha, \dots, \alpha_m$ are the conjugates of α and $\beta_1 = \beta, \dots, \beta_n$ the conjugates of β. Let

$$\gamma = \alpha + a\beta,$$

where $a \in \mathbb{Q}$ is chosen so that the mn numbers

$$\alpha_i + a\beta_j$$

are all distinct. This is certainly possible, since

$$\alpha_i + a\beta_j = \alpha_{i'} + a\beta_{j'} \iff a = \frac{\alpha_{i'} - \alpha_i}{\beta_j - \beta_{j'}}.$$

Thus a has to avoid at most $mn(mn-1)/2$ values.

Since

$$\alpha = \gamma - a\beta,$$

and

$$p(\alpha) = 0,$$

β satisfies the equation

$$p(\gamma - ax) = 0.$$

This is a polynomial equation over the field $k = \mathbb{Q}(\gamma)$.

But β also satisfies the equation

$$q(x) = 0.$$

It follows that β satisfies the equation

$$d(x) = \gcd(p(\gamma - ax), q(x)) = 0.$$

Now

$$(x - \beta) \mid d(x)$$

since β is a root of both polynomials. Also, since

$$d(x) \mid q(x) = (x - \beta_1) \cdots (x - \beta_n),$$

$d(x)$ must be the product of certain of the factors $(x - \beta_j)$. Suppose $(x - \beta_j)$ is one such factor. Then β_j is a root of $p(\gamma - ax)$, ie

$$p(\gamma - a\beta_j) = 0.$$

Thus

$$\gamma - a\beta_j = \alpha_i$$

for some i. Hence

$$\gamma = \alpha_i + a\beta_j.$$

But this implies that $i = 1,\ j = 1$, since we chose a so that the elements

$$\alpha_i + a\beta_j$$

were all distinct.

Thus

$$d(x) = (x - \beta).$$

But if $u(x), v(x) \in k[x]$ then we can compute $\gcd(u(x), v(x))$ by the euclidean algorithm without leaving the field k, ie

$$u(x), v(x) \in k[x] \Longrightarrow \gcd(u(x), v(x)) \in k[x].$$

In particular, in our case

$$x - \beta \in k = \mathbb{Q}(\gamma).$$

But this means that

$$\beta \in \mathbb{Q}(\gamma);$$

and so also

$$\alpha = \gamma - a\beta \in \mathbb{Q}(\gamma).$$

Thus

$$\alpha, \beta \in \mathbb{Q}(\gamma) \Longrightarrow \mathbb{Q}(\alpha, \beta) \subset \mathbb{Q}(\gamma) \subset \mathbb{Q}(\alpha, \beta).$$

Hence

$$\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\gamma).$$

◄

2.4 Algebraic integers

Definition 2.6 *A number $\alpha \in \mathbb{C}$ is said to be an* algebraic integer *if it satisfies a polynomial equation*

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n = 0$$

with integral coefficients $a_i \in \mathbb{Z}$. We denote the set of algebraic integers by $\bar{\mathbb{Z}}$.

Proposition 2.9 *The algebraic integers form a ring $\bar{\mathbb{Z}}$ with*

$$\mathbb{Z} \subset \bar{\mathbb{Z}} \subset \bar{\mathbb{Q}}.$$

Proof ► Evidently

$$\mathbb{Z} \subset \bar{\mathbb{Z}},$$

since $n \in \mathbb{Z}$ satisfies the equation

$$x - n = 0.$$

We have to show that

$$\alpha, \beta \in \bar{\mathbb{Z}} \Longrightarrow \alpha + \beta, \alpha\beta \in \bar{\mathbb{Z}}.$$

Lemma 2.2 *The number $\alpha \in \mathbb{C}$ is an algebraic integer if and only if there exists a finitely-generated (but non-zero) additive subgroup $S \subset \mathbb{C}$ such that*

$$\alpha S \subset S.$$

Proof of Lemma ▷ Suppose $\alpha \in \bar{\mathbb{Z}}$; and suppose the minimal polynomial of α is

$$m(x) = x^d + a_1 x^{d-1} + \cdots + a_d,$$

where $a_1, \ldots, a_d \in \mathbb{Z}$. Let S be the abelian group generated by $1, \alpha, \ldots, \alpha^{d-1}$:

$$S = \langle 1, \alpha, \ldots, \alpha^{d-1} \rangle.$$

Then it is readily verified that

$$\alpha S \subset S.$$

Conversely, suppose S is such a subgroup. ◁

If α is a root of the monic polynomial $f(x)$ then $-\alpha$ is a root of the monic polynomial $f(-x)$. It follows that if α is an algebraic integer then so is $-\alpha$. Thus it is sufficient to show that if α, β are algebraic integers then so are $\alpha + \beta, \alpha\beta$.

Suppose α satisfies the equation

$$f(x) \equiv x^m + a_1 x^{m-1} + \cdots + a_m = 0 \quad (a_1, \ldots, a_m \in \mathbb{Z}),$$

and β the equation

$$g(x) \equiv x^n + b_1 x^{n-1} + \cdots + b_n = 0 \quad (b_1, \ldots, b_n \in \mathbb{Z}).$$

Consider the abelian group (or $\mathbb{Z}$-module)

$$M = \langle \alpha^i \beta^j : 0 \le i < m,\ 0 \le j < n \rangle$$

generated by the mn elements $\alpha^i \beta^j$. Evidently

$$\alpha + \beta, \alpha\beta \in V.$$

As a finitely-generated torsion-free abelian group, M is isomorphic to $\mathbb{Z}^d$ for some d. Moreover M is *noetherian*, ie every increasing sequence of subgroups of M is stationary: if

$$S_1 \subset S_2 \subset S_3 \cdots \subset M$$

then for some N,

$$S_N = S_{N+1} = S_{N+2} = \cdots .$$

Suppose $\theta \in M$. Consider the increasing sequence of subgroups

$$\langle 1 \rangle \subset \langle 1, \theta \rangle \subset \langle 1, \theta, \theta^2 \rangle \subset \cdots .$$

This sequence must become stationary; that is to say, for some N

$$\theta^N \in \langle 1, \theta, \ldots, \theta^{N-1} \rangle.$$

In other words, θ satisfies an equation of the form

$$\theta^N = a_1 \theta^{N-1} + a_2 \theta^{N-2} + \cdots .$$

Thus every $\theta \in M$ is an algebraic integer. In particular $\alpha + \beta$ and $\alpha\beta$ are algebraic integers. ◀

Proposition 2.10 *A rational number $c \in \mathbb{Q}$ is an algebraic integer if and only if it is a rational integer:*

$$\bar{\mathbb{Z}} \cap \mathbb{Q} = \mathbb{Z}.$$

Proof ► Suppose $c = m/n$, where $\gcd(m, n) = 1$; and suppose c satisfies the equation

$$x^d + a_1 x^{d-1} + \cdots + a_d = 0 \quad (a_i \in \mathbb{Z}).$$

Then

$$m^d + a_1 m^{d-1} n + \cdots + a_d n^d = 0.$$

Since n divides every term after the first, it follows that $n \mid m^d$. But that is incompatible with $\gcd(m, n) = 1$, unless $n = 1$, ie $c \in \mathbb{Z}$. ◄

Proposition 2.11 *Every algebraic number α is expressible in the form*

$$\alpha = \frac{\beta}{n},$$

where β is an algebraic integer, and $n \in \mathbb{Z}$.

Proof ► Let the minimal polynomial of α be

$$m(x) = x^d + a_1 x^{d-1} + \cdots + a_d,$$

where $a_1, \ldots, a_d \in \mathbb{Q}$. Let the lcm of the denominators of the a_i be n. Then

$$b_i = n a_i \in \mathbb{Z} \quad (1 \leq i \leq d).$$

Now α satisfies the equation

$$n x^d + b_1 x^{d-1} + \cdots + b_d = 0.$$

It follows that

$$\beta = n\alpha$$

satisfies the equation

$$x^d + b_1 x^{d-1} + (n b_2) x^{d-2} + \cdots + (n^{d-1} b_d = 0.$$

Thus β is an integer, as required. ◄

The following result goes in the opposite direction.

Proposition 2.12 *Suppose α is an algebraic integer. Then we can find an algebraic integer $\beta \neq 0$ such that*

$$\alpha\beta \in \mathbb{Z}.$$

Proof ▸ Let the minimal polynomial of α be

$$m(x) = x^d + a_1 x^{d-1} + \cdots + a_d,$$

where $a_1, \ldots, a_d \in \mathbb{Z}$. Recall that the conjugates of α,

$$\alpha_1 = \alpha, \ldots, \alpha_d$$

are the roots of the minimal equation.

Each of these conjugates is an algebraic integer, since its minimal equation $m(x)$ has integer coefficients. Hence

$$\beta = \alpha_2 \cdots \alpha_d$$

is an algebraic integer; and

$$\alpha\beta = \alpha_1\alpha_2 \cdots \alpha_d = \pm a_d \in \mathbb{Z}.$$

◂

2.5 Units

Definition 2.7 *A number $\alpha \in \mathbb{C}$ is said to be a* unit *if both α and $1/\alpha$ are algebraic integers.*

Any root of unity, ie any number satisfying $x^n = 1$ for some n, is a unit.
But these are not the only units; for example, $\sqrt{2} - 1$ is a unit.
The units form a multiplicative subgroup of $\bar{\mathbb{Q}}^\times$.

2.6 The Integral Basis Theorem

Proposition 2.13 *Suppose A is a number ring. Then we can find $\gamma_1, \ldots, \gamma_d \in A$ such that each $\alpha \in A$ is uniquely expressible in the form*

$$\alpha = c_1\gamma_1 + c_d\gamma_d$$

with $c_1, \ldots, c_d \in \mathbb{Z}$.

In other words, as an additive group

$$A \cong \mathbb{Z}^d.$$

We may say that $\gamma_1, \ldots, \gamma_d$ is a $\mathbb{Z}$-*basis* for A.

Proof ▸ Suppose A is the ring of integers in the number field K. By Proposition **??**,

$$K = \mathbb{Q}(\alpha).$$

By Proposition 2.12,
$$\alpha = \frac{\beta}{m},$$
where $\beta \in \bar{\mathbb{Z}},\ m \in \mathbb{Z}$. Since
$$\mathbb{Q}(\beta) = \mathbb{Q}(\alpha),$$
we may suppose that α is an integer.

Let
$$m(x) = x^d + a_1 x^{d-1} + \cdots + a_d$$
be the minimal polynomial of α; and let
$$\alpha_1 = \alpha, \ldots, \alpha_d$$
be the roots of this polynomial, ie the conjugates of α.

Note that these conjugates satisfy exactly the same set of polynomials over $\mathbb{Q}$; for
$$p(\alpha) = 0 \iff m(x) \mid p(x) \iff p(\alpha_i) = 0.$$

Now suppose $\beta \in A$. Then
$$\beta = b_0 + b_1\alpha + \cdots b_{d-1}\alpha^{d-1},$$
where $b_0, \ldots, b_{d-1} \in \mathbb{Q}$, say
$$\beta = f(\alpha)$$
with $f(x) \in \mathbb{Q}[x]$.

Let
$$\beta_i = b_0 + b_1\alpha_i + \cdots b_{d-1}\alpha_i^{d-1}$$
for $i = 1, \ldots, d$.

Each β_i satisfies the same set of polynomials over $\mathbb{Q}$ as β. for
$$p(\beta) = 0 \iff p(f(\alpha)) = 0 \iff p(f(\alpha_i)) = 0 \iff p(\beta_i) = 0.$$
In particular, each β_i has the same minimal polynomial as β, and so each β_i is an integer.

We may regard the formulae for the β_i as linear equations for the coefficients $b_0, \ldots, b_{d-1}$:
$$\begin{gathered} b_0 + \alpha_1 b_1 + \cdots \alpha^{d-1} b_{d-1} = \beta_1, \\ \ldots \\ b_0 + \alpha_d b_1 + \cdots \alpha_d^{d-1} b_{d-1} = \beta_d. \end{gathered}$$
We can write this as a matrix equation
$$D \begin{pmatrix} b_0 \\ \vdots \\ b_{d-1} \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix}$$

where D is the matrix

$$D = \begin{pmatrix} 1 & \alpha_1 & \dots & \alpha_1^{d-1} \\ \vdots & \dots & \dots & \vdots \\ 1 & \alpha_d & \dots & \alpha_d^{d-1}. \end{pmatrix}$$

By a familiar argument,

$$\det \begin{pmatrix} 1 & x_1 & \dots & x_1^{d-1} \\ \vdots & \dots & \dots & \vdots \\ 1 & x_d & \dots & x_d^{d-1} \end{pmatrix} = \prod_{i<j} (x_i - x_j).$$

(The determinant vanishes whenever $x_i = x_j$ since then two rows are equal. Hence $(x_i - x_j)$ is a factor for each pair i, j; from which the result follows on comparing degrees and leading coefficients.)

Thus

$$\det D = \prod_{i<j} (\alpha_i - \alpha_j).$$

In particular, $\det D$ is an integer.

On solving the equations for $b_0, \dots, b_{d-1}$ by Cramer's rule, we deduce that

$$b_i = \frac{\beta_i}{\det D},$$

where β_i is a co-factor of the matrix D, and so a polynomial in $\alpha_1, \dots, \alpha_d$ with coefficients in $\mathbb{Z}$, and therefore an algebraic integer.

By Proposition 2.12, we can find an integer δ such that

$$\delta \det D = n \in \mathbb{Z},$$

where we may suppose that $n > 0$. Thus each b_i is expressible in the form

$$b_i = \frac{\gamma_i}{n},$$

where

$$\gamma_i \in \bar{\mathbb{Z}} \cap \mathbb{Q} = \mathbb{Z}.$$

In other words, each $\beta \in A$ is expressible in the form

$$\beta = c_o\delta_0 + \dots + c_{d-1}\delta_{d-1},$$

where

$$\delta_i = \frac{\alpha^i}{n}$$

and

$$c_i \in \mathbb{Z} \quad (0 \le i < d).$$

The elements

$$c_o\delta_0 + \dots + c_{d-1}\delta_{d-1} \quad (c_i \in \mathbb{Z})$$

form a finitely-generated and torsion-free abelian group C, of rank d; and A is a subgroup of C of finite index. We need the following standard result from the theory of finitely-generated abelian groups.

Lemma 2.3 *If*

$$S \subset \mathbb{Z}^d$$

is a subgroup of finite index then

$$S \cong \mathbb{Z}^d$$

Proof of Lemma ▷ We have to construct a $\mathbb{Z}$-index for S. We argue by induction on d.

Choose an element

$$e = (e_1, \dots, e_d) \in S$$

with least positive last coordinate e_d. Suppose

$$s = (s_1, \dots, s_d) \in S.$$

Then

$$s_d = qe,$$

or we could find an element of S with smaller last coordinate. Thus

$$s - qe = (t_1, \dots, t_{d-1}, 0).$$

Hence

$$S = \mathbb{Z}e \oplus T,$$

where

$$T = S \cap \mathbb{Z}^{d-1}$$

(identifying Z^{d-1} with the subgroup of $\mathbb{Z}^d$ formed by the d-tuples with last coordinate 0).

The result follows on applying the inductive hypothesis to T. ◁

The Proposition follows on applying the Lemma to

$$A \subset C \cong \mathbb{Z}^d.$$

◀

2.7 Unique factorisation in number rings

As we saw in Chapter 1, a principal ideal domain is a unique factorisation domain. The converse is not true; there is unique factorisation in $\mathbb{Z}[x]$, but the ideal $\langle 2, x\rangle$ is not principal. Our main aim in this Section is to show that the converse *does* hold for number rings A:

$$A \text{ principal ideal domain} \iff A \text{ unique factorisation domain}.$$

We suppose throughout the Section that A is a number ring, ie the ring of integers in a number field K.

Proposition 2.14 *Suppose $\mathfrak{a} \subset A$ is a non-zero ideal. Then the quotient-ring*

$$A/\mathfrak{a}$$

is finite.

Proof ► Take $\alpha \in \mathfrak{a},\ \alpha \neq 0$. By Proposition 1.8, we can find $\beta \in A,\ \beta \neq 0$ such that

$$a = \alpha\beta \in \mathbb{Z}.$$

We may suppose that $a > 0$. Then

$$\langle a \rangle \subset \langle \alpha \rangle \subset \mathfrak{a}.$$

Thus

$$\alpha \equiv \beta \bmod a \Longrightarrow \alpha \equiv \beta \bmod \mathfrak{a}.$$

By Proposition **??**, A has an integral basis $\gamma_1, \ldots, \gamma_d$, ie each $\alpha \in A$ is (uniquely) expressible in the form

$$\alpha = c_1\gamma_1 + c_d\gamma_d$$

with $c_1, \ldots, c_d \in \mathbb{Z}$. It follows that α is congruent $\bmod a$ to one of the numbers

$$r_1\gamma_1 + r_d\gamma_d \quad (0 \leq r_i < a).$$

Thus

$$\|A/\langle a \rangle\| = a^d.$$

Hence

$$\|A/\mathfrak{a}\| \leq a^d.$$

◄

Proposition 2.15 *The number ring A is a unique factorisation domain if and only if it is a principal ideal domain.*

Proof ► We know from Chapter 1 that

$$A \text{ principal ideal domain} \Longrightarrow A \text{ unique factorisation domain}.$$

We have to proce the converse.

Let us suppose therefore that the number ring A is a unique factorisation domain.

Lemma 2.4 *Suppose*

$$\alpha = \epsilon \pi_1^{e_1} \cdots \pi_r^{e_r}, \quad \beta = \epsilon' \pi_1^{f_1} \cdots \pi_r^{f_r}.$$

Let

$$\delta = \pi_1^{\min(e_1,f_1)} \cdots \pi_r^{\min(e_r,f_r)}.$$

Then

$$\delta = \gcd(\alpha, \beta)$$

in the sense that

$$\delta \mid \alpha,\ \delta \mid \beta \quad \textit{and} \quad \delta' \mid \alpha,\ \delta \mid \beta \Longrightarrow \delta' \mid \delta.$$

Proof of Lemma ▷ This follows at once from unique factorisation. ◁

Lemma 2.5 *If*

$$\beta_1 \equiv \beta_2 \bmod \alpha$$

then

$$\gcd(\alpha, \beta_1) = \gcd(\alpha, \beta_2).$$

Proof of Lemma ▷ It is readily verified that if

$$\beta_1 = \beta_2 + \alpha\gamma$$

then

$$\delta \mid \alpha,\ \beta_1 \Longleftrightarrow \delta \mid \alpha,\ \beta_2.$$

◁

We say that α, β are *coprime* if

$$\gcd(\alpha, \beta) = 1.$$

It follows from the Lemma that we may speak of a congruence class $\bar{\beta} \bmod \alpha$ being coprime to α.

Lemma 2.6 *The congruence classes* $\bmod \alpha$ *coprime to* α *form a multiplicative group*

$$(A/\langle\alpha\rangle)^\times.$$

Proof of Lemma ▷ We have

$$\gcd(\alpha, \beta_1\beta_2) = 1 \Longleftrightarrow \gcd(\alpha, \beta_1) = 1,\ \gcd(\alpha, \beta_2) = 1.$$

Thus $(A/\langle\alpha\rangle)^\times$ is closed under multiplication; and if β is coprime to α then the map

$$\bar{\gamma} \mapsto \bar{\beta}\bar{\gamma} : (A/\langle\alpha\rangle)^\times \to (A/\langle\alpha\rangle)^\times$$

is injective, and so surjective since $A/\langle\alpha\rangle$ is finite. Hence $(A/\langle\alpha\rangle)^\times$ is a group. ◁

Lemma 2.7 *Suppose*

$$\gcd(\alpha, \beta) = \delta.$$

Then we can find $u, v \in A$ *such that*

$$\alpha u + \beta v = \delta.$$

Proof of Lemma $\triangleright$ We may suppose, on dividing by δ, that

$$\gcd(\alpha, \beta) = 1,$$

and so

$$\bar{\beta} \in (A/\langle\alpha\rangle)^{\times}.$$

Since this group is finite,

$$\bar{\beta}^n = 1$$

for some $n > 0$. In other words,

$$\beta^n \equiv 1 \bmod \alpha,$$

ie

$$\beta^n = 1 + \alpha\gamma,$$

ie

$$\alpha u + \beta v = 1$$

with $u = -\gamma,\ v = \beta^{n-1}$. $\triangleleft$

We can extend the definition of $\gcd$ to any set (finite or infinite) of numbers

$$\alpha_i \in A \quad (i \in I).$$

and by repeated application of the last Lemma we can find β_i (all but a finite number equal to 0) such that

$$\sum_{i \in I} \alpha_i \beta_i = \gcd_{i \in I}(\alpha_i).$$

Applying this to the ideal $\mathfrak{a}$, let

$$\delta = \gcd_{\alpha \in \mathfrak{a}}(\alpha).$$

Then

$$\delta = \sum \alpha_i \beta_i \in \mathfrak{a};$$

and so

$$\mathfrak{a} = \langle\delta\rangle.$$

◄

Chapter 3

Quadratic Number Fields

3.1 The fields $\mathbb{Q}(\sqrt{m})$

Definition 3.1 *A* quadratic field *is a number field of degree 2.*

Recall that this means the field k has dimension 2 as a vector space over $\mathbb{Q}$:

$$\dim_{\mathbb{Q}} k = 2.$$

Definition 3.2 *The integer $m \in \mathbb{Z}$ is said to be* square-free *if*

$$m = r^2 s \Longrightarrow r = \pm 1.$$

Thus

$$\pm 1, \pm 2, \pm 3, \pm 5, \pm 6, \pm 7, \pm 10, \pm 11, \pm 13, \ldots$$

are square-free.

Proposition 3.1 *Each quadratic field is of the form $\mathbb{Q}(\sqrt{m})$ for a unique square-free integer $m \neq 1$.*

Recall that $\mathbb{Q}(\sqrt{m})$ consists of the numbers

$$x + y\sqrt{m} \quad (x, y \in \mathbb{Q}).$$

Proof ► Suppose k is a quadratic field. Let $\alpha \in k \setminus \mathbb{Q}$. Then $\alpha^2, \alpha, 1$ are linearly dependent over $\mathbb{Q}$, since $\dim_{\mathbb{Q}} k = 2$. In other words, α satisfies a quadratic equation

$$a_0 \alpha^2 + a_1 \alpha + a_2 = 0$$

with $a_0, a_1, a_2 \in \mathbb{Q}$. We may assume that $a_0, a_1, a_2 \in \mathbb{Z}$. Then

$$\alpha = \frac{-a_1 + \sqrt{a_1^2 - 4a_0 a_2}}{2a_0}$$

Thus
$$\sqrt{a_1^2 - 4a_0a_2} = 2a_0\alpha + a_1 \in k.$$
Let
$$a_1^2 - 4a_0a_2 = r^2 m$$
where m is square-free. Then
$$\sqrt{m} = \frac{1}{r}\sqrt{a_1^2 - 4a_0a_2} \in k.$$
Thus
$$\mathbb{Q} \subset \mathbb{Q}(\sqrt{m}) \subset k.$$
Since $\dim_{\mathbb{Q}} k = 2$,
$$k = \mathbb{Q}(\sqrt{m}).$$

To see that different square-free integers m_1, m_2 give rise to different quadratic fields, suppose
$$m_1 \in \mathbb{Q}(\sqrt{m_2}),$$
say
$$m_1 = x + y\sqrt{m_2} \quad (x, y \in \mathbb{Q})$$
Squaring,
$$x_1 = x^2 + m_2 y^2 + 2xy\sqrt{m_2}.$$
Thus either $x = 0$ or $y = 0$ or
$$\sqrt{m_2} \in \mathbb{Q},$$
all of which are absurd. ◂

When we speak of the quadratic field $\mathbb{Q}(\sqrt{m})$ it is understood that m is a square-free integer $\neq 1$.

Definition 3.3 *The quadratic field* $\mathbb{Q}(\sqrt{m})$ *is said to be* real *if* $m > 0$*, and* imaginary *if* $m < 0$.

This is a natural definition since it means that $\mathbb{Q}(\sqrt{m})$ is real if and only if
$$\mathbb{Q}(\sqrt{m}) \subset \mathbb{R}.$$

3.2 Conjugates and norms

Proposition 3.2 *The map*

$$x + y\sqrt{m} \mapsto x - y\sqrt{m}$$

is an automorphism of $\mathbb{Q}(\sqrt{m})$*; and it is the only such automorphism apart from the identity map.*

Proof ▸ The map clearly preserves addition. It also preserves multiplication, since

$$(x + y\sqrt{m})(u + v\sqrt{m} = (xu + yvm) + (xv + yu)\sqrt{m},$$

and so

$$(x - y\sqrt{m})(u - v\sqrt{m} = (xu + yvm) - (xv + yu)\sqrt{m}.$$

Since the map is evidently bijective, it is an automorphism.

Conversely, if θ is an automorphism of $\mathbb{Q}(\sqrt{m})$ then θ preserves the elements of $\mathbb{Q}$; in fact if $\alpha \in \mathbb{Q}(\sqrt{m})$ then

$$\theta(\alpha) = \alpha \Longleftrightarrow \alpha \in \mathbb{Q}.$$

Thus

$$\theta(\sqrt{m})^2 = \theta(m) = m \Longrightarrow \theta(\sqrt{m}) = \pm\sqrt{m},$$

giving the identity automorphism and the automorphism above. ◂

Definition 3.4 *If*

$$\alpha = x + y\sqrt{m} \quad (x, y \in \mathbb{Q})$$

then we write

$$\bar{\alpha} = x - y\sqrt{m} \quad (x, y \in \mathbb{Q})$$

and we call $\bar{\alpha}$ *the* conjugate *of* α*.*

Note that if $\mathbb{Q}(\sqrt{m})$ is imaginary (ie $m < 0$) then the conjugate $\bar{\alpha}$ coincides with the usual complex conjugate.

Definition 3.5 *We define the norm* $\mathcal{N}(\alpha)$ *of* $\alpha \in \mathbb{Q}(\sqrt{m})$ *by*

$$\mathcal{N}(\alpha) = \alpha\bar{\alpha}.$$

Thus if

$$\alpha = x + y\sqrt{m} \quad (x, y \in \mathbb{Q})$$

then

$$\mathcal{N}(\alpha) = (x + y\sqrt{m})(x - y\sqrt{m}) = x^2 - my^2.$$

Proposition 3.3 *1.* $\mathcal{N}(\alpha) \in \mathbb{Q}$*;*

2. $\mathcal{N}(()\alpha = 0 \Longleftrightarrow \alpha = 0$;

3. $\mathcal{N}(\alpha\beta) = \mathcal{N}(\alpha)\mathcal{N}(\beta)$;

4. *If* $a \in \mathbb{Q}$ *then* $\mathcal{N}(a) = a^2$;

5. *If* $m < 0$ *then* $\mathcal{N}(\alpha) \geq 0$.

Proof ► All is clear except perhaps the third part, where

$$\begin{aligned}\mathcal{N}(\alpha\beta) &= (\alpha\beta)(\overline{\alpha\beta}) \\ &= (\alpha\beta)(\bar{\alpha}\bar{\beta}) \\ &= (\alpha\bar{\alpha})(\beta\bar{\beta}) \\ &= \mathcal{N}(\alpha)\mathcal{N}(\beta).\end{aligned}$$

◄

3.3 Integers

Proposition 3.4 *Suppose* $k = \mathbb{Q}(\sqrt{m})$, *where* $m \neq 1$ *is square-free.*

1. *If* $m \not\equiv 1 \bmod 4$ *then the integers in* k *are the numbers*

$$a + b\sqrt{m},$$

   where $a, b \in \mathbb{Z}$.

2. *If* $m \equiv 1 \bmod 4$ *then the integers in* k *are the numbers*

$$\frac{a}{2} + \frac{b}{2}\sqrt{m},$$

   where $a, b \in \mathbb{Z}$ *and*

$$a \equiv b \bmod 2,$$

   ie a, b *are either both even or both odd.*

Proof ► Suppose

$$\alpha = a + b\sqrt{m} \quad (b \in \mathbb{Q})$$

is an integer. Recall that an algebraic number α is an integer if and only if its minimal polynomial has integer coefficients. If $y = 0$ the minimal polynomial of α is $x - a$. Thus $\alpha = a$ is in integer if and only if $a \in \mathbb{Z}$ (as we know of course since $\bar{Z} \cap \mathbb{Q} = \mathbb{Z}$).

If $y \neq 0$ then the minimal polynomial of α is

$$(x - a)^2 - mb^2 = x^2 - 2ax + (a^2 - mb^2).$$

Thus α is an integer if and only if

$$2a \in \mathbb{Z} \quad \text{and} \quad a^2 - mb^2 \in \mathbb{Z}.$$

Suppose $2a = A$, ie

$$a = \frac{A}{2}.$$

Then

$$\begin{aligned} 4a^2 \in \mathbb{Z},\ a^2 - mb^2 \in \mathbb{Z} &\implies 4mb^2 \in \mathbb{Z} \\ &\implies 4b^2 \in \mathbb{Z} \\ &\implies 2b \in \mathbb{Z} \end{aligned}$$

since m is square-free. Thus

$$b = \frac{B}{2},$$

where $B \in \mathbb{Z}$.

Now

$$a^2 - mb^2 = \frac{A^2 - mB^2}{4} \in \mathbb{Z},$$

ie

$$A^2 - mB^2 \equiv 0 \bmod 4.$$

If A is even then

$$2 \mid A \implies 4 \mid A^2 \implies 4 \mid mB^2 \implies 2 \mid B^2 \implies 2 \mid B;$$

and similarly

$$2 \mid B \implies 4 \mid B^2 \implies 4 \mid A^2 \implies 2 \mid A.$$

Thus A, B are either both even, in which case $a, b \in \mathbb{Z}$, or both odd, in which case

$$A^2, B^2 \equiv 1 \bmod 4,$$

so that

$$1 - m \equiv 0 \bmod 4,$$

ie

$$m \equiv 1 \bmod 4.$$

Conversely if $m \equiv 1 \bmod 4$ then

$$\begin{aligned} A, B \text{ odd } &\implies A^2 - mB^2 \equiv 0 \bmod 4 \\ &\implies a^2 - mb^2 \in \mathbb{Z}. \end{aligned}$$

◀

It is sometimes convenient to express the result in the following form.

Corollary 3.1 *Let*

$$\omega = \begin{cases} \sqrt{m} & \text{if } m \not\equiv 1 \bmod 4, \\ \frac{1+\sqrt{m}}{2} & \text{if } m \equiv 1 \bmod 4. \end{cases}$$

Then the integers in $\mathbb{Q}(\sqrt{m})$ *form the ring* $\mathbb{Z}[\omega]$.

Examples:

1. The integers in the gaussian field $\mathbb{Q}(i)$ are the gaussian integers

$$a + bi \qquad (a, b \in \mathbb{Z})$$

2. The integers in $\mathbb{Q}(\sqrt{2})$ are the numbers

$$a + b\sqrt{2} \qquad (a, b \in \mathbb{Z}).$$

3. The integers in $\mathbb{Q}(\sqrt{-3})$ are the numbers

$$a + b\omega \qquad (a, b \in \mathbb{Z})$$

where

$$\omega = \frac{1+\sqrt{-3}}{2}.$$

Proposition 3.5 *If* $\alpha \in \mathbb{Q}(\sqrt{m})$ *is an integer then*

$$\mathcal{N}(\alpha) \in \mathbb{Z}.$$

Proof ► If α is an integer then so is its conjugate $\bar{\alpha}$ (since $\alpha, \bar{\alpha}$ satisfy the same polynomial equations over $\mathbb{Q}$). Hence

$$\mathcal{N}(\alpha) \in \bar{\mathbb{Z}} \cap \mathbb{Q} = \mathbb{Z}.$$

◄

3.4 Units

Proposition 3.6 *An integer* $\epsilon \in \mathbb{Q}(\sqrt{m})$ *is a unit if and only if*

$$\mathcal{N}(\epsilon) = \pm 1.$$

Proof ► Suppose ϵ is a unit, say

$$\epsilon\eta = 1.$$

Then

$$\mathcal{N}(\epsilon)\mathcal{N}(\eta) = \mathcal{N}(1) = 1.$$

Hence

$$\mathcal{N}(\epsilon) = \pm 1.$$

Conversely, suppose

$$\mathcal{N}(\epsilon) = \pm 1,$$

ie

$$\epsilon\bar{\epsilon} = \pm 1.$$

Then

$$\epsilon^{-1} = \pm\bar{\epsilon}$$

is an integer, ie ϵ is a unit. ◄

Proposition 3.7 *An imaginary quadratic number field contains only a finite number of units.*

1. *The units in* $\mathbb{Q}(i)$ *are* $\pm 1, \pm i$*;*
2. *The units in* $\mathbb{Q}(\sqrt{-3})$ *are* $\pm 1, \pm\omega, \pm\omega^2$*, where* $\omega = (1 + \sqrt{-3})/2$*.*
3. *In all other cases the imaginary quadratic number field* $\mathbb{Q}(\sqrt{m})$ *(where* $m < 0$*) has just two units,* ± 1*.*

Proof ► We know of course that ± 1 are always units.

Suppose

$$\epsilon = a + b\sqrt{m}$$

is a unit. Then

$$N)\epsilon) = a^2 + (-m)b^2 = 1$$

by Proposition 3.6. In particular

$$(-m)b^2 \leq 1.$$

If $m \equiv 3 \bmod 4$ then $a, b \in \mathbb{Z}$; and so $b = 0$ unless $m = -1$ in which case $b = \pm 1$ is a solution, giving $a = 0$, ie $\epsilon = \pm i$.

If $m \equiv 1 \bmod 4$ then b may be a half-integer, ie $b = B/2$, and

$$(-m)b^2 = (-m)B^2/4 > 1$$

if $B \neq 0$, unless $m = -3$ and $B = \pm 1$, in which case $A = \pm 1$. Thus we get four additional units in $\mathbb{Q}(\sqrt{-3})$, namely $\pm\omega, \pm\omega^2$. ◄

Proposition 3.8 *Every real quadratic number field $\mathbb{Q}(\sqrt{m})$ (where $m > 0$) contains an infinity of units. More precisely, there is a unique unit $\eta > 1$ such that the units are the numbers*

$$\pm\eta^n \qquad (n \in \mathbb{Z})$$

Proof ► The following exercise in the pigeon-hole principle is due to Kronecker.

Lemma 3.1 *Suppose $\alpha \in \mathbb{R}$. There are an infinity of integers m, n with $m > 0$ such that*

$$|m\alpha - n| < \frac{1}{n}.$$

Proof of Lemma ▷ Let $\{x\}$ denote the fractional part of $x \in \mathbb{R}$. Thus

$$\{x\} = x - [x],$$

where $[x]$ is the integer part of x.

Suppose N is a positive integer. Let us divide $[0, 1)$ into N equal parts:

$$[0, 1/N), [1/N, 2/N), \dots, [(N-1)/N, 1).$$

Consider how the $N + 1$ fractional parts

$$\{0\}, \{\alpha\}, \{2\alpha\}, \dots, \{N\alpha\}$$

fall into these N divisions.

Two of the fractional parts — say $\{r\alpha\}$ and $\{s\alpha\}$, where $r < s$ — must fall into the same division. But then

$$|\{s\alpha\} - \{r\alpha\}| < 1/N,$$

ie

$$|(s\alpha - [s\alpha]) - (r\alpha - [r\alpha])| < N.$$

Let

$$m = s - r,\ n = [s\alpha] - [r\alpha].$$

Then

$$|m\alpha - n| < 1/N \leq 1/m.$$

◁

Lemma 3.2 *There are an infinity of $a, b \in \mathbb{Z}$ such that*

$$|a^2 - b^2 m| < 2\sqrt{m} + 1.$$

Proof of Lemma ▷ We apply Kronecker's Lemma above with $\alpha = \sqrt{m}$. There are an infinity of integers $a, b > 0$ such that

$$|a - b\sqrt{m}| < 1/b.$$

But then

$$a < b\sqrt{m} + 1,$$

and so

$$a + b\sqrt{m} < 2b\sqrt{m} + 1$$

Hence

$$\begin{aligned} |a^2 - b^2 m| &= (a + b\sqrt{m})|a - b\sqrt{m}| \\ &< (2b\sqrt{m} + 1)/b \\ &\leq 2\sqrt{m} + 1. \end{aligned}$$

◁

It follows from this lemma that there are an infinity of integer solutions of

$$a^2 - b^2 m = d$$

for some

$$d < 2\sqrt{m} + 1.$$

But then there must be an infinity of these solutions (a, b) with the same remainders $\bmod d$.

Lemma 3.3 *Suppose*

$$\alpha_1 = a_1 + b_1\sqrt{m},\ \alpha_2 = a_2 + b_2\sqrt{m},$$

where

$$a_1^2 - b_1^2 = d = a_2^2 - b_2^2$$

and

$$a_1 \equiv a_2 \bmod d, \quad b_1 \equiv b_2 \bmod d.$$

Then

$$\frac{\alpha_1}{\alpha_2}$$

is an algebraic integer.

Proof of Lemma ▷ Suppose

$$a_2 = a_1 + mr,\ b_2 = b_1 + ms.$$

Then

$$\alpha_2 = \alpha_1 + d\beta,$$

where

$$\beta = r + s\sqrt{m}.$$

Hence

$$\begin{aligned}\frac{\alpha_1}{\alpha_2} &= \frac{\alpha_1\bar{\alpha}_2}{\alpha_2\bar{\alpha}_2}\\ &= \frac{\alpha_1\bar{\alpha}_2}{d}\\ &= \frac{\alpha_1(\bar{\alpha}_1 + d\bar{\beta})}{d}\\ &= \frac{\alpha_1\bar{\alpha}_1}{d} + \bar{\beta}\\ &= \frac{d}{d} + \beta\\ &= 1 + \beta,\end{aligned}$$

which is an integer. $\triangleleft$

Now suppose $(a_1, b_1), (a_2, b_2)$ are two such solutions. Then

$$\epsilon = \frac{\alpha_1}{\alpha_2}$$

is an integer, and

$$\mathcal{N}(\epsilon) = \frac{\mathcal{N}(\alpha_1)}{\mathcal{N}(\alpha_2)} = \frac{d}{d} = 1.$$

Hence ϵ is a unit, by Proposition 3.6.

Since there are an infinity of integers α satisfying these conditions, we obtain an infinity of units if we fix α_1 and let α_2 vary. In particular there must be a unit

$$\epsilon \neq \pm 1.$$

Just one of the four units

$$\pm\epsilon,\ \pm\epsilon^{-1}$$

must lie in the range $(1, \infty)$. (The others are distributes one each in the ranges $(-\infty, -1)$, $(-1, 0)$ and $(0, 1)$.)

Suppose then that

$$\epsilon = a + b\sqrt{m} > 1.$$

Then

$$|\epsilon^{-1}| < 1,$$

and so

$$\bar{\epsilon} = \pm\epsilon^{-1} \in (-1, 1),$$

ie

$$-1 < a - b\sqrt{m} < 1.$$

Adding these two inequalities,

$$0 < 2a,$$

ie

$$a > 0.$$

On the other hand,

$$\epsilon > \bar{\epsilon} \Longrightarrow b > 0.$$

It follows that there can only be a finite number of units in any range

$$1 < \epsilon \leq c.$$

In particular, if $\epsilon > 1$ is a unit, then there is a smallest unit η in the range

$$1 < \eta \leq \epsilon.$$

Evidently η is the least unit in the range

$$1 < \eta.$$

Now suppose ϵ is a unit $\neq \pm 1$. As we observed, one of the four units $\pm\epsilon, \pm\epsilon^{-1}$ must lie in the range $(1, \infty)$. We can take this in place of ϵ, ie we may assume that

$$\epsilon > 1.$$

Since $\eta^n \to \infty$,

$$\eta^r \leq \epsilon < \eta^{r+1}$$

for some $r \geq 1$. Hence

$$1 \leq \epsilon\eta^{-r} < \eta.$$

Since η is the smallest unit > 1, this implies that

$$\epsilon\eta^{-1} = 1,$$

ie

$$\epsilon = \eta^r.$$

◄

3.5 Unique factorisation

Suppose A is an integral domain. Recall that if A is a *principal ideal domain*, ie each ideal $\mathcal{A} \subset A$ can be generated by a single element a,

$$\mathfrak{a} = \langle a \rangle,$$

then A is a *unique factorisation domain*, ie each $a \in A$ is uniquely expressible — up to order, and equivalence of primes — in the form

$$a = \epsilon \pi_1^{e_1} \cdots \pi_r^{e_r},$$

where ϵ is a unit, and $\pi_1, \ldots, \pi_r$ are inequivalent primes.

We also showed that if A is the ring of integers in an algebraic number field k then the converse is also true, ie

$$A \text{ principal ideal domain } \iff A \text{ unique factorisation domain }.$$

Proposition 3.9 *The ring of integers $\mathbb{Z}[\omega]$ in the quadratic field $\mathbb{Q}(\sqrt{m}$ is a principal ideal domain (and so a unique factorisation domain) if*

$$m = -11, -7, -3, -2, -1, 2, 3, 5, 13.$$

Proof ► We take

$$|\mathcal{N}(\alpha)|$$

as a measure of the size of $\alpha \in \mathbb{Z}[\omega]$.

Lemma 3.4 *Suppose $\alpha, \beta \in \mathbb{Z}[\omega[$, with $\beta \neq 0$. Then there exist $\gamma, \rho \in \mathbb{Z}[\omega]$ such that*

$$\alpha = \beta\gamma + \rho$$

with

$$|\mathcal{N}(\rho)| < |\mathcal{N}(\beta)|.$$

In other words, we can divide α by β, and get a remainder ρ smaller than β.

Proof of Lemma ▷ Let

$$\frac{\alpha}{\beta} = x + y\sqrt{m}$$

where $x, y \in \mathbb{Q}$.

Suppose first that $m \not\equiv 1 \bmod 4$. We can find integers a, b such that

$$|x - a|,\ |y - b| \leq \frac{1}{2}.$$

Let

$$\gamma = a + b\sqrt{m}.$$

Then $\gamma \in \mathbb{Z}[\omega]$; and

$$\frac{\alpha}{\beta} - \gamma = (x - a) + (y - b)\sqrt{m}.$$

Thus

$$\mathcal{N}(\frac{\alpha}{\beta} - \gamma) = (x - a)^2 - m(y - b)^2.$$

If now $m < 0$ then

$$0 \leq \mathcal{N}(\frac{\alpha}{\beta} - \gamma) \leq \frac{1 + m}{4},$$

yielding

$$|\mathcal{N}(\frac{\alpha}{\beta} - \gamma)| < 1$$

if $m = -2$ or -1; while if $m > 0$ then

$$-\frac{m}{4} \leq \mathcal{N}(\frac{\alpha}{\beta} - \gamma) \leq \frac{1}{4},$$

yielding

$$|\mathcal{N}(\frac{\alpha}{\beta} - \gamma)| < 1$$

if $m = 2$ or 3.

On the other hand, if $m \equiv 1 \bmod 4$ then we can choose a, b to be integers or half-integers. Thus we can choose b so that

$$\mathcal{N}(y - b) \leq \frac{1}{4};$$

and then we can choose a so that

$$\mathcal{N}(x - a) \leq \frac{1}{2}.$$

(Note that a must be an integer or half-integer according as b is an integer or half-integer; so we can only choose a to within an integer.)

If $m < 0$ this gives

$$0 \leq \mathcal{N}(\frac{\alpha}{\beta} - \gamma) \leq \frac{4 + m}{16},$$

yielding

$$|\mathcal{N}(\frac{\alpha}{\beta} - \gamma)| < 1$$

if $m = -11, -7$ or -3; while if $m > 0$ then

$$-\frac{m}{16} \leq \mathcal{N}(\frac{\alpha}{\beta} - \gamma) \leq \frac{1}{4},$$

yielding
$$|\mathcal{N}(\frac{\alpha}{\beta} - \gamma)| < 1$$
if $m = 5$ or 13.

Thus in all the cases listed we can find $\gamma \in \mathbb{Z}[\omega]$ such that
$$|\mathcal{N}(\frac{\alpha}{\beta} - \gamma)| < 1$$
Multiplying by β,
$$|\mathcal{N}(\alpha - \beta\gamma)| < |\mathcal{N}(\beta)|,$$
which gives the required result on setting
$$\rho = \alpha - \beta\gamma,$$
ie
$$\alpha = \beta\gamma + \rho.$$
$\triangleleft$

Now suppose $\mathfrak{a} \neq 0$ is an ideal in $\mathbb{Z}[\omega]$. Let $\alpha \in \mathfrak{a}$ ($\alpha \neq 0$) be an element minimising $|\mathcal{N}(\alpha)|$. (Such an element certainly exists, since $|\mathcal{N}(\alpha)|$ is a positive integer.)

Now suppose $\beta \in \mathfrak{a}$. By the lemma we can find $\gamma, \rho \in \mathbb{Z}[\omega]$ such that
$$\beta = \alpha\gamma + \rho$$
with
$$|\mathcal{N}(\rho)| < |\mathcal{N}(\alpha)|.$$
But
$$\rho = \beta - \alpha\gamma \in \mathfrak{a}.$$
Thus by the minimality of $|\mathcal{N}(\alpha)|$,
$$\begin{aligned} \mathcal{N}(\alpha) = 0 &\Longrightarrow \rho = 0 \\ &\Longrightarrow \beta = \alpha\gamma \\ &\Longrightarrow \beta \in \langle \alpha \rangle. \end{aligned}$$
Hence
$$\mathfrak{a} = \langle \alpha \rangle.$$
◀

Remarks:

1. We do not claim that these are the *only* cases in which $\mathbb{Q}(\sqrt{m})$ — or rather the ring of integers in this field — is a unique factorisation domain. There are certainly other m for which it is known to hold; and in fact is not known if the number of such m is finite or infinite. But the result is easily established for the m listed above.

2. On the other hand, unique factorisation fails in many quadratic fields. For example, if $m = -5$ then

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

Now 2 is irreducible in $\mathbb{Z}[\sqrt{5}]$, since

$$a^2 + 5b^2 = 2$$

has no solution in integers. Thus if there were unique factorisation then

$$2 \mid 1 + \sqrt{-5} \quad \text{or} \quad 2 \mid 1 - \sqrt{-5},$$

both of which are absurd.

As an example of a real quadratic field in which unique factorisation fails, consider $m = 10$. We have

$$6 = 2 \cdot 3 = (4 + \sqrt{10})(4 - \sqrt{10})$$

The prime 2 is again irreducible; for

$$a^2 - 10b^2 = \pm 2$$

has no solution in integers, since neither ± 2 is a quadratic residue $\bmod 10$. (The quadratic residues $\bmod 10$ are $0, \pm 1, \pm 4, 5$.) Thus if there were unique factorisation we would have

$$2 \mid 4 + \sqrt{10} \quad \text{or} \quad 2 \mid 4 - \sqrt{10},$$

both of which are absurd.

3.6 The splitting of rational primes

Throughout n this section we shall assume that *the integers $\mathbb{Z}[\omega]$ in $\mathbb{Q}(\sqrt{m})$ form a principal ideal domain* (and so a unique factorisation domain).

Proposition 3.10 *Let $p \in \mathbb{N}$ be a rational prime. Then p either remains a prime in $\mathbb{Z}[\omega]$, or else*

$$p = \pm \pi \bar{\pi},$$

where π is a prime in $Z[\omega]$. In other words, p has either one or two prime factors; and if it has two then these are conjugage.

Proof ► Suppose

$$p = \epsilon \pi_1 \cdots \pi_r.$$

Then

$$\mathcal{N}(\pi_1)\cdots\mathcal{N}(\pi_r) = \mathcal{N}(p) = p^2.$$

Since $\mathcal{N}(\pi_i)$ is an integer $\neq 1$, it follows that either $r = 1$, ie p remains a prime, or else $r = 2$ with

$$\mathcal{N}(\pi_1) = \pm p,\ \mathcal{N}(\pi_2) = \pm p.$$

In this case, writing π for π_1,

$$p = \pm\mathcal{N}(\pi) = \pm\pi\bar{\pi}.$$

◄

We say that p *splits* in $\mathbb{Q}(\sqrt{m})$ in the latter case, ie if p divides into two prime factors in $\mathbb{Z}[\omega]$. We say that p *ramifies* if these two prime factors are equal, ie if

$$p = \epsilon\pi^2,$$

Corollary 3.2 *The rational prime $p \in \mathbb{N}$ splits if and only if there is an integer $\alpha \in \mathbb{Z}[\omega]$ with*

$$\mathcal{N}(\alpha) = \pm p.$$

Proposition 3.11 *Suppose $p \in \mathbb{N}$ is an odd prime with $p \nmid m$. Then p splits in $\mathbb{Q}(\sqrt{m})$ if and only if m is a quadratic residue* $\bmod p$*, ie if and only if*

$$x^2 \equiv m \bmod p$$

for some $x \in \mathbb{Z}$.

Proof ► Suppose

$$x^2 \equiv m \bmod p.$$

Then

$$(x - \sqrt{m})(x + \sqrt{m}) = pq$$

for some $q \in \mathbb{Z}$.

If now p is prime in $\mathbb{Z}[\omega]$ (where it is assumed, we recall, that there is unique factorisation). Then

$$p \mid x - \sqrt{m} \quad \text{or} \quad p \mid x + \sqrt{m},$$

both of which are absurd, since for example

$$\begin{aligned} p \mid x - \sqrt{m} &\Longrightarrow x - \sqrt{m} = p(a + b\sqrt{m}) \\ &\Longrightarrow pb = -1, \end{aligned}$$

where b is (at worst) a half-integer. ◄

It remains to consider two cases, $p \mid m$ and $p = 2$.

Proposition 3.12 *If the rational prime $p \mid m$ then p ramifies in $\mathbb{Q}(\sqrt{m})$.*

Proof ► We have

$$(\sqrt{m})^2 = m = pq,$$

for some $q \in \mathbb{Z}$. If p remains prime then

$$\begin{aligned} p \mid \sqrt{m} &\implies \mathcal{N}(p) \mid \mathcal{N}(\sqrt{m}) \\ &\implies p^2 \mid -m, \end{aligned}$$

which is impossible, since m is square-free.

Hence

$$p = \pm\pi\bar{\pi},$$

and

$$\sqrt{m} = \pi\alpha$$

for some $\alpha \in \mathbb{Z}[\omega]$. Note that α cannot contain $\bar{\pi}$ as a factor, since this would imply that

$$p = \pm\pi\bar{\pi} \mid \sqrt{m},$$

which as we have seen is impossible.

Taking conjugates

$$-\sqrt{m} = \bar{\pi}\bar{\alpha}.$$

Thus

$$\bar{\pi} \mid \sqrt{m}.$$

Since the factorisation of $\sqrt{m}$ is (by assumption) unique,

$$\bar{\pi} \sim \pi,$$

ie p ramifies. ◄

Proposition 3.13 *The rational prime 2 remains prime in $\mathbb{Z}[\omega]$ if and only if*

$$m \equiv 5 \bmod 8.$$

Moreover, 2 ramifies unless

$$m \equiv 1 \bmod 4.$$

Proof ► We have dealt with the case where $2 \mid m$, so we may assume that m is odd.

Suppose first that

$$m \equiv 3 \bmod 4.$$

In this case

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

If 2 does not split then

$$2 \mid 1 - \sqrt{m} \quad \text{or} \quad 2 \mid 1 + \sqrt{m},$$

both of which are absurd.

Thus
$$2 = \pm\pi\bar{\pi},$$
where
$$\pi = a + b\sqrt{m} \quad (a, b \in \mathbb{Z}),$$
say. But then
$$\bar{\pi} = a - b\sqrt{m} = \pi + 2b\sqrt{m}.$$
Since $\pi \mid 2$ is follows that
$$\pi \mid \bar{\pi};$$
and similarly
$$\bar{\pi} \mid \pi.$$
Thus
$$\bar{\pi} = \epsilon\pi,$$
where ϵ is a unit; and so 2 ramifies.

Now suppose
$$m \equiv 1 \bmod 4.$$
Suppose 2 splits, say
$$a^2 - mb^2 = \pm 2,$$
where a, b are integers or half-integers. If $a, b \in \mathbb{Z}$ then
$$a^2 - mb^2 \equiv 0, \pm 1 \bmod 4,$$
since $a^2, b^2 \equiv 0$ or $1 \bmod 4$.

Thus a, b must be half-integers, say $a = A/2,\ b = B/2$, where A, B are odd integers. In this case,
$$A^2 - mB^2 = \pm 8.$$
Hence
$$A^2 - mB^2 \equiv 0 \bmod 8$$
But
$$A^2 \equiv B^2 \equiv 1 \bmod 8,$$
and so
$$A^2 - mB^2 \equiv 1 - m \bmod 8.$$
Thus the equation is insoluble if
$$m \equiv 5 \bmod 8,$$
ie 2 remains prime in this case.

Finally, if
$$m \equiv 1 \bmod 8$$

then
$$\frac{1-\sqrt{m}}{2}\cdot\frac{1+\sqrt{m}}{2}=\frac{1-m}{4}=2q.$$
If 2 does not split then
$$2 \mid \frac{1-\sqrt{m}}{2} \quad \text{or} \quad 2 \mid \frac{1+\sqrt{m}}{2},$$
both of which are absurd.

Suppose
$$2=\pm\pi\bar{\pi},$$
where
$$\pi=\frac{A+B\sqrt{m}}{2},$$
with A, B odd; and
$$\begin{aligned}\bar{\pi}&=\frac{A-B\sqrt{m}}{2}\\ &=\pi-B\sqrt{m}.\end{aligned}$$
Thus
$$\begin{aligned}\pi \mid \bar{\pi} &\Longrightarrow \pi \mid B\sqrt{m}\\ &\Longrightarrow \mathcal{N}(\pi) \mid \mathcal{N}(B\sqrt{m})\\ &\Longrightarrow \pm 2 \mid B^2 m,\end{aligned}$$
which is impossible since B, m are both odd. Hence 2 is unramified in this case. ◀

3.7 Quadratic residues

Definition 3.6 *Suppose p is an odd rational prime; and suppose $a \in \mathbb{Z}$. Then the Legendre symbol is defined by*
$$\left(\frac{a}{p}\right)=\begin{cases}0 & \text{if } p \mid a\\ 1 & \text{if } p \nmid a \text{ and } a \text{ is a quadratic residue } \bmod p\\ -1 & \text{if } a \text{ is a quadratic non-residue } \bmod p\end{cases}$$

Proposition 3.14 *Suppose p is an odd rational prime; and suppose $a, b \in \mathbb{Z}$. Then*
$$\left(\frac{a}{p}\right)\left(\frac{b}{p}\right)=\left(\frac{ab}{p}\right).$$

Proof ► The resul is trivial if $p \mid a$ or $p \mid b$; so we may suppose that $p \nmid a, b$.

Consider the group-homomorphism

$$\theta : (\mathbb{Z}/p)^\times \to (\mathbb{Z}/p)^\times : \bar{x} \mapsto \bar{x}^2.$$

Since

$$\ker \theta = \{\pm 1\}$$

it follows from the First Isomorphism Theorem that

$$|\operatorname{im} \theta| = \frac{p-1}{2},$$

and so

$$(\mathbb{Z}/p)^\times / \operatorname{im} \theta \cong C_2 = \{\pm 1\}.$$

The result follows, since

$$\operatorname{im} \theta = \{\bar{a} \in (\mathbb{Z}/p)^\times : \left(\frac{a}{p}\right) = 1\}.$$

◄

Proposition 3.15 *Suppose p is an odd rational prime; and suppose $a \in \mathbb{Z}$. Then*

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

Proof ► The resul is trivial if $p \mid a$; so we may suppose that $p \nmid a$.

By Lagrange's Theorem (or Fermat's Little Theorem)

$$a^{p-1} \equiv 1 \bmod p.$$

Thus

$$\left(a^{(p-1)/2}\right)^2 \equiv 1 \bmod p;$$

and so

$$a^{(p-1)/2} \equiv \pm 1 \bmod p.$$

Suppose a is a quadratic residue, say

$$a \equiv b^2 \bmod p.$$

Then

$$a^{\frac{p-1}{2}} \equiv b^{p-1} \equiv 1 \bmod p.$$

Thus

$$\left(\frac{a}{p}\right) = 1 \implies a^{\frac{p-1}{2}} \equiv 1 \bmod p.$$

As we saw in the proof of Proposition 3.14, exactly half, ie $\frac{p-1}{2}$ of the numbers $1, 2, \ldots, p-1$ are quadratic residues. On the other hand, the equation

$$x^{\frac{p-1}{2}} - 1 = 0$$

over the field $\mathbb{F}_p = \mathbb{Z}/(p)$ has at most $\frac{p-1}{2}$ roots. It follows that

$$\left(\frac{a}{p}\right) = 1 \Longleftrightarrow a^{\frac{p-1}{2}} \equiv 1 \bmod p;$$

and so

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \bmod p;$$

◀

Corollary 3.3 *If $p \in \mathbb{N}$ is an odd rational prime then*

$$\left(\frac{-1}{p}\right) = \begin{cases} 1 \text{ \textit{if} } p \equiv 1 \bmod 4, \\ -1 \text{ \textit{if} } p \equiv 3 \bmod 4. \end{cases}$$

Proof ► By the Proposition,

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

If

$$p \equiv 1 \bmod 4,$$

say

$$p = 4m + 1,$$

then

$$\frac{p-1}{2} = 2m;$$

while if

$$p \equiv 3 \bmod 4,$$

say

$$p = 4m + 3,$$

then

$$\frac{p-1}{2} = 2m+1.$$

◄

It is sometimes convenient to take the remainder $r \equiv a \bmod p$ in the range

$$-\frac{p}{2} < r < \frac{p}{2}.$$

We may say that a has *negative remainder* $\bmod p$ if

$$-\frac{p}{2} < r < 0.$$

Thus 13 has negative remainder $\bmod 7$, since

$$13 \equiv -1 \bmod 7.$$

Proposition 3.16 *Suppose $p \in \mathbb{N}$ is an odd rational prime; and suppose $p \nmid a$. Then*

$$\left(\frac{a}{p}\right) = (-1)^{\mu},$$

where μ is the number of numbers among

$$1, 2a, \ldots, \frac{p-1}{2}a$$

with negative remainders.

Suppose, for example, $p = 11,\ a = 7$. Then

$$7 \equiv -4,\ 14 \equiv 3,\ 21 \equiv -1,\ 28 \equiv -5,\ 35 \equiv 2 \bmod 11.$$

Thus

$$\mu = 3.$$

Proof ► Suppose

$$1 \le r \le \frac{p-1}{2}.$$

Then just one of the numbers

$$a, 2a, \ldots \frac{p-1}{2}a$$

has remainder $\pm r$.

For suppose

$$ia \equiv r \bmod p, \quad ja \equiv -r \bmod p.$$

Then

$$(i+j)a \equiv 0 \bmod p \Longrightarrow p \mid i+j$$

which is impossible since

$$1 \le i+j \le p-1.$$

It follows (by the Pigeon-Hole Principle) that just one of the congruences

$$ia \equiv \pm r \bmod p \quad (1 \le i \le \frac{p-1}{2})$$

is soluble for each r.

Multiplying together these congruences,

$$a \cdot 2a \cdots \frac{p-1}{2} a \equiv (-1)^{\mu} 1 \cdot 2 \cdots \frac{p-1}{2} \bmod p,$$

ie

$$a^{\frac{p-1}{2}} 1 \cdot 2 \cdots \frac{p-1}{2} \equiv (-1)^{\mu} 1 \cdot 2 \cdots \frac{p-1}{2} \bmod p,$$

and so

$$a^{\frac{p-1}{2}} \equiv (-1)^{\mu} \bmod p.$$

Since

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

by Proposition 3.15, we conclude that

$$\left(\frac{a}{p}\right) \equiv (-1)^{\mu} \bmod p.$$

◄

Proposition 3.17 *If* $p \in \mathbb{N}$ *is an odd rational prime then*

$$\left(\frac{2}{p}\right) = \begin{cases} 1 \text{ if } p \equiv \pm 1 \bmod 8, \\ -1 \text{ if } p \equiv \pm 3 \bmod 8. \end{cases}$$

Proof ► Consider the numbers

$$2, 4, \ldots, p-1.$$

The number $2i$ will have negative remainder if

$$\frac{p}{2} < 2i < p,$$

ie

$$\frac{p}{4} < i < \frac{p}{2}.$$

Thus the μ in Proposition 3.16 is given by

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

We consider $p \bmod 8$. If

$$p \equiv 1 \bmod 8,$$

say

$$p = 8m + 1,$$

then

$$\left[\frac{p}{2}\right] = 4m, \ \left[\frac{p}{4}\right] = 2m,$$

and so

$$\mu = 2m.$$

If

$$p \equiv 3 \bmod 8,$$

say

$$p = 8m + 3,$$

then

$$\left[\frac{p}{2}\right] = 4m + 1, \ \left[\frac{p}{4}\right] = 2m,$$

and so

$$\mu = 2m + 1.$$

If

$$p \equiv 5 \bmod 8,$$

say

$$p = 8m + 5,$$

then

$$\left[\frac{p}{2}\right] = 4m + 2, \ \left[\frac{p}{4}\right] = 2m + 1,$$

and so

$$\mu = 2m + 1.$$

If

$$p \equiv 7 \bmod 8,$$

say

$$p = 8m + 7,$$

then

$$\left[\frac{p}{2}\right] = 4m + 3, \ \left[\frac{p}{4}\right] = 2m + 1,$$

and so

$$\mu = 2m + 2.$$

◄

Corollary 3.4 *If $p \in \mathbb{N}$ is an odd rational prime then*

$$\left(\frac{-2}{p}\right) = \begin{cases} 1 \textit{ if } p \equiv 1 \textit{ or } 3 \bmod 8, \\ -1 \textit{ if } p \equiv 5 \textit{ or } 7 \bmod 8. \end{cases}$$

Proof ► This follows from the Proposition and the Corollary to Proposition 3.15, since

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

by Proposition 3.14. ◄

Proposition 3.18 *If $p \in \mathbb{N}$ is an odd rational prime then*

$$\left(\frac{3}{p}\right) = \begin{cases} 1 \textit{ if } p \equiv \pm 1 \bmod 12, \\ -1 \textit{ if } p \equiv \pm 5 \bmod 12. \end{cases}$$

Proof ► If
$$0 < i < \frac{p}{2}$$
then
$$0 < 3i < \frac{3p}{2}.$$
Thus $3i$ has negative remainder if
$$\frac{p}{2} < 3i < p,$$
ie
$$\frac{p}{6} < i < \frac{p}{3}.$$
Thus
$$\mu = \left[\frac{p}{3}\right] - \left[\frac{p}{6}\right].$$

If
$$p \equiv 1 \bmod 6,$$
say
$$p = 6m + 1,$$
then
$$\left[\frac{p}{3}\right] = 2m, \; \left[\frac{p}{6}\right] = m,$$
and so
$$\mu = m.$$
If
$$p \equiv 5 \bmod 6,$$
say
$$p = 6m + 5,$$
then
$$\left[\frac{p}{3}\right] = 2m + 1, \; \left[\frac{p}{6}\right] = m,$$
and so
$$\mu = m + 1.$$
The result follows. ◄

Corollary 3.5 *If $p \in \mathbb{N}$ is an odd rational prime then*

$$\left(\frac{-3}{p}\right) = \begin{cases} 1 \textit{ if } p \equiv 1 \bmod 6, \\ -1 \textit{ if } p \equiv 5 \bmod 6. \end{cases}$$

Proof ► This follows from the Proposition and the Corollary to Proposition 3.15, since

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

by Proposition 3.14. ◄

Proposition 3.19 *If $p \in \mathbb{N}$ is an odd rational prime then*

$$\left(\frac{5}{p}\right) = \begin{cases} 1 \textit{ if } p \equiv \pm 1 \bmod 10, \\ -1 \textit{ if } p \equiv \pm 3 \bmod 10. \end{cases}$$

Proof ► If

$$0 < i < \frac{p}{2}$$

then

$$0 < 5i < \frac{5p}{2}.$$

Thus $5i$ has negative remainder if

$$\frac{p}{2} < 5i < p \quad \text{or} \quad \frac{3p}{2} < i < 2p,$$

ie

$$\frac{p}{10} < i < \frac{p}{5} \quad \text{or} \quad \frac{3p}{10} < i < \frac{2p}{5}.$$

Thus

$$\mu = \left[\frac{p}{5}\right] - \left[\frac{p}{10}\right] + \left[\frac{2p}{5}\right] - \left[\frac{3p}{10}\right].$$

If

$$p \equiv 1 \bmod 12,$$

say

$$p = 10m + 1,$$

then

$$\left[\frac{p}{5}\right] = 2m, \ \left[\frac{p}{10}\right] = m, \ \left[\frac{2p}{5}\right] = 4m, \ \left[\frac{3p}{10}\right] = 3m,$$

and so

$$\mu = 2m.$$

The other cases are left to the reader. ◄

3.8 Gauss' Law of Quadratic Reciprocity

Proposition 3.16 provides an algorithm for computing the Legendre symbol, as illustrated in Propositions 3.17–3.19, perfectly adequate for our purposes. However, Euler discovered and Gauss proved a remarkable result which makes computation of the symbol childishly simple. This result — The Law of Quadratic Reciprocity — has been called the most beautiful result in Number Theory, so it would be a pity not to mention it, even though — as we said — we do not really need it.

Proposition 3.20 *Suppose $p, q \in \mathbb{N}$ are two distinct odd rational primes. Then*

$$\left(\frac{q}{p}\right)\left(\frac{p}{q}\right) = \begin{cases} -1 \text{ if } p \equiv q \equiv 3 \bmod 4, \\ 1 \text{ otherwise}. \end{cases}$$

Another way of putting this is to say that

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

Proof ► Let

$$S = \{1, 2, \ldots, \frac{p-1}{2}\},\ T = \{1, 2, \ldots, \frac{q-1}{2}\}.$$

We shall choose remainders $\bmod p$ from the set

$$\{-\frac{p}{2} < i < \frac{p}{2}\} = -S \cup \{0\} \cup S,$$

and remainders $\bmod q$ from the set

$$\{-\frac{q}{2} < i < \frac{q}{2}\} = -T \cup \{0\} \cup T.$$

By Gauss' Lemma (Proposition 3.16),

$$\left(\frac{q}{p}\right) = (-1)^{\mu},\ \left(\frac{p}{q}\right) = (-1)^{\nu},$$

where

$$\mu = \|\{i \in S : qi \bmod p \in -S\}\|,\ \nu = \|\{i \in T : pi \bmod q \in -T\}\|.$$

By '$qi \bmod p \in -S$' we mean that there exists a j (necessarily unique) such that

$$qi - pj \in -S.$$

But now we observe that, in this last formula,

$$0 < i < \frac{p}{2} \Longrightarrow 0 < j < \frac{q}{2}.$$

The basic idea of the proof is to associate to each such contribution to μ the 'point' $(i, j) \in S \times T$. Thus

$$\mu = \|\{(i,j) \in S \times T : -\frac{p}{2} < qi - pj < 0\}\|;$$

and similarly

$$\nu = \|\{(i,j) \in S \times T : 0 < qi - pj < \frac{q}{2}\}\|,$$

where we have reversed the order of the inequality on the right so that both formulae are expressed in terms of $(qi - pj)$.

Let us write $[R]$ for the number of integer points in the region $R \subset \mathbb{R}^2$. Then

$$\mu = [R_1],\ \nu = [R_2],$$

where

$$R_1 = \{(x,y) \in R : -\frac{p}{2} < qx - py < 0\},\ R_2 = \{(x,y) \in R : 0 < qx - py < \frac{q}{2}\},$$

and R denotes the rectangle

$$R = \{(x,y) : 0 < x < \frac{p}{2},\ 0 < y < \frac{p}{2}\}.$$

The line

$$qx - py = 0$$

is a diagonal of the rectangle R, and R_1, R_2 are strips above and below the diagonal (Fig 3.8).

This leaves two triangular regions in R,

$$R_3 = \{(x,y) \in R : qx - py < -\frac{p}{2}\},\ R_4 = \{(x,y) \in R : qx - py > \frac{q}{2}\}.$$

We shall show that, surprisingly perhaps, reflection in a central point sends the integer points in these two regions into each other, so that

$$[R_3] = [R_4].$$

Since

$$R = R_1 \cup R_2 \cup R_3 \cup R_4,$$

it will follow that

$$[R_1] + [R_2] + [R_3] + [R_4] = [R] = \frac{p-1}{2}\frac{q-1}{2},$$

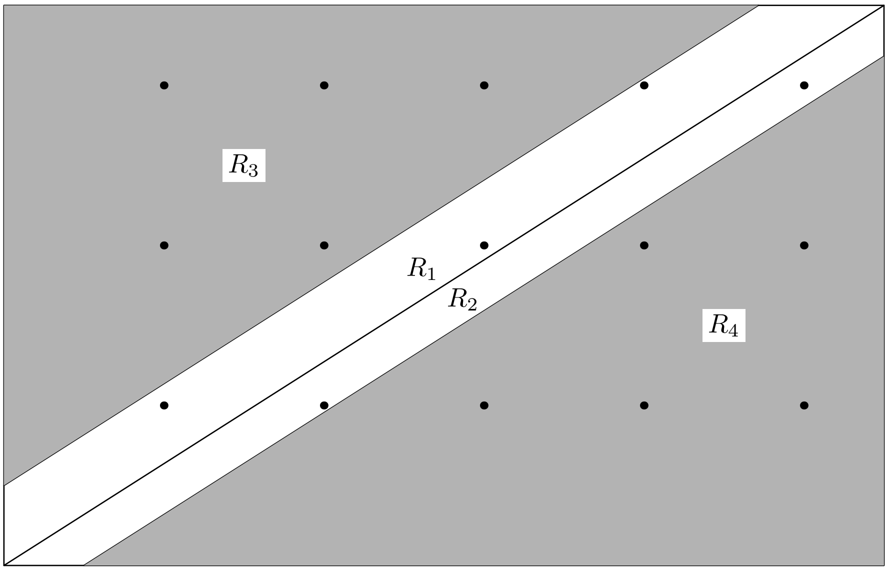

Figure 3.1: $p = 11,\ q = 7$

ie

$$\mu + \nu + [R_3] + [R_4] = \frac{p-1}{2}\frac{q-1}{2}.$$

But if now $[R_3] = [R_4]$ then it will follow that

$$\mu + \nu \equiv \frac{p-1}{2}\frac{q-1}{2} \bmod 2,$$

which is exactly what we have to prove.

It remains to define our central reflection. Note that reflection in the centre $(\frac{p}{4}, \frac{q}{4})$ of the rectangle R will not serve, since this does not send integer points into integer points. For that, we must reflect in a point whose coordinates are integers or half-integers.

We choose this point by "shrinking" the rectangle R to a rectangle bounded by integer points, ie the rectangle

$$R' = \{1 \le x \le \frac{p-1}{2},\ 1 \le y \le \frac{q-1}{2}\}.$$

Now we take P to be the centre of this rectangle, ie

$$P = (\frac{p+1}{4},\ \frac{q+1}{4}).$$

The reflection is then given by

$$(x, y) \mapsto (X, Y) = (\frac{p+1}{-}x, \frac{q+1}{-}y).$$

It is clear that reflection in P will send the integer points of R into themselves. But it is not clear that it will send the integer points in R_3 into those in R_4, and vice versa. To see that, let us shrink these triangles as we shrank the rectangle. If $x, y \in \mathbb{Z}$ then

$$qx - py < -\frac{p}{2} \implies qx - py \leq -\frac{p+1}{2};$$

and similarly

$$qx - py > \frac{q}{2} \implies qx - py \geq \frac{q+1}{2}.$$

Now reflection in P *does* send the two lines

$$qx - py = -\frac{p+1}{2},\ qx - py = \frac{q+1}{2}$$

into each other; for

$$qX - pY = q(p+1-x) - p(q+1-y) = (q-p) - (qx - py),$$

and so

$$qx - py = -\frac{p+1}{2} \iff qX - pY = (q-p) + \frac{p+1}{2} = \frac{q+1}{2}.$$

We conclude that

$$[R_3] = [R_4].$$

Hence

$$[R] = [R_1] + [R_2] + [R_3] + [R_4] \equiv \mu + \nu \bmod 2,$$

and so

$$\mu + \nu \equiv [R] = \frac{p-1}{2}\frac{q-1}{2}.$$

◄

Example: Take $p = 37,\ q = 47$. Then

$$\begin{aligned}
\left(\frac{37}{47}\right) &= \left(\frac{47}{37}\right) \text{ since } 37 \equiv 1 \bmod 4\\
&= \left(\frac{10}{37}\right)\\
&= \left(\frac{2}{37}\right)\left(\frac{5}{37}\right)\\
&= -\left(\frac{5}{37}\right) \text{ since } 37 \equiv -3 \bmod 8\\
&= -\left(\frac{37}{5}\right) \text{ since } 5 \equiv 1 \bmod 4\\
&= -\left(\frac{2}{5}\right)\\
&= -(-1) = 1.
\end{aligned}$$

Thus 37 *is* a quadratic residue $\bmod 47$.

We could have avoided using the result for $\left(\frac{2}{p}\right)$:

$$\begin{aligned}\left(\frac{10}{37}\right) &= \left(\frac{-27}{37}\right)\\ &= \left(\frac{-1}{37}\right)\left(\frac{3}{37}\right)^3\\ &= (-1)^{18}\left(\frac{37}{3}\right)\\ &= \left(\frac{1}{3}\right) = 1.\end{aligned}$$

3.9 Some quadratic fields

We end by applying the results we have established to a small number of quadratic fields.

3.9.1 The gaussian field $\mathbb{Q}(i)$

Proposition 3.21 *1. The integers in $\mathbb{Q}(i)$ are the gaussian integers*

$$a + bi \quad (a, b \in \mathbb{Z})$$

2. *The units in $\mathbb{Z}[i]$ are the numbers*

$$\pm 1, \pm i.$$

3. *The ring of integers $\mathbb{Z}[i]$ is a principal ideal domain (and so a unique factorisation domain).*

4. *The prime 2 ramifies in $\mathbb{Z}[i]$:*

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

The odd prime p splits in $\mathbb{Z}[i]$ if and only if

$$p \equiv 1 \bmod 4,$$

in which case it splits into two conjugate but inequivalent primes:

$$p = \pm\pi\bar{\pi}.$$

Proof ► This follows from Propositions 3.4, 3.7, 3.9, 3.11–3.13, and the Corollary to Proposition 3.15. ◄

Factorisation in the gaussian field $\mathbb{Q}(i)$ gives interesting information on the expression of a number as a sum of two squares.

Proposition 3.22 *An integer $n > 0$ is expressible as a sum of two squares,*

$$n = a^2 + b^2 \quad (a, b \in \mathbb{Z})$$

if and only if each prime $p \equiv 3 \bmod 4$ occurs to an even power in n.

Proof ► Suppose

$$n = a^2 + b^2 = (a + bi)(a - bi).$$

Let

$$a + bi = \epsilon \pi_1^{e_1} \cdots \pi_r^{e_r}.$$

Taking norms,

$$n = \mathcal{N}(a + bi) = \mathcal{N}(\pi_1)^{e_1} \cdots \mathcal{N}(\pi_r)^{e_r}.$$

Suppose

$$p \equiv 3 \bmod 4.$$

Then p remains prime in $\mathbb{Z}[i]$, by Proposition 3.21.

Suppose

$$p^e \parallel a + ib,$$

ie

$$p^e \mid a + ib \quad \text{but} \quad p^{e+1} \nmid a + ib.$$

Then

$$p^e \parallel a - ib,$$

since

$$a + ib = p^e \alpha \Longrightarrow a - ib = p^e \bar{\alpha},$$

on taking conjugates. Hence

$$p^{2e} \parallel n = (a + ib)(a - ib),$$

ie p appears in n with even exponent.

We have shown, incidentally, that if $p \equiv 3 \bmod 4$ then

$$p^{2e} \parallel n = a^2 + b^2 \Longrightarrow p^e \mid a, \; p^e \mid b.$$

In other words, each expression of n as a sum of two squares

$$n = a^2 + b^2$$

is of the form

$$n = (p^e a')^2 + (p^e b')^2,$$

where

$$\frac{n}{p^{2e}} = a'^2 + b'^2.$$

We have shown that each prime $p \equiv 3 \bmod 4$ must occur with even exponent in n. Conversely, suppose that this is so.

Each prime $p \equiv 1 \bmod 4$ splits in $\mathbb{Z}[i]$, by Proposition 3.21, say

$$p = \pi_p \overline{\pi_p}.$$

Also, 2 ramifies in $\mathbb{Z}[i]$:

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

Now suppose

$$n = 2^{e_2} 3^{e_3} 5^{e_5} \cdots ,$$

where $e_3, e_7, e_11, e_19, \dots$ are all even, say

$$p \equiv 3 \bmod 4 \implies e_p = 2f_p.$$

Let

$$\alpha = \alpha_2 \alpha_3 \alpha_5 \cdots ,$$

where

$$\alpha_p = \begin{cases} (1+i)^{e_2} & \text{if } p = 2, \\ \pi_p^{e_p} & \text{if } p \equiv 1 \bmod 4, \\ p^{f_p} & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

Then

$$\mathcal{N}(\alpha_p) = p^{e_p}$$

in all cases, and so

$$\mathcal{N}(\alpha) = \prod_p \mathcal{N}(\alpha_p) = \prod_p p^{e_p} = n.$$

Thus if

$$\alpha = a + bi$$

then

$$n = a^2 + b^2.$$

◀

It's worth noting that this argument actually gives the *number* of ways of expressing n as a sum of two squares, ie the number of solutions of

$$n = a^2 + b^2 \quad (a, b \in \mathbb{Z}).$$

For the number of solutions is the number of integers $\alpha \in \mathbb{Z}[i]$ such that

$$n = \mathcal{N}(()\alpha) = \alpha\bar{\alpha}.$$

Observe that when $p \equiv 1 \bmod 3$ in the argument above we could equally well have taken

$$\alpha_p = \pi^r \bar{\pi}^s$$

for any $r, s \geq 0$ with

$$r + s = e_p.$$

There are just

$$e_p + 1$$

ways of choosing α_p in this way.

It follows from unique factorisation that the choice of the α_p for $p \equiv 1 \bmod 4$ determines α up to a unit, ie the general solution is

$$\alpha = \epsilon(1+i)^{e_2} \prod_{p \equiv 1 \bmod 4} \alpha_p \prod_{p \equiv 3 \bmod 4} p^{f_p}.$$

Since there are four units, $\pm 1,\ \pm i$, we conclude that the number of ways of expressing n as a sum of two sqares is

$$4 \prod_{p \equiv 1 \bmod 4} (e_p + 1).$$

Note that in this calculation, each solution

$$n = a^2 + b^2$$

with

$$0 < a < b$$

gives rise to 8 solutions:

$$n = (\pm a)^2 + (\pm b)^2, \quad n = (\pm b)^2 + (\pm a)^2.$$

To these must be added solutions with $a = 0$ or with $a = b$. The former occurs only if $n = m^2$, giving 4 additional solutions:

$$n = 0^2 + (\pm m)^2 = (\pm m)^2 + 0^2;$$

while the latter occurs only if $n = 2m^2$, again giving 4 additional solutions:

$$n = (\pm m)^2 + (\pm m)^2.$$

We conclude that the number of solutions with $a, b \geq 0$ is

$$\begin{cases} \frac{1}{2}\prod_{p\equiv 1 \bmod 4}(e_p+1) & \text{if } n \neq m^2, 2m^2 \\ \frac{1}{2}\left(\prod_{p\equiv 1 \bmod 4}(e_p+1)+1\right) & \text{if } n = m^2 \text{ or } 2m^2. \end{cases}$$

This is of course assuming that

$$p \equiv 3 \bmod 4 \Longrightarrow 2 \mid e_p,$$

without which there are no solutions.

In particular, each prime $p \equiv 1 \bmod 4$ is uniquely expressible as a sum of two squares

$$n = a^2 + b^2 \quad (0 < a < b),$$

eg

$$53 = 2^2 + 7^2.$$

As another example,

$$108 = 2^2 3^3$$

cannot be expressed as a sum of two squares, since $e_3 = 3$ is odd.

3.9.2 The field $\mathbb{Q}(\sqrt{3})$

Proposition 3.23 *1. The integers in $\mathbb{Q}(\sqrt{3})$ are the numbers*

$$a + b\sqrt{3} \quad (a, b \in \mathbb{Z})$$

2. *The units in $\mathbb{Z}[\sqrt{3}]$ are the numbers*

   $$\pm\eta^n \quad (n \in \mathbb{Z}),$$

   where

   $$\eta = 2 + \sqrt{3}.$$

3. *The ring of integers $\mathbb{Z}[\sqrt{3}]$ is a principal ideal domain (and so a unique factorisation domain).*

4. *The primes 2 and 3 ramify in $\mathbb{Z}[\sqrt{3}]$:*

   $$2 = \eta^{-1}(1+\sqrt{3})^2, \quad 3 = (\sqrt{3})^2.$$

   The odd prime $p \neq 3$ splits in $\mathbb{Z}[\sqrt{3}]$ if and only if

   $$p \equiv \pm 1 \bmod 12,$$

   in which case it splits into two conjugate but inequivalent primes:

   $$p = \pm\pi\bar{\pi}.$$

Proof ► This follows from Propositions 3.4, 3.8, 3.9, 3.11–3.13, and Proposition 3.18. ◄

3.9.3 The field $\mathbb{Q}(\sqrt{5})$

Proposition 3.24 1. *The integers in $\mathbb{Q}(\sqrt{5})$ are the numbers*

$$a + b\omega \quad (a, b \in \mathbb{Z}),$$

where

$$\omega = \frac{1+\sqrt{5}}{2}.$$

2. *The units in $\mathbb{Z}[\sqrt{5}]$ are the numbers*

$$\pm\omega^n \quad (n \in \mathbb{Z}).$$

3. *The ring of integers $\mathbb{Z}[\omega]$ is a principal ideal domain (and so a unique factorisation domain).*

4. *The prime 5 ramifies in $\mathbb{Z}[\omega]$:*

$$5 = (\sqrt{5})^2.$$

The prime $p \neq 5$ splits in $\mathbb{Z}[\omega]$ if and only if

$$p \equiv \pm 1 \bmod 10,$$

in which case it splits into two conjugate but inequivalent primes:

$$p = \pm\pi\bar{\pi}.$$

Proof ► This follows from Propositions 3.4, 3.8, 3.9, 3.11–3.13, and Proposition 3.19. ◄

Chapter 4

Mersenne and Fermat numbers

4.1 Mersenne numbers

Proposition 4.1 *If*

$$n = a^m - 1 \quad (a, m > 1)$$

is prime then

1. $a = 2$;
2. m *is prime.*

Proof ► In the first place,

$$(a - 1) \mid (a^m - 1);$$

so if $a > 2$ then n is certainly not prime.

Suppose $m = rs$, where $r, s > 1$. Evidently

$$(x - 1) \mid (x^s - 1)$$

in $\mathbb{Z}[x]$; explicitly

$$x^s - 1 = (x - 1)(x^{s-1} + x^{s-2} + x^{s-3} + \cdots + 1).$$

Subsitituting $x = a^r$,

$$(a^r - 1) \mid (a^{rs} - 1) = a^m - 1.$$

Thus if $a^m - 1$ is prime then m has no proper factors, ie m is prime. ◄

Definition 4.1 *The numbers*

$$M_p = 2^p - 1,$$

where p is prime, are called Mersenne numbers.

The numbers

$$M_2 = 3,\ M_3 = 7,\ M_5 = 31,\ M_7 = 127$$

are all prime. However,

$$M_{11} = 2047 = 23 \cdot 89.$$

(It should be emphasized that Mersenne never claimed the Mersenne numbers were all prime. He listed the numbers M_p for $p \leq 257$, indicating which were prime, in his view. His list contained several errors.)

The following heuristic argument suggests that there are probably an infinity of Mersenne primes. (Webster's Dictionary defines 'heuristic' as: *providing aid or direction in the solution of a problem but otherwise unjustified or incapable of justification*.)

By the Prime Number Theorem, the probability that a large number n is prime is

$$\approx \frac{1}{\log n}.$$

In this estimate we are including even numbers. Thus the probability that an *odd* number n is prime is

$$\approx \frac{2}{\log n}.$$

Thus the probability that M_p is prime is

$$\approx \frac{2}{p \log 2}.$$

So the expected number of Mersenne primes is

$$\approx \frac{2}{\log 2} \sum \frac{1}{p_n}$$

where p_n is the nth prime.

But — again by the Prime Number Theorem —

$$p_n \approx n \log n.$$

Thus the expected number of Mersenne primes is

$$\approx \frac{2}{\log 2} \sum \frac{1}{n \log n} = \infty,$$

since

$$\sum \frac{1}{n \log n}$$

diverges, eg by comparison with

$$\int^X \frac{1}{x \log x} = \log \log X + C.$$

4.1.1 The Lucas-Lehmer test

Mersenne numbers are important because there is a simple test, announced by Lucas and proved rigorously by Lehmer, for determining whether or not M_p is prime. (There are many *necessary* tests for primality, eg if p is prime then

$$2^p \equiv 2 \bmod p.$$

What is rare is to find a necessary and *sufficient* test for the primality of numbers in a given class, and one which is moreover relatively easy to implement.) For this reason, all recent "record" primes have been Mersenne primes.

We shall give two slightly different versions of the Lucas-Lehmer test. The first is only valid if $p \equiv 3 \bmod 4$, while the second applies to all Mersenne numbers. The two tests are very similar, and equally easy to implement. We are giving the first only because the proof of its validity is rather simpler. So it should be viewed as an introduction to the second, and true, Lucas-Lehmer test.

Both proofs are based on arithmetic in quadratic fields: the first in $\mathbb{Q}(\sqrt{5})$, and the second in $\mathbb{Q}(\sqrt{3})$; and both are based on the following result.

Proposition 4.2 *Suppose α is an integer in the field $\mathbb{Q}(\sqrt{m})$; and suppose P is an odd prime with $P \nmid m$. Then*

$$\alpha^P \equiv \begin{cases} \alpha & \textit{if } \left(\dfrac{P}{m}\right) = 1, \\ \bar{\alpha} & \textit{if } \left(\dfrac{P}{m}\right) = -1. \end{cases}$$

Proof ► Suppose

$$\alpha = a + b\sqrt{m},$$

where a, b are integers if $m \not\equiv 1 \bmod 4$, and half-integers if $m \equiv 1 \bmod 4$.

In fact these cases do not really differ; for 2 is invertible $\bmod P$, so we may consider a as an integer $\bmod P$ if $2a \in \mathbb{Z}$. Thus

$$\alpha^P \equiv a^P + \binom{P}{1} a^{P-1} b\sqrt{m} + \binom{P}{2} a^{P-2} bm + \cdots + b^P m^{\frac{P-1}{2}} \sqrt{m} \bmod P.$$

Now

$$P \mid \binom{P}{r}$$

if $1 \leq r \leq P-1$. Hence

$$\alpha^P \equiv a^P + b^P m^{\frac{P-1}{2}} \sqrt{m} \mod P$$

By Fermat's Little Theorem,

$$a^P \equiv a \bmod P,\ b^P \equiv b \bmod P.$$

Also

$$m^{\frac{P-1}{2}} \equiv \left(\frac{m}{P}\right) \bmod P,$$

by Proposition 3.15. Thus

$$\alpha^P \equiv a + b\left(\frac{P}{m}\right)\sqrt{m} \bmod P,$$

ie

$$\left(\frac{m}{P}\right) = 1 \Longrightarrow \alpha^P \equiv \alpha \bmod P,$$
$$\left(\frac{m}{P}\right) = -1 \Longrightarrow \alpha^P \equiv \bar{\alpha} \bmod P.$$

◄

Corollary 4.1 *For all integers α in $\mathbb{Q}(\sqrt{m}$,*

$$\alpha^{P^2} \equiv \alpha \bmod P.$$

We may regard this as the analogue of Fermat's Little Theorem

$$a^P \equiv a \bmod P$$

for quadratic fields.

There is another way of establishing this result, which we shall sketch briefly. It depends on considering the ring

$$A = \mathbb{Z}[\omega]/(P).$$

formed by the remainders

$$\alpha \bmod P$$

of integers α in $\mathbb{Q}(\sqrt{m})$.

There are P^2 elements in this ring, since each $\alpha \in \mathbb{Z}[\omega]$ is congruent $\bmod P$ to just one of the numbers

$$a + b\sqrt{m}$$

where $a, b \in \mathbb{Z}$ and

$$0 \le a, b < P.$$

There are no nilpotent elements in the ring A if $P \nmid m$; for if $\alpha = a + b\sqrt{m}$ then

$$\begin{aligned} P \mid \alpha^2 &\Longrightarrow P \mid 2ab,\ P \mid a^2 + b^2 m \\ &\Longrightarrow P \mid a, b. \end{aligned}$$

Thus

$$\alpha^2 \equiv 0 \bmod P \Longrightarrow \alpha \equiv 0 \bmod P,$$

from which it follows that, if $n > 0$,

$$\alpha^n \equiv 0 \bmod P \Longrightarrow \alpha \equiv 0 \bmod P,$$

A ring without non-zero nilpotent elements is said to be *semi-simple*. It is not hard to show that *a finite semi-simple commutative ring is a direct sum of fields*.

Now there is just one field (up to isomorphism) containing p^e elements for each prime power p^e, namely the galois field $\mathbf{GF}(p^e)$. It follows that either

1. $\mathbb{Z}[\omega]/(P) \cong \mathbf{GF}(P^2)$; or
2. $\mathbb{Z}[\omega]/(P) \cong \mathbf{GF}(P) \oplus \mathbf{GF}(P)$.

The non-zero elements in $\mathbf{GF}(p^e)$ form a multiplicative group $\mathbf{GF}(p^e)^\times$ with $p^e - 1$ elements. It follows from Legendre's Theorem that

$$a \neq 0 \Longrightarrow a^{p^e - 1} = 1$$

in $\mathbf{GF}(p^e)$. Hence

$$a^{p^e} = a$$

for all $a \in \mathbf{GF}(p^e)$.

Thus in the first case,

$$\alpha^{P^2} \equiv \alpha$$

for all $\alpha \in \mathbb{Z}[\omega]/(P)$; while in the second case we even have

$$\alpha^P \equiv \alpha$$

for all $\alpha \in \mathbb{Z}[\omega]/(P)$, since this holds in each of the constituent fields.

In the first case we can go further. The galois field $\mathbf{GF}(p^e)$ is *of characteristic* p, ie

$$pa = a + \cdots a = 0,$$

for all $ain\mathbf{GF}(p^e)$. Also, the map

$$a \mapsto a^p$$

is an automorphism of $\mathbf{GF}(p^e)$. (This follows by essentially the same argument that we used above to show that $\alpha^P \equiv \alpha$ or $\bar{\alpha}$ above.)

In particular, the map

$$\alpha \mapsto \alpha^P \bmod P$$

is an automorphism of our field

$$\mathbb{Z}[\omega]/(P).$$

On the other hand, the map

$$\alpha \mapsto \bar{\alpha}$$

is also an automorphism of $\mathbb{Z}[\omega]/(P)$, since

$$P \mid \alpha \Longrightarrow P \mid \bar{\alpha}.$$

Moreover, this is the only automorphism of $\mathbb{Z}[\omega]/(P)$ apart from the identity map, since any automorphism must send

$$\sqrt{m} \bmod P \mapsto \pm\sqrt{m} \bmod P.$$

The automorphism

$$\alpha \mapsto \alpha^P \mod P$$

is not the identity map, since the equation

$$x^P - x = 0$$

has at mos P solutions in the field $\mathbb{Z}[\omega]/(P)$. We conclude that

$$\alpha^P \equiv \bar{\alpha} \bmod P.$$

If $\mathbb{Z}[\omega]$ is a principal ideal domain the second case arises if and only if P splits, which by Proposition 3.14 occurs when

$$\left(\frac{m}{P}\right) = 1.$$

Explicitly, if

$$P = \pi_1 \pi_2,$$

then

$$\begin{aligned} \mathbb{Z}[\omega]/(P) &\cong \mathbb{Z}[\omega]/(\pi_1) \oplus \mathbb{Z}[\omega]/(\pi_2) \\ &\cong \mathbf{GF}(P) \oplus \mathbf{GF}(P). \end{aligned}$$

Proposition 4.3 *Suppose* $p \equiv 3 \bmod 4$. *Let the sequence* r_n *be defined by*

$$r_1 = 3, \quad r_{n+1} = r_n^2 - 2.$$

Then M_p *is prime if and only if*

$$M_p \mid r_{p-1}.$$

Proof ► We work in the field $\mathbb{Q}(\sqrt{5})$. By Proposition 3.4, the integers in this field are the numbers

$$a + b\omega \quad (a, b \in \mathbb{Z})$$

where

$$\omega = \frac{1+\sqrt{5}}{2}.$$

By Proposition 3.9, there is unique factorisation in the ring of integers $\mathbb{Z}[\omega]$.

Lemma 4.1 *If r_n is the sequence defined in the Proposition then*

$$r_n = \omega^{2^n} + \omega^{-2^n}$$

for each $n \geq 1$.

Proof of Lemma ▷ Let us set

$$s_n = \omega^{2^n} + \omega^{-2^n}$$

for $n \geq 0$. Then

$$\begin{aligned} s_n^2 &= \left(\omega^{2^n} + \omega^{-2^n}\right)^2 \\ &= \omega^{2^{n+1}} + 2 + \omega^{-2^{n+1}} \\ &= s_{n+1} + 2, \end{aligned}$$

ie

$$s_{n+1} = s_n^2 - 2.$$

Also

$$\begin{aligned} s_0 &= \omega + \omega^{-1} \\ &= \omega - \bar{\omega} \\ &= \sqrt{5}, \end{aligned}$$

and so

$$s_1 = s_0^2 - 2 = 3.$$

We conclude that

$$r_n = s_n = \omega^{2^n} + \omega^{-2^n}$$

for all $n \geq 1$. ◁

Let us suppose first that M_p is prime. Let us write $P = M_p$.

Lemma 4.2 *We have*

$$\left(\frac{5}{P}\right) = -1.$$

Proof of Lemma ▷ Since

$$2^4 \equiv 1 \bmod 5$$

it follows that

$$\begin{aligned} 2^p &\equiv 2^3 \bmod 5 \\ &\equiv 3 \bmod 5. \end{aligned}$$

Hence
$$P = 2^p - 1 \equiv 2 \bmod 5;$$
and so, by Proposition 3.19,
$$\left(\frac{5}{P}\right) = -1.$$
◁

It follows from this Lemma and Proposition 4.2 that
$$\alpha^P \equiv \bar{\alpha} \bmod P$$
for all $\alpha \in \mathbb{Z}[\omega]$. In particular,
$$\omega^P \equiv \bar{\omega} \bmod P.$$
Hence
$$\begin{aligned} \omega^{P+1} &\equiv \omega\bar{\omega} \bmod P \\ &\equiv \mathcal{N}(\omega) \bmod P \end{aligned} \qquad \equiv -1 \bmod P.$$
In other words,
$$\omega^{2^p} \equiv -1 \bmod P.$$
Thus
$$\omega^{2^p} + 1 \equiv 0 \bmod P.$$
Dividing by $\omega^{2^{p-1}}$,
$$\omega^{2^{p-1}} + \omega^{-2^{p-1}} \equiv 0 \bmod P,$$
ie
$$r_{p-1} \equiv 0 \bmod P.$$
Conversely, suppose P is a prime factor of M_p. Then
$$\begin{aligned} M_p \mid r_{p-1} &\implies r_{p-1} \equiv 0 \bmod P \\ &\implies \omega^{2^{p-1}} + \omega^{-2^{p-1}} \equiv 0 \bmod P \\ &\implies \omega^{2^p} + 1 \equiv 0 \bmod P \\ &\implies \omega^{2^p} \equiv -1 \bmod P. \end{aligned}$$
But this implies that the order of $\omega \bmod P$ is 2^{p+1}. For
$$\omega^{2^{p+1}} = (\omega^{2^p})^2 \equiv 1 \bmod P,$$
so if the order of $\omega \bmod P$ is d then
$$d \mid 2^{p+1} \implies d = 2^e$$

for some $e \le p+1$; and if $e \le p$ then

$$\omega^{2^p} \equiv 1 \bmod P.$$

On the other hand, by the Corollary to Proposition 4.2,

$$\omega^{P^2} \equiv \omega \bmod P \Longrightarrow \omega^{P^2-1} \equiv 1 \bmod P.$$

Hence

$$2^{p+1} \mid P^2 - 1 = (P+1)(P-1).$$

Now

$$\gcd(P+1, P-1) = 2.$$

It follows that

$$2^p \mid P+1 \quad \text{or} \quad 2^p \mid P-1.$$

The latter is impossible since

$$2^p > M_p \ge P > P-1;$$

while

$$2^p \mid P+1 \Longrightarrow 2^p \le P+1 \Longrightarrow M_p = 2^p - 1 \le P \Longrightarrow P = M_p.$$

◄

Now for the 'true' Lucas-Lehmer test. As we shall see, the proof is a little harder, which is why we gave the earlier version.

Proposition 4.4 *Let the sequence r_n be defined by*

$$r_1 = 4, \quad r_{n+1} = r_n^2 - 2.$$

Then M_p is prime if and only if

$$M_p \mid r_{p-1}.$$

Proof ► We work in the field $\mathbb{Q}(\sqrt{3})$. By Proposition 3.4, the integers in this field are the numbers

$$a + b\sqrt{3} \quad (a, b \in \mathbb{Z}).$$

By Proposition 3.9, there is unique factorisation in the ring of integers $\mathbb{Z}[\sqrt{3}]$.

We set

$$\eta = 1 + \sqrt{3}, \quad \epsilon = 2 + \sqrt{3}.$$

Lemma 4.3 *The units in $\mathbb{Z}[\sqrt{3}]$ are the numbers*

$$\pm\epsilon^n \quad (n \in \mathbb{N}).$$

Proof of Lemma ▷ It is sufficient, by Proposition 3.8, to show that ϵ is the smallest unit > 1. And from the proof of that Proposition, we need only consider units of the form

$$a + b\sqrt{3}$$

with $a, b \geq 0$.

Thus the only possible units in the range $(1, \epsilon)$ are $\sqrt{3}$ and $1 + \sqrt{3} = \eta$, neither of which is in fact a unit, since

$$\mathcal{N}(\sqrt{3}) = -3, \mathcal{N}(\eta) = -2,$$

whereas a unit must have norm ± 1, by Proposition 3.6. ◁

Lemma 4.4 *If r_n is the sequence defined in the Proposition then*

$$r_n = \epsilon^{2^{n-1}} + \epsilon^{-2^{n-1}}$$

for each $n \geq 1$.

Proof of Lemma ▷ Let us set

$$s_n = \epsilon^{2^{n-1}} + \epsilon^{-2^{n-1}}$$

for $n \geq 1$. Then

$$\begin{aligned} s_n^2 &= \left(\epsilon^{2^{n-1}} + \epsilon^{-2^{n-1}}\right)^2 \\ &= \epsilon^{2^n} + 2 + \epsilon^{-2^n} \\ &= s_{n+1} + 2, \end{aligned}$$

ie

$$s_{n+1} = s_n^2 - 2.$$

Also

$$\begin{aligned} s_1 &= \epsilon + \epsilon^{-1} \\ &= \epsilon + \bar{\epsilon} \\ &= 4. \end{aligned}$$

We conclude that

$$r_n = s_n = \epsilon^{2^{n-1}} + \epsilon^{-2^{n-1}}$$

for all $n \geq 1$. ◁

Suppose first that $P = M_p$ is prime.

Lemma 4.5 *We have*

$$\left(\frac{3}{P}\right) = -1.$$

Proof of Lemma ▷ We have

$$\begin{aligned} M_p &= 2^p - 1 \\ &\equiv (-1)^p - 1 \bmod 3 \\ &\equiv -1 - 1 \bmod 3 \\ &\equiv 1 \bmod 3; \end{aligned}$$

while

$$M_p \equiv -1 \bmod 4.$$

By the Chinese Remainder Theorem there is just one remainder $\bmod 12$ with these remainders $\bmod 3$ and $\bmod 4$; and that is $7 \equiv -5 \bmod 12$. For any odd prime p,

$$M_p \equiv 7 \bmod 12$$

Hence

$$\left(\frac{3}{P}\right) = -1.$$

by Proposition 3.18, ◁

It follows from this Lemma and Proposition 4.2 that

$$\alpha^P \equiv \bar{\alpha} \bmod P$$

for all $\alpha \in \mathbb{Z}[\sqrt{3}]$. In particular,

$$\epsilon^P \equiv \bar{\epsilon} \bmod P.$$

Hence

$$\begin{aligned} \epsilon^{P+1} &\equiv \epsilon\bar{\epsilon} \bmod P \\ &\equiv \mathcal{N}(\epsilon) \bmod P \qquad \equiv 1 \bmod P. \end{aligned}$$

In other words,

$$\epsilon^{2^p} \equiv 1 \bmod P.$$

It follows that

$$\epsilon^{2^{p-1}} \equiv \pm \bmod P.$$

We want to show that in fact

$$\epsilon^{2^{p-1}} \equiv -1 \bmod P.$$

This is where things get a little trickier than in the first version of the Lucas-Lehmer test. In effect, we need a number with negative norm. To this end we introduce

$$\eta = 1 + \sqrt{3}.$$

Lemma 4.6 *1.* $\mathcal{N}(\eta) = -2$.

2. $\eta^2 = 2\epsilon$.

Proof of Lemma ▷ This is a matter of simple verification:

$$\mathcal{N}(\eta) = 1 - 3 = -2,$$

while

$$\begin{aligned}\eta^2 &= (1+\sqrt{3})^2 \\ &= 4 + 2\sqrt{3} \\ &= 2\epsilon.\end{aligned}$$

◁

By Proposition /refMersenneLemma,

$$\eta^P \equiv \bar{\eta} \bmod P,$$

and so

$$\eta^{P+1} \equiv \eta\bar{\eta} - 2 \bmod P,$$

ie

$$\eta^{2^p} \equiv -2 \bmod P.$$

By the Lemma, this can be written

$$(2\epsilon)^{2^{p-1}} \equiv -2 \bmod P,$$

ie

$$2^{2^{p-1}}\epsilon^{2^{p-1}} \equiv -2 \bmod P,$$

But by Proposition 3.14,

$$\begin{aligned}2^{\frac{P-1}{2}} = 2^{2^{p-1}-1} &\equiv \left(\frac{2}{P}\right) \bmod P \\ &\equiv 1 \bmod P,\end{aligned}$$

by Proposition 3.17, since

$$P = 2^p - 1 \equiv -1 \bmod 8.$$

Thus

$$2^{2^{p-1}} \equiv 2 \bmod P$$

and so

$$2\epsilon^{2^{p-1}} \equiv -2 \bmod P.$$

Hence

$$\epsilon^{2^{p-1}} \equiv -1 \bmod P.$$

Thus

$$\epsilon^{2^{p-1}} + 1 \equiv 0 \bmod P.$$

Dividing by $\epsilon^{2^{p-2}}$,

$$\epsilon^{2^{p-2}} + \epsilon^{-2^{p-2}} \equiv 0 \bmod P,$$

ie

$$r_{p-1} \equiv 0 \bmod P.$$

Conversely, suppose P is a prime factor of M_p. Then

$$\begin{aligned} M_p \mid r_{p-1} &\implies r_{p-1} \equiv 0 \bmod P \\ &\implies \epsilon^{2^{p-2}} + \epsilon^{-2^{p-2}} \equiv 0 \bmod P \\ &\implies \epsilon^{2^{p-1}} + 1 \equiv 0 \bmod P \\ &\implies \epsilon^{2^{p-1}} \equiv -1 \bmod P. \end{aligned}$$

But (by the argument we used in the proof of the first Lucas-Lehmer test) this implies that the order of $\epsilon \bmod P$ is 2^p.

On the other hand, by the Corollary to Proposition 4.2,

$$\epsilon^{P^2} \equiv \epsilon \bmod P \implies \epsilon^{P^2-1} \equiv 1 \bmod P.$$

Hence

$$2^p \mid P^2 - 1 = (P+1)(P-1).$$

Now

$$\gcd(P+1, P-1) = 2.$$

It follows that

$$2^{p-1} \mid P+1 \quad \text{or} \quad 2^{p-1} \mid P-1.$$

In either case,

$$\begin{aligned} 2^{p-1} \leq P+1 &\implies P \geq 2^{p-1} - 1 = \frac{M_p - 1}{2} \\ &\implies P \geq \frac{M_p}{3} \\ &\implies \frac{M_p}{P} < 3. \end{aligned}$$

Since M_p is odd, this implies that

$$P = M_p,$$

ie M_p is prime. ◀

4.1.2 Perfect numbers

Mersenne numbers are also of interest because of their intimate connection with *perfect* numbers.

Definition 4.2 *For $n \in \mathbb{N},\ n > 0$ we denote the number of divisors of n by $d(n)$, and the sum of these divisors by $\sigma(n)$.*

Example: Since 12 has divisors $1, 2, 3, 4, 6, 12$,

$$d(12) = 6,\ \sigma(12) = 28.$$

Definition 4.3 *The number $n \in \mathbb{N}$ is said to be* perfect *if*

$$\sigma(n) = 2n,$$

ie if n is the sum of its proper divisors.

Example: The number 6 is perfect, since

$$6 = 1 + 2 + 3.$$

Proposition 4.5 *If*

$$M_p = 2^p - 1$$

is a Mersenne prime then

$$2^{p-1}(2^p - 1)$$

is perfect.

Conversely, every even *perfect number is of this form.*

Proof ► In number theory, a function $f(n)$ defined on $\{n \in \mathbb{N} : n > 0\}$ is said to be *multiplicative* if

$$\gcd(m, n) = 1 \Longrightarrow f(mn) = f(m)f(n).$$

If the function $f(n)$ is multiplicative, and

$$n = p_1^{e_1} \cdots p_r^{e_r}$$

then

$$f(n) = f(p_1^{e_1}) \cdots f(p_r^{e_r}).$$

Thus the function $f(n)$ is completely determined by its value $f(p^e)$ for prime powers.

Lemma 4.7 *The functions $d(n)$ and $\sigma(n)$ are both multiplicative.*

Proof of Lemma ▷ Suppose $\gcd(m, n) = 1$; and suppose

$$d \mid mn.$$

Then d is uniquely expressible in the form

$$d = d_1 d_2 \qquad (d_1 \mid m,\ d_2 \mid n).$$

In fact

$$d_1 = \gcd(d, m),\ d_2 = \gcd(d, n).$$

It follows that

$$d(mn) = d(m)d(n);$$

and

$$\begin{aligned}\sigma(mn) &= \sum_{d \mid mn} d \\ &= \sum_{d_1 \mid m} d_1 \sum_{d_2 \mid n} d_2 \\ &= \sigma(m)\sigma(n).\end{aligned}$$

◁

Now suppose

$$n = 2^{p-1} M_p$$

where M_p is prime. Since M_p is odd,

$$\gcd(2^{p-1}, M_p) = 1.$$

Hence

$$\sigma(n) = \sigma(2^{p-1})\sigma(M_p).$$

If P is prime then evidently

$$\sigma(P) = 1 + P.$$

On the other hand,

$$\sigma(P^e) = 1 + P + P^2 + \cdots + P^e = \frac{P^{e+1} - 1}{P - 1}.$$

In particular,

$$\sigma(2^e) = 2^{e+1} - 1.$$

Thus

$$\sigma(2^{p-1}) = 2^p - 1 = M_p,$$

while

$$\sigma(M_p) = M_p + 1 = 2^p.$$

We conclude that

$$\sigma(n) = 2^p M_p = 2n.$$

Conversely, suppose n is an even perfect number. We can write n (uniquely) in the form

$$n = 2^e m$$

where m is odd. Since 2^e and m are coprime,

$$\sigma(n) = \sigma(2^e)\sigma(m) = (2^{e+1} - 1)\sigma(m).$$

On the other hand, if n is perfect then

$$\sigma(n) = 2n = 2^{e+1}m.$$

Thus

$$\frac{2^{e+1} - 1}{2^{e+1}} = \frac{m}{\sigma(m)}.$$

The numerator and denominator on the left are coprime. Hence

$$m = d(2^{e+1} - 1),\ \sigma(m) = d2^{e+1},$$

for some $d \in \mathbb{N}$.

If $d > 1$ then m has at least the factors $1, d, m$. Thus

$$\sigma(m) \geq 1 + d + m = 1 + d2^{e+1},$$

contradicting the value for $\sigma(m)$ we derived earlier.

It follows that $d = 1$. But then

$$\sigma(m) = 2^{e+1} = m + 1.$$

Thus the only factors of m are 1 and m, ie

$$m = 2^{e+1} - 1 = M_{e+1}$$

is prime. Setting $e + 1 = p$, we conclude that

$$n = 2^{p-1} M_p,$$

where M_p is prime. ◄

It is an unsolved problem whether or not there are any *odd* perfect numbers.

The first 4 even perfect numbers are

$$2^1 M_2 = 6,\ 2^2 M_3 = 28,\ 2^4 M_5 = 496,\ 2^6 M_7 = 8128.$$

(In fact these are the first 4 perfect numbers, since it is known that any odd perfect number must have at least 300 digits!)

4.2 Fermat numbers

Proposition 4.6 *If*

$$n = a^m + 1 \quad (a, m > 1)$$

is prime then

1. $a2$ *is even;*
2. $m = 2^e$.

Proof ► If a is odd then n is even and > 2, and so not prime.

Suppose m has an odd factor, say

$$m = rs,$$

where r is odd. Since $x^r + 1 = 0$ when $x = -1$, it follows by the Remainder Theorem that

$$(x + 1) \mid (x^r + 1).$$

Explicitly,

$$x^r + 1 = (x + 1)(x^{r-1} - x^{r-2} + \cdots - x + 1).$$

Substituting $x = y^s$,

$$(y^s + 1) \mid (y^m + 1)$$

in $\mathbb{Z}[x]$. Setting $y = a$,

$$(a^s + 1) \mid (a^{rs} + 1) = (a^m + 1).$$

In particular, $a^m + 1$ is not prime.

Thus if $a^m + 1$ *is* prime then m cannot have any odd factors. In other words,

$$m = 2^e.$$

◄

Definition 4.4 *The numbers*

$$F_n = 2^{2^n} + 1 \qquad (n = 0, 1, 2, \dots)$$

are called Fermat numbers.

Fermat hypothesized — he didn't claim to have a proof — that all the numbers

$$F_0, F_1, F_2, \dots$$

are prime. In fact this is true for

$$F_0 = 3,\ F_1 = 5,\ F_2 = 17,\ F_3 = 257,\ F_4 = 65537.$$

However, Euler showed in 1747 that

$$F_5 = 2^{32} + 1 = 4294967297$$

is composite. In fact, no Fermat prime beyond F_4 has been found.

The heuristic argument we used above to suggest that the number of Mersenne primes is probably infinite now suggests that the number of Fermat primes is probably finite.

For by the Prime Number Theorem, the probability of F_n being prime is

$$\begin{aligned} &\approx 2/\log F_n \\ &\approx 2 \cdot 2^{-n}. \end{aligned}$$

Thus the expected number of Fermat primes is

$$2 \approx \sum 2^{-n} = 4 < \infty.$$

This argument assumes that the Fermat numbers are "independent", as far as primality is concerned. It might be argued that our next result shows that this is not so. However, the Fermat numbers are so sparse that this does not really affect our heuristic argument.

Proposition 4.7 *The Fermat numbers are coprime, ie*

$$\gcd(F_m, F_n) = 1$$

if $m \neq n$.

Proof ► Suppose

$$\gcd(F_m, F_n) > 1.$$

Then we can find a prime p (which must be odd) such that

$$p \mid F_m, \ p \mid F_n.$$

Now the numbers $\{1, 2, \ldots, p-1\}$ form a group $(\mathbb{Z}/p)^{\times}$ under multiplication $\bmod p$. Since $p \mid F_m$,

$$2^{2^m} \equiv -1 \bmod p.$$

It follows that the order of $2 \bmod p$ (ie the order of 2 in $(\mathbb{Z}/p)^{\times}$) is exactly 2^{m+1}. For certainly

$$2^{2^{m+1}} = (2^{2^m})^2 \equiv 1 \bmod p;$$

and so the order of 2 divides 2^{m+1}, ie it is 2^e for some $e \leq m+1$. But if $e \leq m$ then

$$2^{2^m} \equiv 1 \bmod p,$$

whereas we just saw that the left hand side was $\equiv -1 \bmod p$. We conclude that the order must be 2^{m+1}.

But by the same token, the order is also 2^{n+1}. This is a contradiction, unless $m = n$. ◄

We can use this result to give a second proof of Euclid's Theorem that there are an infinity of primes.

Proof ► Each Fermat number F_n has at least one prime divisor, say q_n. But by the last Proposition, the primes

$$q_0, q_1, q_2, \ldots$$

are all distinct. ◄

We end with a kind of pale imitation of the Lucas-Lehmer test, but now applied to Fermat numbers.

Proposition 4.8 *The Fermat number*

$$F_n = 2^{2^n} + 1$$

is prime if and only if

$$3^{\frac{F_n - 1}{2}} \equiv -1 \bmod F_n.$$

Proof ► Suppose $P = F_n$ is prime.

Lemma 4.8 *We have*

$$F_n \equiv 5 \bmod 12.$$

Proof of Lemma ▷ Evidently

$$F_n \equiv 1 \bmod 4;$$

while

$$\begin{aligned} F_n &\equiv (-1)^{2^n} + 1 \bmod 3 \\ &\equiv 2 \bmod 3. \end{aligned}$$

By the Chinese Remainder Theorem these two congruences determine $F_n \bmod 12$; and observation shows that

$$F_n \equiv 5 \bmod 12.$$

◁

It follows from this Lemma, and Proposition 3.18, that

$$\left(\frac{3}{P}\right) = -1.$$

Hence

$$3^{\frac{P-1}{2}} \equiv -1 \bmod P$$

by Proposition 3.14.

Conversely, suppose

$$3^{\frac{F_n-1}{2}} \equiv -1 \bmod F_n;$$

and suppose P is a prime factor of F_n. Then

$$3^{\frac{F_n-1}{2}} \equiv -1 \bmod P,$$

ie

$$3^{2^{2^n}-1} \equiv -1 \bmod P.$$

It follows (as in the proof of the Lucas-Lehmer theorems) that the order of $3 \bmod P$ is

$$2^{2^n}.$$

But by Fermat's Little Theorem,

$$3^{P-1} \equiv 1 \bmod P.$$

Hence

$$2^{2^n} \mid P - 1,$$

ie

$$F_n - 1 \mid P - 1.$$

Since $P \mid F_n$ this implies that

$$F_n = P,$$

ie F_n is prime. ◀

This test is more-or-less useless, even for quite small n, since it will take an inordinate time to compute the power, even working modulo F_n. However, it does give a short proof — which we leave to the reader — that F_5 is composite.

It may be worth noting why this test is simpler than its Mersenne analogue. In the case of Mersenne primes $P = M_p$ we had to introduce quadratic fields because the analogue of Fermat's Little Theorem,

$$\alpha^{P^2-1} \equiv 1 \bmod P,$$

then allowed us to find elements of order $P+1 = 2^p$. In the case of Fermat primes $P = F_n$ Fermat's Little Theorem

$$a^{P-1} = a^{2^{2^n}} \equiv 1 \bmod P$$

suffices.