UNIVERSITY OF DUBLIN
TRINITY COLLEGE
SCHOOL OF MATHEMATICS

2006 Course 4281
Prime Numbers

Timothy Murphy

Bernhard Riemann 1826–1866

Chapter 1

From Euclid to Gauss

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 $p > 1$ and 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 ▶

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

Proof ▶ 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. ◀

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. But by the lemma, N has a prime factor p, which must therefore differ 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

$$p_n \leq x \iff \pi(x) \geq n.$$

In the present case, for example, we deduce that

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

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

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

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.)

In the second part of the course we shall prove the *Prime Number Theorem*, which 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 $1/\log n$. Thus roughly 1 in 11.5 numbers around 10^6 are prime; while roughly 1 in 23 around 10^{12} are prime.

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 $m, n \in \mathbb{N}$, $m \neq 0$. Then there exist unique $q.r \in \mathbb{N}$ such that*

$$n = qm + r, \quad 0 \leq r < m.$$

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

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

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,

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

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

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

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 $x' - x,\ y' - y$ 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.

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 The Fundamental Theorem of Arithmetic

Proposition 1.3. *(Euclid's Lemma) Suppose $p \in \mathbb{N}$ is a prime number; and suppose $a, b \in \mathbb{Z}$. Then*

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

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, 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

$$n = p q_1 q_2 \cdots q_s.$$

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 $n \leq N$ 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 $(1 \leq i \leq m)$,

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

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

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

So our hypothesis implies that

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

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 $x \geq 3$, 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 $n \leq N$ if

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

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

Our proof gives the bound

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

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 $m \geq 2$. 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 $n \leq N$. Since

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

the number of numbers of the above form is

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

Thus we shall reach a contradiction when

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

ie

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

This gives us the bound

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

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

1.4 Fermat numbers

Numbers of the form $2^m \pm 1$ (or more generally $a^m \pm 1$) have an honoured place in the history of prime number theory, and continue to be of relevance, because there are special tests for determining their primality, and these lead to the "discovery" of enormously large primes.

Proposition 1.4. *Suppose*

$$n = a^e + 1,$$

where $a > 1$, $e > 1$. *If* n *is prime then* a *is even, and*

$$e = 2^m$$

for some m.

Proof ▶ If a is odd then n is even and > 2, and so not prime.

Suppose e has an odd factor, say

$$e = rs,$$

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 $x = y^s$,

$$(y^s + 1) \mid (y^n + 1).$$

Setting $y = a$,

$$(a^s + 1) \mid (a^{rs} + 1) = (a^n + 1).$$

In particular, $a^n + 1$ is not prime.

Thus if $a^n + 1$ *is* prime then n cannot have any odd factors. In other words,

$$n = 2^m.$$

◀

Definition 1.5. *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

$$F_0, F_1, F_2, \ldots$$

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.

We shall see later that there is a simple test for the primality of F_n, which allows us to conclude at once F_5 is composite.

There is a kind of argument — it is certainly not a proof — that the number of Fermat primes is finite. It runs as follows. By the Prime Number Theorem, the probability of F_n being prime is approximately

$$1/\log F_n \approx 2^{-n}.$$

Thus the expected number of Fermat primes is approximately

$$\sum 2^{-n} = 2 < \infty.$$

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.

Proposition 1.5. *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 mod$p$. Since $p \mid F_m$,

$$2^{2^m} \equiv -1 \bmod p.$$

It follows that the order of 2 mod 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. ◀

1.5 Mersenne numbers

Proposition 1.6. *Suppose*

$$n = a^e - 1,$$

where $a > 1$, $e > 1$. *If* n *is prime then* $a = 2$ *and* p *is prime.*

Proof ▶ In the first place,

$$(a-1) \mid (a^e - 1);$$

so if $a > 2$ then n is certainly not prime.

Suppose $n = rs$, where $r, s > 1$.

Since $x^r - 1 = 0$ when $x = 1$, it follows (from the Remainder Theorem) that

$$(x-1) \mid (x^r - 1).$$

Explicitly,

$$x^r - 1 = (x-1)(x^{r-1} + x^{r-2} + x^{r-3} + \cdots + 1).$$

Subsitituting x^s for x,

$$(x^s - 1) \mid (x^{rs} - 1) = (x^e - 1).$$

Setting $x = a$,

$$(a^s - 1) \mid (a^n - 1).$$

Thus if $a^n - 1$ is prime then n has no proper factors, ie n is prime. ◀

Definition 1.6. *The numbers*

$$F_n = 2^p - 1,$$

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.)

We shall provide an algorithm for determining whether or not the Mersenne M_p is prime. This is in fact the source of all the recent "record" primes.

1.6 Perfect numbers

Mersenne numbers are also of interest because of their intimate connection with *perfect* numbers.

Definition 1.7. *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 1.8. *A function $f(n)$ defined on $\{n \in \mathbb{N} : n > 0\}$ is said to be* multiplicative *if*

$$\gcd(m, n) = 1 \implies f(mn) = f(m)f(n).$$

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.

Multiplicative functions will play an important rôle in our later work.

Proposition 1.7. *The functions $d(n)$ and $\sigma(n)$ are both multiplicative.*

Proof ▶ 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

$$\begin{aligned}\sigma(mn) &= \sum_{d\mid mn} d \\ &= \sum_{d_1\mid m} d_1 \sum_{d_2\mid n} d_2 \\ &= \sigma(m)\sigma(n).\end{aligned}$$

◀

Definition 1.9. *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 1.8. *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 ▶ 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!)

Chapter 2

Gaussian integers

Although our principal object of study remains the "classic" primes $2, 3, 5, 7, \ldots$ in $\mathbb{N}$, it is both interesting and instructive to consider primality in a wider context.

Let A be an *integral domain*, ie a commutative ring (with 1) having no zero divisors, ie

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

If $a, b \in A$, we say that b divides a, and write $b \mid a$, if

$$a = bc$$

for some $cinA$.

Definition 2.1. *Suppose A is an integral domain. An element $u \in A$ is said to be a* unit *if it is invertible, ie there exists a $v \in A$ such that*

$$uv = 1.$$

In other words, u is a unit if $u \mid 1$. The units in A form a commutative group $A^\times$. For example,

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

Definition 2.2. *Suppose A is an integral domain. We say that $a, b \in A$ are* equivalent*, and we write $a \sim b$, if*

$$b = au$$

for some unit $u \in A^\times$.

In general we do not distinguish between equivalent divisors; for if $a \sim b$ then a and b divide exactly the same elements.

Definition 2.3. *Suppose A is an integral domain. The element $p \in A$ is said to be prime if it is not a unit, but every divisor of p is equivalent to 1 or to p itself:*

$$a \mid p \implies a \sim 1 \text{ or } a \sim p.$$

In $\mathbb{Z}$, for example, $\pm 2, \pm 3, \pm 5, \ldots$ are pairs of equivalent primes; and every number $n \in \mathbb{Z},\ n \neq 0$ is uniquely expressible in the form

$$n = \epsilon 2^{e_2} 3^{e_3} \ldots,$$

where $\epsilon = \pm 1$. (We adopt the convention that an empty product has value 1; so $n = 1$ is included as the product of 0 primes.)

2.1 Gaussian numbers

Definition 2.4. *A* Gaussian integer *is a complex number of the form*

$$z = m + ni \qquad (m, n \in \mathbb{Z}).$$

We denote the Gaussian integers by $\mathbb{Z}[i]$.

A Gaussian rational *is a complex number of the form*

$$z = x + yi \qquad (x, y \in \mathbb{Q}).$$

We denote the Gaussian rationals by $\mathbb{Q}(i)$.

Proposition 2.1. *The Gaussian integers* $\mathbb{Z}[i]$ *form an integral domain.*

The Gaussian rationals $\mathbb{Q}(i)$ *form a field,*

Proof ▶ It is clear that $\mathbb{Z}[i]$ and $\mathbb{Q}(i)$ are closed under addition and multiplication. To see that $\mathbb{Q}(i)$ is closed under division by non-zero elements, note that if $z = x + yi, w = X + Yi$ then

$$\begin{aligned} \frac{z}{w} &= \frac{x + yi}{X + Yi} \\ &= \frac{(x + yi)(X - iY)}{(X + Yi)(X - iY)} \\ &= \frac{xX - yY}{X^2 + Y^2} + \frac{yX - xY}{X^2 + Y^2} i. \end{aligned}$$

◀

Recall that we can always extend a integral domain A to its *field of fractions* F, in exactly the same way that we extend $\mathbb{Z}$ to $\mathbb{Q}$. Thus each element of F is expressible as a/b where $a, b \in A$ with $b \neq 0$; and

$$a/b = c/d \iff ad = bc.$$

In particular, we can identify the Gaussian rationals $\mathbb{Q}(i)$ with the field of fractions of the Gaussian integers $\mathbb{Z}[i]$.

Each Gaussian rational $z = x + iy \in \mathbb{Q}[i]$ has a *conjugate*

$$\bar{z} = x - iy$$

in $\mathbb{Q}[i]$. The map $z \mapsto \bar{z}$ is an *automorphism* of $\mathbb{Q}[i]$, sending the Gaussian integers $\mathbb{Z}[i]$ into themselves. Moreover, if $z = a + bi \in \mathbb{Z}]i]$ then

$$|z|^2 = z\bar{z} = a^2 + b^2 \in \mathbb{N}.$$

Proposition 2.2. *There are just 4 units in $\mathbb{Z}[i]$, namely $\pm 1, \pm i$.*

Proof ▶

Lemma 2. *The number*

$$u = a + bi \in \mathbb{Z}[i]$$

is a unit if and only if

$$u\bar{u} = a^2 + b^2 = 1.$$

Proof ▶ If

$$\begin{aligned} p &= a^2 + b^2 \\ &= (a + bi)(a - bi) \end{aligned}$$

we have an explicit factoring of p.

Conversely, suppose p splits, say

$$p = \pi_1 \pi_2.$$

Then

$$p^2 = |\pi_1|^2 |\pi_2|^2.$$

It follows that

$$|\pi_1|^2 = p = |\pi_2|^2$$

Thus if $\pi_1 = a + bi$,

$$p = a^2 + b^2.$$

◀

If $a, b \in \mathbb{Z}$,

$$a^2 + b^2 = 1 \implies a = 0,\ b = \pm 1 \text{ or } a = \pm 1,\ b = 0,$$

giving the 4 units $\pm 1, \pm i$. ◀

Proposition 2.3. *Every Gaussian integer $z = a + ib$ factorises into primes (modulo the units).*

Proof ▶ We prove this by induction on $|z|^2$. It is certainly true if $|z|^2 = 1$, since then z is a unit, as we have seen.

Suppose $|z|^2 > 1$. If z is prime, there is nothing to prove. If not, then

$$z = st$$

where neither s nor t is a unit. But then

$$|s|^2, |t|^2 > 1 \implies |s|^2, |t|^2 < |z|^2.$$

According to our inductive hypothesis, both s and t can be expressed as a product of primes, and these combine to give an expression for $z = st$ as a product of primes. ◀

Remark. This does not, of course, prove that the expression for z as a product of primes is *unique*. We shall establish that shortly.

Proposition 2.4. *A prime number $p \in \mathbb{N}$ splits into at most 2 prime factors in $\mathbb{Z}[i]$.*

Proof ▶ Suppose

$$p = \pi_1 \pi_2 \cdots \pi_r.$$

Then

$$|\pi_1|^2 |\pi_2|^2 \cdots |\pi_r|^2 = p^2.$$

It follows, by the Unique Factorisation Theorem for $\mathbb{N}$, that

$$|\pi_i|^2 = 1$$

for all but 1 or 2 of the i. (In fact, either one of them takes the value p^2, and the rest are 1; or two of them take the value p, and the rest are 1.)

But we have seen that if $|z|^2 = 1$ then z is a unit, and not a prime. Hence $r \leq 2$. ◀

If $\pi = a + ib \in \mathbb{Z}[i]$ is prime then so is its conjugate $\bar{\pi} = a - ib$; for any factorisation of π gives a factorisation of $\bar{pi}$ and vice versa:

$$\pi = zw \implies \bar{\pi} = \bar{z}\bar{\pi}.$$

Thus the primes in $\mathbb{Z}[i]$ divide into two classes: self-conjugate primes, for which

$$\bar{\pi} \sim \pi,$$

and conjugate prime-pairs $\pi, \bar{\pi}$.

Suppose $\pi = a + ib$ is self-conjugate, ie

$$\bar{\pi} \sim \pi.$$

Then

$$a - ib = \epsilon(a + ib)$$

where $\epsilon \in \{\pm 1, \pm i\}$.

If $a = 0$ or $b = 0$ then $\pi \sim p \in \mathbb{N}$. It is clear that p is prime number (in the classic sense), since any factorisation of p in $\mathbb{N}$ is *a fortiori* a factorisation of π in $\mathbb{Z}[i]$.

If however $a, b \neq 0$ then

$$\bar{\pi} = a - ib \neq \pm(a + ib);$$

and so

$$\bar{pi} \sim \pi \implies (a - ib) = \pm i(a + ib) \implies b = \pm a.$$

But in this case $a \mid \pi$; so if π is prime then $a = \pm 1$, and then $b = \pm 1$. Thus

$$\pi = \pm 1 \pm i.$$

In fact these are all equivalent:

$$1 + i \sim 1 - i \sim -1 + i \sim -1 - i.$$

Thus there is *just one self-conjugate prime*, apart from those in $\mathbb{N}$, namely $\pi = 1 + i$.

This prime is a factor, in fact a double factor, of 2:

$$2 = -i(1 + i)^2 \sim \pi^2,$$

where $\pi = 1 + i$.

Thus 2 becomes a prime-square in $\mathbb{Z}[i]$ (up to equivalence), while each odd prime p must either remain prime in $\mathbb{Z}[i]$, or else split into 2 distinct (and conjugate) primes.

Proposition 2.5. *The prime number $n \in \mathbb{N}$ splits in $\mathbb{Z}[i]$ if and only if it is expressible in the form*

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

Proof ▶ If $p = a^2 + b^2$ then we have an explicit factorisation:

$$p = (a + ib)(a - ib).$$

Conversely, suppose p splits in $\mathbb{Z}[i]$. We have seen that it must split into 2 conjugate primes:

$$p \sim \pi\bar{\pi}.$$

Let $\pi = a + ib$. Then

$$p = \epsilon(a^2 + b^2).$$

Since both p and $a^2 + b^2$ are positive integers, it follows that $\epsilon = 1$, ie

$$p = a^2 + b^2.$$

◀

It remains to determine which odd primes split, and which remain inviolate in $\mathbb{Z}[i]$.

Note too that we have left another question open. We have not shown that every prime π in $\mathbb{Z}[i]$ arises in this way, as a factor of some prime number $p \in \mathbb{N}$.

Above all, we have not shown that prime factorisation in $\mathbb{Z}[i]$ is *unique*. We establish this by showing that the euclidean algorithm (suitably modified) can still be applied in $\mathbb{Z}[i]$.

2.2 The Euclidean Algorithm for $\mathbb{Z}[i]$

Suppose $z, w \in \mathbb{Z}[i]$. It is not immediately obvious that how to divide z by w "with remainder". However, we can accomplish this as follows. Let

$$z/w = x + iy$$

where $x, y \in \mathbb{Q}$. Choose the closest integers $a, b \in \mathbb{Z}$ to x, y. To remove ambiguity, let

$$x - 1/2 \leq a < x + 1/2, \quad y - 1/2 \leq b < b + 1/2.$$

Set

$$q = a + bi;$$

and let

$$r = z - qw.$$

Evidently

$$q, r \in \mathbb{Z}[i].$$

We have

$$r/w = z/w - q = s + it,$$

where

$$s = x - a,\ t = y - b.$$

Now

$$|s| \leq 1/2,\ |t| \leq 1/2;$$

and so

$$|r/w|^2 = s^2 + t^2 \leq (1/2)^2 + (1/2)^2 = 1/2.$$

In other words,

$$|r|^2 \leq |w|^2/2.$$

Thus we have established

Proposition 2.6. *Given $z, w \in \mathbb{Z}[i]$ with $w \neq 0$ we can find $q, r \in \mathbb{Z}[i]$ such that*

$$z = qw + r, \quad |r|^2 < |q|^2.$$

This is sufficient to allow us to carry the Euclidean Algorithm over to $\mathbb{Z}[i]$.

Example. Let

$$z = 7 + 3i,\ w = 3 - 5i.$$

Since

$$|z|^2 = 58,\ |w|^2 = 34$$

we begin by dividing z by w:

$$\begin{aligned}\frac{z}{w} &= \frac{7+3i}{3-5i} \\ &= \frac{(7+3i)(3+5i)}{(3-5i)(3+5i)} \\ &= \frac{6+44i}{34}.\end{aligned}$$

The closest gaussian integer is

$$q_0 = i;$$

and then the remainder is

$$r_0 = z - q_0 w = 2.$$

Now

$$\frac{w}{r_0} = \frac{3}{2} - \frac{5}{2}i.$$

The closest gaussian integer is

$$q_1 = 1 - 3i,$$

giving

$$\begin{aligned}r_1 &= w - q_1 r_0 \\ &= 1 + i.\end{aligned}$$

Continuing in the same fashion,

$$\begin{aligned}\frac{r_0}{r_1} &= \frac{2}{1+i} \\ &= \frac{2(1-i)}{(1+i)(1-i} \\ &= 1 - i.\end{aligned}$$

Since this lies in $\mathbb{Z}[i]$ we have an exact division, with remainder 0:

$$r_0 = q_2 r_1$$

with $q_2 = 1 - i$.

We conclude that

$$\gcd(z, w) = 1 - i,$$

the last non-zero remainder.

Note that since the gcd is only defined up to a unit, we could equally well say that

$$\gcd(z, w) = i(1 - i) = 1 + i.$$

This extension of the Euclidean Algorithm to $\mathbb{Z}[i]$ allows us to assert

Theorem 2.1. *(The Fundamental Theorem of Arithmetic for the Gaussian integers $\mathbb{Z}[i]$) Each number $z = a + bi \in \mathbb{Z}[i]$ is equivalent to a product of primes:*

$$z = \epsilon\pi_1\pi_2\cdots\pi_r,$$

where ϵ is a unit. Moreover the primes π_i are uniquely defined up to order and multiplication by units.

Note that if $r > 0$, ie z is not itself a unit, then we can absorb ϵ into one of the primes, and express z in the form

$$z = \pi_1\pi_2\cdots\pi_r.$$

2.3 The primes in $\mathbb{Z}[i]$

Proposition 2.7. *Each prime $\pi \in \mathbb{Z}[i]$ divides a unique prime number $p \in \mathbb{N}$.*

Proof ▶ Suppose $\pi = a + bi$. Then

$$\pi \mid \pi\bar{\pi} = |\pi|^2 = a^2 + b^2.$$

Let

$$a^2 + b^2 = p_1p_2\cdots p_s$$

in $\mathbb{N}$. Then π must divide one of the primes p_j, by the Fundamental Theorem for $\mathbb{Z}[i]$. This prime is unique; for suppose

$$\pi \mid p,\ \pi \mid q$$

where p, q are distinct primes. Then

$$\gcd(p, q) = 1.$$

Hence we can find $x, y \in \mathbb{Z}$ such that

$$px + qy = 1.$$

But then

$$\pi \mid p, q \implies \pi \mid 1,$$

ie π is a unit, contrary to the definition of a prime. ◀

Proposition 2.8. *A prime number $p \in \mathbb{N}$ splits into at most 2 prime factors in $\mathbb{Z}[i]$.*

Proof ▶ Suppose

$$p = \pi_1 \pi_2 \cdots \pi_r.$$

Then

$$p^2 = |p|^2 = |\pi_1|^2 |\pi_2|^2 \cdots |\pi_r|^2.$$

Since

$$|\pi_i|^2 > 1$$

for each i, it follows from the Fundamental Theorem for $\mathbb{N}$ that there are at most 2 terms on the right. ◀

Proposition 2.9. *The prime number $p \in \mathbb{N}$ splits in $\mathbb{Z}[i]$ if and only if p is expressible as a sum of two squares:*

$$p = a^2 + b^2 \quad (a, b \in \mathbb{N}).$$

Proof ▶ If p is of this form then

$$p = (a + ib)(a - ib)$$

is an explicit split.

Conversely, suppose p splits, say

$$p = \pi_1 \pi_2.$$

(We can absorb any unit into π_1.) Then

$$p^2 = |p|^2 = |\pi_1|^2 |\pi_2|^2.$$

It follows from the Fundamental Theorem for $\mathbb{N}$ that

$$|\pi_1|^2 = p = |\pi_2|^2.$$

But then, if $\pi_1 = a + bi$,

$$p = |\pi_1|^2 = a^2 + b^2.$$

◀

2.4 Quadratic residues

It is convenient at this point to introduce the notion of a *quadratic residue*, which will play a central rôle in much of our work.

Definition 2.5. *Suppose $p \in \mathbb{N}$ is a prime number; and suppose $a \in \mathbb{Z}$, $p \nmid a$. Then we say that a is a* quadratic residue mod$p$*, and we write*

$$\left(\frac{a}{p}\right) = 1,$$

if we can find $b \in \mathbb{Z}$ *such that*

$$a \equiv b^2 \bmod p.$$

If there is no such b *then we say that* p *is a* quadratic non-residue*, and we write*

$$\left(\frac{a}{p}\right) = -1.$$

We call $\left(\frac{a}{p}\right)$ the *Legendre symbol* of a mod p. We shall sometimes write

$$\left(\frac{a}{p}\right) = 0$$

if $p \mid a$.

Example. Suppose $p = 7$. Then

$$1^2 \equiv 1,\ 2^2 \equiv 4,\ 3^2 \equiv 2.$$

It follows that

$$4^2 \equiv (-3)^2 = 3^2,\ 5^2 \equiv (-2)^2 = 2^2,\ 6^2 \equiv (-1)^2 = 1^2.$$

Thus we conclude that there are just 3 quadratic residues mod7, namely $1, 2, 4$, ie

$$\left(\frac{1}{7}\right) = 1,\ \left(\frac{2}{7}\right) = 1,\ \left(\frac{3}{7}\right) = -1,\ \left(\frac{4}{7}\right) = 1,\ \left(\frac{5}{7}\right) = -1,\ \left(\frac{6}{7}\right) = -1.$$

Suppose $p \in \mathbb{N}$ is a prime number. Then the residues $\{1, 2, \ldots, p-1\}$ mod$p$ form a multiplicative group, $(\mathbb{Z}/p)^\times$; and the map $a \mapsto a^2$ defines a homomorphism

$$\theta : (\mathbb{Z}/p)^\times \to (\mathbb{Z}/p)^\times.$$

The set Q of quadratic residues is the image of this map:

$$Q = \operatorname{im} \theta.$$

Proposition 2.10. *Suppose* $p \in \mathbb{N}$ *is an odd prime number. Then just half of the numbers* $1, 2, \ldots, p-1$ *are quadratic residues, and half are quadratic non-residues. Moreover if* $p \nmid a, b$ *then*

$$\left(\frac{a}{p}\right)\left(\frac{b}{p}\right) = \left(\frac{ab}{p}\right).$$

Proof ▶ The kernel of the homomorphism θ is

$$\begin{aligned}\ker\theta &= \{a : a^2 = 1\}\\ &= \{\pm 1\}.\end{aligned}$$

For

$$\begin{aligned}a^2 \equiv 1 &\implies p \mid (a^2-1)\\ &\implies p \mid (a-1)(a+1)\\ &\implies p \mid (a-1) \text{ or } p \mid (a+1)\\ &\implies a \equiv 1 \text{ or } a \equiv -1.\end{aligned}$$

By the First Isomorphism Theorem for finite groups,

$$|\ker\theta||\mathrm{im}\,\theta| = |(\mathbb{Z}/p)^\times| = p-1.$$

Thus

$$|\mathrm{im}\,\theta| = (p-1)/2,$$

ie just half the elements of $(\mathbb{Z}/p)^\times$ are quadratic residues, and half are quadratic non-residues.

The quadratic residues, as we have seen, form a subgroup $Q = \mathrm{im}\,\theta$, of index 2 in $(\mathbb{Z}/p)^\times$. It follows that

$$a, b \in Q \implies ab \in Q;$$

and

$$a \notin Q,\ b \in Q \implies ab \notin Q,$$

since $b, ab \in Q$ implies that $a = (ab)b^{-1} \in Q$. Thus if $a \notin Q$, the bijection

$$(\mathbb{Z}/p)^\times \to (\mathbb{Z}/p)^\times : x \mapsto ax$$

maps Q into $(\mathbb{Z}/p)^\times \setminus Q$; and therefore it maps $(\mathbb{Z}/p)^\times \setminus Q$ into Q, ie

$$a, b \notin Q \implies ab \in Q.$$

In other words,

$$\left(\frac{a}{p}\right)\left(\frac{b}{p}\right) = \left(\frac{ab}{p}\right).$$

◀

This last result is simply a consequence of the fact that Q is a subgroup of index 2; if S is a subgroup of G of index 2 (even if G is non-commutative) then we have a unique homomorphism

$$\epsilon : G \to \{\pm 1\}$$

such that

$$\epsilon(g) = 1 \iff g \in S.$$

Theorem 2.2. *(Gauss's Lemma) Suppose $p \in \mathbb{N}$ is an odd prime number; and suppose $p \nmid a$. Then*

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

Proof ▶ By Lagrange's Theorem,

$$a^{p-1} \equiv 1.$$

Thus

$$\left(a^{(p-1)/2}\right)^2 = 1.$$

It follows that

$$a^{(p-1)/2} \equiv \pm 1.$$

Now if a is a quadratic residue, say $a \equiv b^2$ then

$$a^{(p-1)/2} \equiv b^{p-1} \equiv 1.$$

We can re-word this as follows: the quadratic residues are all roots of the polynomial

$$x^{(p-1)/2} - 1$$

over the finite field $\mathbb{Z}/(p)$ formed by the residues modulo p. But a polynomial of degree d has at most d roots. It follows that the quadratic residues are *all* the roots of this polynomial. Thus if a is *not* a quadratic residue we must have

$$a^{(p-1)/2} \equiv -1 \bmod p.$$

We have shown therefore that

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

for all a coprime to p. ◀

Corollary 2.1. *Suppose $p \in \mathbb{N}$ is an odd prime number. Then*

$$\left(\frac{-1}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \bmod 4 \\ -1 & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

Proof ▶ By Gauss's Lemma

$$\begin{aligned} \left(\frac{-1}{p}\right) &\equiv (-1)^{(p-1)/2} \\ &= \begin{cases} 1 & \text{if } p \equiv 1 \bmod 4 \\ -1 & \text{if } p \equiv 3 \bmod 4. \end{cases} \end{aligned}$$

◀

2.5 Primes in $\mathbb{Z}[i]$ again

We can now determine which prime numbers $p \in \mathbb{N}$ split in $\mathbb{Z}[i]$, and which remain prime.

Dealing with $p = 2$ first, we observe that

$$2 = (1+i)(1-i) \sim (1+i)^2,$$

since $(1-i) = -i(1+i) \sim (1+i)$. Thus 2 splits into two *equal* primes in $\mathbb{Z}[i]$. (We say that 2 is *ramified* in $\mathbb{Z}[i]$.)

It remains to deal with the odd primes.

Theorem 2.3. *Suppose $p \in \mathbb{N}$ is an odd prime number. Then p remains prime in*
$Z[i]$ if $p \equiv 3 \bmod 4$; while if $p \equiv 1 \bmod 4$ then p splits into two distinct but conjugate prime factors:

$$p = \pi\bar{\pi}.$$

Proof ▶ It is easy to see that for all $a \in \mathbb{Z}$,

$$a^2 \equiv 0 \text{ or } 1 \bmod 4.$$

It follows that

$$a^2 + b^2 \equiv 0, 1 \text{ or } 2 \bmod 4.$$

In particular,

$$a^2 + b^2 \not\equiv 3 \bmod 4.$$

But we have seen that if p splits in $\mathbb{Z}[i]$ then $p = a^2 + b^2$. It follows that p cannot split if $p \equiv 3 \bmod 4$.

It remains to consider the case $p \equiv 1 \bmod 4$. We have seen that -1 is a quadratic residue in this case, say

$$x^2 + 1 \equiv 0 \bmod p.$$

We may assume that $0 < x < p$. (In fact we could take $0 < x < p/2$, since $-x$ is a solution if x is.) It follows that

$$(x+i)(x-i) = x^2 + 1 = pm,$$

where $0 < m < p$. But now suppose p does not split. Since there is unique factorisation in $\mathbb{Z}[i]$ it follows that

$$p \mid (x+i) \text{ or } p \mid (x-i).$$

But then

$$p \mid (x+i) \implies (x+i) = pz \implies (x-i) = p\bar{z} \implies p \mid (x-i),$$

and vice versa. Thus

$$p^2 \mid pm,$$

which is absurd. We conclude that p must split. ◀

2.6 Representation of a number as a sum of two squares

Proposition 2.11. *Suppose $n \in \mathbb{N},\ n > 0$. Then n is expressible as a sum of two squares,*

$$n = a^2 + b^2,$$

if and only if each prime number $p \equiv 3 \bmod 4$ divides n to an even power:

$$p \equiv 3 \bmod 4 \implies p^{2e} \parallel n.$$

Furthermore, if $p \equiv 3 \bmod 4$ then

$$p^{2e} \parallel n \implies p^e \mid a,\ p^e \mid b.$$

(Recall that $p^e \parallel n$ means that $p^e \mid n$ but $p^{e+1} \nmid n$.)

Proof ▶ Suppose

$$n = a^2 + b^2 = (a + bi)(a - bi);$$

and suppose $p \equiv 3 \bmod 4$. Since p is prime in $\mathbb{Z}[i]$. it follows that

$$p \mid n \implies p \mid (a + bi) \text{ or } p \mid (a - bi).$$

But

$$p^e \parallel (a + ib) \implies p^e \parallel (a - ib) \implies p^{2e} \parallel n.$$

Thus each $p \equiv 3 \bmod 4$ must divide n to an even power.

Note too that

$$\begin{aligned} p^e \mid a + bi,\ p^e \mid a - bi &\implies p^e \mid 2a,\ p^e \mid 2bi, \\ &\implies p^e \mid a,\ p^e \mid b. \end{aligned}$$

Conversely, suppose each prime number $p \equiv 3 \bmod 4$ divides n to an even power. Then we can express n in the form

$$n = PQ^2,$$

where P is a product of 2's and prime numbers $p \equiv 1 \bmod 4$, say

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

and Q is a product of prime numbers $p \equiv 3 \bmod 4$.

We can split each p_j, say

$$p_j = (c_j + d_j i)(c_j - d_j i).$$

Let

$$c + di = \prod (c_j + d_j i).$$

Then

$$c^2 + d^2 = P,$$

and

$$n = (Qc)^2 + (Qd)^2,$$

as required. ◀

Example. We cannot express $15 = 3 \cdot 5$ as a sum of two squares, since the prime $3 \equiv 3 \bmod 4$ occurs to an odd power in n.

On the other hand,

$$1000 = 2^3 5^3$$

is expressible as a sum of 2 squares, since it has no prime factor $\equiv 3 \bmod 4$. Our argument gives a concrete solution, namely

$$\begin{aligned} a + bi &= (1+i)^3(1+2i)^3 \\ &= 2i(1+i)(-3+4i)(1+2i) \\ &= 2i(1+i)(-11-2i) \\ &= 2i(-9-13i) \\ &= 26 - 18i, \end{aligned}$$

giving

$$1000 = 26^2 + 18^2.$$

We can get other solutions by splitting 5^3 differently, eg

$$\begin{aligned} a + bi &= (1+i)^3(1+2i)^2(1-2i) \\ &= 2i(1+i)(1+2i)5 \\ &= 10i(-1+3i) \\ &= 10(-3-i), \end{aligned}$$

giving

$$1000 = 30^2 + 10^2.$$

Note that splitting 2^3 differently in the same way does *not* give a genuinely new solution, since $(1-i) \sim (1+i)$.

Taking this argument a little further, we can determine exactly how many ways there are of expressing n as a sum of two squares.

Definition 2.6. *We denote by $r(n)$ the number of ways of expressing n as a sum of two squares,*

$$n = a^2 + b^2 \quad (a, b \in \mathbb{Z}),$$

where we count all *solutions separately.*

By this we mean, for example, that $r(1) = 4$ since

$$4 = 0^2 + (\pm 2)^2 = (\pm 2)^2 + 0^2.$$

Similarly, $r(5) = 8$ since

$$5 = (\pm 1)^2 + (\pm 2)^2 = (\pm 2)^2 + (\pm 1)^2.$$

Proposition 2.12. *Suppose*

$$n = 2^{e_2} \prod_{p \equiv 1 \bmod 4} p^{e_p} \prod_{p \equiv 3 \bmod 4} p^{2f_p}.$$

Then

$$r(n) = 4 \prod_{p \equiv 1 \bmod 4} (e_p + 1).$$

Proof ▶ Since

$$n = a^2 + b^2 = (a + ib)(a - ib)$$

the number of solutions, $r(n)$, is just the number of ways of factorising n in $\mathbb{Z}[i]$.

For each prime $p \equiv 1 \bmod 4$ let us choose a definite factorisation

$$p = \pi_p \bar{\pi}_p.$$

By the Fundamental Theorem for $\mathbb{Z}[i]$, $a + bi$ can be expressed uniquely in the form

$$a + bi = \epsilon \sigma_1^{e_1} \sigma_2^{e_2} \cdots \sigma_3^{e_3},$$

where ϵ is a unit, and each σ is either (1_i), a prime number $p \equiv 3 \bmod 4$ or one of the factors π_p or $\bar{\pi}_p$ of a prime number $p \equiv 1 \bmod 4$.

Now we see that the only options we have in constructing $a + ib$ are in the choice of one of the 4 units for ϵ, and a choice between the factors π_p and $\bar{\pi}_p$ for the prime numbers $p \equiv 1 \bmod 4$ dividing n.

The choices for these prime numbers are independent, and we can treat them separately. Suppose then that $p \equiv 1 \bmod 4$, and suppose

$$p^e \parallel n.$$

Writing $\pi, \bar{\pi}$ for $\pi_p, \bar{\pi}_p$, we have to divide the factors $\pi^e \bar{p}\bar{i}^e$ between $a + bi$ and $a - bi$.

In fact it is sufficient to determine how the factors π^e are divided between $a + bi$ and $a - bi$; for

$$\pi^f \parallel a + bi,\ \pi^{e-f} \parallel a - bi \implies \bar{p}\bar{i}^{e-f} \parallel a + bi,\ \bar{p}\bar{i}^f \parallel a - bi.$$

Thus we have just $e + 1$ choices, namely $f = 0, 1, \ldots, e$.

We conclude that

$$r(n) = 4 \prod_{p \equiv 1 \bmod 4} (e_p + 1).$$

◀

Chapter 3

Primality in $\mathbb{Z}[\omega]$

3.1 The ring $\mathbb{Z}[\omega]$

Let

$$\omega^3 = 1,\ \omega \neq 1.$$

(We may take

$$\omega = e^{2\pi/3} = (-1 + \sqrt{3}i)/2$$

if we like to think of ω as a complex number.)

Since

$$x^3 - 1 = (x-1)(x^2 + x + 1)$$

we have

$$\omega^2 + \omega + 1 = 0.$$

This has the two roots ω and ω^2. Thus the complex conjugate of ω is

$$\bar{\omega} = \omega^2.$$

More generally, if

$$z = a + b\omega$$

then

$$\bar{z} = a + b\omega^2;$$

and so

$$\begin{aligned} |z|^2 &= z\bar{z} \\ &= (a + b\omega)(a + b\omega^2) \\ &= a^2 + ab(\omega + \omega^2) + b^2 \\ &= a^2 - ab + b^2, \end{aligned}$$

since $\omega + \omega^2 = -1$.

Note that

$$a^2 - ab + b^2 = (a - b/2)^2 + 3b^2/4$$

is *positive-definite*, ie

$$a^2 - ab + b^2 \geq 0 \text{ and } a^2 - ab + b^2 = 0 \iff a = b = 0.$$

Thus

$$z \in \mathbb{Z}[\omega] \implies N(z) \in \mathbb{N};$$

and

$$N(z) = 0 \iff z = 0.$$

Definition 3.1. *We denote by $\mathbb{Z}[\omega]$ the set of numbers of the form*

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

and by $\mathbb{Q}(\omega)$ the set of numbers of the form

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

Proposition 3.1. *$\mathbb{Z}[\omega]$ is an integral domain, and $\mathbb{Q}(\omega)$ is its field of fractions.*

Proof ▶ Evidently $\mathbb{Z}[\omega]$ and $\mathbb{Q}(\omega)$ are both closed under addition.

Suppose

$$z = a + b\omega,\ w = A + B\omega.$$

Then

$$\begin{aligned} zw &= aA + (aB + bA)\omega + bB\omega^2 \\ &= (aA - bB) + (aB + bA - bB)\omega, \end{aligned}$$

since $\omega^2 = -1 - \omega$. Thus $\mathbb{Z}[\omega]$ and $\mathbb{Q}(\omega)$ are both closed under multiplication. It follows that $\mathbb{Z}[\omega]$ is a ring (in fact an integral domain).

Also

$$\begin{aligned} \frac{z}{w} &= \frac{a + b\omega}{A + B\omega} \\ &= \frac{(a + b\omega)(A + B\omega^2)}{(A + B\omega)(A + B\omega^2)} \\ &= \frac{(aA + bB - aB) + (bA - aB)\omega}{A^2 - AB + B^2}, \end{aligned}$$

so $\mathbb{Q}(\omega)$ is closed under division by non-zero elements. Thus $\mathbb{Q}(\omega)$ is a field. ◀

Proposition 3.2. *There are just 6 units in $\mathbb{Z}[\omega]$: $\pm 1, \pm\omega, \pm\omega^2$.*

Proof ▶

Lemma 3. *The number*

$$u = a + b\omega \in \mathbb{Z}[\omega]$$

is a unit if and only if

$$u\bar{u} = a^2 - ab + b^2 = 1.$$

Proof ▶ If $u\bar{u} = 1$ then u is certainly a unit with inverse $\bar{u}$. Conversely, suppose u is a unit, say

$$uv = 1.$$

Then

$$|u|^2|v|^2 = 1 \implies |u|^2 = 1.$$

◀

By the Lemma, $u = a + b\omega$ is a unit if and only if

$$a^2 - ab + b^2 = 1.$$

In this case,

$$4a^2 - 4ab + 4b^2 = (2a - b)^2 + 3b^2 = 4.$$

Thus $b = 0, \pm 1$; and similarly $a = 0, \pm 1$. Thus any unit must lie in the set

$$\{\pm 1, \pm\omega, \pm 1 \pm \omega\}.$$

We have

$$1 + \omega = -\omega^2, \ -1 - \omega = \omega^2;$$

while

$$z = \pm(1 - \omega) \implies z\bar{z} = 1 + 1 + 1 = 3.$$

We conclude that the only units are the 6 given. ◀

3.2 The Euclidean algorithm in $\mathbb{Z}[\omega]$

We can extend the Euclidean algorithm to $\mathbb{Z}[\omega]$ in exactly the same way that we extended it to $\mathbb{Z}[i]$.

Thus suppose $z, w \in \mathbb{Z}[\omega]$. Let

$$z/w = x + y\omega,$$

where $x, y \in \mathbb{Q}$. Let a, b be the closest integers to x, y, say

$$x - 1/2 \leq a < x + 1/2, \ y - 1/2 \leq b < y + 1/2.$$

Set

$$q = a + b\omega.$$

Then

$$z/w - q = s + t\omega$$

where

$$|s|, |t| \leq 1/2.$$

It follows that

$$\begin{aligned} |z/w - q|^2 &= |s + t\omega|^2 \\ &= s^2 - st + t^2 \\ &\leq (1/2)^2 + (1/2)^2 + (1/2)^2 = 3/4. \end{aligned}$$

Thus if we set

$$r = z - qw,$$

then

$$|r|^2 \leq \frac{3}{4}|q|^2.$$

In particular, we have established

Proposition 3.3. *Suppose $z, w \in \mathbb{Z}[\omega]$, with $w \neq 0$. Then we can find $q, r \in \mathbb{Z}[\omega]$ such that*

$$z = qw + r, \quad |r|^2 < |q|^2.$$

This proposition allows us to implement the Euclidean algorithm; and as a consequence we have the following

Theorem 3.1. *(The Fundamental Theorem of Arithmetic for $\mathbb{Z}[\omega]$) Each number $z \in \mathbb{Z}[\omega]$ can be factorised into prime factors; and the factorisation is unique up to the order of the factors, and multiplication by units.*

3.3 Quadratic residues revisited

Suppose p is an odd prime. It is convenient to choose the residues mod$p$ from the set

$$-p/2 < r < p/2.$$

We can then divide the set of residues

$$R = \{-(p-1)/2, \ldots, -1, 0, 1, \ldots, (p-1)/2\}$$

into 3 disjoint subsets

$$R = -S \cup \{0\} \cup S,$$

where

$$S = \{1, \ldots, (p-1)/2\}, \quad -S = \{-1, \ldots, -(p-1)/2\}.$$

Suppose $a \in \mathbb{Z},\ p \nmid a$. Consider

$$aS = \{a, 2a, \ldots, (p-1)a/2\}.$$

We can divide the residues ia mod p into 2 sets; those in S and those in $-S$.

Proposition 3.4. *Suppose $a \in \mathbb{Z}$, $p \nmid a$. Then*

$$\left(\frac{a}{p}\right) = (-1)^{\mu},$$

where μ is the number of residues ia $(0 < i < p/2)$ lying in the subset $-S$.

In other words, μ is the number of numbers in the range $1 \le i \le (p-1)/2$ such that the least positive remainder of $ai \bmod p$ is $\ge (p+1)/2$.

Proof ▶ By Gauss's Lemma,

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

Lemma 4. *For each $r \in S$ there is exactly one $i \in S$ such that*

$$ai = \pm r \bmod p.$$

Proof ▶ Suppose both remainders $\pm r \bmod p$ appear in aS, say

$$as \equiv r,\ at \equiv -r.$$

Then

$$a(s+t) \equiv 0.$$

But that is impossible since $0 < s + t < p$ (and $p \nmid a$).

Thus the $(p-1)/2$ elements aS are distributed among the $(p-1)/2$ "pigeon-holes" $\{\pm r\}$, with at most one in each. It follows that there is exactly one in each. ◀

On multiplying together the elements of aS,

$$\prod_{i \in S} ai = (-1)^{\mu} \prod_{i \in S} i,$$

where μ is the number of elements of aS in $-S$. On the other hand,

$$\prod_{i \in S} ai = a^{(p-1)/2} \prod_{i \in S} i,$$

We conclude that

$$\left(\frac{a}{p}\right) \equiv (-1)^{(p-1)/2} \equiv (-1)^{\mu} \bmod p.$$

◀

Corollary 3.1. *Suppose $a > 0$. Then*

$$\left(\frac{a}{p}\right) \equiv (-1)^{\mu} \bmod p,$$

where

$$\mu = \begin{cases} [p/a] - [p/2a] + [3p/a] - \cdots - [(a-1)p/2a] + [p/2] & \text{if } a \text{ is even} \\ [p/a] - [p/2a] + [3p/a] - \cdots - [(a-2)p/2a] + [(a-1)p/2a] & \text{if } a \text{ is odd} \end{cases}$$

(Here $[x]$, as usual, denotes the greatest integer $\leq x$.)

Proof ▶ Suppose $ia \bmod p \in -S$, say

$$ia = mp + r,$$

where $p/2 < r < p$. In other words,

$$mp + p/2 < ia < (m+1)p,$$

ie

$$(m + 1/2)p/a < i < (m+1)p/a.$$

The number of numbers i in this range is

$$[(m+1)p/a] - [(m+1/2)p/a].$$

For $m = 0, 1, 2, \ldots, [a/2]$ this gives

$$[p/a] - [p/2a],\ [2p/a] - [3p/2a],\ [3p/a] - [5p/2a],\ \ldots,$$

ending with $[p/2] - [(a-1)p/2a]$ if a is even, and $[(a-1)p/2a] - [(a-2)p/2a]$ if a is odd. ◀

Corollary 3.2. *Suppose p is prime, $p \neq 2, 3$. Then*

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \bmod 12 \\ -1 & \text{if } p \equiv \pm 5 \bmod 12. \end{cases}$$

Proof ▶ By the Proposition,

$$\left(\frac{3}{p}\right) = [p/3] - [p/6].$$

Consider the function

$$f(n) = [n/3] - [n/6],$$

for $n \in \mathbb{N}$. If n is increased by 12, then $[n/3]$ increases by 4 and $[n/6]$ by 2. Thus the parity of n (ie whether n is even or odd) does not change. It

follows that it is sufficient to consider n in the range $0 \leq n < 12$; and since we are only interested in the value of $f(p)$ for primes p we need only consider $n = 1, 5, 7, 11$. But

$$\begin{aligned} f(1) &= 0 - 0 = 0, \\ f(5) &= 1 - 0 = 1, \\ f(7) &= 2 - 1 = 1, \\ f(11) &= 3 - 1 = 2. \end{aligned}$$

We conclude that 3 is a quadratic residue mod$p$ if $p \equiv 1$ or 11 mod 12, ie if $p \equiv \pm 1$ mod 12; and 3 is a quadratic non-residue if $p \equiv 5$ or 7 mod 12, ie if $p \equiv \pm 5$ mod 12. ◀

Proposition 3.5. *Suppose $p \in \mathbb{N}$ is an odd prime; and suppose $a, b \in \mathbb{Z}$, with $p \nmid a, b$. Then*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

Proof ▶ By Gauss' Lemma,

$$\begin{aligned} \left(\frac{ab}{p}\right) &\equiv (ab)^{(p-1)/2} \bmod p \\ &= (a)^{(p-1)/2}(b)^{(p-1)/2} \\ &\equiv \left(\frac{a}{p}\right)\left(\frac{b}{p}\right). \end{aligned}$$

Since each side has value ± 1, the result follows. ◀

Corollary 3.3. *Suppose p is a prime, $p \neq 2, 3$. Then*

$$\left(\frac{-3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \bmod 3 \\ -1 & \text{if } p \equiv \pm 2 \bmod 3. \end{cases}$$

Proof ▶ Since

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

this is just a matter of combining the results for

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} = \begin{cases} 1 & \text{if } p \equiv 1 \bmod 4 \\ -1 & \text{if } p \equiv 3 \bmod 4 \end{cases}$$

with the result for $\left(\frac{3}{p}\right)$ above. We conclude that

$$\left(\frac{-3}{p}\right) = \begin{cases} 1 \times 1 = 1 & \text{if } p \equiv 1 \bmod 12 \\ 1 \times -1 = -1 & \text{if } p \equiv 5 \bmod 12 \\ -1 \times -1 = 1 & \text{if } p \equiv 7 \bmod 12 \\ -1 \times 1 = -1 & \text{if } p \equiv 11 \bmod 12 \end{cases}$$

which is equivalent to the stated result. ◀

3.4 The primes in $\mathbb{Z}[\omega]$

Proposition 3.6. *A prime number $p \in \mathbb{N}$ splits into at most 2 factors in $\mathbb{Z}[\omega]$.*

Proof ▶ Suppose

$$p = \pi_1 \pi_2 \cdots \pi_r.$$

Then

$$p^2 = |\pi_1|^2 |\pi_2|^2 \cdots |\pi_r|^2.$$

Since $|\pi|^2 > 1$ for any prime π, it follows from the Fundamental Theorem in $\mathbb{N}$ that there are at most 2 prime numbers on the right. ◀

Proposition 3.7. *The prime number $p \in \mathbb{N}$ splits in $\mathbb{Z}[\omega]$ if and only if p is expressible in the form*

$$p = a^2 + ab + b^2$$

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

Proof ▶ If

$$\begin{aligned} p &= a^2 + ab + b^2 \\ &= (a - b\omega)(a - b\omega^2) \end{aligned}$$

we have an explicit factoring of p.

Conversely, suppose p splits, say

$$p = \pi_1 \pi_2.$$

Then

$$p^2 = |\pi_1|^2 |\pi_2|^2.$$

It follows that

$$|\pi_1|^2 = p = |\pi_2|^2$$

Thus if $\pi_1 = a + b\omega$,

$$\begin{aligned} p &= a^2 - ab + b^2 \\ &= a^2 + aB + B^2, \end{aligned}$$

with $B = -b$. ◀

Proposition 3.8. *Every prime π in $\mathbb{Z}[\omega]$ is a factor of a unique prime number $p \in \mathbb{N}$.*

Proof ▶ Let

$$N(\pi) = \pi \bar{pi} = p_1 \ldots p_r.$$

By the Fundamental Theorem for $\mathbb{Z}[\omega]$, π must divide one of the factors $p_1, \ldots, p_r$ on the right.

On the other hand, suppose $\pi \mid p, q$ where p, q are distinct prime numbers. There exist $x, y \in \mathbb{Z}$ such that

$$px + qy = 1.$$

It follows that

$$\pi \mid 1,$$

which is absurd. ◀

Evidently 3 splits into 2 equivalent factors in $\mathbb{Z}[w]$:

$$3 = -(\sqrt{-3})^2 = -\eta^2,$$

where

$$\eta = 1 + 2\omega.$$

We say that 3 is *ramified* in $\mathbb{Z}[\omega]$.

Proposition 3.9. *Suppose p is a prime number, $p \neq 2, 3$. Then p splits in $\mathbb{Z}[\omega]$ if and only if $p \equiv 1 \bmod 3$.*

Proof ▶ Suppose p splits in $\mathbb{Z}[\omega]$. By Proposition 3.7,

$$p = a^2 + ab + b^2$$

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

$$a^2 + ab + b^2 \equiv 0 \bmod p,$$

with $p \nmid a, b$. Hence

$$4a^2 + 4ab + 4b^2 \equiv 0, \mathrm{mod} p$$

ie

$$(2a + b)^2 + 3b^2 \equiv 0 \bmod p.$$

Hence

$$\left(\frac{2a + b}{b}\right)^2 + 3 \equiv 0 \bmod p.$$

(Note that we can treat b^{-1} as an integer mod$p$. In effect b has an inverse mod$p$, say $bb' \equiv 1 \bmod p$, and we may regard b^{-1} as an alias for b'.)

It follows that -3 is a quadratic residue mod$p$:

$$\left(\frac{-3}{p}\right) = 1.$$

As we have seen, the condition for this is that $p \equiv 1 \bmod 3$. So p certainly does *not* split if $p \equiv 2 \bmod 3$.

Conversely, suppose $p \equiv 1 \bmod 3$; and suppose p does *not* split in $\mathbb{Z}[\omega]$. As we just saw, -3 is a quadratic residue mod$p$, ie we can find $x \in \mathbb{Z}$ such that

$$x^2 + 3 \equiv 0 \bmod p,$$

say

$$x^2 + 3 = pm,$$

for some $m \in \mathbb{N}$.

We can re-write this as

$$(x + \sqrt{-3})(x - \sqrt{-3}) = pm.$$

By the Fundamental Theorem for $\mathbb{Z}[\omega]$, since p (by hypothesis) remains prime in this ring,

$$p \mid x + \sqrt{-3},$$

ie

$$x + \sqrt{-3} = p(a + b\omega).$$

Comparing the coefficients of $\sqrt{-3}$ on each side,

$$1 = pb/2$$

which is absurd.

Therefore p must split; and the proof is complete. ◀

Proposition 3.10. *Suppose p is a prime number $\equiv 1 \bmod 3$. Then there exist unique integers a, b with $0 < a < b$ such that*

$$p = a^2 + ab + b^2.$$

Proof ▶ We know that p splits in this case, say

$$p = \pi\bar{\pi}.$$

If

$$\pi = a - b\omega$$

then (a, b) will be a solution to the equation; and every solution will arise from a factor of p in this way. Thus there are just 12 solutions; if π is one solution these are:

$$\pm\pi, \pm\omega\pi, \pm\omega^2\pi, \pm\bar{\pi}, \pm\omega\bar{\pi}, \pm\omega^2\bar{\pi}.$$

Thus the solution (a, b) gives rise to the solutions:

$$\begin{array}{ll}
\pm\pi = \pm(a - b\omega) & \mapsto \pm(a, b), \\
\pm\omega\pi = \pm(a\omega - b\omega^2) = \pm(b + (a + b)\omega) & \mapsto \pm(b, -a - b), \\
\pm\omega^2\pi = \pm(a\omega^2 - b) = \pm(-a - b - a\omega) & \mapsto \pm(-a - b, a), \\
\pm\bar{\pi} = \pm(a - b\omega^2) = (a + b + b\omega) & \mapsto \pm(a + b, -b), \\
\pm\omega\bar{\pi} = \pm(a\omega - b) = \pm(-b + a\omega) & \mapsto \pm(-b, -a), \\
\pm\omega^2\bar{\pi} = \pm(a\omega^2 - b\omega) = \pm(-a - (a + b)\omega) & \mapsto \pm(-a, a + b).
\end{array}$$

Let (a, b) be a solution with minimal $|a|$. Since (b, a) is also a solution we have $|a| \leq |b|$. In fact, since $b = \pm a \implies a^2 \mid p$, we must have $|a| < |b|$. Also, since $(-a, -b)$ is a solution we may assume that $a > 0$. If now $b < 0$ then $b < -a$ and so the solution $(-a - b, a)$ has both 'coordinates' > 0. Thus, after swapping the coordinates if necessary, we have a solution (a, b) with $0 < a < b$.

On the other hand, it is readily verified that if (a, b) is such a solution then none of the 11 other solutions has the same property. Ergo there is just one such solution. ◀

Example. Suppose $p = 37$. Since $37 \equiv 1 \bmod 3$ we can find $a, b \in \mathbb{Z}$ with $0 < a < b$ such that

$$a^2 + ab + b^2 = 37.$$

Since $a < b$,

$$3a^2 < 37.$$

Thus $a = 1, 2$ or 3. A brief inspection yields the solution

$$a = 2,\ b = 5.$$

3.5 Fermat's Last Theorem for exponent 3

Theorem 3.2. *The equation*

$$x^3 + y^3 + z^3 = 0$$

has no solutions in $\mathbb{Z}[\omega]$ *except* $x = y = z = 0$.

A fortiori *it has no solutions in the integers* $\mathbb{Z}$ *except* $x = y = z = 0$.

Idea of proof. As the details of our proof are rather intricate, it may help if we give a very crude outline of the argument. There are two different ideas involved.

1. We can write the equation in the form

$$(x + y)(x + \omega y)(x + \omega^2 y) = (-z)^3.$$

Now suppose the 3 factors on the left were "pairwise coprime", ie no 2 of them had any factor in common. Then we would conclude that each of them was a cube, up to a multiple, say

$$x + y = \epsilon X^3,\ x + \omega y = \epsilon' Y^3,\ x + \omega^2 y = \epsilon'' Z^3,$$

where $\epsilon, \epsilon', \epsilon''$ are units.

2. We have
$$(x + y) + \omega(x + \omega y) + \omega^2(x + \omega^2 y) = 0.$$
Thus (using the result above)
$$X^3 + \epsilon_1 Y^3 + \epsilon_2 Z^3 = 0.$$
where ϵ_1, ϵ_2 are units.

Now this is an equation of the same form as — perhaps a bit more general than — the equation we started with, and we have a much smaller solution (approximately the cube root of the previous one).

This is the crux of Fermat's "method of infinite descent"; if every solution leads to a smaller solution then we will reach a contradiction.

We have over-simplified in describing the first idea above; it turns out that $x + y,\ x + \omega y,\ x + \omega^2 y$ actually *do* have a factor in common, but one which we can easily deal with.

Proof ▶ We may assume that x, y, z have no factor in common. Since any prime factor of two of these is necessarily a factor of the third, this implies that x, y, z are pairwise coprime:

$$\gcd(x, y) = \gcd(x, z) = \gcd(y, z) = 1.$$

Let
$$\pi = 1 - \omega.$$
Then π is a prime in $\mathbb{Z}[\omega]$, since
$$|\pi| = 3;$$
and
$$3 = -\omega^2 \pi^2 \sim \pi^2.$$

Lemma 5. *There are just 3 residue classes* $\bmod \pi$*, represented by* $0, \pm 1$.

Proof ▶ It is easy to see that these 3 elements are not congruent $\bmod \pi$. For example,
$$-1 \equiv 1 \bmod \pi \implies \pi \mid 2,$$
which is impossible since $\pi \mid 3$ and $\gcd(2, 3) = 1$.

Since $\pi \mid 3$,

$$x \equiv y \bmod 3 \implies x \equiv y \bmod \pi.$$

There are 9 residue class mod3 in $\mathbb{Z}[\omega]$, represented by

$$a + b\omega \quad (a, b \in \{0, \pm 1\}).$$

(We could equally well have taken $a, b \in \{0, 1, 2\}$.)

It is sufficient therefore to show that the 6 elements

$$a + \omega,\ a - \omega \quad (a \in \{0, \pm 1\})$$

are each congruent modπ to one of $\{0, \pm 1\}$. That is straightforward. For example,

$$\omega \equiv 1 \bmod \pi$$

since $1 - \omega = \pi$. And

$$1 + \omega = -\omega^2 \equiv -1 \bmod \pi$$

since $1 - \omega^2 = -\omega^2(1 - \omega)$, while

$$-1 + \omega = -\pi \equiv 0 \bmod \pi.$$

Taking the negations of these 3 congruences,

$$-\omega \equiv -1,\ -1 - \omega \equiv 1,\ 1 - \omega \equiv 0.$$

We conclude that there are just 3 residue classes, represented by $0, \pm 1$. ◀

Lemma 6. *Suppose $x \in \mathbb{Z}[\omega]$. Then*

$$x \equiv 1 \bmod \pi \implies x^3 \equiv 1 \bmod \pi^4;$$

and

$$x \equiv -1 \bmod \pi \implies x^3 \equiv -1 \bmod \pi^4.$$

Proof ▶ Suppose $x \equiv 1 \bmod \pi$, ie

$$x = 1 + \pi a$$

for some $a \in \mathbb{Z}[\omega]$. Then

$$x^3 = 1 + 3\pi a + 3\pi^2 a^2 + \pi^3 a^3.$$

It follows at once that

$$x^3 \equiv 1 \bmod \pi^3,$$

since the last 3 terms on the right are all divisible by π^3. To improve this result, we note that

$$\pi^4 \mid 3\pi^2,$$

since $\pi^2 \mid 3$. We are left with

$$\begin{aligned} 3\pi a + \pi^3 a^3 &= -\omega^2 \pi^3 a + \pi^3 a^3 \\ &= (-\omega^2 + a^2) a\pi^3 \\ &= (a - \omega)(a + \omega) a\pi^3. \end{aligned}$$

But now it is readily verified that

$$\pi \mid (a - \omega)(a + \omega)a$$

if $a = 0, \pm 1$. It follows that

$$\pi^4 \mid (3\pi a + \pi^3 a^3)$$

in all cases; and so

$$x \equiv 1 \bmod \pi \implies x^3 \equiv 1 \bmod \pi^4.$$

Taking the negative of this,

$$x \equiv -1 \bmod \pi \implies x^3 \equiv -1 \bmod \pi^4.$$

◀

Lemma 7. *Suppose*

$$x^3 + y^3 + z^3 = 0,$$

where x, y, z are pairwise coprime. Then x, y, z must have remainders $0, \pm 1 \bmod \pi$ in some order.

Proof ▶ Evidently

$$x^3 + y^3 + z^3 \equiv 0 \bmod \pi^3.$$

But each of x^3, y^3, z^3 is congruent $\bmod \pi^3$ to $0, \pm 1$. The remainders are not all 0, since x, y, z are coprime. Thus the only way they can sum to 0 is if they are $0, 1, -1$ in some order. ◀

We may suppose without loss of generality that

$$x \equiv 1 \bmod \pi,\ y \equiv -1 \bmod \pi,\ z \equiv 0 \bmod \pi.$$

Lemma 8. *In fact*

$$\pi \mid z \implies \pi^2 \mid z;$$

and so

$$\pi^6 \mid z^3.$$

Proof ▶ By Lemma 6,

$$x^3 \equiv 1 \bmod \pi^4,\ y^3 \equiv -1 \bmod \pi^4 \implies z^3 \equiv 0 \bmod \pi^4.$$

But

$$\pi^4 \mid z^3 \implies \pi^2 \mid z \implies \pi^6 \mid z^3.$$

◀

It follows that we can write our equation

$$x^3 + y^3 = \pi^6 t^3,$$

where $t = -z/\pi^2$.

We shall prove that the more general equation

$$x^3 + y^3 = \epsilon\pi^6 t^3,$$

where $\pi \nmid x, y$ and ϵ is a unit, has no solution in $\mathbb{Z}[\omega]$ apart from $x = y = t = 0$.

We may assume here that $\epsilon \in \{1, \omega, \omega^2\}$, since the factor -1, if present, can be absorbed into t.

We can factorize the cubic on the left:

$$(x + y)(x + \omega y)(x + \omega^2 y) = \epsilon\pi^6 t^3.$$

Lemma 9. *The 3 factors on the left have just the factor π in common:*

$$\gcd(x + y,\ x + \omega y) = \gcd(x + y,\ x + \omega^2 y) = \gcd(x + \omega y,\ x + \omega^2 y) = \pi.$$

Proof ▶ We have

$$d \mid (x + y),\ d \mid (x + \omega y) \implies d \mid (x + y) - (x + \omega y) \implies d \mid \pi y.$$

Similarly

$$d \mid (x + y),\ d \mid (x + \omega y) \implies d \mid \omega(x + y) - (x + \omega y) \implies d \mid \pi x.$$

Hence

$$d \mid (x + y),\ d \mid (x + \omega y) \implies d \mid \pi \gcd(x, y) \implies d \mid \pi.$$

Thus

$$\gcd(x + y,\ x + \omega y) = 1 \text{ or } \pi;$$

and the same result holds true for the other 2 pairs of factors.

Our argument also shows that

$$\pi \mid (x + y) \iff \pi \mid (x + \omega y) \iff \pi \mid (x + \omega^2 y).$$

Since one at least of these factors is divisible by π (as their product is divisible by π^6) it follows that all are divisible by π, two of them being exactly divisible by π and the other being divisible by at least π^4. ◀

After replacing y by ωy or $\omega^2 y$, if necessary, we may suppose that

$$\pi^4 \mid (x+y),\ \pi \parallel (x+\omega y),\ \pi \parallel (x+\omega^2 y).$$

It follows from the equation

$$(x+y)(x+\omega y)(x+\omega^2 y) = \pi^6 t^3$$

that the three factors on the left are expressible in the form

$$x+y = \epsilon_1 \pi Z^3,\ x+\omega y = \epsilon_2 \pi X^3,\ x+\omega^2 y = \epsilon_3 \pi Y^3,$$

where $\epsilon_1, \epsilon_2, \epsilon_3$ are units, and $X, Y, Z \in \mathbb{Z}[\omega]$ with $\pi \mid Z$ but $\pi \nmid X, Y$. But

$$(x+y) + (x+\omega y) + (x+\omega^2 y) = 0.$$

It follows that

$$X^3 + \epsilon Y^3 = \epsilon' Z^3$$

where ϵ, ϵ' are units. Since we could absorb a factor -1 into the cube, we may assume that $\epsilon, \epsilon' \in \{1, \omega, \omega^2\}$.

Now $X^3, Y^3 \equiv \pm 1 \bmod \pi^3$ (in fact $\bmod \pi^4$, but we don't need that here). Since $\pi^3 \mid Z^3$ it follows that

$$\pm 1 \pm \epsilon \equiv 0 \bmod \pi^3.$$

It follows from this that $\epsilon = 1$, since none of the other combinations $\pm 1 \pm \omega,\ \pm 1 \pm \omega^2$ is divisible by π^3.

Thus our equation takes the form

$$X^3 + Y^3 = \epsilon Z^3 \qquad (\pi \nmid X, Y,\ \pi \mid Z),$$

where $\epsilon \in \{1, \omega, \omega^2\}$. This differs from our original equation in having a unit on the right. However, this does not affect our argument, which only involved factorisation of the left-hand side.

We conclude that any solution of one of these three equations (corresponding to the three possible values of ϵ) leads to another solution of one of the equations. Moreover, this new solution is smaller than the one we started with, if we measure the size of a solution (x, y, z) by $\max(|x|, |y|, |z|)$. For

$$|X|^3,\ |Y|^3,\ |Z|^3 \le 2\max(|x|, |y|, |z|);$$

and so

$$\max(|X|, |Y|, |Z|) \le (2\max(|x|, |y|, |z|))^{1/3}.$$

Since $(2N)^{1/3} < N$ for $N \ge 2$, while $\pi \mid z \implies |z| \ge 3$, we conclude that — starting from any solution — iteration of our "descent" is bound to lead to a contradiction. Hence there is no solution. ◀

3.6 Pythagorean triples

Our proof of Fermat's Last Theorem for $n = 3$ used factorization in $\mathbb{Z}[\omega]$, as well as Fermat's "Method of Infinite Descent". The proof in the case $n = 4$ only uses the latter, and so may illustrate the method more clearly.

As a preparation for this, we consider the Pythagorean equation

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

We know of course that this has integral solutions, eg $(3, 4, 5)$. We want to find all solutions.

We may assume that

$$\gcd(x, y, z) = 1.$$

One of x, y must be even, and the other odd; for if both were odd we would have

$$x^2 + y^2 \equiv 1 + 1 = 2 \bmod 4;$$

but

$$z^2 \equiv 0 \text{ or } 1 \bmod 4$$

for all $z \in \mathbb{Z}$.

Definition 3.2. *A solution of Pythagoras' equation*

$$x^2 + y^2 = z^2,$$

with $x, y, z \in \mathbb{N}$, $\gcd(x, y, z) = 1$ *and* x *odd, is called a* Pythagorean triple.

Proposition 3.11. *Suppose* (x, y, z) *is a Pythagorean triple* (x, y, z). *Then there exist unique* $u, v \in \mathbb{N}$ *such that*

$$x = u^2 - v^2,\ y = 2uv,\ z = u^2 + v^2.$$

Furthermore $\gcd(u, v) = 1$, *one of* u, v *is odd and the other even, and* $u > v$. *Conversely, each such pair* $u, v \in \mathbb{N}$ *defines a Pythagorean triple* (x, y, z).

Proof ▶ Suppose (x, y, z) is a Pythagorean triple. Let $y = 2Y$. Then

$$4Y^2 = z^2 - x^2 = (z + x)(z - x).$$

Since x, z are both odd, $x + y, x - y$ are both even.

We claim that

$$\gcd(z + x, z - x) = 2.$$

For

$$d \mid z + x,\ d \mid z - x \implies d \mid 2z, d \mid 2x \implies d \mid 2\gcd(z, x) = 2.$$

Now suppose $p \in \mathbb{N}$ is a prime number. Let $p^e \parallel Y$. Then it follows that

$$p^{2e} \parallel (z+x)/2,\ p \nmid (z-x)/2 \text{ or } p^{2e} \parallel (z-x)/2,\ p \nmid (z+x)/2.$$

Thus, taking u to be the product of the p^e in the first case, and v to be the product of the p^e in the second case,

$$Y = uv,\ (z+x)/2 = u^2,\ (z-x)/2 = v^2;$$

and so

$$y = 2uv,\ z = u^2 + v^2,\ x = u^2 - v^2.$$

Since $u^2 = (z+x)/2,\ v^2 = (z-x)/2$, the numbers $u, v \in \mathbb{N}$ are uniquely defined by (x, y, z). Also $\gcd(u, v) = 1$, since

$$d \mid u, v \implies d^2 \mid x, y, z.$$

If u, v were both odd then x, y, z would all be even; hence one is odd and the other even. Finally, $u \geq v$ and therefore, since one is even and one odd, $u > v$.

Conversely, it is readily verified that

$$(u^2 - v^2)^2 + (2uv)^2 = (u^2 + v^2)^2;$$

so any pair $u, v \in \mathbb{N}$ with $u \geq v$ will give a solution of Pythagoras' equation. Moreover, if one of u, v is odd and the other even then x, z are odd. Finally, if $\gcd(u, v) = 1$ then $\gcd(x, y, z) = 1$. For suppose $p \mid x, y, z$ for some prime number p. Then p is odd (since x is odd) and so

$$\begin{aligned} p \mid x, z &\implies p \mid (x+z)/2, (x-z)/2 \\ &\implies p \mid u^2, v^2 \\ &\implies p \mid u, v. \end{aligned}$$

◀

Example. The pair $(u, v) = (2, 1)$ gives the Pythagorean triple

$$x = u^2 - v^2 = 3,\ y = 2uv = 4,\ z = u^2 + v^2 = 5.$$

Similarly

$$\begin{aligned} (3, 2) &\mapsto (5, 12, 13), \\ (4, 1) &\mapsto (15, 8, 17), \\ (4, 3) &\mapsto (7, 24, 25), \end{aligned}$$

and so on.

3.7 Fermat's Last Theorem for exponent 4

This was, as far as is known, the only case proved by Fermat.

Theorem 3.3. *The equation*

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

has no solutions in $\mathbb{Z}$ *except* $x = y = z = 0$.

Proof ▶ We shall prove a slightly more general result, namely that the equation

$$x^4 + y^4 = z^2$$

has no solutions in $\mathbb{Z}$ except $x = y = z = 0$.

We may suppose that $x, y, z \in \mathbb{N}$, and that x, y, z are pairwise coprime. One of x, y must be even, and one odd, on considering remainders mod4. Let us suppose that x is odd, and y even. Then (x^2, y^2, z) is a Pythagorean triple, and so there exists unique $u, v \in \mathbb{N}$ such that

$$x^2 = u^2 - v^2,\ y^2 = 2uv,\ z = u^2 + v^2.$$

Moreover, $\gcd(u, v) = 1$.

Since $x^2 \equiv 1 \bmod 4$, u must be odd and v even, say

$$v = 2v'.$$

Then

$$y^2 = 4uv',$$

with $\gcd(u, v') = 1$. But this implies that

$$u = s^2,\ v' = t^2,$$

where $s, t \in \mathbb{N}$ with $\gcd(s, t) = 1$. Thus

$$x^2 = s^4 - 4t^4,$$

ie

$$x^2 + 4t^4 = s^4.$$

This is a little like our original equation. We apply the same idea, but now to the Pythagorean triple $(x, 2t^2, s^2)$. By Proposition 3.11, there exist unique U, V such that

$$x = U^2 - V^2,\ 2t^2 = 2UV,\ s^2 = U^2 + V^2.$$

Since $t^2 = UV$ and $\gcd(U, V) = 1$,

$$U = X^2,\ V = Y^2$$

for $X, Y \in \mathbb{N}$. Thus

$$s^2 = X^4 + Y^4.$$

This is identical to our original equation, with X, Y, s in place of x, y, z. Thus each solution (x, y, z) of the equation gives rise to a (much) smaller solution.

To see how much smaller, let us take z as a measure of the size of the solution. Then

$$s = \sqrt{u} < \sqrt[4]{z}.$$

Clearly if we iterate this "descent" we shall rapidly reach a contradiction.

◀

3.8 Algebraic numbers and algebraic integers

We have skated around one issue in the discussion above. Since $\omega = (-1 + \sqrt{-3})/2$,

$$\mathbb{Q}(\omega) = \mathbb{Q}(\sqrt{-3}).$$

However,

$$\mathbb{Z}[\sqrt{-3}] \subset \mathbb{Z}[\omega] \text{ but } \mathbb{Z}[\sqrt{-3}] \neq \mathbb{Z}[\omega],$$

since ω is not expressible in the form

$$\omega = a + b\sqrt{-3}$$

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

Why then did we choose $\mathbb{Z}[\omega]$ for our ring of integers rather than $\mathbb{Z}[\sqrt{-3}]$? It is easy to see that if we had chosen the latter we would have lost unique factorisation. For consider

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

We have

$$z\bar{z} = 4|\omega|^2 = 4 = 2 \cdot 2.$$

However, z is 'prime' in $\mathbb{Z}[\sqrt{-3}]$; for if it factored, say

$$z = uv,$$

where u, v are not units then we must have

$$|u|^2 = 2 = |v|^2.$$

But if $u = a + b\sqrt{-3}$ (with $a, b \in \mathbb{Z}$) then

$$|u|^2 = a^2 + 3b^2 \neq 2.$$

It turns out that the solution to this 'problem' is simple; we must take all the *algebraic integers* in $\mathbb{Q}(\omega)$ as our ring of integers.

Definition 3.3. *The number $\alpha \in \mathbb{C}$ is said to be* algebraic *if it satisfies a polynomial equation*

$$f(x) = x^n + c_1 x^{n-1} + \cdots + c_n = 0$$

with $c_1, \ldots, c_n \in \mathbb{Q}$. We denote the set of algebraic numbers by $\bar{\mathbb{Q}}$.

The number α is said to be an algebraic integer *if it satisfies a polynomial equation*

$$f(x) = x^n + c_1 x^{n-1} + \cdots + c_n = 0$$

with $c_1, \ldots, c_n \in \mathbb{Z}$. We denote the set of algebraic integers by $\bar{\mathbb{Z}}$.

Proposition 3.12. *The algebraic integers $\bar{\mathbb{Z}}$ form an integral domain. The algebraic numbers $\bar{\mathbb{Q}}$ form a field, the field of fractions of $\bar{\mathbb{Z}}$. Moreover,*

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

Proof ▶ The last part is trivial; if $\alpha \in \mathbb{Q}$ then α satisfies the equation

$$x - \alpha = 0,$$

and so $\alpha \in \bar{\mathbb{Q}}$. Similarly if $\alpha \in \mathbb{Z}$ then $\alpha \in \bar{\mathbb{Z}}$.

We show next that $\bar{\mathbb{Q}}$ is a field. The complex numbers $\mathbb{C}$ form an infinite-dimensional vector space over $\mathbb{Q}$. We are interested in finite-dimensional subspaces of this vector space.

Lemma 10. *The number $\alpha \in \mathbb{C}$ is algebraic if and only if we can find a finite-dimensional vector subspace $V \subset \mathbb{C}$ over $\mathbb{Q}$ such that*

$$\alpha V \subset V.$$

Proof ▶ Suppose first that $\alpha \in \bar{\mathbb{Q}}$, say

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

Let

$$V = \langle 1, \alpha, \alpha^2, \ldots, \alpha^{n-1} \rangle$$

be the vector space spanned by $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$. Then it is easy to see that

$$\alpha V \subset V;$$

for

$$\alpha \cdot 1 = \alpha,\ \alpha \cdot \alpha = \alpha^2,\ \ldots, \alpha \cdot \alpha^{n-1} = \alpha^n = -a_1 \alpha^{n-1} - \cdots - a_n.$$

Conversely, suppose

$$\alpha V \subset V,$$

where V is a finite-dimensional subspace of $\mathbb{C}$. Let $e_1, \ldots, e_n$ be a basis for V. Then

$$\alpha e_i = a_{i1}e_1 + \cdots a_{in}e_n \qquad (1 \leq i \leq n).$$

It follows that

$$\det(\alpha I - A) = 0,$$

where A is the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ & \ldots & & \\ a_{n1} & a_{n2} & \ldots & a_{nn}. \end{bmatrix}$$

Thus α satisfies the polynomial equation

$$\det(xI - A) = 0,$$

with coefficients in $\mathbb{Q}$. Hence $\alpha \in \bar{\mathbb{Q}}$. ◀

Now suppose $\alpha, \beta \in \bar{\mathbb{Q}}$, say

$$\alpha^m + a_1\alpha^{m-1} + \cdots + a_m = 0,\ \beta^m + b_1\beta^{m-1} + \cdots + b_n = 0.$$

Let

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

be the vector space (over $\mathbb{Q}$) spanned by the numbers $\alpha^i\beta^j$. Then

$$\alpha V \subset V,\ \beta V \subset V.$$

It follows that

$$(\alpha \pm \beta)V \subset V,\ (\alpha\beta)V \subset V.$$

Hence

$$\alpha \pm \beta, \alpha\beta \in \bar{\mathbb{Q}}.$$

Suppose $\alpha \neq 0$. The sequence of decreasing vector spaces

$$V \supset \alpha V \supset \alpha^2 V \ldots$$

must be stationary, say

$$\alpha^{r+1}V = \alpha^r V.$$

Let $U = \alpha^r V$. Then

$$\alpha U = U,$$

and so

$$\alpha^{-1}U = U.$$

Hence $\alpha^{-1} \in \bar{\mathbb{Q}}$.

We have shown therefore that $\bar{\mathbb{Q}}$ is a field. We show next that $\bar{\mathbb{Z}}$ is a ring. The proof is superficially similar, using finitely-generated additive subgroups of $\mathbb{C}$ in place of finite-dimensional vector subspaces.

Lemma 11. *The number $\alpha \in \mathbb{C}$ is an algebraic integer if and only if we can find a finitely-generated subgroup $A \subset \mathbb{C}$ such that*

$$\alpha A \subset A.$$

Proof ▶ Suppose first that $\alpha \in \bar{\mathbb{Z}}$, say

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

where $a_1, \ldots, a_n \in \mathbb{Z}$. Let

$$A = \langle 1, \alpha, \alpha^2, \ldots, \alpha^{n-1} \rangle$$

be the abelian group generated by $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$, ie the set of numbers of the form

$$\theta = a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1},$$

where $a_0, a_1, \ldots, a_n \in \mathbb{Z}$.

Then we see as before that

$$\alpha A \subset A;$$

Conversely, suppose

$$\alpha A \subset A,$$

where A is a finitely-generated subgroup of $\mathbb{C}$. Let $g_1, \ldots, g_n$ be generators of A. Then

$$\alpha g_i = a_{i1}g_1 + \cdots a_{in}g_n \qquad (1 \le i \le n).$$

It follows as before that α satisfies the polynomial equation

$$\det(xI - A) = 0,$$

with coefficients now in $\mathbb{Z}$. Hence $\alpha \in \bar{\mathbb{Z}}$. ◀

Now suppose $\alpha, \beta \in \bar{\mathbb{Z}}$, say

$$\alpha^m + a_1\alpha^{m-1} + \cdots + a_m = 0,\ \beta^m + b_1\beta^{m-1} + \cdots + b_n = 0.$$

Let

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

be the abelian group generated by the numbers $\alpha^i\beta^j$. Then

$$\alpha V \subset V,\ \beta V \subset V.$$

It follows that

$$(\alpha \pm \beta)V \subset V,\ (\alpha\beta)V \subset V.$$

Hence

$$\alpha \pm \beta, \alpha\beta \in \bar{\mathbb{Z}},$$

and so $\bar{\mathbb{Z}}$ is a ring.

It remains to show that $\bar{\mathbb{Q}}$ is the field of fractions of $\bar{\mathbb{Z}}$, ie that every $z \in \bar{\mathbb{Q}}$ is expressible in the form

$$z = \frac{u}{v}$$

with $u, v \in \bar{\mathbb{Z}}$. In fact we shall prove the following stronger result.

Lemma 12. *Each algebraic number $\alpha \in \bar{\mathbb{Q}}$ is expressible in the form*

$$\alpha = \frac{\beta}{d}$$

where $\beta \in \bar{\mathbb{Z}}$, $d \in \mathbb{N}$.

Proof ▶ Suppose α satisfies the equation

$$\alpha^n + a_1\alpha^{n-1} + \cdots + a_n = 0.$$

On multiplying this by the product, d say, of the denominators of the a_i, we can write it in the form

$$d\alpha^n + b_1\alpha^{n-1} + \cdots + b_0 = 0,$$

where $b_1, \ldots, b_n \in \mathbb{Z}$. But then

$$\beta = d\alpha$$

satisfies the equation

$$\beta^n + db_1\beta^{n-1} + \cdots + d^n b_0 = 0,$$

and so $\beta \in \bar{\mathbb{Z}}$. ◀

◀

The following result — although we shall make no use of it — shows that $\bar{\mathbb{Q}}$ and $\bar{\mathbb{Z}}$ are both in a sense *complete*.

Proposition 3.13. *The algebraic numbers $\bar{\mathbb{Q}}$ are* algebraically closed*, ie if $\alpha \in \mathbb{C}$ satisfies the equation*

$$\alpha^n + a_1\alpha^{n-1} + \cdots + a_n = 0,$$

with $a_1, \ldots, a_n \in \bar{\mathbb{Q}}$, then $\alpha \in \bar{\mathbb{Q}}$.

Similarly, the algebraic integers $\bar{\mathbb{Z}}$ are integrally closed*, ie if $\alpha \in \mathbb{C}$ satisfies the equation*

$$\alpha^n + a_1\alpha^{n-1} + \cdots + a_n = 0,$$

with $a_1, \ldots, a_n \in \bar{\mathbb{Z}}$, then $\alpha \in \bar{\mathbb{Z}}$.

Proof ▶ Suppose c_i satisfies an equation of degree d_i over $\mathbb{Q}$. Let V be the vector space spanned by the elements

$$c_1^{j_1} \cdots c_n^{j_n} \qquad (0 \le j_1 < d_1, \ldots, 0 \le j_n < d_n).$$

Then it is readily verified that

$$\alpha V \subset V;$$

and so α is algebraic.

Similarly, if the c_i are algebraic integers, let A be the abelian group generated by the same set of elements. Then

$$\alpha A \subset A;$$

and so α is an algebraic integer. ◀

Finally, we introduce the general concept of a number field, and show that it is not perhaps as general as one might fear.

Definition 3.4. *A number field is a subfield $k \subset \mathbb{C}$ which is of finite dimension as a vector space over $\mathbb{Q}$.*

Proposition 3.14. *Every number field k is generated over $\mathbb{Q}$ by a single algebraic number α:*

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

Each element of k is expressible as a polynomial in α, and is algebraic.

Proof ▶ First of all, if $\alpha \in k$ then α is algebraic by Lemma 10, since

$$\alpha k \subset k.$$

To see that k is generated by a single element, it is sufficient to show that if $\alpha, \beta \in \bar{\mathbb{Q}}$ then

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

for some element θ.

To see this, suppose α, β satisfy the equations

$$f(\alpha) = \alpha^m + a_1\alpha^{m-1} + \cdots + a_m = 0,\ g(\beta) = \beta^n + b_1\beta^{n-1} + \cdots + b_n = 0.$$

Let the roots of these equations be

$$\alpha = \alpha_1, \ldots, \alpha_m,\ \beta = \beta_1, \ldots, \beta_n.$$

Now let

$$\theta = \alpha + c\beta,$$

where $c \in \mathbb{Q}$ is chosen so that the numbers $\alpha_i + c\beta_j$ are all distinct, ie

$$\alpha_i + c\beta_j = \alpha_{i'} + c\beta_{j'} \implies i = i', j = j'.$$

(This is certainly possible since we only have to avoid a finite number of numbers.)

We shall show that

$$\alpha, \beta \in \mathbb{Q}(\theta),$$

from which it will follow that $\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\theta)$, as required. The proof is deceptively simple.

Since

$$\alpha = \theta - c\beta,$$

β is a root of the equation

$$f(\theta - cx) = 0$$

over $\mathbb{Q}(\theta)$. But β is also a root of the equation

$$g(x) = 0.$$

These two equations have only the root β in common; for suppose a root of the second equation, say β_j, satisfies the first. Then

$$f(\theta - c\beta_j) = 0,$$

and so

$$\theta - c\beta_j = \alpha_i,$$

ie

$$\alpha_1 + c\beta_1 = \alpha_i + c\beta_j,$$

which is only possible if $i = j = 1$ by our choice of c.

It follows that the gcd of these two polynomials over $\mathbb{Q}(\theta)$ is $x - \beta$. Consequently, $\beta \in \mathbb{Q}(\theta)$; and so also $\alpha \in \mathbb{Q}(\theta)$, ie

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

◀

Algebraic number theory is the study of these number fields $k = \mathbb{Q}(\alpha)$. As *integers* in the number field k we take the algebraic integers in k, ie the ring

$$k \cap \bar{\mathbb{Z}}.$$

Since we introduced the concept of an algebraic integer in order to explain why we took $\mathbb{Z}[\omega]$ rather than $\mathbb{Z}[\sqrt{-3}]$ as our integers in $\mathbb{Q}(\omega) = \mathbb{Q}(\sqrt{-3})$, we should show that these are indeed the algebraic integers in this field.

Proposition 3.15. *The algebraic integers in $\mathbb{Q}(\omega)$ are just the numbers*

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

Proof ▶ Certainly

$$\omega \in \bar{\mathbb{Z}},$$

since ω satisfies $x^3 - 1 = 0$. Hence

$$a + b\omega \in \bar{\mathbb{Z}}$$

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

Conversely suppose

$$z = x + y \in \bar{\mathbb{Z}}$$

where $x, y \in \mathbb{Q}$. Since z and its complex conjugate $\bar{z}$ satisfy the same polynomial equations over $\mathbb{Z}$, it follows that

$$\bar{z} = x + y\omega^2 \in \bar{\mathbb{Z}}.$$

Hence

$$z + \bar{z} = 2x - y \in \bar{\mathbb{Z}} \cap \mathbb{Q} = \mathbb{Z}.$$

Similarly

$$\omega^2 z + \omega\bar{z} = -x + 2y \in \bar{\mathbb{Z}} \cap \mathbb{Q} = \mathbb{Z}.$$

It follows that

$$3x \in \mathbb{Z},\ 3y \in \mathbb{Z},\ x + y \in \mathbb{Z}.$$

In other words,

$$a = m/3,\ b = n/3,$$

where $m, n \in \mathbb{Z}$ and $3 \mid (m + n)$.

Thus if there were any algebraic integer of the form $x + y\omega$, where x and/or y is non-integral, then it would follow — on subtraction of $a + b\omega$ for suitable $a, b \in \mathbb{Z}$ — that

$$z = \frac{1}{3} - \frac{1}{3}\omega = \frac{\pi}{3} \in \bar{\mathbb{Z}}.$$

But that is impossible, since

$$z\bar{z} = \frac{1}{3} \notin \mathbb{Z}.$$

◀

Chapter 4

Primality in $\mathbb{Z}[\sqrt{2}]$

4.1 The ring $\mathbb{Z}[\sqrt{2}]$

Definition 4.1. *We denote by $\mathbb{Q}(\sqrt{2})$ the set of numbers of the form*

$$a + b\sqrt{2} \qquad (a, b \in \mathbb{Q});$$

and by $\mathbb{Z}[\sqrt{2}]$ the set of numbers of the form

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

Proposition 4.1. *$\mathbb{Z}[\sqrt{2}]$ is the ring of algebraic integers in the number field $\mathbb{Q}(\sqrt{2})$:*

$$\mathbb{Q}(\sqrt{2}) \cap \bar{\mathbb{Z}} = \mathbb{Z}[\sqrt{2}].$$

Proof ▶ First we show that $\mathbb{Q}(\sqrt{2})$ is a field, and $\mathbb{Z}[\sqrt{2}]$ a ring. Each is evidently closed under addition. Suppose

$$z = a + b\sqrt{2},\ w = A + B\sqrt{2}.$$

Then

$$zw = (aA + 2bB) + (aB + bA)\sqrt{2}.$$

Thus $\mathbb{Z}[\sqrt{2}]$ and $\mathbb{Q}(\sqrt{2})$ are both closed under multiplication. It follows that $\mathbb{Z}[\sqrt{2}]$ is a ring (in fact an integral domain).

Also

$$\begin{aligned} z/w &= \frac{a + b\sqrt{2}}{A + B\sqrt{2}} \\ &= \frac{(a + b\sqrt{2})(A - B\sqrt{2})}{(A + B\sqrt{2})(A - B\sqrt{2})} \\ &= \frac{(aA - 2bB) + (bA - aB)\sqrt{2}}{A^2 - 2B^2}, \end{aligned}$$

so $\mathbb{Q}(\sqrt{2})$ is closed under division by non-zero elements. Thus $\mathbb{Q}(\sqrt{2})$ is a field; and it is clearly the field of fractions of $\mathbb{Z}[\sqrt{2}]$.

It remains to show that $\mathbb{Z}[\sqrt{2}]$ is the integer ring of $\mathbb{Q}(\sqrt{2})$. The minimal polynomial of

$$z = a + b\sqrt{2}$$

is

$$(x-a)^2 = 2b^2,$$

ie

$$x^2 - 2ax + (a^2 - 2b^2) = 0.$$

It follows that $z \in \bar{\mathbb{Z}}$, ie z is an algebraic integer, if and only if

$$2a \in \mathbb{Z} \text{ and } a^2 - 2b^2 = n \in \mathbb{Z}.$$

Let $2a = a'$. Then

$$a'^2 - 8b^2 = 4n.$$

It follows that

$$(4b)^2 = 2a'^2 - 8n \in \mathbb{Z};$$

and so

$$4b \in \mathbb{Z},$$

say $4b = b''$. Then

$$2a'^2 - b''^2 = 8n.$$

Hence b'' is even, say $b'' = 2b'$, ie $2b' = b$. Thus

$$a'^2 - 2b'^2 = 4n.$$

Hence a' is even, ie $a \in \mathbb{Z}$. But now

$$2a^2 - b'^2 = 2n,$$

and so b' is even, ie $b \in \mathbb{Z}$. Thus $a, b \in \mathbb{Z}$, ie $z \in \mathbb{Z}[\sqrt{2}]$. ◀

Proposition 4.2. *The map*

$$z = a + b\sqrt{2} \mapsto \tilde{z} = a - b\sqrt{2} : \mathbb{Q}(\sqrt{2}) \to \mathbb{Q}(\sqrt{2})$$

is an automorphism, which sends $\mathbb{Z}[\sqrt{2}]$ into itself.

Definition 4.2. *We define the* norm *of $z = x + y\sqrt{2}$ to be*

$$N(z) = z\tilde{z} = a - 2b^2.$$

Proposition 4.3. *If $z, u \in \mathbb{Q}(\sqrt{2})$ then*

$$N(zu) = N(z)N(u);$$

and

$$N(z) = 0 \iff z = 0.$$

Furthermore, if $z \in \mathbb{Z}[\sqrt{2}]$ then

$$N(z) \in \mathbb{Z}.$$

Proof ▶ For the last part,

$$N(zu) = (zu)\widetilde{zu} = zu\tilde{z}\tilde{u} = (z\tilde{z})(u\tilde{u}) = N(u)N(v).$$

◀

4.2 Integers in $\mathbb{Q}(\sqrt{m})$

We have looked at three quadratic number fields — the 'purely imaginary' fields $\mathbb{Q}(i)$ and $\mathbb{Q}(\omega) = \mathbb{Q}(\sqrt{-3})$ and the real field $\mathbb{Q}(\sqrt{2})$.

In the first and last cases we found that the algebraic integers were just what might be expected: the ring $\mathbb{Z}[\sqrt{m}]$ (where $m = -1$ in the first case and $m = 2$ in the last case), consisting of all numbers of the form

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

However, in the second case we found that there were more integers than that; the integer ring was $\mathbb{Z}[\omega]$, consisting of all numbers of the form

$$\frac{1}{2}(c + d\sqrt{m}),$$

where now $c, d \in \mathbb{Z}$ with $c \equiv d \bmod 2$.

It is convenient at this point to determine the integer ring in the general quadratic number field $\mathbb{Q}(\sqrt{m})$, where we may suppose without loss of generality that $m \in \mathbb{Z}$ is square-free, since $\mathbb{Q}(\sqrt{d^2 m}) = \mathbb{Q}(\sqrt{m})$.

We omit proofs where these are essentially identical to those in the special cases considered earlier.

Proposition 4.4. *Suppose $m \in \mathbb{Z}$ is square-free, $m \neq 0, 1$. Then the numbers*

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

form a field $\mathbb{Q}(\sqrt{m})$; and the map

$$z = a + b\sqrt{m} \mapsto \tilde{z} = a - b\sqrt{m}$$

is an automorphism of $\mathbb{Q}(\sqrt{m})$ whose fixed points are the rationals $\mathbb{Q}$.

Definition 4.3. *If $z = a + b\sqrt{m} \in \mathbb{Q}(\sqrt{m})$, we set*

$$N(z) = z\tilde{z} = a^2 - mb^2.$$

Proposition 4.5. *If $z, w \in \mathbb{Q}(\sqrt{m})$ then*

$$N(zw) = N(z)N(w);$$

and

$$N(z) = 0 \iff z = 0.$$

Proposition 4.6. *Suppose $m \in \mathbb{Z}$ is square-free, $m \neq 0, 1$. If $m \not\equiv 1 \bmod 4$ then the integers in $\mathbb{Q}(\sqrt{m})$ are the numbers*

$$a + b\sqrt{m} \quad (a, b \in \mathbb{Z}).$$

If $m \equiv 1 \bmod 4$ then the integers in $\mathbb{Q}(\sqrt{m})$ are the numbers

$$\frac{1}{2}(c + d\sqrt{m}) \quad (c, d \in \mathbb{Z},\ c \equiv d \bmod 2).$$

Proof ▶ The minimal polynomial of $z = a + b\sqrt{m}$ is

$$(x - a)^2 = mb^2,$$

ie

$$x^2 - 2ax + (a^2 - mb^2) = 0.$$

It follows that $z \in \mathbb{Z}$, ie z is an algebraic integer, if and only if

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

Let $2a = a'$. Then

$$4mb^2 = a'^2 - 4n \in \mathbb{Z}.$$

Lemma 13. *If m is square-free and $x \in \mathbb{Q}$ then*

$$mx^2 \in \mathbb{Z} \implies x \in \mathbb{Z}.$$

Proof ▶ Suppose

$$x = \frac{r}{s}$$

in its lowest terms, ie $\gcd(r, s) = 1$. Then

$$mr^2 = ns^2.$$

Since $\gcd(r^2, s^2) = 1$ it follows that

$$s^2 \mid m.$$

Since m is square-free, this implies that $s = 1$, ie $x \in \mathbb{Z}$. ◀

By the Lemma,

$$m(2b)^2 \in \mathbb{Z} \implies 2b \in \mathbb{Z}.$$

Let $2b = b'$. Then

$$a'^2 - mb'^2 = 4n.$$

If a' is even then so is b', and vice versa. If a', b' are both odd, then

$$a'^2 \equiv 1 \equiv b'^2 \bmod 4$$

and so

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

Conversely, if $m \equiv 1 \bmod 4$ and a', b' are both odd, then

$$a'^2 - mb'^2 \equiv 0 \bmod 4,$$

and so

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

Thus the minimal polynomial of z has integral coefficients, and so $z \in \bar{\mathbb{Z}}$. ◀

4.3 The units in $\mathbb{Z}[\sqrt{2}]$

The main difference between real and imaginary quadratic number fields is that the former have an infinity of units.

Proposition 4.7. *The number*

$$u = a + b\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$$

is a unit if and only if

$$N(u) = a^2 - 2b^2 = \pm 1.$$

Proof ▶ If $N(u) = u\tilde{u} = \pm 1$ then u is certainly a unit with inverse $\pm\tilde{u}$. Conversely, suppose u is a unit, say

$$uv = 1.$$

Then

$$N(u)^2 N(v)^2 = 1 \implies N(u)^2 = \pm 1.$$

◀

Proposition 4.8. *Let*

$$\epsilon = 1 + \sqrt{2}.$$

Then the units in $\mathbb{Z}[\sqrt{2}]$ *are the numbers*

$$\pm\epsilon^n \quad (n \in \mathbb{Z}).$$

Proof ▶ Certainly ϵ is a unit, since

$$N(\epsilon) = -1.$$

Hence $\pm\epsilon^n$ are units for all $n \in \mathbb{Z}$.

Lemma 14. *Suppose $\eta = a + b\sqrt{2}$ is a unit. Then*

$$\eta > 1 \iff a > 0,\ b > 0.$$

Proof ▶ Suppose η is a unit $\neq \pm 1$. Then the 4 units

$$\{\pm\eta, \pm\eta^{-1}\}$$

fall into the 4 regions

$$x < -1,\ -1 < x < 0,\ 0 < x < 1,\ 1 < x.$$

For example, if $0 < \eta < 1$ then $-\eta^{-1} < -1,\ -1 < -\eta < 0,\ 1 < \eta^{-1}$; and so on.

Also, if $\eta = a + b\sqrt{2}$ then

$$\{\pm\eta, \pm\eta^{-1}\} = \{\pm a \pm b\sqrt{2}\},$$

It follows that the largest of these 4 units is the one with positive a, b; and this must be the one in the region $1 < x$. ◀

Corollary 4.1. *If η is a unit then*

$$\eta > 1 \implies \eta \geq \epsilon,$$

ie ϵ is the smallest unit > 1.

Since $\epsilon > 1$, the sequence ϵ^n is monotone increasing, with

$$\epsilon^n \to \infty \text{ as } n \to \infty.$$

Thus if $\eta > 1$ is a unit then we can find $n \in \mathbb{N}$ such that

$$\epsilon^n \leq \eta < \epsilon^{n+1}.$$

But then

$$1 \leq \epsilon^{-n}\eta < \epsilon.$$

Hence, from the Corollary above,

$$\epsilon^{-n}\eta = 1,$$

ie

$$\eta = \epsilon^n.$$

It follows that the units in $0 < x < 1$ are ϵ^n with $n < 0$; and similarly the units in $x < 0$ are $-\epsilon^n$. ◀

4.4 The units in $\mathbb{Q}(\sqrt{m})$

The structure of the unit-group in the general real quadratic field $\mathbb{Q}(\sqrt{m})$ (more precisely, in the integer-ring of this field) is exactly the same as in the case $\mathbb{Q}(\sqrt{2})$ considered above. The only difficulty is to show that there actually exist units apart from ± 1.

It is convenient to show at the same time the simpler result that the imaginary quadratic fields $\mathbb{Q}(\sqrt{-m})$ only contain a finite number of roots; so for the moment we allow m to be positive or negative.

Proposition 4.9. *Suppose $m \in \mathbb{N}$ is square-free, $m \neq 0, 1$; and suppose*

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

is an integer in $\mathbb{Q}(\sqrt{m})$. Then η is a unit if and only if

$$N(\eta) = a^2 - mb^2 = \pm 1.$$

Where a proof is identical to that in the special case $m = 2$ — as here — it is omitted.

Proposition 4.10. *Suppose $m \in \mathbb{N}$ is square-free, $m < 0$. Then the number of units in $\mathbb{Q}(\sqrt{m})$ is*

$$\begin{cases} 4 & \text{if } m = -1, \\ 6 & \text{if } m = -3, \\ 2 & \text{otherwise.} \end{cases}$$

Proof ▶ We may suppose that $m \neq -1, -3$, since we have already considered these cases.

If $\eta = a + b\sqrt{m}$ is a unit then, by the last Proposition,

$$a^2 + (-m)b^2 = 1.$$

Half-integers can only occur if $m \equiv 1 \bmod 4$, which implies that either $m = -3$ (which we have already considered) or else $m \leq -7$, in which case $(-m)b^2 > 1$.

Thus we may suppose that $a, b \in \mathbb{Z}$; and then it is clear that $b = 0$ unless $m = -1$ (which we have already considered) and so $\eta = \pm 1$. ◀

Now let us turn to the real case.

Proposition 4.11. *Suppose $m \in \mathbb{N}$ is square-free, $m > 1$. If $\eta = a + b\sqrt{m}$ is a unit then*

$$\eta > 1 \iff a > 0,\ b > 0.$$

Corollary 4.2. *If there are any units in $\mathbb{Q}(\sqrt{m})$ apart from ± 1 then there is a least such unit $\epsilon > 1$.*

Proof ▶ This follows at once, since a, b are integers or half-integers. ◀

Definition 4.4. *We call the least unit* > 1 *(if it exists) the* fundamental unit *in the real quadratic number field* $\mathbb{Q}(\sqrt{m})$.

Proposition 4.12. *Suppose* $m \in \mathbb{N}$ *is square-free,* $m > 1$. *Then if there exists any units in* $\mathbb{Q}(\sqrt{m})$ *apart from* ± 1, *there is a fundamental unit* ϵ; *and the units are just*

$$\pm\epsilon^n \quad (n \in \mathbb{Z}).$$

We shall show that there *are* always units in a real quadratic field apart from ± 1, so the fundamental unit ϵ *is* always defined, and the unit-group takes the form $\{\pm\epsilon^n\}$.

More precisely, we shall show that *Pell's equation*

$$x^2 - my^2 = 1$$

always has an infinity of solutions $x, y \in \mathbb{Z}$.

4.4.1 Approximation of irrationals

Suppose

$$x^2 - my^2 = 1,$$

where $x, y \in \mathbb{N}$. We can write this equation

$$(x - \sqrt{m}y)(x + \sqrt{m}y) = 1.$$

Thus

$$\begin{aligned} |x - \sqrt{m}y| &= \frac{1}{x + \sqrt{m}y} \\ &< \frac{1}{x}, \end{aligned}$$

and so

$$|\sqrt{m} - \frac{y}{x}| < \frac{1}{x^2}.$$

We may say that y/x is a *quadratic approximation* to $\sqrt{m}$.

Thus solutions to Pell's equation are related to approximations to $\sqrt{m}$. A subtle application of the Pigeon Hole Principle shows that every irrational has quadratic approximations.

Theorem 4.1. *(Kronecker's Theorem) Suppose* α *is irrational, ie* $\alpha \in \mathbb{R} \setminus \mathbb{Q}$; *and suppose* $N \in \mathbb{N},\ N > 0$. *Then there exist* $m, n \in \mathbb{Z}$ *with* $1 \leq n \leq N$ *such that*

$$|n\alpha - m| < \frac{1}{N}.$$

Proof ▶ Let $\{x\}$ denote the fractional part of $x \in \mathbb{R}$, so that

$$x = [x] + \{x\}.$$

Thus

$$0 \leq \{x\} < 1.$$

Let us $[0, 1)$ into N equal parts:

$$0 \leq x < \frac{1}{N},\ \frac{1}{N} \leq x < \frac{2}{N},\ \ldots \frac{N-1}{N} \leq x < 1.$$

Consider the $N + 1$ fractional parts $\{i\alpha\}$ $(0 \leq i \leq N)$. By the Pigeon Hole Principle, two of these must fall into the same subdivision, say

$$\frac{r}{N} \leq \{i\alpha\}, \{j\alpha\} < \frac{r+1}{N},$$

where, say, $0 \leq i < j \leq N$. Then

$$|\{j\alpha\} - \{i\alpha\}| < \frac{1}{N}$$

ie

$$|j\alpha - [j\alpha] - (i\alpha - [i\alpha])| < \frac{1}{N}$$

ie

$$|n\alpha - m| < \frac{1}{N},$$

where

$$n = j - i,\ m = [j\alpha] - [i\alpha].$$

◀

Corollary 4.3. *There exist an infinity of rationals $\frac{y}{x}$ with*

$$|\alpha - \frac{y}{x}| < \frac{1}{x^2}.$$

The following result, although not strictly necessary for our purposes, shows that one cannot do better than this, at least for 'quadratic surds' $a + \sqrt{m}b$.

Proposition 4.13. *(Liouville's Theorem) Suppose α is an algebraic number of degree $n > 1$. Then there is a constant $C = C(\alpha) > 0$ such that*

$$|\alpha - \frac{y}{x}| \geq \frac{C}{y^n}.$$

Proof ▶ Let

$$m(t) = c_0 t^n + c_1 t^{n-1} + \cdots + c_n \quad (c_0, c_1, \ldots, c_n \in \mathbb{Z})$$

be the minimal polynomial for α. If now $x, y \in \mathbb{Z}$ then

$$x^n m\left(\frac{y}{x}\right) = c_0 y^n + c_1 y^{n-1} x + \cdots + c_n x^m \in \mathbb{Z}.$$

Thus

$$|x^n m\left(\frac{y}{x}\right)| \geq 1,$$

ie

$$|m\left(\frac{y}{x}\right)| \geq \frac{1}{x^n},$$

if $x > 0$.

But now consider the function

$$f(t) = \frac{m(t)}{t - \alpha}.$$

Since $(t - \alpha) \mid m(t)$, this function is bounded on any finite interval, say

$$|\frac{m(t)}{t - \alpha}| \leq c \text{ if } |t - \alpha| \leq 1$$

for some $c > 0$. Hence

$$\begin{aligned} |\alpha - \frac{y}{x}| &\geq c^{-1} |m\left(\frac{y}{x}\right)| \\ &\geq \frac{1}{cy^n} \end{aligned}$$

provided $|\alpha - \frac{y}{x}| \leq 1$. These two inequalities "pull in the same direction"; they say that if y/x is an approximation to α then it is not too good an approximation. We conclude that

$$|\alpha - \frac{y}{x}| \geq \frac{C}{x^n}$$

where $C = \max(c^{-1}, 1)$. ◀

Corollary 4.4. *If α is an irrational quadratic surd then there exists a $C > 0$ such that*

$$|\alpha - \frac{y}{x}| \geq \frac{C}{y^2}.$$

For quadratic surds, the two bounds are of the same order; there exist constants $0 < C_1 \leq C_2$ such that

$$|\alpha - \frac{y}{x}| \geq \frac{C_1}{y^2},$$

but there exist an infinity of rationals y/x such that

$$|\alpha - \frac{y}{x}| \leq \frac{C_2}{y^2}.$$

A very difficult theorem, due to Roth, states that for any algebraic number α and any $\epsilon > 0$ there exists a constant C such that

$$|\alpha - \frac{y}{x}| \geq \frac{C_1}{y^{2+\epsilon}}.$$

Roughly speaking, we cannot hope for greater than quadratic approximation to an algebraic number.

4.4.2 Pell's equation

Theorem 4.2. *(Pell's Theorem) Suppose $m \in \mathbb{N}$ is square-free, $m > 1$. Then the equation*

$$x^2 - my^2 = 1$$

has an infinity of solutions with $x, y \in \mathbb{Z}$.

Proof ▶ It is sufficient to find one solution (x, y) with $y > 0$; for then if we set

$$X + \sqrt{m}Y = (x + \sqrt{m}y)^n$$

(X, Y) will be a solution, since

$$N(X + \sqrt{m}Y) = N(x + \sqrt{m})^n = 1.$$

By Kronecker's Theorem, we can find an infinity of $x, y \in \mathbb{Z}$ such that

$$|\sqrt{m} - \frac{x}{y}| < \frac{1}{y^2}.$$

But then

$$\begin{aligned} |\sqrt{m} + \frac{x}{y}| &\leq |\frac{x}{y} - \sqrt{m}| + 2\sqrt{m} \\ &\leq 2\sqrt{m} + \frac{1}{y^2} \\ &\leq 3\sqrt{m}. \end{aligned}$$

Thus

$$|m - \frac{x^2}{y^2}| < \frac{3\sqrt{m}}{y^2}$$

ie

$$|x^2 - my^2| < 3\sqrt{m}.$$

Thus there are an infinity of integers x, y satisfying this inequality, where we may also assume that $\gcd(x, y) = 1$. It follows that for some integer $c \in (-3\sqrt{m}, 3\sqrt{m})$ there are an infinity of solutions of

$$x^2 - my^2 = c$$

with $\gcd(x, y) = 1$.

Suppose $(x, y),\ (X, Y)$ are two such solutions, Consider

$$\begin{aligned}\frac{x + \sqrt{m}y}{X + \sqrt{m}Y} &= \frac{(x + \sqrt{m}y)(X + \sqrt{m}Y)}{c} \\ &= \frac{xX - myY}{c} + \sqrt{m}\frac{Xy - Yx}{c} \\ &= u + \sqrt{m}v,\end{aligned}$$

say. Then

$$N(u + \sqrt{m}v) = \frac{N(x + \sqrt{m}y}{N(X + \sqrt{m}Y} = \frac{c}{c} = 1,$$

ie

$$u^2 - mv^2 = 1.$$

Although $u, v \in \mathbb{Q}$, in general $u, v \notin \mathbb{Z}$. We have to show that sometimes, at least, this is so.

Since

$$v \in \mathbb{Z} \implies u^2 = 1 + mv^2 \in \mathbb{Z} \implies u \in \mathbb{Z},$$

it is sufficient to ensure that $v \in \mathbb{Z}$, ie

$$xY - yX \equiv 0 \bmod c.$$

Since $x^2 - my^2 = c$ and $\gcd(x, y) = 1$,

$$\gcd(y, c) = 1.$$

Similarly

$$\gcd(Y, c) = 1.$$

It follows that we can write the condition for v (and therefore also u) to be integral in the form

$$\frac{x}{y} \equiv \frac{X}{Y} \bmod c.$$

But since there are at most $|c|$ residues mod$c$, and since there are an infinity of solutions (x, y), we can find an infinite number having the same residue x/y mod c. Any two of these $(x, y), (X, Y)$ will give a solution (u, v) of $u^2 - mv^2 = 1$ with $u, v \in \mathbb{Z}$ (and $v \neq 0$). ◀

Corollary 4.5. *Suppose $m \in \mathbb{N}$ is square-free, $m > 1$. Then there exist an infinity of units in $\mathbb{Q}(\sqrt{m})$.*

As we have seen, this implies that there is a smallest unit $\epsilon < 1$; and the units in $\mathbb{Q}(\sqrt{m})$ (or more precisely, in the corresponding ring of integers) are just

$$\pm\epsilon^n \quad (n \in \mathbb{Z}).$$

4.5 The Euclidean algorithm in $\mathbb{Z}[\sqrt{2}]$

Suppose $z, w \in \mathbb{Z}[\sqrt{2}]$. Let

$$z/w = x + y\sqrt{2}$$

where $x, y \in \mathbb{Q}$. Let a, b be the closest integers to x, y, say

$$s = x - 1/2 \leq a < x + 1/2,\ t = y - 1/2 \leq b < y + 1/2.$$

Set

$$q = a + b\sqrt{2}$$

Then

$$z/w - q = s + t\sqrt{2},$$

and

$$N(z/w - q)^2 = s^2 - 2t^2.$$

It follows that

$$|N(z/w - q)| \leq \frac{1}{2},$$

ie

$$|N(z - wq)| \leq \frac{1}{2}|N(w)|.$$

In particular, we have established

Proposition 4.14. *Suppose $z, w \in \mathbb{Z}[\sqrt{2}]$, with $w \neq 0$. Then we can find $q, r \in \mathbb{Z}[\sqrt{2}]$ such that*

$$z = qw + r, \quad |N(r)| < |N(w)|.$$

This proposition allows us to implement the Euclidean algorithm; and as a consequence we have

Theorem 4.3. *(The Fundamental Theorem of Arithmetic for $\mathbb{Z}[\sqrt{2}]$) Each number $z \in \mathbb{Z}[\sqrt{2}]$ can be factorised into prime factors; and the factorisation is unique up to the order of the factors, and multiplication by units.*

4.6 The primes in $\mathbb{Z}[\sqrt{2}]$

Proposition 4.15. *A prime number $p \in \mathbb{N}$ splits into at most 2 factors in $\mathbb{Z}[\sqrt{2}]$.*

We omit proofs where they are essentially identical — as here — to cases considered earlier.

Proposition 4.16. *The prime number $p \in \mathbb{N}$ splits in $\mathbb{Z}[\sqrt{2}]$ if and only if p is expressible in the form*

$$p = a^2 - 2b^2$$

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

Proof ▶ If

$$\begin{aligned} p &= a^2 - 2b^2 \\ &= (a - b\sqrt{2})(a - b\sqrt{2}^2) \end{aligned}$$

we have an explicit factoring of p.

Conversely, suppose p splits, say

$$p = \pi_1 \pi_2.$$

Then

$$N(p_1)N(p_2) = N(p) = p^2.$$

It follows that

$$N(\pi_1) = \pm p = N(\pi_2).$$

If $\epsilon = 1 + \sqrt{2}$ then

$$N(\epsilon) = -1.$$

Thus if $N(p_1) = -p$ then

$$N(\epsilon p_1) = p,$$

and so we may assume that

$$N(p_1) = p.$$

Thus if $p_1 = a + b\sqrt{2}$ then

$$p = a^2 - 2b^2.$$

◀

Since

$$2 = (\sqrt{2})^2,$$

the prime number 2 *ramifies* in $\mathbb{Z}[\sqrt{2}]$, ie splits into 2 equal primes.

Proposition 4.17. *Suppose $p \in \mathbb{N}$ is an odd prime. Then p splits in $\mathbb{Z}[\sqrt{2}]$ if and only if*

$$\left(\frac{2}{p}\right) = 1.$$

Proof ▶ If p splits, then as we have seen we can find $a, b \in \mathbb{Z}$ such that

$$a^2 - 2b^2 = p.$$

Evidently $p \nmid b$ since

$$p \mid b \implies p \mid a \implies p^2 \mid p.$$

Thus b has an inverse $b^{-1} \bmod p$; and

$$(ab^{-1})^2 \equiv 2 \bmod p.$$

Thus

$$\left(\frac{2}{p}\right) = 1.$$

Suppose conversely that this is so. Then we can find $a \in (0, p)$ such that

$$a^2 \equiv 2 \bmod p,$$

ie

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

ie

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

Suppose p does *not* split. Then by the Fundamental Theorem, p divides one of the factors on the left, say

$$a + \sqrt{2} = pz,$$

where $z = c + d\sqrt{2}$ with $c, d \in \mathbb{Z}$. But then comparing the coefficients of $\sqrt{2}$,

$$1 = pd,$$

which is absurd.

We conclude that p splits if and only if $\left(\frac{2}{p}\right) = 1$, ie 2 is a quadratic residue *bmodp*. ◀

Proposition 4.18. *Suppose $p \in \mathbb{N}$ is an odd prime number. 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 ▶ We showed in the last Chapter that

$$\left(\frac{a}{p}\right) = (-1)^{\mu},$$

where

$$\mu = -[p/2a] + [p/a] - [3p/2a] + \cdots + [rp/a],$$

with $r = [a/2]$.

In the present case, where $a = 2$, this reduces to

$$\mu = -[p/4] + [p/2].$$

For $n \in \mathbb{N}$, let

$$f(n) = -[n/4] + [n/2].$$

If n is increased by 8, then $[n/4]$ is increased by 2, and $[n/2]$ by 4. It follows that

$$f(n+8) = f(n) + 2.$$

In particular, the parity of $f(n)$ (which is all that concerns us) is unchanged if n is increased by 8. In other words,

$$m \equiv n \bmod 8 \implies f(m) \equiv f(n) \bmod 2.$$

Thus it is only necessary to compute $f(n)$ for $0 \leq n < 8$; and since we are only concerned with the value when n is prime, it is sufficient to consider the values $1, 3, 5, 7$. Since

$$\begin{aligned} f(1) &= -0 + 0 = 0, \\ f(3) &= -0 + 1 = 1, \\ f(5) &= -1 + 2 = 1, \\ f(7) &= -1 + 3 = 2, \end{aligned}$$

we conclude that μ is even, ie $\left(\frac{2}{p}\right) = 1$, if $n \equiv 1, 7 \bmod 8$, or in other words, if $n \equiv \pm 1 \bmod 8$; and μ is odd, ie $\left(\frac{2}{p}\right) = 1$, if $n \equiv 3, 5 \bmod 8$, or in other words if $n \equiv \pm 3 \bmod 8$. ◀

To summarize,

Corollary 4.6. *The prime number $p \in \mathbb{N}$ splits in $\mathbb{Z}[\sqrt{2}]$ if $p = 2$ (in which case it splits into two equal factors) or $p \equiv \pm 1 \bmod 8$ (in which case it splits into two different factors). If $p \equiv \pm 3 \bmod 8$ it does not split.*

Example. The Diophantine equation

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

has no solution with $x, y \in \mathbb{Z}$ since 3 does not split in $\mathbb{Z}[\sqrt{2}]$. On the other hand,

$$x^2 - 2y^2 = 7$$

does have a solution. Inspection gives the solution $x = 3,\ y = 1$, corresponding to the factorization

$$7 = (3 + \sqrt{2})(3 - \sqrt{2}).$$

It follows from the Fundamental Theorem that the general solution is given by

$$x + \sqrt{2}y = \pm\epsilon^n(3 \pm \sqrt{2}),$$

where

$$\epsilon = 1 + \sqrt{2},$$

and n must be *even* to ensure that $x^2 - 2y^2 = +1$ rather than -1.

Taking $n = 2$, for example, gives the solution

$$\begin{aligned} x + \sqrt{2}y &= (1 + \sqrt{2})^2(3 - \sqrt{2}) \\ &= (3 + 2\sqrt{2})(3 - \sqrt{2}) \\ &= 5 + 3\sqrt{2}, \end{aligned}$$

ie $x = 5,\ y = 3$.

Chapter 5

Arithmetic in $\mathbb{Q}[\sqrt{5}]$

We have already established the basic facts about this field. In particular, since $5 \equiv 1 \bmod 4$ the integers in the field are the numbers

$$\frac{1}{2}(c + d\sqrt{5}) \quad (c, d \in \mathbb{Z},\ c \equiv d \bmod 2).$$

Let

$$\omega = \frac{1}{2}(1 + \sqrt{5}).$$

Then the integers are just the numbers of the form

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

In other words,

$$\mathbb{Q}(\sqrt{5}) \cap \bar{\mathbb{Z}} = \mathbb{Z}[\omega].$$

Note that ω is a unit; for

$$N(\omega) = \omega\tilde{\omega} = -1,$$

since $\omega, \tilde{\omega}$ are the roots of

$$x^2 - x - 1 = 0.$$

In fact ω is the fundamental unit; for if $\eta = a + b\sqrt{5}$ is a unit then

$$\begin{aligned} \eta > 1 &\implies a, b > 0 \\ &\implies a, b \geq \frac{1}{2} \\ &\implies \eta \geq \frac{1}{2}(1 + \sqrt{5}) = \omega. \end{aligned}$$

Thus the units in $\mathbb{Q}(\sqrt{5})$ are

$$\pm\omega^n \quad (n \in \mathbb{Z}).$$

5.1 The Fundamental Theorem

We extend the Euclidean Algorithm to $\mathbb{Z}[\omega]$ in the usual way, although now we have two slightly different routes we can follow.

First of all, given $z, w \in \mathbb{Z}/[\omega]$ we can express z/w in the form

$$z/w = x + y\omega,$$

where $x, y \in \mathbb{Q}$. Let a, b be the closest integers to x, y, say

$$x = a + s,\ y = b + t,$$

where $0 \leq |s|, |t| < \frac{1}{2}$; and set

$$q = a + b\omega,\ r = z - qw.$$

Then

$$\begin{aligned} N(r/w) &= N(z/w - q) \\ &= N(s + t\omega) \\ &= (s + t\omega)(s + \tilde{\omega}) \\ &= s^2 + st - t^2. \end{aligned}$$

It follows that

$$|N(r/w)| < \frac{3}{4}.$$

ie

$$|N(r)| < \frac{3}{4}|N(w)|.$$

Thus we have established

Proposition 5.1. *Suppose $z, w \in \mathbb{Z}[\omega]$ with $w \neq 0$. Then there exist $q, r \in \mathbb{Z}[\omega]$ such that*

$$z = qw + r,\ |N(r)| < |N(w)|.$$

We can get a slightly sharper result if we express z/w in the form

$$z/w = x + y\sqrt{5}.$$

Now choose d to be the nearest *half-integer* (or integer) to y, say

$$y = d + t,\ |t| < \frac{1}{4}.$$

Next we choose c to be the nearest integer or half-integer to x, according as d was an integer or half-integer, say

$$x = c + s,\ |s| < \frac{1}{2}.$$

If now we set

$$q = c + d\sqrt{5},\ z - qw = r,$$

then

$$N(r/q) = N(s + t\sqrt{5}) = s^2 - 5t^2.$$

It follows that

$$-\frac{5}{16} < N(r/q) < \frac{1}{4};$$

and so

$$|N(r/q)| < \frac{5}{16},$$

quite an improvement on the $\frac{3}{4}$ we obtained before.

But whichever way we go, we derive

Theorem 5.1. *(The Fundamental Theorem of Arithmetic for $\mathbb{Q}(\sqrt{5})$) Each number $z \in \mathbb{Z}[\omega]$ can be factorised into prime factors; and the factorisation is unique up to the order of the factors, and multiplication by units.*

5.2 The primes in $\mathbb{Z}[\omega]$

Proposition 5.2. *A prime number $p \in \mathbb{N}$ splits into at most 2 factors in $\mathbb{Z}[\omega]$.*

We omit proofs that are identical to those considered earlier.

Proposition 5.3. *Suppose $p \in \mathbb{N}$ is a prime number, $p \neq 2, 5$. Then p splits in $\mathbb{Z}[\omega]$ if and only if*

$$\left(\frac{5}{p}\right) = 1.$$

Proof ▶ Suppose p splits, say

$$p = \pi_1\pi_2.$$

Then

$$N(p_1)N(p_2) = N(p) = p^2.$$

It follows that

$$N(\pi_1) = \pm p = N(\pi_2).$$

Let

$$\pi_1 = \frac{1}{2}(a + b\sqrt{5}) \quad (a, b \in \mathbb{Z}).$$

Then

$$N(\pi_1) = \frac{1}{4}(a^2 - 5b^2) = \pm p,$$

ie

$$a^2 - 5b^2 = \pm 4p.$$

Evidently $p \nmid b$. Hence

$$(ab^{-1})^2 - 5 \equiv 0 \bmod p;$$

and so

$$\left(\frac{5}{p}\right) = 1.$$

Conversely, suppose that is so. Then we can find $a \in (0, p)$ such that

$$a^2 \equiv 5 \bmod p,$$

ie

$$a^2 - 5 = pm,$$

ie

$$(a - \sqrt{5})(a + \sqrt{5}) = pm.$$

If p does *not* split then by the Fundamental Theorem it must divide one of the factors on the left, say

$$a + \sqrt{5} = p(c + d\omega),$$

where c, d are integers or half-integers. But then comparing the coefficients of $\sqrt{5}$,

$$1 = pd,$$

which is absurd.

We conclude that p splits if and only if $\left(\frac{5}{p}\right) = 1$, ie 5 is a quadratic residue *bmodp*. ◀

It remains to consider the cases $p = 2, 5$.

The prime 5 is evidently ramified in $\mathbb{Z}[\omega]$:

$$5 = (\sqrt{5})^2.$$

By contrast, the prime 2 does not split. For if it did then from above there would be integers a, b such that

$$a^2 - 5b^2 = \pm 8.$$

Evidently a, b are both even, or both odd. If they are both odd, then

$$a^2 \equiv b^2 \equiv 1 \bmod 8,$$

and so

$$a^2 - 5b^2 \equiv 4 \bmod 8.$$

On the other hand, if a, b are both even, say $a = 2c,\ b = 2d$ then

$$c^2 - 5d^2 = 2.$$

Evidently c, d must both be odd. But then

$$c^2 \equiv d^2 \equiv 1 \bmod 4,$$

and so

$$c^2 - 5d^2 \equiv 0 \bmod 4.$$

5.3 Gauss' Quadratic Reciprocity Law

To determine which prime numbers $p \in \mathbb{N}$ split in $\mathbb{Z}[\omega]$, we need to computer $\left(\frac{5}{p}\right)$. We could do this in the same way that we computed $\left(\frac{2}{p}\right)$. using the formula

$$\left(\frac{5}{p}\right) = (-1)^{\mu},$$

where

$$\mu = -[p/10] + [p/5] - [3p/10] + [2p/5].$$

Since it is easy to show that the parity of

$$f(n) = -[n/10] + [n/5] - [3n/10] + [2n/5]$$

only depends on n mod 10, it is sufficient to consider the cases $n = 1, 3, 7, 9$.

However, there is a much simpler way, once we have established what has been called "the most beautiful result in the whole of number theory".

Theorem 5.2. *(The Law of Quadratic Reciprocity) Suppose $p, q \in \mathbb{N}$ are odd primes. Then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}} = \begin{cases} -1 & \text{if } p \equiv q \equiv 3 \bmod 4 \\ 1 & \text{otherwise.} \end{cases}$$

Proof ▶ Let

$$S = \{1, 2, \ldots, \frac{p-1}{2}\},\ T = \{1, 2, \ldots, \frac{q-1}{2}\}.$$

We shall choose remainders mod$p$ from the set

$$\{-\frac{p}{2} < i < \frac{p}{2}\} = -S \cup \{0\} \cup S,$$

and remainders mod$q$ from the set

$$\{-\frac{q}{2} < i < \frac{q}{2}\} = -T \cup \{0\} \cup T.$$

By Gauss' Lemma,

$$\left(\frac{q}{p}\right) = (-1)^{\mu}, \ \left(\frac{p}{q}\right) = (-1)^{\nu},$$

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,

$$0 < i < \frac{p}{2} \implies 0 < j < \frac{q}{2}.$$

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

$$R = \{(x, y) : 0 < x < \frac{p}{2}, \ 0 < y < \frac{p}{2}\}.$$

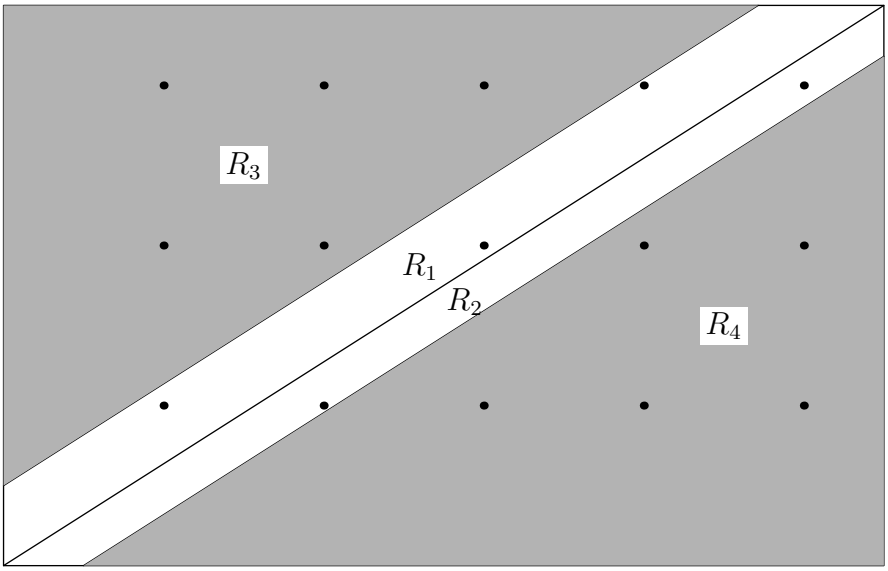

Figure 5.1: $p = 11,\ q = 7$

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 5.3).

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

$$[R_1] + [R_2] + [R_3] + [R_4] = [R] = \frac{p-1}{2}\frac{q-1}{2},$$

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

$$R' = \{1 \leq x \leq \frac{p-1}{2},\ 1 \leq y \leq \frac{q-1}{2}\}.$$

Now we take P to be the centre of this rectangle, ie

$$P = (\frac{p+1}{2},\ \frac{q+1}{2}).$$

The reflection is then given by

$$(x, y) \mapsto (X, Y) = (p + 1 - x, q + 1 - y).$$

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 mod47.

We could have avoided using the result for $\left(\frac{2}{p}\right)$:

$$\begin{aligned}\left(\frac{10}{37}\right) &= \left(\frac{-27}{37}\right)\\ &= \left(\frac{-1}{37}\right)\left(\frac{3}{37}\right)^3\\ &= (-1)^{18}\left(\frac{37}{3}\right)\\ &= \left(\frac{1}{3}\right) = 1.\end{aligned}$$

Proposition 5.4. *Suppose $p \in \mathbb{N}$ is a prime number. Then*

$$p \begin{cases} \textit{splits into 2 equal primes} & \textit{if } p = 5\\ \textit{splits into 2 distinct primes} & \textit{if } p \equiv \pm 15\\ \textit{does not split} & \textit{if } p \equiv \pm 25.\end{cases}$$

Proof ▶ We know that if $p \neq 2, 5$ then its splitting depends on the value of $\left(\frac{5}{p}\right)$. But by the Quadratic Reciprocity Law, since $5 \equiv 1 \bmod 4$,

$$\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right) = \begin{cases}1 & \text{if } p \equiv \pm 1 \bmod 5\\ -1 & \text{if } p \equiv \pm 3 \bmod 5.\end{cases}$$

We note that if $p = 2$ then p does not split, and $p \equiv 2 \bmod 5$, so the result is also valid in this case.

It remains to show that if $p \equiv 1 \bmod 5$ then p splits into *distinct* factors.

◀

5.4 Fibonacci numbers

Definition 5.1. *The Fibonacci sequence u_n is defined by the linear recursion relation*

$$u_{n+2} = u_{n+1} + u_n \quad (n \in \mathbb{N}),$$

with initial values

$$u_0 = 1,\ u_1 = 1.$$

Thus the sequence runs

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, \ldots$$

We recall that there is a simple formula for the solution to a linear recursion relation like this.

Proposition 5.5. *Suppose the sequence $a_n (n \in \mathbb{N})$ satisfies the linear recursion relation*

$$a_{n+2} = Aa_{n+1} + Ba_n \quad (n \in \mathbb{N}),$$

where A, B are constants; and suppose the equation

$$x^2 - Ax - B = 0$$

has distinct roots λ, μ. Then there exist constants C, D such that

$$a_n = C\lambda^n + D\mu^n$$

for all $n \in \mathbb{N}$.

Proof ▶ It is easy to see that $a_n = \lambda^n$ satisfies the relation, since

$$a_{n+2} - Aa_{n+1} - Ba_n = \lambda^n(\lambda^2 - A\lambda - B) = 0.$$

Similarly $a_n = \mu^n$ satisfies the equation; and so, since the equation is linear, does

$$a_n = C\lambda^n + D\mu^n$$

for any C, D.

But it is clear that the sequence is completely determined once a_0, a_1 are given. Thus if we choose A, B to satisfy

$$A + B = a_0,\ A\lambda + B\mu = a_1$$

then our sequence must coincide with the given one. Since $\lambda \neq \mu$ these equations have a simultaneous solution; so the result follows. ◀

If the "auxiliary equation"

$$x^2 - Ax - B = 0$$

has equal roots λ, λ then the general solution takes the form

$$a_n = \lambda^n(A + Bn).$$

The theory extends without difficulty to the general linear recursion relation

$$a_{n+r} = A_1 a_{n+r-1} + A_2 a_{n+r-2} + \cdots + A_r a_n \quad (n \in \mathbb{N}).$$

If the auxiliary equation

$$x^r = A_1 x^{r-1} + A_2 x^{r-2} + \cdots + A_r$$

has distinct roots $\lambda_1, \ldots, \lambda_r$, then the general solution (by the same argument as in the case $r = 2$) is

$$a_n = C_1\lambda_1^n + \cdots + C_r\lambda_r^n.$$

If a root, λ_1 say, is repeated s times then $\lambda_1^n, \lambda_2^n, \ldots, \lambda_s^n$ are replaced by $\lambda_1^n, n\lambda_1^n, \ldots, n^{s-1}\lambda_1^n$, ie λ_1^n is multiplied by a general polynomial in n of degree $s - 1$.

We shall not give a formal proof of this, since we shall only meet the simplest case in the Proposition above, where $r = 2$ and the roots are distinct.

Proposition 5.6. *The Fibonnaci numbers are given by*

$$u_n = \frac{\omega^n - \tilde{\omega}^n}{\omega - \tilde{\omega}} = \begin{cases} \frac{\omega^n - \omega^{-n}}{\omega + \omega^{-1}} & \text{if } n \text{ is even,} \\ \frac{\omega^n + \omega^{-n}}{\omega + \omega^{-1}} & \text{if } n \text{ is odd.} \end{cases}$$

Proof ▶ The auxiliary equation is

$$x^2 - x - 1 = 0,$$

which has roots

$$\omega = \frac{1}{2}(1 + \sqrt{5}),\ \tilde{\omega} = \frac{1}{2}(1 - \sqrt{5}).$$

Since

$$N(\omega) = \omega\tilde{\omega} = -1,$$

we can equally well write the roots as

$$\omega, -\omega^{-1}.$$

The last Proposition tells us that the sequence is given by

$$u_n = A\omega^n + B\tilde{\omega}^n = A\omega^n + (-1)^n\omega^{-n}.$$

For $n = 0, 1$ this gives

$$u_0 = 0 = A + B,\ u_1 = 1 = A\omega + B\tilde{\omega}.$$

Thus

$$A = \frac{1}{\omega - \tilde{\omega}} = -B;$$

and so

$$u_n = \frac{\omega^n - \tilde{\omega}^n}{\omega - \tilde{\omega}}.$$

◀

Proposition 5.7. *Suppose* $m, n \in \mathbb{N}$. *Let* $\gcd(m, n) = d$. *Then*

$$\gcd(u_m, u_n) = u_d.$$

Proof ▶ We prove the result by induction on $\max(m, n)$. We may suppose that $m \leq n$. The result is trivial if $m = n$ or if $m = 0$. Thus we may suppose that

$$0 < m < n.$$

Since $\gcd(m, n - m) = \gcd(m, n)$ it is sufficient to show that

$$\gcd(u_m, u_{n-m}) = \gcd(u_m, u_n).$$

Also, since

$$u_n = \frac{\omega^{-n}}{\omega + \omega^{-1}}(\omega^2 n \pm 1),$$

according as n is odd or even, we may replace u_n by

$$w_n = \omega^{2n} \pm 1;$$

For ω^{-n} is a unit, and so will not affect the gcd, while the factor

$$\frac{1}{\omega + \omega^{-1}}$$

occurs throughout, and so can be eliminated.

Accordingly, we have to show that

$$\gcd(w_m, w_n) = \gcd(w_m, w_{n-m}).$$

There are 4 cases to consider, according as m, n are even or odd.

Suppose m, n are both odd. Then $m - n$ is even, and

$$\begin{aligned}
\gcd(w_m, w_n) &= \gcd(\omega^{2m} + 1, \omega^{2n} + 1) \\
&= \gcd(\omega^{2m} + 1, \omega^{2n} - \omega^{2m}) \\
&= \gcd(\omega^{2m} + 1, \omega^{2m}(\omega^{2(n-m)} - 1)) \\
&= \gcd(\omega^{2m} + 1, \omega^{2(n-m)} - 1) \\
&= \gcd(w_m, w_{n-m}),
\end{aligned}$$

where we have used the fact that $\gcd(a,b) = \gcd(a, b-a)$.

The other 3 cases all follow in the same way. Thus if m is odd and n is even then $m-n$ is odd, and

$$\begin{aligned}\gcd(w_m, w_n) &= \gcd(\omega^{2m}+1, \omega^{2n}-1)\\ &= \gcd(\omega^{2m}+1, \omega^{2n}+\omega^{2m})\\ &= \gcd(\omega^{2m}+1, \omega^{2m}(\omega^{2(n-m)}+1))\\ &= \gcd(\omega^{2m}+1, \omega^{2(n-m)}+1)\\ &= \gcd(w_m, w_{n-m}),\end{aligned}$$

where now we have used the fact that $\gcd(a,b) = \gcd(a, b+a)$. ◀

Corollary 5.1. *If $m \mid n$ then*

$$u_m \mid u_n.$$

Corollary 5.2. *If u_n is prime then either $n = 4$ or n is prime.*

Of course we are not saying that u_p is prime if p is prime; we leave it as an exercise to find the first counter-example ....

Proposition 5.8. *Suppose $p \in \mathbb{N}$ is a prime number, $p \neq 5$. Then*

$$\begin{cases} p \mid u_p & \text{if } p = 1\\ p \mid u_{p-1} & \text{if } p \equiv \pm 1 \bmod 5\\ p \mid u_{p+1} & \text{if } p \equiv \pm 2 \bmod 5.\end{cases}$$

Proof ▶ Note first that by the Quadratic Reciprocity Law,

$$\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \bmod 5\\ -1 & \text{if } p \equiv \pm 2 \bmod 5,\end{cases}$$

since 1 and $4 \equiv -1$ are quadratic residues mod5, while 2 and $3 \equiv -2$ are not.

Lemma 15. *Suppose $z \in \mathbb{Z}[\omega]$. Then*

$$z^p \equiv \begin{cases} z \bmod p & \text{if } p \equiv \pm 1 \bmod 5,\\ \bar{z} \bmod p & \text{if } p \equiv \pm 2 \bmod 5\end{cases}$$

Proof ▶ Suppose first that $p \neq 2$. Let

$$z = c + d\sqrt{5}$$

where c, d are integers or half-integers. By the binomial theorem,

$$z^p = c^p + \binom{1}{p} c^{p-1} d\sqrt{5} + \cdots + d^p(\sqrt{5})^p.$$

The binomial coefficients in the middle are all divisible by p. Also, by Fermat's Little Theorem (or Lagrange's Theorem),

$$c^p \equiv c \bmod p,\ d^p \equiv d \bmod p.$$

(Note the since we are excluding $p = 2$ this will still be true even if c, d are half-integers.) Hence

$$z^p \equiv c + d5^{\frac{p-1}{2}}\sqrt{5} \bmod p.$$

By Gauss' Lemma,

$$5^{\frac{p-1}{2}} \equiv \left(\frac{5}{p}\right) \bmod p.$$

Thus if $p \equiv \pm 1 \bmod 5$ then $\left(\frac{5}{p}\right) = 1$ and so

$$z^p \equiv c + d\sqrt{5} = z \bmod p;$$

while if $p \equiv \pm 2 \bmod 5$ then $\left(\frac{5}{p}\right) = -1$ and so

$$z^p \equiv c - d\sqrt{5} = \tilde{z} \bmod p.$$

The result still holds if $p = 2$; for if $z = a + b\omega$ then

$$\begin{aligned} z^2 &\equiv a^2 + b^2\omega^2 \bmod 2 \\ &\equiv a + b(\omega + 1) \bmod 2 \\ &\equiv a + b(-\omega + 1) \bmod 2 \\ &equiva + b\tilde{\omega} \bmod 2 \\ &\tilde{z} \bmod 2. \end{aligned}$$

◀

Now suppose $p \equiv \pm 1 \bmod 5$. Then by the Lemma

$$\omega^p \equiv \omega \bmod p$$

Since ω is a unit, it follows that

$$\omega^{p-1} \equiv 1 \bmod p.$$

Similarly

$$\tilde{\omega}^{p-1} \equiv 1 \bmod p,$$

and so

$$u_{p-1} = \frac{\omega^{p-1} - \tilde{\omega}^{p-1}}{\omega - om\tilde{e}ga} \qquad \equiv \frac{1-1}{\omega - \tilde{\omega}} = 0.$$

On the other hand, if $p \equiv \pm 2 \bmod 5$ then by the Lemma

$$\omega^p \equiv \tilde{\omega} \bmod p.$$

Hence

$$\omega^{p+1} \equiv \omega\tilde{\omega} = -1 \bmod p.$$

Similarly,

$$\tilde{\omega}^{p+1} \equiv= -1 \bmod p;$$

and so

$$u_{p+1} = \frac{\omega^{p+1} - \tilde{\omega}^{p+1}}{\omega - om\tilde{e}ga} \qquad \equiv \frac{-1+1}{\omega - \tilde{\omega}} = 0.$$

◀

It's worth looking a little more closely at the congruences

$$z^p \equiv \begin{cases} z \bmod p & \text{if } p \equiv \pm 1 \bmod 5 \\ \tilde{z} \bmod p & \text{if } p \equiv \pm 2 \bmod 5. \end{cases}$$

In the first case, p splits in $\mathbb{Z}[\omega]$, say

$$p = \pi_1\pi_2.$$

Thus if $z, w \in \mathbb{Z}[\omega]$ the congruence

$$z \equiv w \bmod p$$

splits into the two congruences

$$z \equiv w \bmod \pi_1,\ z \equiv w \bmod \pi_2.$$

The Chinese Remainder Theorem holds equally well for the primes π_1, π_2; that is, given any remainders $a \bmod \pi_1$ and $b \bmod \pi_2$, there exists a unique remainder $c \bmod p = \pi_1\pi_2$ with

$$c \equiv a \bmod \pi_1, c \equiv b \bmod \pi_2.$$

This follows, for example, on applying the Euclidean algorithm (which as we have seen can still be used in $\mathbb{Z}[\omega]$) to π_1, π_2.

Algebraically, there is a ring-isomorphism

$$\Theta : \mathbb{Z}[\omega]/(p) \to \mathbb{Z}[\omega]/(\pi_1) \times \mathbb{Z}[\omega]/(\pi_1)$$

sending

$$z \bmod p \mapsto (z \bmod \pi_1,\ z \bmod \pi_2).$$

Now there are p^2 remainders mod$p$ in $\mathbb{Z}[\omega]$, namely

$$i + j\omega \quad (0 \le i < p, 0 \le j < p).$$

Since $\pi_2 \sim \tilde{\pi_1}$, it is clear that the number of remainders modπ_1 and modπ_2 are equal. It follows that the number of remainders modπ_1 is p:

$$\|\mathbb{Z}[\omega]/(\pi_1)\| = \|\mathbb{Z}[\omega]/(\pi_1)\| = p.$$

But since π_1, π_2 are primes, each of the rings $\mathbb{Z}[\omega]/(\pi_1)$, $\mathbb{Z}[\omega]/(\pi_2)$ are in fact fields; and in particular, the non-zero elements form a multiplicative group of order $p-1$. It follows therefore by Lagrange's Theorem that if $\pi_1 \nmid z$ then

$$z^{p-1} \equiv 1 \bmod \pi_1.$$

Thus Fermat's Little Theorem still holds:

$$z^p \equiv z \bmod \pi_1$$

for all $z \in \mathbb{Z}[\omega]$. Similarly

$$z^p \equiv z \bmod \pi_2$$

for all $z \in \mathbb{Z}[\omega]$. It follows that

$$z^p \equiv z \bmod p$$

for all $z \in \mathbb{Z}[\omega]$, as indeed we saw.

Now suppose $p \equiv \pm 2 \bmod 5$. In that case p does *not* split in $\mathbb{Z}[\omega]$. It follows that the residue-ring $\mathbb{Z}[\omega]/(p)$ — which still contains p^2 elements — is a field. In particular, the non-zero elements form a group of order $p^2 - 1$; and so, by Lagrange's Theorem,

$$z^{p^2-1} \equiv 1 \bmod p$$

if $p \nmid z$; or

$$z^{p^2} \equiv z \bmod p$$

for all $z \in \mathbb{Z}[\omega]$.

The map

$$\Phi : z \mapsto z^p \bmod p$$

is an automorphism of the field $\mathbb{Z}[\omega]/(p)$ since

$$(z+w)^p \equiv z^p + w^p \bmod p, \ (zw)^p \equiv z^p w^p \bmod p.$$

Moreover, this automorphism is of order 2, since

$$\Phi^2(z) = (z^p)^p = z^{p^2} = z.$$

But the automorphism $z \mapsto \tilde{z}$ of $\mathbb{Z}[\omega]$ — also of order 2 — induces an automorphism of $\mathbb{Z}[\omega]/(p)$.

But it is easy to see that $\mathbb{Z}[\omega]/(p)$ has only 1 non-trivial automorphism Ψ. For such an automorphism necessarily sends ω into a root of

$$x^2 - x - 1 = 0,$$

ie $\Psi(\omega)$ is either ω, in which case Ψ is trivial; or else $\Psi(\omega) = \tilde{\omega}$.

It follows that

$$z^p \equiv \tilde{z} \bmod p,$$

as indeed we found, by a different route.

5.5 The primality of Fermat numbers

In the next section we shall give a necessary and sufficient condition for the Mersenne number M_p to be prime, the proof of whose validity makes use of factorisation in $\mathbb{Z}[\omega]$.

As an introduction, we give a simpler condition — but along much the same lines — for the primality of the Fermat number

$$F_n = 2^{2^n} + 1.$$

Proposition 5.9. *The Fermat number F_n is prime if and only if*

$$5^{2^{2^n-1}} \equiv -1 \bmod F_n.$$

Proof ▶ Suppose $n \geq 2$. Then

$$2^n \equiv 0 \bmod 4.$$

Since $2^4 \equiv 1 \bmod 5$, it follows that

$$2^{2^n} \equiv 1 \bmod 5.$$

Thus

$$F_n \equiv 2 \bmod 5.$$

Suppose F_n is prime, say $P = F_n$. By the Quadratic Reciprocity Law,

$$\left(\frac{5}{P}\right) = \left(\frac{P}{5}\right) = \left(\frac{2}{5}\right) = -1.$$

Hence, by Gauss' Lemma,

$$5^{\frac{P-1}{2}} \equiv -1 \bmod P,$$

ie

$$5^{2^{2^n-1}} \equiv -1 \bmod P.$$

Conversely, suppose this is so; and suppose p is a prime factor of F_n.

Lemma 16. *Suppose $a \in (\mathbb{Z}/p)^\times$; and suppose*

$$a^{2^r} \equiv -1 \bmod p.$$

Then the order of a in this group is 2^{r+1}.

Proof ▶ Since

$$a^{2^{r+1}} \equiv (-1)^2 = 1 \bmod p,$$

the order of a certainly divides 2^{r+1}, ie it is 2^i for some $i \leq r+1$. But if $i \leq r$ then

$$a^{2^i} \equiv 1 \implies a^{2^r} = (a^{2^i})^{2^{r-i}} \equiv 1,$$

whereas by hypothesis

$$a^{2^r} \equiv -1 \bmod p.$$

◀

Now

$$5^{2^{2^n-1}} \equiv -1 \bmod F_n \implies 5^{2^{2^n-1}} \equiv -1 \bmod p.$$

It follows by the Lemma that the order of 5 in the group $(\mathbb{Z}/p)^\times$ is exactly 2^{2^n}. But by Lagrange's Theorem, the order of an element divides the order of the group, which in this case is $p-1$. Hence

$$2^{2^n} \mid p-1.$$

In particular,

$$2^{2^n} \leq p-1,$$

ie

$$p \geq 2^{2^n} + 1 = F_n.$$

Hence F_n is prime. ◀

5.6 The primality of Mersenne numbers

Recall that

$$M_p = 2^p - 1,$$

where p is prime.

Proposition 5.10. *Suppose $p \equiv 3 \bmod 4$. Then M_p is prime if and only if*

$$\omega^{2^p} \equiv -1 \bmod M_p.$$

Proof ▶ Suppose M_p is prime. Since $2^4 \equiv 1 \bmod 5$,

$$\begin{aligned} p \equiv 3 \bmod 4 &\implies 2^p \equiv 2^3 \bmod 5 \\ &\implies M_p \equiv 2^3 - 1 \bmod 5 \\ &\implies M_p \equiv 2 \bmod 5. \end{aligned}$$

Accordingly, M_p does not split in $\mathbb{Z}[\omega]$. It follows therefore that — writing P for M_p —

$$z^P \equiv \tilde{z} \bmod P$$

for all $z \in \mathbb{Z}[\omega]$. In particular,

$$\omega^P \equiv \tilde{\omega} \bmod P.$$

Thus

$$\omega^{P+1} \equiv \tilde{\omega} = -1 \bmod P,$$

ie

$$\omega^{2^p} \equiv -1 \bmod M_p,$$

as required.

Conversely, suppose this is so; and suppose M_p is composite. Then we can find a prime factor q of M_p, with

$$q \leq \sqrt{M_p}.$$

Now

$$\omega^{2^p} \equiv -1 \bmod M_p \implies \omega^{2^p} \equiv -1 \bmod q.$$

It follows that the order of ω in the group $\mathbb{Z}[\omega]/(q)$ is exactly 2^{p+1}. Since this group is of order $q^2 - 1$, it follows by Lagrange's Theorem that

$$2^{p+1} \mid q^2 - 1.$$

Hence

$$2^{p+1} \leq q^2 - 1,$$

ie

$$q^2 \geq 2^{p+1} + 1 > M_p,$$

ie

$$q > \sqrt{M_p},$$

contrary to assumption. ◀

We can express this result in a form not involving algebraic numbers, by introducing the sequence of *Lucas numbers*.

Definition 5.2. *The Lucas sequence* $v_n (n \in \mathbb{N})$ *is defined by the same linear recurrence relation*

$$v_{n+2} = v_{n+1} + v_n \quad (n \in \mathbb{N})$$

as the Fibonnaci numbers, but with initial values

$$v_0 = 2,\ v_1 = 1.$$

Proposition 5.11. *The Lucas numbers are given by*

$$v_n = \omega^n + \tilde{\omega}^n = \begin{cases} \omega^n + \omega^{-n} & \text{if } n \text{ is even} \\ \omega^n - \omega^{-n} & \text{if } n \text{ is odd.} \end{cases}$$

Proof ▶ We know that

$$v_n = C\omega^n + D\tilde{\omega}^n.$$

Taking $C = D = 1$ gives

$$v_0 = 1 + 1 = 2,\ v_1 = \omega + \tilde{\omega} = 1,$$

as required. ◀

The Lucas sequence runs

$$2, 1, 3, 4, 7, 11, 18, 29, 47, 76, 123, 199, \ldots.$$

Definition 5.3. *For* $n \geq 1$ *we set*

$$r_n = v_{2^n}.$$

Proposition 5.12. *The sequence* r_n *satisfies the non-linear recurrence relation*

$$r_{n+1} = r_n^2 - 2 \quad (n \geq 1),$$

with initial value $r_1 = 3$.

Proof ▶ If m is even then

$$v_m = \omega^m + \omega^{-m}.$$

Hence

$$\begin{aligned} v_m^2 &= (\omega^m + \omega^{-m})^2 \\ &= \omega^{2m} + 2 + \omega^{-2m} \\ &= v_{2m} + 2. \end{aligned}$$

◀

Proposition 5.13. *Suppose $p \equiv 3 \bmod 4$. Then M_p is prime if and only if*

$$r_{p-1} \equiv 0 \bmod M_p,$$

where r_n is the sequence defined above.

Proof ► We know that M_p is prime if and only if

$$\omega^{2^p} \equiv -1 \bmod M_p,$$

ie

$$\omega^{2^{p-1}} \equiv -\omega^{2^{p-1}} \bmod M_p$$

ie

$$r_{p-1} = \omega^{2^{p-1}} + omega^{2^{p-1}} \equiv 0 \bmod M_p.$$

◄

Example. Consider the Mersenne number

$$M_7 = 2^7 - 1 = 127.$$

We have to determine $r_6 \bmod 127$.

Note that we work mod$p$ throughout. If p is large it would be foolish to try to compute r_{p-1} completely, Since r_n is roughly speaking squared each time n increases by 1, which M_n is roughly speaking multiplied by 2, it is clear that r_{p-1} will be vastly larger than M_p for large p.

We have

$$\begin{aligned}
r_1 &= 3,\\
r_2 &= 3^2 - 2 = 7,\\
r_3 &= 7^2 - 2 = 47,\\
r_4 &= 47^2 - 2\\
&= 47 \cdot 48 - 49\\
&= (47 \cdot 3) \cdot 16 - 49\\
&\equiv 14 \cdot 16 - 49 \bmod 127\\
&\equiv 173 \bmod 127\\
&\equiv 48 \bmod 127, r_5 && \equiv 48^2 - 2 \bmod 127\\
&\equiv (47^2 - 2) + 95 \bmod 127\\
&\equiv 48 + 95 \bmod 127\\
&\equiv 16 \bmod 127, r_6 && \equiv 16^2 - 2 \bmod 127\\
&\equiv 2 \cdot 128 - 2 \bmod 127\\
&\equiv 0 \bmod 127.
\end{aligned}$$

We conclude that $M_7 = 127$ is prime.

This test only determines the primality of M_p when $p \equiv 3 \bmod 4$. We shall give a similar, but slightly more complicated, proof in the next Chapter which works for all p.

Chapter 6

Arithmetic in $\mathbb{Q}[\sqrt{3}]$

We have already established the basic facts about this field. In particular, since $3 \not\equiv 1 \bmod 4$ the integers are the numbers

$$a + b\sqrt{3} \quad (a, b \in \mathbb{Z}).$$

ie the integer-ring is $\mathbb{Z}[\sqrt{3}]$.

The fundamental unit, by inspection, is

$$\omega = 2 + \sqrt{3};$$

and so the general unit is

$$\pm\omega^n \quad (n \in \mathbb{Z}).$$

Note that

$$N(\omega) = \omega\tilde{\omega} = 2^2 - 3 \cdot 1^2 = 1.$$

It follows that

$$N(\epsilon) = 1$$

for all units, unlike in $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{5})$ where the fundamental unit had norm -1.

6.1 The Fundamental Theorem

We extend the Euclidean Algorithm to $\mathbb{Z}[\omega]$ in the usual way; given $z, w \in \mathbb{Z}[\sqrt{3}]$ we determine

$$z/w = x + y\sqrt{3},$$

take a, b to be the closest integers to x, y, say

$$x = a + s,\ y = b + t,\ 0 \leq |s|, |t| < \frac{1}{2};$$

and set

$$q = a + b\omega,\ r = z - qw.$$

Then

$$\begin{aligned} N(r/w) &= N(z/w - q) \\ &= N(s + t\sqrt{3}) \\ &= s^2 - 3t^2. \end{aligned}$$

It follows that

$$-\frac{3}{4} < N(r/w) < \frac{1}{4};$$

and so

$$|N(r)| < \frac{3}{4}|N(w)|.$$

Thus we have established

Proposition 6.1. *Suppose $z, w \in \mathbb{Z}[\sqrt{3}]$ with $w \neq 0$. Then there exist $q, r \in \mathbb{Z}[\sqrt{3}]$ such that*

$$z = qw + r,\ |N(r)| < |N(w)|.$$

From this we derive the Fundamental Theorem for $\mathbb{Q}(\sqrt{3})$.

Theorem 6.1. *Each number $z \in \mathbb{Z}[\sqrt{3}]$ can be factorised into prime factors; and the factorisation is unique up to the order of the factors, and multiplication by units.*

6.2 The primes in $\mathbb{Z}[\sqrt{3}]$

Proposition 6.2. *A prime number $p \in \mathbb{N}$ splits into at most 2 factors in $\mathbb{Z}[\omega]$.*

As usual, we omit proofs given earlier.

Proposition 6.3. *Suppose $p \in \mathbb{N}$ is a prime number, $p \neq 2, 3$. Then $p \in \mathbb{N}$ splits in $\mathbb{Z}[\sqrt{3}]$ if and only*

$$\left(\frac{3}{p}\right) = 1.$$

Proof ▶ Suppose p splits, say

$$p = \pi_1 \pi_2.$$

Let

$$\pi = a + b\sqrt{3}.$$

Then

$$N(\pi_1) = a^2 - 3b^2 = \pm p.$$

Hence

$$(ab^{-1})^2 \equiv 3 \bmod p,$$

and so

$$\left(\frac{3}{p}\right) = 1.$$

Conversely, suppose this is so. Then we can find $a \in (0, p)$ such that

$$a^2 \equiv 3 \bmod p,$$

ie

$$a^2 - 3 = pm,$$

ie

$$(a - \sqrt{3})(a + \sqrt{3}) = pm.$$

If p does *not* split then by the Fundamental Theorem it must divide one of the factors on the left, say

$$a + \sqrt{5} = p(c + d\sqrt{5}),$$

and so $p \mid 1$, which is absurd. ◀

Proposition 6.4. *Suppose $p \in \mathbb{N}$ is a prime number. Then*

$$p \begin{cases} \textit{splits into two equivalent factors} & \textit{if } p = 2, 3 \\ \textit{splits into two non-equivalent factors} & \textit{if } p \equiv \pm 1 \bmod 12 \\ \textit{does not split} & \textit{if } p \equiv \pm 5 \bmod 12 \end{cases}$$

Proof ▶ Evidently,

$$3 = (\sqrt{3})^2,$$

ie the prime number 3 splits into 2 equal primes in $\mathbb{Z}[\sqrt{3}]$.

The prime number 2 is also ramified; for if

$$\eta = 1 + \sqrt{3},$$

then

$$\eta^2 = 2(2 + \sqrt{3}) = 2\omega.$$

Thus

$$2 \sim \eta^2.$$

Now suppose $p \neq 2, 3$. From the previous Proposition, we have to determine $\left(\frac{3}{p}\right)$. By Gauss' Law of Quadratic Reciprocity,

$$\left(\frac{3}{p}\right) = \begin{cases} \left(\frac{p}{3}\right) & \text{if } p \equiv 1 \bmod 4 \\ -\left(\frac{p}{3}\right) & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

On the other hand,

$$\left(\frac{p}{3}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \bmod 3 \\ -1 & \text{if } p \equiv 2 \bmod 3. \end{cases}$$

We conclude that

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \bmod 12 \\ -1 & \text{if } p \equiv \pm 5 \bmod 12, \end{cases}$$

from which the result follows. ◀

6.3 The primality of Mersenne numbers

Proposition 6.5. *The Mersenne number $M_p = 2^p - 1$, $p > 2$ is prime if and only if*

$$\omega^{2^{p-1}} \equiv -1 \bmod M_p,$$

where

$$\omega = 2 + \sqrt{3}.$$

Proof ▶ Suppose M_p is prime, say $P = M_p$.

Then

$$\begin{aligned} P &\equiv (-1)^p - 1 \bmod 3 \\ &\equiv 1 \bmod 3. \end{aligned}$$

On the other hand,

$$P \equiv 0 - 1 = -1 \bmod 4.$$

Combining these,

$$P \equiv 7 \bmod 12.$$

It follows that

$$\left(\frac{3}{P}\right) = -1;$$

and P does not split in $\mathbb{Z}[\sqrt{3}]$.

Suppose

$$z = a + b\sqrt{3} \in \mathbb{Z}[\sqrt{3}],$$

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

$$\begin{aligned} z^P &\equiv a^P + b^P 3^{\frac{P}{2}} \bmod P \\ &\equiv a + b\, 3^{\frac{P-1}{2}}\sqrt{3} \bmod P \\ &\equiv a + b\left(\frac{3}{P}\right)\sqrt{3} \bmod P \\ &\equiv a - b\sqrt{3} \bmod P \\ &\equiv \tilde{z} \bmod P. \end{aligned}$$

Hence

$$a^{P+1} \equiv a^{2^p} \equiv z\tilde{z} = N(z) \bmod P.$$

In particular,

$$\omega^{2^p} \equiv N(\omega) = 1 \bmod P.$$

It follows that

$$\omega^{2^{p-1}} \equiv \pm 1 \bmod P.$$

To apply the same argument that we used in the previous test (based on computations in $\mathbb{Q}(\sqrt{5})$) we need to show that

$$\omega^{2^{p-1}} \equiv -1 \bmod P.$$

This we can prove as follows. If

$$\eta = 1 + \sqrt{3}$$

then

$$\eta^2 = 2\omega,$$

as we have already seen. Raising this to the 2^{p-1}th power,

$$\begin{aligned} (2\omega)^{2^{p-1}} &= \eta^{2^p} \\ &= \eta^{P+1} \\ &\equiv N(\eta) \bmod P \\ &= -2. \end{aligned}$$

On the other hand,

$$\begin{aligned} 2^{2^{p-1}} &= 2 \cdot 2^{\frac{P-1}{2}} \\ &\equiv 2\left(\frac{2}{P}\right) \\ &= 2, \end{aligned}$$

since

$$P \equiv -1 \bmod 8 \implies \left(\frac{2}{P}\right) = 1.$$

It follows that

$$2\omega^{2^{p-1}} \equiv -2 \bmod P,$$

ie

$$\omega^{2^{p-1}} \equiv -1 \bmod M_p.$$

The converse argument is identical to the previous test. If M_p is composite then it must have a prime factor q with

$$q^2 \leq M_p < 2^p.$$

But

$$\omega^{2^{p-1}} \equiv -1 \bmod M_p \implies \omega^{2^{p-1}} \equiv -1 \bmod q.$$

It follows that the order of ω in $(\mathbb{Z}[\sqrt{5}]/q)^\times$ is exactly 2^p. But we know that this group has $< q^2$ elements. It follows by Lagrange's Theorem that

$$2^p < q^2,$$

contrary to our hypothesis about q. ◀

As before, we can re-state this result so that it does not explicitly use algebraic numbers.

Definition 6.1. *The Lucas-Lehmer sequence w_n is defined for $n \in \mathbb{N}$ by*

$$w_n = \omega^n + \omega^{-n},$$

where $\omega = 2 + \sqrt{3}$.

Proposition 6.6. *The sequence w_n is defined by the linear recurrence relation*

$$w_{n+1} = 4w_n - w_{n-1} \quad (n \in \mathbb{N})$$

with initial values $w_0 = 2,\ w_1 = 4$.

Proof ▶ This follows at once from the fact that $\omega, \omega^{-1} = \tilde{\omega}$ are roots of

$$(x-2)^2 = 3,$$

ie

$$x^2 - 4x + 1 = 0.$$

◀

The sequence starts

$$2, 4, 14, 52, 194, 724, \ldots.$$

Proposition 6.7. *Let $p > 2$ be a prime number. The Mersenne number $M_p = 2^p - 1$ is prime if and only if*

$$r_{p-1} \equiv 0 \bmod M_p,$$

where r_n $(n > 0)$ is the sequence defined by the non-linear recurrence relation

$$r_{n+1} = r_n^2 - 2,$$

with initial value $r_1 = 4$.

Proof ▶ The proof is as before, except that now we set

$$r_n = w_{2^{n-1}} = \omega^{2^{n-1}} + \omega^{-2^{n-1}}.$$

As before,

$$\begin{aligned} r_n^2 &= (\omega^{2^{n-1}} + \omega^{-2^{n-1}})^2 \\ &= \omega^{2^n} + 2 + \omega^{-2^n} \\ &= r_{n+1} + 2, \end{aligned}$$

From the previous Proposition,

$$\begin{aligned} M_p \text{ prime} &\iff \omega^{2^{p-1}} \equiv -1 \bmod M_p \\ &\iff \omega^{2^{p-2}} \equiv -\omega^{-2^{p-2}} \bmod M_p \\ &\iff \omega^{2^{p-2}} + \omega^{-2^{p-2}} \equiv 0 \bmod M_p \\ &\iff r_{p-1} \equiv 0 \bmod M_p. \end{aligned}$$

◀

Chapter 7

Primality Testing

7.1 Complexity

Factorising a number $n \in \mathbb{N}$ is thought to be "hard" or "intractable", while determining whether n is prime or not is known to be "tractable".

A problem is said to be *tractable* if it can be completed in *polynomial time* in terms of the *length* of the input.

The length ℓ of $n \in \mathbb{N}$ is defined to be the number of bits in n. Thus

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

Since we are only concerned with the order of the quantities involved, we may take $\ell = \log_2 n$.

Accordingly, a problem about $n \in \mathbb{N}$ — such as whether n is prime or not — is said to be tractable if we can find an algorithm which determines the primality of n in $\leq p(\ell)$ steps, where $p(x)$ is a polynomial.

What do we mean by a "step"? To define this precisely, we would have to introduce the notion of a *universal Turing machine*. Such a machine operates at discrete "moments" $t = 0, 1, 2, \ldots$, and we take t as a measure of the 'time'.

Since one universal Turing machine can 'emulate' another, the times computed by two universal machines will differ by less than a constant; so certainly if one universal machine completes the computation in polynomial time then so will the other. The notion of tractability is thus well-defined.

7.2 The Fermat test

Proposition 7.1. *The number n is prime if and only if*

$$a^{n-1} \equiv 1 \bmod n$$

for $a = 1, 2, \ldots, n-1$.

Proof ▶ Suppose n is prime. Then the numbers $a = 1, 2, \ldots, n-1$ form a group $(\mathbb{Z}/n)^\times$ under multiplication mod$n$. Since this group contains $n-1$ elements it follows from Lagrange's Theorem for finite groups that

$$a^{n-1} \equiv 1 \bmod n.$$

Conversely, suppose n is composite, say $p \mid n$. Then

$$p^{n-1} \equiv 0 \bmod p,$$

and so

$$p^{n-1} \not\equiv 1 \bmod p.$$

A fortiori,

$$p^{n-1} \not\equiv 1 \bmod n.$$

◀

Example. Suppose $n = 21$. Taking $a = 2$,

$$2^5 = 32 \equiv 11 \bmod 21.$$

Hence

$$\begin{aligned} 2^{20} &\equiv 11^2 = 121 \\ &\equiv 16 \bmod 21. \end{aligned}$$

Thus 21 has "failed the Fermat test", with *witness* $a = 2$.

If the only Fermat witnesses to the non-primality of n are the factors of n then the Fermat test is tantamount to finding a factor of n, which we believe to be intractable. So if such numbers exist, the Fermat test must be rejected as a practicable test for primality.

Definition 7.1. *A* Carmichael number *is a number $n \in \mathbb{N}$ which is composite (non-prime) and > 1 but which has the property that*

$$a^{n-1} \equiv 1 \bmod n.$$

for all a coprime to n, ie with $\gcd(a, n) = 1$.

Example. The smallest Carmichael number is $561 = 3 \cdot 11 \cdot 17$. To see that this *is* a Carmichael number, note that 560 is divisible by $3-1$, by $11-1$ and by $17-1$. Thus if $\gcd(a, 561) = 1$ then $\gcd(a, 3) = 1$, $\gcd(a, 11) = 1$ and $\gcd(a, 17) = 1$; and so, by Fermat's Little Theorem,

$$\begin{aligned} a^2 \equiv 1 \bmod 3 &\implies a^{560} \equiv 1 \bmod 3 \\ a^{10} \equiv 1 \bmod 11 &\implies a^{560} \equiv 1 \bmod 11 \\ a^{16} \equiv 1 \bmod 17 &\implies a^{560} \equiv 1 \bmod 17. \end{aligned}$$

It follows that

$$a^{560} \equiv 1 \bmod 561.$$

It has recently been shown that there are an infinity of Carmichael numbers. We must therefore reject the Fermat test for primality.

7.3 The Jacobi symbol

Although we have rejected the Fermat test, a simple modification of it, using Gauss' Lemma, is more plausible. Before describing this, we must extend the definition of the Legendre symbol $\left(\frac{a}{p}\right)$, which we recall is only defined if $p \in \mathbb{N}$ is a prime number (and $p \nmid a$).

Definition 7.2. *Suppose $n \in \mathbb{N}$ is odd. Let*

$$n = p_1 \cdots p_r$$

where the p_i are prime (but not necessarily distinct). Then we define the Jacobi symbol $\left(\frac{m}{n}\right)$ for $m \in \mathbb{Z}$ by

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

if $\gcd(m, n) = 1$, and 0 otherwise.

Note the when the Legendre symbol and the Jacobi symbol are both defined, ie when n is an odd prime and $n \nmid m$, both take the same value; so there is no ambiguity in using the same symbol.

Example.

$$\begin{aligned}\left(\frac{17}{21}\right) &= \left(\frac{17}{3}\right)\left(\frac{17}{7}\right) \\ &= \left(\frac{2}{3}\right)\left(\frac{3}{7}\right) \\ &= -1 \cdot -1 \\ &= 1.\end{aligned}$$

It is important to note that this does *not* imply that 17 is a quadratic residue mod21. In fact this is certainly not the case, since it would imply that 17 was a quadratic residue mod3 and mod7, both of which are false.

The Jacobi symbol (as opposed to the Legendre symbol) is best seen simply as a computational tool, without great significance on its own.

Proposition 7.2. *1.* $a \equiv b \bmod n \implies \left(\frac{a}{n}\right) = \left(\frac{b}{n}\right)$.

2. $\left(\frac{ab}{n}\right) = \left(\frac{a}{n}\right)\left(\frac{b}{n}\right)$.

3. If $m \in \mathbb{N}$ is odd then

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

Proof ▶ The first two parts are immediate consequences of the corresponding results for the Legendre symbol.

For the third part, suppose

$$m = q_1 \cdots q_s,$$

where the q_j are prime (but not necessarily distinct). Then

$$\begin{aligned}\left(\frac{m}{n}\right)\left(\frac{n}{m}\right) &= \prod_{i,j}\left(\frac{p_i}{q_j}\right)\left(\frac{q_j}{p_i}\right)\\ &= (-1)^{\sum_{i,j}\frac{p_i-1}{2}\frac{q_j-1}{2}}\\ &= (-1)^{\sum_i \frac{p_i-1}{2}\sum_j \frac{q_j-1}{2}}.\end{aligned}$$

Thus the result will follow if we can show that

$$n = \prod_i p_i \implies \sum_i \frac{p_i - 1}{2} \equiv \frac{n-1}{2} \bmod 2.$$

Since

$$\frac{n-1}{2} \equiv \begin{cases} 0 \bmod 2 \text{ if } n \equiv 1 \bmod 4 \\ 1 \bmod 2 \text{ if } n \equiv -1 \bmod 4 \end{cases},$$

this simply states that $n \equiv \pm 1 \bmod 4$ according as the number of factors $p_i \equiv -1 \bmod 4$ is even or odd, which is more or less self-evident.

Alternatively, the result follows on repeated application of the fact that if a, b are odd then

$$\frac{a-1}{2} + \frac{b-1}{2} \equiv \frac{ab-1}{2} \bmod 2$$

since

$$\frac{ab-1}{2} - \frac{a-1}{2} - \frac{b-1}{2} = \frac{(a-1)(b-1)}{2} \equiv 0 \bmod 2.$$

◀

7.4 Congruences to composite moduli

To establish our primality test, we need to consider what happens if the number is *not* prime; and for that we need to consider congruences to composite moduli $n = p_1^{e_1} \cdots p_r^{e_r}$.

The study of congruences to composite moduli divides into two parts: firstly, the reduction to congruences modulo prime powers p^e, using the Chinese Remainder Theorem; and secondly, the reduction to congruences modulo p, which can usually be accomplished by Hensel's Lemma.

7.4.1 The Chinese Remainder Theorem

Suppose $n \in \mathbb{N}$. Recall that $\mathbb{Z}/(n)$ denotes the ring formed by the remainders (or residues) mod$m$.

Now suppose $m, n \in \mathbb{N}$. Since

$$x \equiv y \bmod mn \implies x \equiv y \bmod n$$

there is a natural map

$$\mathbb{Z}/(mn) \to \mathbb{Z}/(n);$$

and it is easy to see that this map is a ring-homomorphism. Similarly there is a homomorphism $\mathbb{Z}/(mn) \to \mathbb{Z}/(m)$; and these two homomorphisms combine to define the homomorphism

$$\Theta : \mathbb{Z}/(mn) \to \mathbb{Z}/(m) \times \mathbb{Z}/(n)$$

under which

$$x \bmod mn \mapsto (x \bmod m, x \bmod n).$$

(Recall that the product of two rings A, B is the ring defined on the cartesian product $A \times B$ by setting

$$(a, b) + (a', b') = (a + a', b + b'),\ (a, b)(a', b') = (aa', bb').$$

This product-ring has zero $(0, 0)$ and identity $(1, 1)$, assuming that A, B are rings with 1.)

For example, if $m = 2$ and $n = 6$ then the homomorphism Θ sends

$$\begin{array}{lllllll} 0 \mapsto (1,1), & 1 \mapsto (1,1), & 2 \mapsto (0,2), & 3 \mapsto (1,3), & 4 \mapsto (0,4), & 5 \mapsto (1,5), \\ 6 \mapsto (0,0), & 7 \mapsto (1,1), & 8 \mapsto (0,2), & 9 \mapsto (1,3), & 10 \mapsto (0,4), & 11 \mapsto (1,5). \end{array}$$

Proposition 7.3. *(The Chinese Remainder Theorem) Suppose $m, n \in \mathbb{N}$; and suppose* $\gcd(m, n) = 1$. *Then the ring homomorphism*

$$\Theta : \mathbb{Z}/(mn) \to \mathbb{Z}/(m) \times \mathbb{Z}/(n)$$

is an isomorphism.

Proof ▶ Since the two rings $\mathbb{Z}/(mn)$ and $\mathbb{Z}/(m) \times \mathbb{Z}/(n)$ both contain mn elements, to prove that Θ is bijective — and therefore an isomorphism — it is sufficient to show that Θ is injective.

This in turn will follow if we show that $\ker \Theta = \{0\}$. But

$$\begin{aligned} x \in \ker \Theta &\implies x \bmod m = 0 \text{ and } x \bmod n = 0 \\ &\implies x \bmod mn = 0, \end{aligned}$$

since $\gcd(m, n) = 1$. ◀

Corollary 7.1. *For all $a, b \in \mathbb{Z}$ the simultaneous congruences*

$$x \equiv a \bmod m,\ x \equiv b \bmod n$$

have a unique solution $x \bmod mn$.

Example. Suppose $m = 7,\ n = 25;\ a = 2,\ b = 4$, ie we are trying to solve the simultaneous congruences

$$x \equiv 2 \bmod 7,\ x \equiv 4 \bmod 25.$$

Perhaps the simplest way to solve this problem is to find u such that

$$u \equiv 1 \bmod 7,\ u \equiv 0 \bmod 25,$$

and v such that

$$v \equiv 0 \bmod 7,\ v \equiv 1 \bmod 25.$$

Then

$$x = 2u + 4v$$

will be a solution to the original problem.

We can find u, v in this case by inspection. Thus $u = 50$ solves the first "auxiliary equation", while $v = -49$ satisfies the second. Hence

$$x_0 = 2 \cdot 50 - 4 \cdot 49 = -96.$$

is a solution to the original problem.

If x is any other solution then

$$x - x_0 \equiv 0 \bmod m,\ x - x_0 \equiv 0 \bmod n.$$

Since $\gcd(m, n) = 1$, it follows that

$$mn \mid (x - x_0).$$

Thus the general solution to the simultaneous congruences is

$$x = x_0 + tmn = -96 + 175t$$

for $t \in \mathbb{Z}$. In particular, there is a unique solution with $0 \leq x < 175$, namely $x = 79$.

Our proof of the Chinese Remainder Theorem does not provide a practical way of finding a solution (to simultaneous congruences to coprime moduli). Fortunately, the Euclidean Algorithm fills this gap.

Recall that if $\gcd(m, n) = 1$ then we can always find $y, z \in \mathbb{Z}$ such that

$$my + nz = 1;$$

in fact y, z appear as 'bye-products' when we apply the Euclidean Algorithm to m, n. But now we note that

$$nz \equiv 1 \bmod m, \ nz \equiv 0 \bmod n.$$

Thus $u = nz$ is a solution to our first auxiliary equation, while $v = my$ is a solution to our second. So

$$x = anz + bmy$$

is a particular solution to the original congruences.

In this case, the Euclidean Algorithm yields

$$\begin{aligned} 25 &= 3 \cdot 7 + 4 \\ 7 &= 4 \cdot 1 + 3 \\ 4 &= 1 \cdot 3 + 1. \end{aligned}$$

Working backwards,

$$\begin{aligned} 1 &= 4 - 3 \\ &= 4 - (7 - 4) \\ &= 2 \cdot 4 - 7 \\ &= 2(25 - 3 \cdot 7) - 7 \\ &= 2 \cdot 25 - 7 \cdot 7, \end{aligned}$$

giving $u = 50, \ v = -49$, as before.

Corollary 7.2. *Suppose $n_1, \ldots, n_r \in \mathbb{N}$ are pairwise-coprime; and suppose $a_1, \ldots, a_r \in \mathbb{Z}$. Then there is a unique $x \bmod n_1 \cdots n_r$ such that*

$$x \equiv a_i \bmod n_i \qquad (1 \leq i \leq r).$$

This can either be derived by repeated application of the previous Corollary, or by proving directly that the homomorphism

$$\mathbb{Z}/(n_1 \cdots n_r) \to \mathbb{Z}/(n_1) \times \cdots \times \mathbb{Z}/(n_r)$$

is an isomorphism.

Now suppose $f(x) \in \mathbb{Z}[x]$ is a polynomial with integral coefficients; and suppose $n = n_1 \cdots n_r$, where the n_i are pairwise coprime. Then

$$f(x) \equiv 0 \bmod n \iff f(x) \equiv 0 \bmod n_i \quad (1 \leq i \leq r).$$

Thus if we can solve each of the congruences

$$f(x_i) \equiv 0 \bmod n_i$$

separately, then by the Chinese Remainder Theorem we can find $x \in \mathbb{Z}$ such that $x \equiv x_i \bmod n_i$ for $1 \leq i \leq r$; and this x will satisfy $f(x) \equiv 0 \bmod n$. Conversely every solution of this congruence will arise in this way from solutions of the congruences modulo $n_1, \ldots, n_r$.

7.4.2 The multiplicative group $\mathrm{mod} n$

In practice we shall be more concerned with the group $(\mathbb{Z}/n)^\times$ than with the ring $\mathbb{Z}/(n)$. Recall that if A is a ring with 1 then the invertible elements in A form a multiplicative group $A^\times$. In particular, the invertible elements in $\mathbb{Z}/(n)$ form the group $(\mathbb{Z}/n)^\times$.

Recall too that $a \in \mathbb{Z}$ is invertible mod$n$ if and only if $\gcd(a, n) = 1$. If a is invertible, say $ab \equiv 1 \bmod n$ then evidently $\gcd(a, n) = 1$. Conversely, if $\gcd(a, n) = 1$ then the Euclidean Algorithm gives $x, y \in \mathbb{Z}$ such that

$$ax + ny = 1;$$

and x is the inverse of a modulo n.

Proposition 7.4. *Suppose $m, n \in \mathbb{N}$; and suppose $\gcd(m, n) = 1$. Then there is a group isomorphism*

$$(\mathbb{Z}/mn)^\times = (\mathbb{Z}/m)^\times \times (\mathbb{Z}/n)^\times.$$

Proof ▶ This follows at once from the last Proposition, since

$$\gcd(x, mn) = 1 \iff \gcd(x, m) = 1 \text{ and } \gcd(x, n) = 1.$$

◀

Example. Taking $m = 4,\ n = 3$,

$$\begin{aligned}(\mathbb{Z}/12)^\times &= (\mathbb{Z}/4)^\times \times (\mathbb{Z}/3)^\times \\ &= C_2 \times C_2.\end{aligned}$$

Corollary 7.3. *Suppose $n_1, \ldots, n_r$ are pairwise coprime. Then there is a group isomorphism*

$$(\mathbb{Z}/n_1 \cdots n_r)^\times \to (\mathbb{Z}/n_1)^\times \times \cdots \times (\mathbb{Z}/n_r)^\times.$$

7.5 Hensel's Lemma

Proposition 7.5. *Suppose $p \in \mathbb{N}$ is a prime number; suppose $f(x) \in \mathbb{Z}[x]$ is a polynomial with integral coefficients; suppose*

$$f(a) \equiv 0 \bmod p^e;$$

and suppose

$$f'(a) \not\equiv 0 \bmod p.$$

If $p > 2$ and $e \geq 1$, or if $p = 2$ and $e \geq 2$, then there exists a unique $x \bmod p^{e+1}$ such that

1. $f(x) \equiv 0 \bmod p^{e+1}$,

2. $x \equiv a \bmod p^e$.

In other words, a solution modulo p^e can be extended, uniquely, to a solution modulo p^{e+1}, provided only that the derivative $f'(a)$ does not vanish modulo p. (We might say that $f(x)$ was *non-singular* at $a \bmod p$.)

Proof ▶ If $x \equiv a \bmod p^e$ then

$$x = a + p^e y.$$

We expand $f(x)$ using Taylor's Theorem:

$$f(x) = f(a) + p^e y f'(a) + \frac{p^{2e}y^2}{2!} f''(a) + \cdots .$$

We shall show that each term on the right after the second is $\equiv 0 \bmod p^{e+1}$. Thus we have to find y such that

$$f(a) + p^e y f'(a) \equiv 0 \bmod p^{e+1}.$$

By hypothesis, $f(a) \equiv 0 \bmod p^e$, say

$$f(a) = p^e b.$$

Therefore we have to solve the congruence

$$p^e b \equiv p^e y f'(a) \bmod p^{e+1},$$

which is equivalent to

$$b \equiv y f'(a) \bmod p.$$

Since $f'(a) \not\equiv 0 \bmod p$ this has the unique solution (modulo p)

$$y \equiv f'(a)^{-1} b \bmod p.$$

Since the value of $x \bmod p^{e+1}$ is determined by the value of $y \bmod p$, the result follows.

It remains to prove that the terms on the right after the second vanish $\bmod p^{e+1}$. This will follow if we show that

$$p^{e+1} \mid \frac{p^{re}}{r!}$$

for $r = 2, 3, \ldots$.

Lemma 17. *Suppose* $p^f \parallel r!$. *Then*

$$f = \left[\frac{r}{p}\right] + \left[\frac{r}{p^2}\right] + \left[\frac{r}{p^3}\right] + \cdots .$$

Proof ▶ The number of numbers in $1, \ldots, r$ divisible by p is $[r/p]$. Similarly the number divisible by p^2 is $[r/p^2]$, the number divisible by p^3 is $[r/p^3]$, etc. Thus if $p^i \parallel j$, where $1 \leq j \leq i$, then i 1's will be contributed to the terms on the right, as required. ◀

Since

$$\begin{aligned}\left[\frac{r}{p}\right] + \left[\frac{r}{p^2}\right] + \cdots &< \frac{r}{p} + \frac{r}{p^2} + \cdots \\ &= \frac{r}{p-1},\end{aligned}$$

the result will hold if

$$e + 1 \leq r\left(e - \frac{1}{p-1}\right).$$

A fortiori, it will hold if

$$e + 1 \leq r(e-1),$$

ie

$$(r-1)(e-1) \geq 2.$$

Thus the result will certainly hold if $r \geq 3$ and $e \geq 2$.

On the other hand, the result always holds if $2 \leq r < p$ since then $p \nmid r!$ and the result reduces to

$$e + 1 \leq re.$$

Similarly, if $p \leq r < p^2$ then $p \parallel r!$ and the result reduces to

$$e + 1 \leq re - 1,$$

ie

$$e(r-1) \geq 2,$$

which is certainly true in this case.

Remembering that the case $p = 2,\ e = 1$ was excluded, it only remains to consider the case $p \geq 3,\ e = 1,\ r \geq p^2$. But in this case

$$\begin{aligned}r\left(e - \frac{1}{p-1}\right) &\geq 9\left(1 - \frac{1}{2}\right) \\ &> 2 = e + 1.\end{aligned}$$

◀

As an illustration of Hensel's Lemma, let us consider quadratic residues modulo p^2, where p is an odd prime number.

Suppose $a \in \mathbb{Z},\ p \nmid a$. If $a \in \mathbb{Z}$ is a quadratic non-residue mod$p$, then a *fortiori* a is a quadratic non-residue modp^2.

On the other hand, if a is a quadratic residue mod$p$ then we can apply Hensel's Lemma to the polynomial

$$f(x) = x^2 - a.$$

If x is a solution of this then $p \nmid x$, and so

$$p \nmid f'(x) = 2x.$$

It follows that x extends to a unique solution modp^2.

In particular, a is a quadratic residue mod p^2 if and only if it is a quadratic residue mod$p$.

It follows that just half, that is $p(p-1)/2$, of the elements of $(\mathbb{Z}/p)^\times$ are quadratic residues.

7.6 The Solovay-Strassen test for primality

Theorem 7.1. *Suppose $n \in \mathbb{N}$ is odd, $n \geq 3$. Then n is prime if and only if*

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

for all a coprime to n.

Proof ▶ If n is prime then it follows from Gauss' Lemma that

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

for all a coprime to n, since the Jacobi symbol reduces to the Legendre symbol in this case.

Suppose conversely that n is composite, say

$$n = p_1^{e_1} \cdots p_r^{e_r};$$

and suppose

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

for all a coprime to n. We must show that this leads to a contradiction.

Suppose first that some $e_i \geq 2$, ie $p^2 \mid n$ for some prime p. Our hypothesis implies that

$$x^{\frac{n-1}{2}} \equiv \pm 1 \bmod p^2$$

for all x coprime to n. It follows that this holds for all x coprime to p; for given x_0 coprime to p we can find x such that

$$x \equiv \begin{cases} x_0 \bmod p \\ 1 \bmod p_i \ (p_i \neq p), \end{cases}$$

by the Chinese Remainder Theorem; and x is then coprime to n.

But consider the polynomial

$$f(x) = x^{\frac{n-1}{2}} - 1.$$

Differentiating,

$$\begin{aligned} f'(x) &= \frac{n-1}{2} x^{\frac{n-3}{2}} \\ &\equiv -\frac{1}{2} x^{\frac{n-3}{n}} \bmod p, \end{aligned}$$

since $p \mid n$. Thus

$$p \nmid x \implies f'(x) \not\equiv 0 \bmod p.$$

Therefore, by Hensel's Lemma, the number of solutions of

$$x^{\frac{n-1}{2}} \equiv 1 \bmod p^2$$

is the same as the number of solutions mod$p$.

Similarly the number of solutions of

$$x^{\frac{n-1}{2}} \equiv -1 \bmod p^2$$

is the same as the number of solutions mod$p$.

It follows that the number of solutions of

$$x^{\frac{n-1}{2}} \equiv \pm 1 \bmod p^2$$

is at most $p-1$, leaving at least $p(p-1) - (p-1) = (p-1)^2$ residues modp^2 coprime to p for which

$$x^{\frac{n-1}{2}} \not\equiv \pm 1 \bmod p^2,$$

contrary to our hypothesis.

We conclude that n must be square-free, ie

$$n = p_1 \cdots p_r,$$

where the primes $p_1, \ldots, p_r$ are distinct.

Suppose p is one of these primes. By hypothesis

$$x^{\frac{n-1}{2}} \equiv \pm 1 \bmod p$$

for all x coprime to n. As we saw above, this implies that the result holds for all x coprime to p.

If x is a quadratic residue mod$p$, say $x \equiv y^2 \bmod p$, then

$$x^{\frac{n-1}{2}} \equiv \left(y^{\frac{n-1}{2}}\right)^2 \equiv (\pm 1)^2 = 1 \bmod p$$

Thus

$$\left(\frac{x}{p}\right) = 1 \implies x^{\frac{n-1}{2}} \equiv 1 \bmod p.$$

Accordingly, at least half (ie $\frac{p-1}{2}$) of the elements of $(\mathbb{Z}/p)^\times$ satisfy

$$x^{\frac{n-1}{2}} \equiv 1 \bmod p.$$

But

$$S = \{x \in (\mathbb{Z}/p)^\times : x^{\frac{n-1}{2}} \equiv 1 \bmod p\}$$

is a subgroup of $(\mathbb{Z}/p)^\times$. It follows by Lagrange's Theorem that either S is the whole group $(\mathbb{Z}/p)^\times$; or else S is the group of quadratic residues, in which case

$$x^{\frac{n-1}{2}} \equiv \left(\frac{x}{p}\right) \bmod p.$$

The first case is impossible; for we can certainly find x such that $\left(\frac{x}{n}\right) = -1$, eg by choosing a quadratic non-residue $x_1 \bmod p_1$, and taking

$$x \equiv \begin{cases} x_1 \bmod p_1 \\ 1 \bmod p_i \ (i > 1). \end{cases}$$

But then

$$\begin{aligned} \left(\frac{x}{n}\right) &= \left(\frac{x}{p_1}\right) \cdots \left(\frac{x}{p_r}\right) \\ &= -1 \cdot 1 \cdots 1 = -1; \end{aligned}$$

and therefore

$$x^{\frac{n-1}{2}} \equiv \left(\frac{x}{n}\right) = -1 \bmod n,$$

so that

$$x^{\frac{n-1}{2}} \equiv \left(\frac{x}{n}\right) = -1 \bmod p.$$

We have established therefore that

$$x^{\frac{n-1}{2}} \equiv \left(\frac{x}{p}\right) \bmod p.$$

But now let us choose quadratic non-residues $x_1 \bmod p_1$ and $x_2 \bmod p_2$. By the Chinese Remainder Theorem we can find x such that

$$x \equiv \begin{cases} x_1 \bmod p_1 \\ x_2 \bmod p_2 \\ 1 \bmod p_i \ (i > 2). \end{cases}$$

Then

$$\begin{aligned} \left(\frac{x}{n}\right) &= \left(\frac{x}{p_1}\right) \cdots \left(\frac{x}{p_r}\right) \\ &= -1 \cdot -1 \cdot 1 \cdots 1 \\ &= 1. \end{aligned}$$

By hypothesis, therefore,

$$x^{\frac{n-1}{2}} \equiv 1 \bmod n.$$

However, we have seen that

$$\begin{aligned} x^{\frac{n-1}{2}} &\equiv x_1^{\frac{n-1}{2}} \bmod p \\ &\equiv -1 \bmod p. \end{aligned}$$

Since these two congruences are contradictory, we conclude that our hypothesis is untenable, and n is prime. ◀

Example. Suppose $n = 21$. Testing with 2, we have

$$2^5 \equiv 10 \bmod 21.$$

Hence

$$2^10 \equiv 100 \equiv -5 \bmod 21.$$

Thus

$$2^{\frac{n-1}{2}} \not\equiv \pm 1,$$

and so 2 is a witness that 21 is composite.

It may not be apparent that this test has any advantage over the Fermat test. But consider the subset

$$S = \{x \in (\mathbb{Z}/n)^\times : x^{\frac{n-1}{2}} \equiv \left(\frac{x}{n}\right) \bmod n\}.$$

This set is a *subgroup* of $(\mathbb{Z}/n)^\times$. Since we have shown that $S \neq (\mathbb{Z}/n)^\times$, it follows from Lagrange's Theorem that

$$\|S\| \leq \frac{1}{2}\|(\mathbb{Z}/n)^\times\|.$$

Thus at least half the elements of $(\mathbb{Z}/n)^\times$ will witness that n is composite. (In practice it is likely to be much higher, since there is no reason to suppose that

$$x^{\frac{n-1}{2}} \equiv \pm 1 \bmod n,$$

let alone that the left-hand side takes the correct value from these two.)

In fact one can show the if the *Extended Riemann Hypothesis* holds (we shall see precisely what that means in the second part of the course) then given any proper $S \subset (\mathbb{Z}/n)^\times$ there exists an element $a \notin S$ with $2 \leq a \leq 2\log^2 n$.

It follows from this that *if* the Extended Riemann Hypothesis holds (as almost everyone believes) then the Solovay-Strassen test determines the primality or otherwise of n in polynomial time. Accordingly, with this proviso the problem of primality is tractable.

7.7 Elliptic curve tests

Although perfectly practicable, the Solovay-Strassen test is no longer the primality test of choice. (We gave it because it fitted in well with the rest of the course.)

The most popular methods today depend on the properties of elliptic curves over finite fields. The following remarks are merely intended to give a taste of the ideas involved, and do not contain any exact results.

Recall that an elliptic curve $\mathcal{E}$ over a field k of characterstic $\neq 2$ is defined by an equation

$$y^2 = x^3 + ax^2 + bx + c,$$

subject only to the condition that the cubic polynomial on the right must have distinct roots. (A slightly more general equation is required if $\operatorname{char} k = 2$.) If one adds a point "at infinity" then each line meets the curve in exactly 3 points (possibly coincident in the case of tangents). It is not difficult to see that one can define an additive group structure on $\mathcal{E}$ such that

$$P + Q + R = 0 \iff P, Q, R \text{ are collinear.}$$

All this holds equally well if k is a *finite* field of characteristic $\neq 2$.

The Solovay-Strassen test (and the Fermat test) were based on the group $(\mathbb{Z}/n)^\times$. In elliptic curve primality testing we use the group $\mathcal{E}$ instead. Apart from that, the basic ideas are very similar. Thus one can show that the group $\mathcal{E}$ over a finite field has at most 2 cyclic components:

$$\mathcal{E} = \mathcal{C}_\lceil \times \mathcal{C}_\rceil \quad (\rceil \mid \lceil).$$

It follows that $\mathcal{E}$ has an element of order $\geq \sqrt{m}$, where $m = \|\mathcal{E}\|$ is the number of points on the curve.

If we are given a number n we work on the assumption that n is prime. (If it is not, and the computation therefore breaks down at some point, then we know that n is composite.) If we can find a point of order $\geq \sqrt{m}$ then it follows (in much the same way as our tests of Mersenne numbers above) that n must be prime.

These elliptic curve methods have found many applications apart from primality testing. The newest forms of encryption use the same idea; and so too do various — unsuccessful to date — attempts to break the factorization problem (ie to show that factorization is tractable).

Chapter 2

P-adic numbers

2.1 Valuations

Definition 2.1. *A* valuation *on a field k is a map*

$$x \mapsto \|x\| : k \to \mathbb{R}$$

such that

1. $\|x\| \geq 0$ *and* $\|x\| = 0 \iff x = 0$;
2. $\|x + y\| \leq \|x\| + \|y\|$;
3. $\|xy\| = \|x\|\|y\|$;
4. $\|x\| \neq 1$ *for some* $x \neq 0$.

We sometimes use the term *valued field* for a field k together with a valuation $\|\cdot\|$ on k.

Proposition 2.1. 1. $\|1\| = 1$;

2. $\|-1\| = 1$;
3. $\|-x\| = \|x\|$.

Proof ▶. 1. This follows from $1^2 = 1$;

2. Similarly, this follows from $(-1)^2 = 1$;
3. $\|-x\| = \|-1\|\|x\|\|x\|$.

◀

Examples:

1. The absolute value $|x|$ defines valuations on $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$.

2. Suppose k is a field. Recall that $k(x)$ denotes the field of rational functions
$$f(x) = \frac{u(x)}{v(x)},$$
where $u(x), v(x) \in k[x]$ are polynomials.

   If $f(x)$ is not identically zero then we can write
$$f(x) = x^n \, \frac{r(x)}{s(x)},$$
where $r(0), s(0) \neq 0$ (ie $x \nmid r(x), s(x)$).

   It is readily verifed that
$$\|f(x)\| = 2^{-n}$$
defines a valuation on $k(x)$.

   Thus $\|f(x)\|$ is determined by the order of the pole (or zero) at $x = 0$.

   The choice of 2 was arbitrary. We could equally well have set $\|f(x)\| = e^{-n}$. We shall return to this point (or place) shortly.

   More generally, for any $a \in k$ we can define a norm $\|f(x)\|_a$ on $k(x)$ by setting
$$\|f(x)\|_a = 2^{-n}$$
if n is the order of the pole (or zero) at $x = a$.

3. We can define another norm on $k(x)$ by setting
$$\|u(x)/v(x)\|_\infty = \deg u(x) - \deg v(x).$$
   We can think of this as the 'norm at infinity' since
$$\|f(x)\|_\infty = \|f(1/x)\|_0.$$

Each non-zero rational
$$r = \frac{n}{d}$$
can be written as
$$r = p^e \, \frac{u}{v},$$
where $p \nmid u, v$. We may say that
$$p^e \mid\mid r.$$

Recall that if p is a prime and $n \in \mathbb{Z}$ then we write
$$p^e \mid\mid n \text{ if } p^e \mid n \text{ but } p^{e+1} \nmid n.$$
We can extend this to $\mathbb{Q}$ by setting
$$p^e \mid\mid r = \frac{n}{d} \text{ if } r = p^e \, \frac{u}{v} \quad (p \nmid u, v).$$

Definition 2.2. *Let p be a prime. Suppose $r \in \mathbb{Q},\ r \neq 0$. If*

$$p^e \mid\mid r$$

then we set

$$\|x\|_p = p^{-e}.$$

We call $\|\cdot\|_p$ the p-adic valuation *on $\mathbb{Q}$.*

Proposition 2.2. *The p-adic valuation is indeed a valuation of $\mathbb{Q}$.*

Proof ▶. If

$$p^e \mid\mid r, \quad p^f \mid\mid s$$

then

$$p^{e+f} \mid\mid rs$$

while

$$p^{\min(e.f)} \mid r + s.$$

◀

We sometimes denote the absolute valuation on $\mathbb{Q}$ by

$$\|x\|_\infty = |x|.$$

However, the p-adic valuations $\|x\|_p$ differ in one important way from the absolute valuation $\|x\|_\infty$; they satisfy a much stronger triangle inequality.

Proposition 2.3. *If $r, s \in \mathbb{Q}$ then*

$$\|r + s\| \leq \max(\|r\|, \|s\|)$$

Proof ▶. Suppose

$$\|r\|_p = e, \quad \|s\|_p = f,$$

ie

$$p^{-e} \mid\mid r, \quad p^{-f} \mid\mid s.$$

Then

$$p^{\min(-e,-f)} = p^{-\max(e,f)} \mid r + s,$$

and so

$$\|r + s\|_p \leq \max(e, f).$$

◀

Definition 2.3. *The valuation $\|x\|$ is said to be* non-archimedean *if*

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

for all x, y. If this is not so the valuation is said to be archimedean.

Evidently the p-adic valuation on $\mathbb{Q}$ is non-archimedean, while the absolute value is archimedean.

The term "ultrametric" is sometimes used for a non-archimedean valuation.

For any field k. there is a unique ring-homomorphism

$$\mathbb{Z} \to k$$

If $n \in \mathbb{Z}$ we write $n \in k$ for the image of n under this homomorphism.

Proposition 2.4. *The valuation $\|\cdot\|$ on k is archimedean if and only if*

$$\|n\| > 1$$

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

Proof ▶. We have to show that if

$$\|n\| \leq 1$$

for all $n \in \mathbb{Z}$ then the valuation is non-archimedean.

Suppose $x, y \in k$. Then

$$(x+y)^n = x^n + c_1 x^{n-1} y + \cdots + y^n,$$

where

$$c_i = \binom{n}{i} \in \mathbb{Z} \implies \|c_i\| \leq 1.$$

Thus

$$\begin{aligned} \|x+y\|^n &= \|(x+y)^n\| \\ &= \|x^n + c_1 x^{n-1} y + \cdots + y^n\| \\ &\leq \|x^n\| + \|x^{n-1} y\| + \cdots + \|y^n\| \\ &= \|x\|^n + \|x\|^{n-1}\|y\| + \cdots + \|y\|^n \\ &\leq (n+1)\max(\|x\|, \|y\|)^n. \end{aligned}$$

Hence

$$\|x+y\| \leq (n+1)^{1/n} \max(\|x\|, \|y\|).$$

Since $(n+1)^{1/n} \to 1$ as $n \to \infty$ (as one can see by taking logarithms),

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

◀

Corollary 2.1. *A valuation on a number field k restricts to a valuation on $\mathbb{Q}$; and the valuation is archimedean or non-archimedean according as the restriction is archimedean or non-archimedean.*

Proof ▶. All is immediate, except perhaps that a valuation on k might become trivial on $\mathbb{Q}$, ie $\|c\| = 1$ for all $c \in \mathbb{Q}$.

Suppose that is so. Then the valuation on k must be non-archimedean. Suppose $\|\alpha\| \neq 1$ for $\alpha \in k,\ \alpha \neq 0$. Taking α^{-1} in place of α, if necessary, we may assume that $\|\alpha\| > 1$.

Since α is an algebraic number it satisfies some equation

$$\alpha^n + c_1\alpha^{n-1} + \cdots + c_n = 0,$$

with $c_i \in \mathbb{Q}$. Since $\|c_i\| = 1$,

$$\begin{aligned}\|\alpha\|^n &= \|c_1\alpha^{n-1} + \cdots + c_n\| \\ &\leq \max(\|\alpha\|^{n-1}, \|\alpha\|^{n-1}, \ldots, 1) \\ &= \|\alpha\|^{n-1},\end{aligned}$$

whence

$$\|\alpha\| \leq 1,$$

contrary to assumption. ◀

2.2 Places

A valuation on k defines a metric

$$d(x, y) = \|x - y\|;$$

and this in turn defines a topology on k.

Definition 2.4. *Two valuations on k are said to be* equivalent *if they define the same topology.*

An equivalence class of valuations is called a place.

Proposition 2.5. *The valuations $\|\cdot\|_1,\ \|\cdot\|_2$ are equivalent if and only if*

$$\|x\|_2 = \|x\|_1^\rho$$

for some $\rho > 0$.

Proof ▶. It is evident the valuations will be equivalent if they satisfy such a relation.

Conversely, suppose the valuations are equivalent. With any valuation,

$$x^n \to 0 \iff \|x\| < 1.$$

Thus, since the topologies are the same,

$$\|x\|_1 < 1 \iff \|x\|_2 < 1.$$

Hence, taking x/y in place of x,

$$\|x\|_1 < \|y\|_1 \iff \|x\|_2 < \|y\|_2.$$

We have to show, in effect, that

$$\frac{\log\|x\|_1}{\log\|x\|_2}$$

is constant, ie

$$\frac{\log\|x\|_1}{\log\|y\|_1} = \frac{\log\|x\|_2}{\log\|y\|_2}$$

for all $x, y \neq 0$.

It is sufficient to prove this when $\|x\|_1, \|y\|_1 > 1$. Take a high power x^n, and suppose

$$\|y\|_1^m \leq \|x\|_1^n \leq \|y\|_1^{m+1}.$$

Then

$$\|y\|_2^m \leq \|x\|_2^n \leq \|y\|_2^{m+1}.$$

Taking logs,

$$\frac{m}{n} \leq \frac{\log\|x\|_1}{\log\|y\|_1}, \frac{\log\|x\|_2}{\log\|y\|_2} \leq \frac{m+1}{n}$$

Since this is true for arbitrarily large n,

$$\frac{\log\|x\|_1}{\log\|y\|_1} = \frac{\log\|x\|_2}{\log\|y\|_2},$$

as required. ◀

Note that we do not assert that if $\|x\|$ is a valuation on k then so is $\|x\|^\rho$. This is true if $0 < \rho < 1$, but is not true in general; for example, $|x|^2$ does not satisfy the triangle inequality in $\mathbb{R}$. All we are saying is that if we have two equivalent valuations then they must be related in this way.

2.3 Places in $\mathbb{Q}$

Theorem 2.1. *A valuation on $\mathbb{Q}$ is equivalent either to a p-adic valuations $\|\cdot\|_p$ or to the absolute valuation $|\cdot|$.*

Proof ▶. Suppose first that $\|\cdot\|$ is a non-archimedean valuation on $\mathbb{Q}$, so that

$$\|n\| \leq 1$$

for all $n \in \mathbb{Z}$.

We must have $\|n\| < 1$ for some $n \neq 0$; for otherwise we would have $\|x\| = 1$ for all non-zero $x = m/n$. Let

$$n = \pm p_1^{e_1} \cdots p_n^{e_n}.$$

Then $\|p_i\| < 1$ for some i.

Set $p = p_i$; and suppose q is another prime. Then we can find $u, v \in \mathbb{Z}$ such that

$$up + vq = 1.$$

It follows that $\|q\| = 1$, since otherwise $\|1\| < 1$.

But now we see that $\|n\|$ depends only on the power p^e of p dividing n:

$$\|n\| = \|p\|^e;$$

from which it follows that $\|\cdot\|$ is equivalent to the p-adic valuation $\|\cdot\|_p$.

Now suppose $\|\cdot\|$ is archimedean. We want to show that

$$\|x\| = |x|^\rho$$

for some ρ.

It is sufficient to prove this for all $a \in \mathbb{N}$. This is equivalent to showing that

$$\frac{\|a\|}{\|b\|} = \frac{|a|}{|b|}$$

for all integers $a, b > 1$.

Take a high power b^f of b; and suppose

$$a^e \leq b^f < a^{e+1}.$$

Then

$$e \log a \leq f \log b < (e+1) \log a$$

ie

$$\frac{e}{f} \leq \frac{\log b}{\log a} \leq \frac{e+1}{f}.$$

Now let us express b^f to base a, say

$$b^f = a^e + c_1 a^{e-1} + \cdots + c_r,$$

where

$$0 \leq c_i < a \quad (1 \leq i \leq r).$$

It follows that

$$\begin{aligned} \|b\|^f &\leq \|a\|^e + \|c_1\|\|a\|^{e-1} + \cdots + \|c_r\| \\ &\leq C\left(\|a\|^e + \|a\|^{e-1} + ...1\right), \end{aligned}$$

where

$$C = \max(\|1\|, \|2\|, \ldots, \|r-1\|).$$

If $\|a\| \leq 1$ this gives

$$\|b\|^f \leq C(e+1).$$

Thus

$$\|b\| \leq \left(C(e+1)^{1/f}\right)$$

As $f \to \infty$,

$$\leq \left(C(e+1)^{1/f}\right) \to 1,$$

since

$$e \leq \frac{\log b}{\log a} f.$$

It follows that

$$\|b\| \leq 1.$$

Since this is true for all b, the valuation is non-archimedean, contrary to hypothesis. We conclude that

$$\|a\| > 1$$

for all $a > 1$.

Now the inequality above yields

$$\|b\|^f \leq C(e+1)\|a\|^e$$

ie

$$f \log\|b\| \leq e \log\|a\| + \log C(e+1).$$

Thus

$$\begin{aligned} \frac{\log\|b\|}{\log\|a\|} &\leq \frac{e}{f} + \frac{\log C(e+1)}{f \log\|a\|} \\ &\leq \frac{\log b}{\log a} + \frac{\log C(e+1)}{f \log\|a\|}. \end{aligned}$$

As before, the last term $\to 0$ as $f \to \infty$. Hence

$$\frac{\log\|b\|}{\log\|a\|} \leq \frac{\log b}{\log a}.$$

Similarly,

$$\frac{\log\|a\|}{\log\|b\|} \leq \frac{\log a}{\log b}.$$

Thus

$$\frac{\log\|b\|}{\log\|a\|} = \frac{\log b}{\log a},$$

as required. ◀

We have shown, accordingly, that there is a place in $\mathbb{Q}$ corresponding to each prime p, together with a place corresponding to the absolute valuation, which we denote by ∞. In general, the places in a number field corresponding to archimedean valuations are said to be infinite.

2.4 P-adic numbers

The reals $\mathbb{R}$ can be constructed from the rationals $\mathbb{Q}$ by *completing* the latter with respect to the valuation $|x|$. In this construction each Cauchy sequence

$$\{x_i \in \mathbb{Q} : |x_i - x_j| \to 0 \text{ as } i, j \to \infty\}$$

defines a real number, with 2 sequences defining the same number if $|x_i - y_i| \to 0$.

(There are 2 very different ways of constructing $\mathbb{R}$ from $\mathbb{Q}$: by completing $\mathbb{Q}$, as above; or alternatively, by the use of *Dedekind sections*. In this each real number corresponds to a partition of $\mathbb{Q}$ into 2 subsets L, R where

$$l \in L,\ r \in R \implies l < r.$$

The construction by completion is much more general, since it applies to any metric space; while the alternative construction uses the fact that $\mathbb{Q}$ is an *ordered* field. John Conway, in *On Numbers and Games,* has generalized Dedekind sections to give an extraordinary construction of rationals, reals and infinite and infinitesimal numbers, starting 'from nothing'. Knuth has given a popular account of Conway numbers in *Surreal Numbers*.)

We can complete $\mathbb{Q}$ with respect to the p-adic valuation in just the same way. The resulting field is called *the field of p-adic numbers*, and is denoted by $\mathbb{Q}_p$. We can identify $x \in \mathbb{Q}$ with the Cauchy sequence $(x, x, x, \dots)$. Thus

$$\mathbb{Q} \subset \mathbb{Q}_p.$$

To bring out the parallel with the reals, we sometimes write

$$\mathbb{R} = \mathbb{Q}_\infty.$$

The numbers $x \in \mathbb{Q}_p$ with $\|x\|_p \leq 1$ are called *p-adic integers*. The p-adic integers form a ring, denoted by $\mathbb{Z}_p$. For if $x, y \in \mathbb{Z}_p$ then by property (3) above,

$$\|x + y\|_p \leq \max(\|x\|_p, \|y\|_p) \leq 1,$$

and so $x + y \in \mathbb{Z}_p$. Similarly, by property (1),

$$\|xy\|_p = \|x\|_p \|y\|_p \leq 1,$$

and so $xy \in \mathbb{Z}_p$.

Evidently

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

More generally,

$$x = \frac{m}{n} \in \mathbb{Z}_p$$

if $p \nmid n$. (We sometimes say that a rational number x of this form is *p-integral*.) In other words,

$$\mathbb{Q} \cap \mathbb{Z}_p = \{\frac{m}{n} : p \nmid n\}.$$

Evidently the p-integral numbers form a sub-ring of $\mathbb{Q}$.

The p-adic numbers are in many ways simpler than real numbers, as the following result suggests.

Proposition 2.6. *The series*

$$\sum a_n$$

in $\mathbb{Q}_p$ converges if and only if

$$a_n \to 0 \text{ as } n \to \infty.$$

Proposition 2.7. *Each element $x \in \mathbb{Z}_p$ is uniquely expressible in the form*

$$x = c_0 + c_1 p + c_2 p^2 + \cdots$$

with $c_i \in \{0, 1, \ldots, p-1\}$.

More generally, each element $x \in \mathbb{Q}_p$ is uniquely expressible in the form

$$x = c_{-i} p^{-i} + c_{-i+1} p^{-i+1} + \cdots + c_0 + c_1 p + \cdots \quad (0 \le c_i < p).$$

We can think of this as the p-adic analogue of the decimal expansion of a real number $x \in \mathbb{R}$.

Suppose for example $p = 3$. Let us express $1/2 \in \mathbb{Q}_3$ in standard form. The first step is to determine if

$$\frac{1}{2} \equiv 0, 1 \text{ or } 2 \bmod 3.$$

In fact $2^2 \equiv 1 \bmod 3$; and so

$$\frac{1}{2} \equiv 2 \bmod 3.$$

Next

$$\frac{1}{3}\left(\frac{1}{2} - 2\right) = -\frac{1}{2} \equiv 1 \bmod 3$$

ie

$$\frac{1}{2} - 2 \equiv 1 \cdot 3 \bmod 3^2.$$

Thus

$$\frac{1}{2} \equiv 2 + 1 \cdot 3 \bmod 3^2$$

For the next step,

$$\frac{1}{3}\left(-\frac{1}{2}-1\right) = -\frac{1}{2} \equiv 1 \bmod 3$$

giving

$$\frac{1}{2} \equiv 2 + 1 \cdot 3 + 1 \cdot 3^2 \bmod 3^3$$

It is clear that this pattern will be repeated indefinitely. Thus

$$\frac{1}{2} = 2 + 3 + 3^2 + 3^3 + \cdots .$$

To check this,

$$\begin{aligned} 2 + 3 + 3^2 + \cdots &= 1 + (1 + 3 + 3^2 + \cdots) \\ &= 1 + \frac{1}{1-3} \\ &= 1 - \frac{1}{2} \\ &= \frac{1}{2}. \end{aligned}$$

As another illustration, let us expand $3/5 \in \mathbb{Q}_7$. We have

$$\begin{aligned} \frac{3}{5} &\equiv 2 \bmod 7 \\ \frac{1}{7}\left(\frac{3}{5}-2\right) = -\frac{1}{5} &\equiv 4 \bmod 7 \\ \frac{1}{7}\left(-\frac{1}{5}-4\right) = -\frac{3}{5} &\equiv 5 \bmod 7 \\ \frac{1}{7}\left(-\frac{3}{5}-5\right) = -\frac{4}{5} &\equiv 2 \bmod 7 \\ \frac{1}{7}\left(-\frac{4}{5}-2\right) = -\frac{2}{5} &\equiv 1 \bmod 7 \\ \frac{1}{7}\left(-\frac{2}{5}-1\right) = -\frac{1}{5} &\equiv 4 \bmod 7 \end{aligned}$$

We have entered a loop; and so (in $\mathbb{Q}_7$)

$$\frac{3}{5} = 2 + 4 \cdot 7 + 5 \cdot 7^2 + 2 \cdot 7^3 + 1 \cdot 7^4 + 4 \cdot 7^5 + 5 \cdot 7^6 + \cdots$$

Checking,

$$\begin{aligned} 1 + \left(1 + 4 \cdot 7 + 5 \cdot 7^2 + 2 \cdot 7\right) \frac{1}{1 - 7^4} &= 1 - \frac{960}{2400} \\ &= 1 - \frac{2}{5} \\ &= \frac{3}{5}. \end{aligned}$$

It is not difficult to see that a number $x \in \mathbb{Q}_p$ has a recurring p-adic expansion if and only if it is rational (as is true of decimals).

Let $x \in \mathbb{Z}_p$. Suppose $\|x\|_p = 1$. Then

$$x = c + yp,$$

where $0 < c < p$ and $y \in \mathbb{Z}_p$. Suppose first that $c = 1$, ie

$$x = 1 + yp.$$

Then x is invertible in $\mathbb{Z}_p$, with

$$x^{-1} = 1 - yp + y^2p^2 - y^3p^3 + \cdots .$$

Even if $c \neq 1$ we can find d such that

$$dc \equiv 1 \bmod p.$$

Then

$$dx \equiv dc \equiv 1 \bmod p,$$

say

$$dx = 1 + py,$$

and so x is again invertible in $\mathbb{Z}_p$, with

$$x^{-1} = d\left(1 - yp + y^2p^2 - \cdots\right).$$

Thus the elements $x \in \mathbb{Z}_p$ with $\|x\|_p = 1$ are all *units* in $\mathbb{Z}_p$, ie they have inverses in $\mathbb{Z}_p$; and all such units are of this form. These units form the multiplicative group

$$\mathbb{Z}_p^\times = \{x \in \mathbb{Z}_p : \|x\|_p = 1\}.$$

2.5 The product formula

Proposition 2.8. *Suppose $\alpha \in \mathbb{Q}$, $\alpha \neq 0$. Then*

$$\|\alpha\|_p = 1$$

for almost all places p, ie for all but a finite number of p; and

$$\prod_p \|\alpha\|_p = 1,$$

where the product extends over all the places in $\mathbb{Q}$.

Appendix A

The Structure of Finite Abelian Groups

A.1 The p-components

Proposition A.1. *Suppose A is an abelian group. For each prime p, the elements of order p^n in A for some $n \in N$ form a subgroup*

$$A_p = \{a \in A : p^n a = 0 \text{ for some } n \in \mathbb{N}\}.$$

Proof ▶. Suppose $a, b \in A_p$. Then

$$p^m a = 0,\ p^n b = 0,$$

for some m, n. Hence

$$p^{m+n}(a+b) = 0,$$

and so $a + b \in A_p$. ◀

Definition A.1. *We call A_p the* p-component *of A.*

Proposition A.2. *If A is an abelian group of order n then $A_p = 0$ unless $p \mid n$; and A is the direct sum of the A_p for $p \mid n$:*

$$A = \oplus_p A_p.$$

Proof ▶. Suppose

$$n = p_1^{e_1} \cdots p_r^{e_r}.$$

Let

$$m_i = n/p_i^{e_i} = p_1^{e_1} \cdots p_{i-1}^{e_{i-1}} p_{i+1}^{e_{i+1}} \cdots p_r^{e_r}.$$

Then

$$p_i^{e_i} m_i = n;$$

and therefore

$$p_i^{e_i}(m_i a) = na = 0$$

for all $a \in A$. Thus

$$m_i a \in A_p.$$

Now $\gcd(m_1, \ldots, m_r) = 1$. Therefore we can find $u_1, \ldots, u_r \in \mathbb{Z}$ such that

$$m_1 u_1 + \cdots + m_r u_r = 1.$$

Then

$$m_1 u_1 a + \cdots + m_r u_r a = a,$$

ie

$$a = a_1 + \cdots + a_r,$$

where

$$a_i = m_i n_i a \in A_p.$$

Hence A is the sum of the subgroups A_p.

To see that this sum is direct, suppose

$$a_1 + \cdots + a_r = 0,$$

where $a_i \in A_{p_i}$. Suppose

$$p_i^{e_i} a_i = 0.$$

Let

$$m_i = p_1^{e_1} \cdots p_{i-1}^{e_{i-1}} p_{i+1}^{e_{i+1}} \cdots p_r^{e_r}.$$

Then

$$m_i a_j = 0 \text{ for } i \neq j.$$

Thus (multiplying the given relation by m_i),

$$m_i a_i = 0.$$

But $\gcd(m_i, p_i^{e_i}) = 1$. Hence we can find m, n such that

$$m m_i + n p_i^{e_i} = 1.$$

But then

$$a_i = m(m_i a_i) + n(p_i^{e_i} a_i) = 0.$$

We conclude that A is the direct sum of its p-components A_p. ◀

A.2 Abelian p-groups

A group G (not necessarily abelian, or even finite) is said to be a p-group if every element $g \in G$ has order p^e for some e.

If G is finite this is the same as saying that $\|G\| = p^e$ for some e. That follows from Sylow's Theorem; if a different prime q divides $\|G\|$ then G has a subgroup of order q^f whose elements will have order q^r.

If G is abelian this is easier to establish, by induction on $\|G\|$. For if we take any subgroup $S \subset G$ then the elements of both S and G/S will have orders of the form p^e.

Theorem A.1. *Suppose A is a finite abelian p-group. Then A can be expressed as a direct sum of cyclic p-groups:*

$$A = \mathbb{Z}/(p^{e_1}) \oplus \cdots \oplus \mathbb{Z}/(p^{e_r}).$$

Moreover the powers $p^{e_1}, \ldots, p^{e_r}$ are uniquely determined by A.

Proof ▶. We argue by induction on $\|A\| = p^n$. We may assume therefore that the result holds for the subgroup

$$pA = \{pa : a \in A\}.$$

For pA is stricty smaller than A, since

$$pA = A \implies p^n A = A,$$

while we know from Lagrange's Theorem that $p^n A = 0$.

Suppose

$$pA = \langle pa_1 \rangle \oplus \cdots \oplus \langle pa_r \rangle.$$

Then the sum

$$\langle a_1 \rangle + \cdots + \langle a_r \rangle = B,$$

say, is direct. For suppose

$$n_1 a_1 + \cdots + n_r a_r = 0.$$

If $p \mid n_1, \ldots, n_r$, say $n_i = pm_i$, then we can write the relation in the form

$$m_1(pa_1) + \cdots + m_r(pa_r) = 0,$$

whence $m_i p a_i = n_i a_i = 0$ for all i.

On the other hand, if p does not divide all the n_i then

$$n_1(pa_1) + \cdots + n_r(pa_r) = 0,$$

and so $p n_i a_i = 0$ for all i. But if $p \nmid n_i$ this implies that $pa_i = 0$. (For the order of a_i is a power of p, say p^e; and $p^e \mid n_i p$ implies that $e \leq 1$.) But this

contradicts our choice of pa_i as a generator of a direct summand of pA. Thus the subgroup $B \subset A$ is expressed as a direct sum

$$B = \langle a_1 \rangle \oplus \cdots \oplus \langle a_r \rangle.$$

Let

$$K = \{a \in A : pa = 0\}.$$

Then

$$A = B + K.$$

For suppose $a \in A$. Then $pa \in pA$, and so

$$pa = n_1(pa_1) + \cdots + n_r(pa_r)$$

for some $n_1, \ldots, n_r \in \mathbb{Z}$. Thus

$$p(a - n_1a_1 - \cdots - n_ra_r) = 0,$$

and so

$$a - n_1a_1 - \cdots - n_ra_r = k \in K.$$

Hence

$$a = (n_1a_1 + \cdots + n_ra_r) + k \in B + K.$$

If $B = A$ then all is done. If not, then $K \not\subset B$, and so we can find $k_1 \in K, k_1 \notin B$. Now the sum

$$B_1 = B + \langle k_1 \rangle$$

is direct. For $\langle k_1 \rangle$ is a cyclic group of order p, and so has no proper subgroups. Thus

$$B \cap \langle k_1 \rangle = \{0\},$$

and so

$$B_1 = B \oplus \langle k_1 \rangle$$

If now $B_1 = A$ we are done. If not we can repeat the construction, by choosing $k_2 \in K, k_2 \notin B_1$. As before, this gives us a direct sum

$$B_2 = B_1 \oplus \langle k_2 \rangle = B \oplus \langle k_1 \rangle \oplus \langle k_2 \rangle.$$

Continuing in this way, the construction must end after a finite number of steps (since A is finite):

$$\begin{aligned} A = B_s &= B \oplus \langle k_1 \rangle \oplus \cdots \oplus \langle k_s \rangle \\ &= \langle a_1 \rangle \oplus \cdots \oplus \langle a_r \rangle \oplus \langle k_1 \rangle \oplus \cdots \oplus \langle k_s \rangle. \end{aligned}$$

It remains to show that the powers $p^{e_1}, \ldots, p^{e_r}$ are uniquely determined by A. This follows easily by induction. For if A has the form given in the theorem then

$$pA = \mathbb{Z}/(p^{e_1-1}) \oplus \cdots \oplus \mathbb{Z}/(p^{e_r-1}).$$

Thus if $e > 1$ then $\mathbb{Z}/(p^e)$ occurs as often in A as $\mathbb{Z}/(p^{e-1})$ does in pA. It only remains to deal with the factors $\mathbb{Z}/(p)$. But the number of these is now determined by the order $\|A\|$ of the group. ◀

Remark. It is important to note that if we think of A as a direct sum of cyclic *subgroups*, then the orders of these subgroups are uniquely determined, by the theorem; but *the actual subgroups themselves are not in general uniquely determined.*

For example, if

$$A = \mathbb{Z}/(p) \oplus \mathbb{Z}/(p)$$

then every non-zero element of A is of order p. Thus if we take any $a \neq 0$, and then any $b \notin \langle a \rangle$ we will have

$$A = \langle a \rangle \oplus \langle b \rangle.$$

In fact it is not hard to see that the component subgroups are *never* uniquely determined, unless A is a cyclic p-group (in which case there is only one summand)

To see this, it is sufficient to consider the case of 2 summands:

$$A = \mathbb{Z}/(p^e) \oplus \mathbb{Z}/(p^f).$$

We may suppose that $e \geq f$. Let a_1, a_2 be the generators of the 2 summands. Then it is easy to see that we could equally well take $a_1' = a_1 + a_2$ in place of a_1:

$$A = \langle a_1 + a_2 \rangle \oplus \langle a_2 \rangle.$$

For certainly these elements $a_1 + a_2, a_2$ generate the group; and the sum must be direct, since otherwise there would not be enough terms $m_1 a_1' + m_2 a_2$ to give all the p^{e+f} elements in A.

A.3 The Structure Theorem

Putting together the results of the last two sections, we derive the Structure Theorem for Finite Abelian Groups.

Theorem A.2. *Every finite abelian group A is expressible as a direct sum of cyclic groups of prime-power order:*

$$A = \mathbb{Z}/(p^{e_1}) \oplus \cdots \oplus \mathbb{Z}/(p^{e_s}).$$

Moreover the prime-powers $p_1^{e_1}, \ldots, p_s^{e_s}$ are uniquely determined by A.

Proof ▶. We first split the group into its p-components:

$$A = A_{p_1} \oplus A_{p_r}.$$

Then we can split each component A_p into cyclic p-group, as we have just seen. ◀

Remark. The splitting of A into its components A_p *is* unique, since A_p contains all the elements of order p_e.

But as we have seen, the splitting of A_p into cyclic summands is unique only if A_p is cyclic.

Thus the splitting of A is unique if (and only if) each component A_p is cyclic. As we shall see in the next Section, this is the case if and only if A itself is cyclic.

A.4 Cyclic groups

Proposition A.3. *If* $\gcd(m.n) = 1$ *then*

$$\mathbb{Z}/(m) \oplus \mathbb{Z}/(n) \cong \mathbb{Z}/(mn).$$

Proof ▶. This is just a re-statement of the Chinese Remainder Theorem.

For any m, n, the natural group-homomorphisms

$$\mathbb{Z}/(mn) \to \mathbb{Z}/(m),\ \mathbb{Z}/(mn) \to \mathbb{Z}/(n)$$

combine to give a homomorphism

$$\Theta : \mathbb{Z}/(mn) \to \mathbb{Z}/(m) \oplus \mathbb{Z}/(n).$$

If $\gcd(m, n) = 1$ then Θ is injective, since

$$a \bmod m = 0,\ a \bmod n = 0 \implies a \bmod mn = 0.$$

Since the groups on each side have the same order mn it follows that Θ is bijective, ie an isomorphism. ◀

The converse is also true.

Proposition A.4. *If* $\gcd(m.n) > 1$ *then* $\mathbb{Z}/(m) \oplus \mathbb{Z}/(n)$ *is not cyclic.*

Proof ▶. Suppose $\gcd(m, n) = d$. Let

$$m = dm',\ n = dn'.$$

Then

$$mn/d = mn' = m'n.$$

Thus each element $a \in \mathbb{Z}/(m) \oplus \mathbb{Z}/(n)$ satisfies

$$(mn/d)\ a = 0.$$

So there is no element of order mn, and the direct sum is not cyclic. ◀

Proposition A.5. *Every subgroup of a cyclic group is cyclic.*

Proof ▶. Suppose

$$S \subset A = \langle a \rangle,$$

where a is of order n.

Let d be the smallest integer ≥ 1 such that

$$da \in S.$$

We assert that da must generate S:

$$S = \langle da \rangle.$$

For suppose

$$s = ma \in S.$$

Divide m by d,

$$m = qd + r,$$

where $0 \leq r < d$. Then

$$ra = s - q(da) \in S.$$

Hence $r = 0$, from the definition of d, ie

$$s = q(da).$$

◀

Theorem A.3. *The finite abelian group A is cyclic if and only if each component A_p is cyclic.*

Proof ▶. If A is cyclic then so is each component A_p by the last Proposition.

On the other hand, if each component is cyclic, then so is their direct sum A, by Proposition A.3. ◀

A.5 Concluding remarks

A.5.1 Finitely-generated abelian groups

The Structure Theorem above extends to *finitely-generated* abelian groups.

Such a group splits into a direct sum of cyclic subgroups of prime-power order (as before), together with a number of copies of the additive group $\mathbb{Z}$.

Furthermore, the prime-powers that arise are uniquely determined by the group, as also are the number of copies of $\mathbb{Z}$.

The proof depends on the fact that the elements of finite order in a finitely-generated abelian group A form a subgroup T, the *torsion subgroup* of A. The quotient-group A/T is *torsion-free*, ie it has no elements of finite order except 0.

Now one can show that a finitely-generated and torsion-free abelian group is necessarily the direct sum $\mathbb{Z}^r$ of a number of copies of Z. Moreover, A splits into a direct sum

$$A = \mathbb{Z}^r \oplus T.$$

The number r of copies of Z is called the *rank* of A.

A.5.2 Abelian groups and modules

Recall that a module M over a ring R is defined in exactly the same way as a vector space V over a field k. The only difference is we do not assume that a non-zero scalar $\lambda \in R$ has an inverse.

An abelian group A can be regarded as a $\mathbb{Z}$-module, that is, a module over the ring of integers, where we regard na as multiplication of a by the scalar n.

Thus an abelian group can be regarded either as a particular kind of group, or as a particular kind of module. The second point of view is probably the more appropriate one in most cases.

In particular, the Structure Theorem of Finite Abelian Groups expounded above does not extend in any natural way to non-commutative groups. It *does* however extend more or less unchanged to modules over *Principal Ideal Domains*.

In particular, it extends to modules over the ring $k[x]$ of polynomials with coefficients in a field k.

One important application of this is to the Jordan Form for a linear transformation $t : V \to V$ of a finite-dimensional vector space V over $\mathbb{C}$ (or equivalently, of a square matrix T over $\mathbb{C}$).

We can regard V as a module over the ring $\mathbb{C}[x]$, with the polynomial $p(x)$ acting on V by

$$(p, v) \mapsto p(T)\ v.$$

The Structure Theorem shows that V splits into "cyclic" components belonging to the different eigenvalues λ of T (which correspond to the different primes p). These cyclic components define the sub-matrices into which the matrix (or rather, a matrix similar to T) splits.