A Course of Elementary Number Theory

R. C. Vaughan

Pennsylvania State University

21st May 2022

Contents

Preface

This book is based on courses given for nearly fifty years starting in 1972 at Imperial College London and Penn State University, and initially modeled on courses attended by the author when an undergraduate at University College, London and given by Professors J. H. H. Chalk and G. L. Watson. It contains typically enough material for about thirty six hours of presentations and nine to twelve hours of problem solving and tutorials. All the exercises have been used at least once for homework or the basis of examination questions.

The material in the last chapter or two might be considered to be somewhat biased towards analytic number theory, which is hardly surprising since that has been the main thrust of the author's research. Moreover it can be mentioned that research in analytic number theory has increased in intensity over the last couple of decades and two of the Millennium Problems are related to this field. However the only prerequisite is knowledge of basic college algebra, calculus and some facility with the manipulation of formulæ. This author prefers to avoid as much jargon as possible and generally avoids clouding the issue with constant reference to concepts from abstract algebra. It would also be remiss not to point out that much of modern abstract algebra can trace its origins to questions in elementary number theory.

One word of warning. This is a subject which demands proofs, and it would be wise to also have some facility with constructing simple proofs in good English. If one wishes to understand the reasons for a particular phenomenon this can often only be seen by understanding why the proof works.

Chapter 1

Introduction

1.1 The integers

Number theory in its most basic form is the study of the set of *integers*

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$$

and its important subset

$$\mathbb{N} = \{1, 2, 3, \ldots\},$$

the set of positive integers, sometimes called the *natural numbers*. They have all kinds of amazing and beautiful properties. The usual rules of arithmetic apply, and can be deduced from a set of axioms. If you multiply any two members of $\mathbb{Z}$ you get another one. Likewise for $\mathbb{N}$, or if you subtract one member of $\mathbb{Z}$ from another, e.g.

$$173 - 192 = -19.$$

But this last fails for $\mathbb{N}$.

You can do other standard things, such as

$$x(y + z) = xy + xz$$

and

$$xy = yx$$

is always true.

Here is an interesting question. Let me start by listing some numbers

$$\begin{aligned}
1 &= 0^2 + 1^2 \\
2 &= 1^2 + 1^2 \\
3 &= ? \\
4 &= 0^2 + 2^2 \\
5 &= 1^2 + 2^2 \\
6 &= ? \\
7 &= ? \\
13 &= 2^2 + 3^2 \\
19 &= ? \\
21 &= ? \\
41 &= 4^2 + 5^2 \\
45 &= 3^2 + 6^2
\end{aligned}$$

Fermat figured out exactly which numbers are the sums of two squares and which are not, and the first published proof is by Euler. Fermat probably had a proof, at least when n is prime, but did not publish it. It looks as though an odd prime number p is the sum of two squares if and only if it leaves the remainder 1 on division by 4. Check some examples yourself. Then wonder how you might prove it!

Which brings me to an important point. This is a *proofs* based course. The proofs will be mostly short and simple, but they are necessary, and as a general principle understanding the proof usually reveals the underlying structure which is the reason why the theorem is true. There is also an instructive example due to J. E. Littlewood in 1912.

Let $\pi(x)$ denote the number of prime numbers not exceeding x. Gauss had suggested that

$$\int_0^x \frac{dt}{\log t}$$

should be a good approximation to $\pi(x)$

$$\pi(x) \sim \mathrm{li}(x).$$

For all values of x for which $\pi(x)$ has been calculated it has been found that

$$\pi(x) < \mathrm{li}(x).$$

There is a table of values in §8.1 which illustrates this for various values of x out to 10^{27}. But nevertheless Littlewood in 1914 showed that there are infinitely many values of x for which

$$\pi(x) > \mathrm{li}(x).$$

We now believe that the first sign change occurs when

$$x \approx 1.387162 \times 10^{316} \tag{1.1}$$

well beyond what can be calculated directly. For many years it was only known that the first sign change occurs for *some* x satisfying

$$x < 10^{10^{10^{964}}}.$$

The number on the right was computed by Skewes. G. H. Hardy once wrote that this is probably the largest number which has ever had any *practical* (my emphasis) value. But still even now the only way of establishing this is by a proper mathematical proof.

1.2 Divisibility

We start with some definitions. We need some concept of divisibility and factorization. Given two integers a and b we say that a divides b, if there is a third integer c such that

$$ac = b$$

and we write

$$a|b.$$

Example 1.1. *If $a|b$ and $b|c$, then $a|c$.*

Proof. There are d and e so that $b = ad$ and $c = be$. Hence $a(de) = (ad)e = be = c$ and de is an integer. □

There are some facts which are useful. For any a we have $0a = 0$, and if $ab = 1$, then $a = \pm 1$ and $b = \pm 1$ (with the same sign in each case). Also if $a \neq 0$ and $ac = ad$, then $c = d$.

Definition 1.1. *A member of $\mathbb{N}$ greater than 1 which is only divisible by 1 and itself is called a prime number.*

By the way we will use the letter p routinely to denote a prime number.

Example 1.2. *101 is a prime number.*

Proof. How to prove this? Well obviously one only needs to check for divisors d with $1 < d < 100$. Moreover if d is a divisor, then there is an e so that $de = 101$, and one of d, e is $\leq \sqrt{101}$ so we only need to check out to 10. Oh, and really we only need to check the primes $2, 3, 5, 7$. Moreover 2 and 5 are obviously not divisors and 3 is easily checked by the usual rule, so only 7 needs any checking, and this leaves the remainder 3, not 0. □

Since we are dealing with simple proofs for facts about $\mathbb{N}$ there is one proof method which is very important. This is the principle of induction. It is actually embedded into the definition of $\mathbb{N}$. That is, we have $1 \in \mathbb{N}$ and it is the least member and given any $n \in \mathbb{N}$ the next member is $n+1$. In this way one sees that $\mathbb{N}$ is *defined* inductively.

A statement which is provably equivalent is the well-ordering principle which says that any non-empty set of integers which is bounded below has a minimal element.

Theorem 1.1. *Every member of* $\mathbb{N}$ *is a product of prime numbers.*

Proof. 1 is an "empty product" of primes, so the case $n = 1$ holds. Suppose that we have proved the result for every m with $m \leq n$. If $n+1$ is prime we are done. Suppose $n+1$ is not prime. Then there is an a with $a|n+1$ and $1 < a < n+1$. Then also $1 < \frac{n+1}{a} < n+1$. But then on the inductive hypothesis both a and $\frac{n+1}{a}$ are products of primes. □

We can use this to deduce

Theorem 1.2 (*Euclid*)**.** *There are infinitely many primes.*

Proof. We argue by contradiction. Suppose there are only a finite number of primes. Call them $p_1, p_2, \ldots, p_n$ and consider the number

$$m = p_1 p_2 \ldots p_n + 1.$$

Since we already know some primes it is clear that $m > 1$. Hence it is a product of primes, and in particular there is a prime p which divides m. But p is one of the primes $p_1, p_2, \ldots, p_n$ so $p|m - p_1 p_2 \ldots p_n = 1$. But 1 is not divisible by any prime. So our assumption must have been false. □

Hardy cites this proof as an example of beauty in mathematics.

Here is an example which we will use multiple times during some of our simple proofs.

Example 1.3 (*Dirichlet's box principle*)**.** *Suppose that we have n boxes and a collection of* $n+1$ *objects and we put the objects into boxes at random. Then one box will contain at least two objects.*

Proof. The case $n = 1$ is obvious (I hope). Suppose the n-th case is already proven and now we have $n+1$ boxes and $n+2$ objects. We argue by contradiction. Put the objects into the boxes at random and suppose that no box would have two objects in it. However even so at least one box would have one object in it. Remove that box. Now we have placed $n+1$ objects in the n remaining boxes and we have a contradiction to the case already proven. □

1.2.1 Exercises

Divisibility and Factorisation

1. Let $a, b, c \in \mathbb{Z}$. Prove each of the following.
   (i) $a|a$.
   (ii) If $a|b$ and $b|a$, then $a = \pm b$.
   (iii) If $a|b$ and $b|c$, then $a|c$.
   (iv) If $ac|bc$ and $c \neq 0$, then $a|b$.
   (v) If $a|b$, then $ac|bc$.
   (vi) If $a|b$ and $a|c$, then $a|bx + cy$ for all $x, y \in \mathbb{Z}$.

2. The Fibonacci sequence (1202) is defined iteratively by $F_1 = F_2 = 1$, $F_{n+1} = F_n + F_{n-1}$ $(n = 2, 3, \ldots)$. Show that if m, $n \in \mathbb{N}$ satisfy $m|F_n$ and $m|F_{n+1}$, then $m = 1$.

3. Prove that if n is odd, then $8|n^2 - 1$.

4. (i) Show that if m and n are integers of the form $4k + 1$, then so is mn.
   (ii) Show that if $m, n \in \mathbb{N}$, and mn is of the form $4k - 1$, then so is one of m and n.
   (iii) Show that every number of the form $4k - 1$ has a prime factor of this form.
   (iv) Show that there are infinitely many primes of the form $4k - 1$.

5. (i) Show that if m and n are integers of the form $6k + 1$, then so is mn.
   (ii) Show that if l and m are of the form $6k + 1$, then so is lm.
   (iii) Show that if lm id of the form $6k - 1$, then either l is of this form or m is.
   (iv) Show that if $n \in \mathbb{N}$ and n is of the form $6k - 1$, then there is a prime number p such that $p|n$ and p is of the form $6k - 1$.
   (v) Show that there are infinitely many primes of the form $6k - 1$.

6. Show that if p is a prime number and $1 \leq j \leq p - 1$, then p divides the binomial coefficient $\binom{p}{j}$.

7. Show that $n|(n - 1)!$ for all composite $n > 4$.

8. Prove that if $2^m + 1$ is an odd prime, then there is an $n \in \mathbb{N}$ such that $m = 2^n$. These are the Fermat primes. Fermat thought that all numbers of the form $2^{2^n} + 1$ are prime. Show that $641|2^{2^5} + 1$.

1.3 The fundamental theorem of arithmetic

We now come to something very important

Theorem 1.3 (The division algorithm)**.** *Suppose that $a \in \mathbb{Z}$ and $d \in \mathbb{N}$. Then there are unique q, $r \in \mathbb{Z}$ such that*

$$a = dq + r, \quad 0 \leq r < d.$$

The number q is called the quotient and r the remainder. By the way, it is exactly this which one uses when one performs long division.

Example 1.4. *Try dividing* 17 *into* 192837465 *by the method you were taught at primary school.*

Proof. To prove the theorem we introduce the following subset of the integers

$$\mathcal{D} = \{a - dx : x \in \mathbb{Z}\}.$$

If $a \geq 0$, then $a - d(-1) \in \mathcal{D}$ and $a - d(-1) = a + d > 0$, and if $a < 0$, then $a - d(a-1) = (d-1)(-a) + d > 0$. Hence $\mathcal{D}$ contains non-negative integers. Let $\mathcal{D}^* = \mathcal{D} \cap \mathbb{N}$. Then $\mathcal{D}^*$ is bounded below and non-empty, so by the well-ordering principle it has a minimum. Let r denote this minimum, and let q be the corresponding value of x. Then we have

$$a = dq + r, \quad 0 \leq r.$$

Moreover if we would have $r \geq d$, then

$$a = d(q+1) + (r - d)$$

gives another solution, but with the remainder $r - d < r$ contradicting the minimality of r. Hence

$$r < d$$

as required.

To prove the uniqueness observe that if we have a second solution

$$a = dq' + r', \quad 0 \leq r' < d$$

then $0 = a - a = (dq' + r') - (dq + r) = d(q' - q) + (r' - r)$. Moreover if $q' \neq q$, then we would have $d \leq d|q' - q| = |r' - r| < d$ which is impossible, so $q' = q$ and $r' = r$. □

We will make frequent use of the division algorithm as well as the next theorem.

Theorem 1.4. *Given two integers* a *and* b*, not both* 0*, define*

$$\mathcal{D}(a,b) = \{ax + by : x \in \mathbb{Z}, y \in \mathbb{Z}\}.$$

Then $\mathcal{D}(a,b)$ *has positive elements. Let* (a,b) *denote the least positive element. Then* (a,b) *has the properties*

(i) $(a,b)|a$,

(ii) $(a,b)|b$,

(iii) if the integer c *satisfies* $c|a$ *and* $c|b$*, then* $c|(a,b)$.

Definition 1.2. *The number* (a,b) *is called the greatest common divisor of* a *and* b*, often abbreviated to GCD. The symbol* (a,b) *has many uses in mathematics, so to be clear one sometimes writes*

$$GCD(a,b).$$

Proof of Theorem 1.4. If a is positive, then so is $a.1 + b.0$. Likewise if b is positive. If a is negative, then $a(-1) + b.0$ is positive, and again likewise if b is negative. The only remaining case is $a = b = 0$ which is expressly excluded. Thus $\mathcal{D}(a, b)$ does indeed have positive elements. Thus (a, b) exists. Suppose (i) is false. By the division algorithm we have

$$a = (a, b)q + r$$

with $0 \leq r < (a, b)$. But the falsity of (i) means that $0 < r$. Thus

$$r = a - (a, b)q = a - (ax + by)q$$

for some integers x and y. Hence

$$r = a(1 - xq) + b(-yq).$$

Since $0 < r < (a, b)$ this contradicts the minimality of (a, b).

Likewise for (ii). Now suppose $c|a$ and $c|b$, so that $a = cu$ and $b = cv$ for some integers u and v. Then

$$(a, b) = ax + by = cux + cvy = c(ux + vy)$$

so (iii) holds. □

The GCD has some interesting properties. Here is one

Example 1.5. *We have*

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

To see this observe that if $d = \left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right)$, *then* $d|\frac{a}{(a,b)}$ *and* $d|\frac{b}{(a,b)}$, *and hence* $d(a, b)|a$ *and* $d(a, b)|b$. *But then* $d(a, b)|(a, b)$ *and so* $d|1$, *whence* $d = 1$.

Here is another

Example 1.6. *Suppose that* a *and* b *are not both* 0. *Then for any integer* x *we have* $(a + bx, b) = (a, b)$. *Here is a proof. First of all* $(a, b)|a$ *and* $(a, b)|b$, *so* $(a, b)|a + bx$. *Hence* $(a, b)|(a + bx, b)$. *On the other hand* $(a + bx, b)|a + bx$ *and* $(a + bx, b)|b$ *so that* $(a + bx)|a + bx - bx = a$. *Hence* $(a + bx, b)|(a, b)|(a + bx, b)$ *and so* $(a, b) = (a + bx, b)$.

Here is yet another

Example 1.7. *Suppose that* $(a, b) = 1$ *and* $ax = by$. *Then there is a* z *such that* $x = bz$, $y = az$. *It suffices to show that* $b|x$, *for then the conclusion follows on taking* $z = x/b$. *To see this observe that there are* u *and* v *so that* $au + bv = (a, b) = 1$. *Hence* $x = aux + bvx = byu + bvx = b(yu + vx)$ *and so* $b|x$.

Following from the previous theorem we immediately have the following

Corollary 1.5. *Suppose that a and b are integers not both 0. Then there are integers x and y such that*

$$(a, b) = ax + by.$$

Later we will look at a way of finding suitable x and y in examples. As it stands the theorem gives no constructive way of finding them. It is a pure existence proof.

As a first application we establish

Theorem 1.6 (Euclid)**.** *Suppose that p is a prime number, and a and b are integers such that $p|ab$. Then either $p|a$ or $p|b$.*

You might think this is obvious, but look at the following

Example 1.8. *Consider the set $\mathcal{A}$ of integers of the form $4k + 1$. If you multiply two of them together, e.g. $(4k_1+1)(4k_2+1) = 16k_1k_2+4k_2+4k_1+1 = 4(4k_1k_2+k_1+k_2)+1$ you get another integer of the same kind. So they have "closure" under multiplication. We can define a "prime" p in this system if it is only divisible by 1 and itself in the system. Here is a list of "primes" in $\mathcal{A}$.*

$$5, 9, 13, 17, 21, 29, 33, 37, 41, 49 \ldots$$

Note that 9 is one because 3 is not in the system. Likewise 21 and 49 because 3 and 7 are not in the system. Also the "prime" factorisation of 45 is 5×9. Now look at 441. We have

$$441 = 9 \times 49 = 21^2.$$

Wait a minute, in this system, factorisation is not unique!. If you look at the above theorem, it must be false in the system $\mathcal{A}$ because we have $21|9 \times 49$ but 21 does not divide 9 or 49!

What is the difference between $\mathbb{Z}$ and $\mathcal{A}$? Well $\mathbb{Z}$ has an additive structure and $\mathcal{A}$ does not. Add two members of $\mathbb{Z}$ and you get another one. Add two members of $\mathcal{A}$ and you get a number which leaves the remainder 2 on division by 4, so is not in $\mathcal{A}$. Amazingly we have to use the additive structure to get something fundamental about the multiplicative structure. This is of huge significance and underpins some of the most fundamental questions in mathematics.

Proof. If a or b are 0, then the result is obvious. Thus we may suppose that $ab \neq 0$. Suppose that $p \nmid a$. We know from the previous theorem that there are x and y so that $(a, p) = ax + py$ and that $(a, p)|p$ and $(a, p)|a$. Since p is prime we must have $(a, p) = 1$ or p. But we are supposing that $p \nmid a$ so $(a, p) \neq p$, i.e. $(a, p) = 1$. Hence

$$1 = ax + py.$$

But then

$$b = abx + pby$$

and since $p|ab$ we have $p|b$ as required. □

We can use this to establish the following

Theorem 1.7. *Suppose that $p, p_1, p_2, \ldots, p_r$ are prime numbers and*

$$p|p_1p_2\ldots p_r.$$

Then $p = p_j$ for some j.

Proof. We can prove this by induction on r. The case $r = 1$ is immediate from the definition of prime. Suppose we have established the r-th case and that we have $p|p_1p_2\ldots p_{r+1}$. Then by the previous theorem we have $p|p_{r+1}$ or $p|p_1p_2\ldots p_r$. In the first case we must have $p = p_{r+1}$. In the second by the inductive hypothesis we must have $p = p_j$ for some j with $1 \leq j \leq r$. □

Theorem 1.8 (The Fundamental Theorem of Arithmetic)**.** *Factorization into prime numbers is unique apart from the order of the factors. More precisely if a is a non-zero integer and $a \neq \pm 1$, then*

$$a = (\pm 1)p_1p_2\ldots p_r$$

for some $r \geq 1$ and prime numbers $p_1, \ldots, p_r$, and r and the choice of sign is unique and the primes p_j are unique apart from their ordering.

Proof. It is clear that we may suppose that $a > 0$, and hence that $a \geq 2$. We saw in the very first theorem that a will be a product of r primes, say

$$a = p_1p_2\ldots p_r$$

with $r \geq 1$. We prove the result by induction on r. Suppose $r = 1$ and it is another product of primes

$$a = p_1'\ldots p_s'$$

where $s \geq 1$. Then $p_1'|p_1$ and so $p_1' = p_1$ and $p_2'\ldots p_s' = 1$, whence $s = 1$ also. Now suppose that the result holds for some $r \geq 1$ and we have a product of $r + 1$ primes, and and as before

$$a = p_1p_2\ldots p_{r+1} = p_1'\ldots p_s'.$$

Then we see from the previous theorem that $p_1' = p_j$ for some j and then

$$p_2'\ldots p_s' = p_1p_2\ldots p_{r+1}/p_j$$

and we can apply the inductive hypothesis to obtain the desired conclusion. □

There are various other properties of GCDs which can now be described. Suppose a and b are positive integers. Then by the previous theorem we can write

$$a = p_1^{r_1}\ldots p_k^{r_k}, \quad b = p_1^{s_1}\ldots p_k^{s_k}$$

where the $p_1, \dots p_k$ are the different primes in the factorization of a and b and we allow the possibility that the exponents r_j and s_j may be zero. Then it can be checked easily that

$$(a, b) = p_1^{\min(r_1,s_1)} \dots p_k^{\min(r_k,s_k)}$$

and this could be taken as the definition of GCD.

Definition 1.3. *We can also introduce here the least common multiple LCM*

$$[a, b] = \frac{ab}{(a, b)}$$

and this could also be defined by

$$[a, b] = p_1^{\max(r_1,s_1)} \dots p_k^{\max(r_k,s_k)}.$$

1.3.1 Exercises

1. Suppose that $l, m, n \in \mathbb{N}$. Prove that $(lm, ln) = l(m, n)$. 2. The squarefree numbers are the natural numbers which have no repeated prime factors, e.g 6, 105. Note that 1 is the only natural number which is both squarefree and a perfect square. Prove that every $n \in \mathbb{N}$ can be written uniquely as the product of a perfect square and a squarefree number.

3. Let $a, b, c \in \mathbb{Z}$ with a and b not both zero. Prove each of the following.
   (i) If $(a, b) = 1$ and $a|bc$, then $a|c$.
   (ii) $\left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right) = 1$.
   (iii) $(a, b) = (a + cb, b)$.

4. Show that if $(a, b) = 1$, then $(a - b, a + b) = 1$ or 2. Exactly when is the value 2?

5. Show that if $ad - bc = \pm 1$, then $(a + b, c + d) = 1$.

6. Suppose that a, $b \in \mathbb{N}$. Prove that $(a, b)[a, b] = ab$.

7. Let $a \in \mathbb{N}$ and $b \in \mathbb{Z}$. Prove that the equations $(x, y) = a$ and $xy = b$ can be solved simultaneously in integers x and y if and only if $a^2|b$.

8. Prove that if $m \in \mathbb{N}$ and $n \in \mathbb{N}$, then there are integers a, b such that $(a, b) = m$ and $[a, b] = n$ if and only if $m|n$.

9. Let $a, b, c, d \in \mathbb{Z}$ with ab and cd not both 0. Prove that

$$(ab, cd) = (a, c)(b, d) \left(\frac{a}{(a, c)}, \frac{d}{(b, d)}\right) \left(\frac{c}{(a, c)}, \frac{b}{(b, d)}\right).$$

10. Prove that there are no positive integers a, b, n with $n > 1$ such that

$$(a^n - b^n)|(a^n + b^n).$$

1.4 Notes

§1 The usual approach to the definition of $\mathbb{N}$ and $\mathbb{Z}$ is to assume $\mathbb{N}$ satisfies a version of the Peano axioms

N1. 1 is a natural number.

N2. Every natural number has a successor which is also a natural number.

N3. 1 is not the successor of any natural number.

N4. If the successor of x equals the successor of y then $x = y$

N5. **Induction Axiom.** If a statement $S(n)$ is true for $n = 1$, and if for each $n \in \mathbb{N}$ the truth of $S(n)$ implies the truth for the successor of n, then the statement is true for every $n \in \mathbb{N}$.

There is an increasing tendency to include 0 in $\mathbb{N}$ and make it play the rôle of 1 in the above axioms, and then define 1 to be the successor of 0. Perhaps the most satisfying way of defining $\mathbb{N}$ is due to Von Neumann.

One can also axiomatise $\mathbb{Z}$ by supposing that there are two operations $+$ and $\times$ and an order relationship $<$ on pairs of elements of $\mathbb{Z}$ such that for every $a, b, c \in \mathbb{Z}$ we have

Z1 Closure. $a + b \in \mathbb{Z}$, $a \times b \in \mathbb{Z}$.

Z2 Associativity. $a + (b + c) = (a + b) + c$, $a \times (b \times c) = (a \times b) \times c$.

Z3 Commutativity. $a + b = b + a$, $a \times b = b \times a$.

Z4 Identities. There are elements 0 and $1 \in \mathbb{Z}$ such that $a + 0 = a$, $a \times 1 = a$.

Z5 Inverse. Given $a \in \mathbb{Z}$ there is an element $(-a) \in \mathbb{Z}$ such that $a + (-a) = 0$.

Z6 Distributivity. $a \times (b + c) = (a \times b) + (a \times c)$ and $(a + b) \times c = (a \times c) + (b \times c)$.

Z7 No zero divisors. If $a \times b = 0$, then $a = 0$ or $b = 0$.

Z8 Order. Exactly one of $a < b$, $a = b$, $b < a$ holds.

Z9 Order $+$. If $a < b$, then $a + c < b + c$.

Z10 Order $\times$. If $a < b$ and $0 < c$, then $a \times c < b \times c$.

By dividing the ordered pairs $(m, n) \in \mathbb{N}^2$ into equivalence classes by putting in the same class those (m, n), (m', n') for which $m + n' = m' + n$ one can construct $\mathbb{Z}$ from $\mathbb{N}$. One can then spend considerable effort deducing all the usual rules of arithmetic from these axioms. For more details see the Wikipedia articles on Natural Numbers and Integers.

Littlewood's theorem is in J. E. Littlewood, J. E. (1914). "Sur la distribution des nombres premiers", Comptes Rendus, 158, 1869–1872. The number (1.1) is computed in D. Stoll, P. Demichel (2011), "The impact of $\zeta(s)$ complex zeros on $\pi(x)$ for $x < 10^{10^{13}}$", Mathematics of Computation, 80 (276), 2381–2394. Skewes work is in S. Skewes (1933), "On the difference $\pi(x) - \mathrm{li}(x)$", Journal of the London Mathematical Society, 8, 277–283 and S. Skewes (1955), "On the difference $\pi(x) - \mathrm{li}(x)$ (II)", Proceedings of the London Mathematical Society, 5, 48–70.

§2 The division algorithm is in Euclid, Book VII, Proposition 1.

§3 The fundamental theorem of arithmetic in special cases is buried in Euclid Book VII and Book IX.

Chapter 2

Euclid's algorithm

2.1 Euclid's algorithm

The question arises. We know that given integers a, b not both 0, there are integers x and y so that

$$(a, b) = ax + by. \tag{2.1}$$

How do we find x and y? Euclid (or at least someone in the Pythagorean school - Euclid was just the amanuensis) solved this problem more than 2000 years ago. Moreover their solution gives a very efficient algorithm and it is still the basis for many numerical methods in arithmetical applications. For example in factorisation routines.

We may certainly suppose that $b > 0$ since multiplying by (-1) does not change the (a, b) - we can replace y by $-y$. For convenience of notation put $r_0 = b$, $r_{-1} = a$, Now apply the division algorithm iteratively as follows

$$\begin{aligned}
r_{-1} &= r_0 q_1 + r_1, \quad 0 < r_1 \leq r_0, \\
r_0 &= r_1 q_2 + r_2, \quad 0 < r_2 < r_1, \\
r_1 &= r_2 q_3 + r_3, \quad 0 < r_3 < r_2, \\
&\dots \\
r_{s-3} &= r_{s-2} q_{s-1} + r_{s-1}, \quad 0 < r_{s-1} < r_{s-2}, \\
r_{s-2} &= r_{s-1} q_s.
\end{aligned}$$

That is, we stop the moment that there is a remainder equal to 0. This could be r_1 if $b|a$, for example, although the way it is written out above it is as if s is at least 3. The important point is that because $r_j < r_{j-1}$, sooner or later we must have a zero remainder. By the way, the algorithm has a neater appearance if we take $r_0 = b$ and $r_{-1} = a$.

Euclid proved that $(a, b) = r_{s-1}$. This is easy to see. First of all we know that $(a, b)|a$ and $(a, b)|b$. Thus from the first line we have $(a, b)|r_1$. Repeating this argument we get that successively $(a, b)|r_j$ for $j = 2, 3, \dots, s-1$. On the other hand, starting at the bottom line $r_{s-1}|r_{s-2}$, $r_{s-1}|r_{s-3}$ and so on until we have $r_{s-1}|b$ and $r_{s-1}|a$. Recall that this means

that $r_{s-1}|(a,b)$. Thus we have just proved that

$$r_{s-1}|(a,b), \quad (a,b)|r_{s-1}$$

and so $r_{s-1} = (a,b)$.

Example 2.1. *Let* $a = 10678$, $b = 42$

$$\begin{aligned} 10678 &= 42 \times 254 + 10 \\ 42 &= 10 \times 4 + 2 \\ 10 &= 2 \times 5. \end{aligned}$$

Thus $(10678, 42) = 2$.

But how to compute the x and y in $(a,b) = ax + by$? We could just work backwards through the algorithm using back substitution, but this is tedious and computationally wasteful since it requires all our calculations to be stored. A simpler way is as follows. Define $x_{-1} = 1$, $y_{-1} = 0$, $x_0 = 0$, $y_0 = 1$ and then lay the calculations out as follows.

$$\begin{array}{lll} r_{-1} = r_0 q_1 + r_1, & x_1 = x_{-1} - q_1 x_0, & y_1 = y_{-1} - q_1 y_0 \\ r_0 = r_1 q_2 + r_2, & x_2 = x_0 - q_2 x_1, & y_2 = y_0 - q_2 y_1 \\ r_1 = r_2 q_3 + r_3, & x_3 = x_1 - q_3 x_2, & y_3 = y_1 - q_3 y_2 \\ \vdots & \vdots & \vdots \\ r_{s-3} = r_{s-2} q_{s-1} + r_{s-1}, & x_{s-1} = x_{s-3} - q_{s-1} x_{s-2}, & y_{s-1} = y_{s-3} - q_{s-1} y_{s-2} \\ r_{s-2} = r_{s-1} q_s. & & \end{array}$$

Now the claim is that we have $x = x_{s-1}$, $y = y_{s-1}$. More generally we have

$$r_j = ax_j + by_j \tag{2.2}$$

and again this can be proved by induction. First, by construction we have

$$r_{-1} = ax_{-1} + by_{-1}, \quad r_0 = ax_0 + by_0.$$

Suppose we have established (2.2) for all $j \leq k$. Then

$$\begin{aligned} r_{k+1} &= r_{k-1} - q_{k+1} r_k \\ &= (ax_{k-1} + gy_{k-1}) - q_k(ax_k + by_k) \\ &= ax_{k+1} + by_{k+1}. \end{aligned}$$

In particular

$$(a,b) = r_{s-1} = ax_{s-1} + by_{s-1}.$$

Hence laying out the example above in this expanded form we have

$$r_{-1} = 10678,\ r_0 = 42,\ x_{-1} = 1,\ x_0 = 0,\ y_{-1} = 0,\ y_0 = 1,$$

$$\begin{array}{lll} 10678 = 42 \times 254 + 10, & x_1 = 1 - 254 \times 0 = 1, & y_1 = 0 - 1 \times 254 = -254 \\ 42 = 10 \times 4 + 2, & x_2 = 0 - 4 \times 1 = -4, & y_2 = 1 - 4 \times (-254) = 1017 \\ 10 = 2 \times 5. & & \end{array}$$

$$(10678, 42) = 2 = 10678 \times (-4) + 42 \times (1017).$$

It is also possible to set this up using matrices. Lay out the sequences in rows

$$\begin{array}{lll} r_{-1}, & x_{-1}, & y_{-1} \\ r_0, & x_0, & y_0 \\ \vdots & \vdots & \vdots \end{array}$$

Now proceed to compute each successive row as follows. If the s-th row is the last one to be computed, calculate $q_s = \lfloor r_{s-1}/r_s \rfloor$. Then take the last two rows computed and pre multiply by $(1, -q_s)$

$$(1, -q_s) \begin{pmatrix} r_{s-1}, & x_{s-1}, & y_{s-1} \\ r_s, & x_s, & y_s \end{pmatrix} = (r_{s+1}, x_{s+1}, y_{s+1})$$

to obtain the $s+1$-st row.

Example 2.2. *Let $a = 4343$, $b = 973$. We can lay this out as follows*

$$\begin{array}{rrrr} & 4343 & 1 & 0 \\ 4 & 973 & 0 & 1 \\ 2 & 451 & 1 & -4 \\ 6 & 71 & -2 & 9 \\ 2 & 25 & 13 & -58 \\ 1 & 21 & -28 & 125 \\ 5 & 4 & 41 & -183 \\ & 1 & -233 & 1040 \end{array}$$

Thus $(4343, 973) = 1 = (-233)4343 + (1040)973$.

2.1.1 Linear Diophantine Equations

We can use this to find the complete solution in integers to linear diophantine equations of the kind

$$ax + by = c.$$

Here a, b, c are integers and we wish to find all integers x and y which satisfy this. There are some obvious necessary conditions. First of all if $a = b = 0$, then it is not soluble unless $c = 0$ and then it is soluble by any x and y, which is not very interesting.

Thus it makes sense to suppose that one of a or b is non-zero. Then since (a, b) divides the left hand side, we can only have solutions if $(a, b)|c$. If we choose x and y so that $ax + by = (a, b)$, then we have

$$a(xc/(a, b)) + b(yc/(a, b)) = (ax + by)c/(a, b) = c$$

so we certainly have a solution of our equation. Call it x_0, y_0. Now consider any other solution. Then

$$ax + by - ax_0 - by_0 = c - c = 0.$$

Thus

$$a(x - x_0) = b(y_0 - y).$$

Hence

$$\frac{a}{(a, b)}(x - x_0) = \frac{b}{(a, b)}(y_0 - y).$$

Then since

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

we have by an earlier example that $y_0 - y = z\frac{a}{(a,b)}$ and $x - x_0 = z\frac{b}{(a,b)}$ for some integer z. But any x and y of this form give a solution, so we have found the complete solution set.

Theorem 2.1. *Suppose that a and b are not both 0 and $(a, b)|c$. Suppose further that $ax_0 + by_0 = c$. Then every solution of*

$$ax + by = c$$

is given by

$$x = x_0 + z\frac{b}{(a, b)}, \quad y = y_0 - z\frac{a}{(a, b)}$$

where z is any integer.

One can see here that the solutions x all leave the same remainder on division by $\frac{b}{(a,b)}$ and likewise for y on division by $\frac{a}{(a,b)}$. This suggests that there may be a useful way of classifying integers.

2.1.2 Exercises

1. Find integers x and y such that $182x + 1155y = (182, 1155)$.

2. Find all pairs of integers x and y such that $922x + 2163y = 7$.

3. Find all pairs of integers x and y such that $812x + 2013y = 5$.

4. Find $(1819, 3587)$, and find the complete solution in integers x and y to $1819x+3587y = (1819, 3587)$.

5. Find integers x and y such that $1547x + 2197y = (1547, 2197)$.

6. Find integers m and n so that

$$4709m + 6188n = (4709, 6188).$$

7. Let $n_1, n_2, \ldots, n_s \in \mathbb{Z}$. Define the greatest common divisor d of $n_1, n_2, \ldots, n_s$ and prove that there exist integers $m_1, m_2, \ldots, m_s$ such that $n_1m_1 + n_2m_2 + \cdots + n_sm_s = d$.

8. Discuss the solubility of $a_1x_1 + a_2x_2 + \cdots + a_sx_s = c$ in integers.

2.2 Notes

The equation (2.1) is called Bézout's identity, and is in É. Bézout (1779), Théorie générale des équations algébriques, Paris, Ph.-D. Pierres. Euclid's algorithm is in Book VII, Propositions 1 and 2.

Chapter 3

Congruences and Residue Classes

3.1 Residue Classes

We now introduce a topic that was first developed by Gauss. We will hear a lot about Gauss in this course.

Definition 3.1. *Let $m \in \mathbb{N}$ and define the residue class $\overline{r}$ modulo m by*

$$\overline{r} = \{x \in \mathbb{Z} : m|(x - r)\}.$$

By the division algorithm every integer is in one of the residue classes

$$\overline{0}, \overline{1}, \ldots, \overline{m-1}.$$

This is often called a complete system of residues modulo m.

The remarkable thing is that we can perform arithmetic on the residue classes just as if they were numbers.

The residue class $\overline{0}$ behaves like the number 0. The reason is that $\overline{0}$ just consists of the integral multiples of m and adding any one of them to an element of the residue class $\overline{r}$ does not change the remainder. Thus for any r

$$\overline{0} + \overline{r} = \overline{r} = \overline{r} + \overline{0}.$$

Suppose that we are given any two residue classes $\overline{r}$ and $\overline{s}$ modulo m. Let t be the remainder of $r+s$ on division by m. Then every element of $\overline{r}$ and $\overline{s}$ are of the form $r+mx$ and $s+mx$ and we know that $r+s = t+mz$ for some z. Thus $r+mx+s+my = t+m(z+x+y)$ is in $\overline{t}$, and it is readily seen that the converse is true. Thus it makes sense to write $\overline{r}+\overline{s} = \overline{t}$, and then we have $\overline{r}+\overline{s} = \overline{s}+\overline{r}$.

One can also check that

$$\overline{r} + \overline{-r} = \overline{0}.$$

In connection with this there is a notation that was introduced by Gauss.

Definition 3.2. *Let $m \in \mathbb{N}$. If two integers x and y satisfy $m|x-y$, then we write*

$$x \equiv y \pmod{m}$$

and we say that x is congruent to y modulo m.

Here are some of the properties of congruences.

$$x \equiv x \pmod{m},$$

$$x \equiv y \pmod{m} \text{ iff } y \equiv x \pmod{m},$$

$$x \equiv y \pmod{m}, y \equiv z \pmod{m} \text{ implies } x \equiv z \pmod{m}.$$

These say that the relationship $\equiv$ is reflexive, symmetric and transitive. Thus congruences modulo m partition the integers into equivalence classes. I leave their proofs as an exercise.

One can also check the following

If $x \equiv y \pmod{m}$ and $z \equiv t \pmod{m}$, then $x + z \equiv y + t \pmod{m}$ and $xz \equiv yt \pmod{m}$.

If $x \equiv y \pmod{m}$, then for any $n \in \mathbb{N}$, $x^n \equiv y^n \pmod{m}$ (use induction on n).

If f is a polynomial with integer coefficients, and $x \equiv y \pmod{m}$, then $f(x) \equiv f(y) \pmod{m}$.

Wait a minute, this means that one can use congruences just like doing arithmetic on the integers!

Here is a very useful result that begins to tell us something about the structure that we have just created.

Theorem 3.1. *Suppose that $m \in \mathbb{N}$, $k \in \mathbb{Z}$, $(k, m) = 1$ and*

$$\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_m$$

forms a complete set of residues modulo m. Then so does

$$\overline{ka_1}, \overline{ka_2}, \ldots, \overline{ka_m}.$$

Proof. Since we have m residue classes, we need only check that they are disjoint. Consider any two of them, $\overline{ka_i}$ and $\overline{ka_j}$. Let $ka_i + mx$ and $ka_j + my$ be typical members of each class. If they were the same integer, than $ka_i + mx = ka_j + my$, so that $k(a_i - a_j) = m(y - x)$. But then $m|k(a_i - a_j)$ and since $(k, m) = 1$ we would have $m|a_i - a_j$ so $\overline{a}_i$ and $\overline{a}_j$ would be identical residue classes, which would contradict them being part of a complete system. □

An important rôle is played by the residue classes r modulo m with $(r, m) = 1$. In connection with this we introduce an important arithmetical function ϕ, called Euler's function.

Definition 3.3. *A real or complex valued function defined on* $\mathbb{N}$ *is called an arithmetical function.*

Definition 3.4. *Euler's function* $\phi(n)$ *is defined to be the number of* $x \in \mathbb{N}$ *with* $1 \leq x \leq n$ *and* $(x, n) = 1$.

Example 3.1. *Since* $(1, 1) = 1$ *we have* $\phi(1) = 1$.

If p *is prime, then the* x *with* $1 \leq x \leq p - 1$ *satisfy* $(x, p) = 1$, *but* $(p, p) = p \neq 1$. *Hence* $\phi(p) = p - 1$.

The numbers x *with* $1 \leq x \leq 30$ *and* $(x, 30) = 1$ *are*

$$1, 7, 11, 13, 17, 19, 23, 29,$$

so $\phi(3) = 8$.

Definition 3.5. *A set of* $\phi(m)$ *distinct residue classes* $\overline{r}$ *modulo* m *with* $(r, m) = 1$ *is called a reduced set of residues modulo* m.

One way of thinking about this is to start from a complete set of fractions with denominator m in the interval $(0, 1]$

$$\frac{1}{m}, \frac{2}{m}, \ldots, \frac{m}{m}.$$

Now remove just the ones whose numerator has a common factor $d > 1$ with m. What is left are the $\phi(m)$ *reduced* fractions with denominator m.

Suppose instead of removing the non reduced ones we just write them in their lowest form. Then for each divisor k of m we obtain all the reduced fractions with denominator k, In fact we just proved the following.

Theorem 3.2. *For each* $m \in \mathbb{N}$ *we have*

$$\sum_{k|m} \phi(k) = m.$$

Example 3.2. *We have* $\phi(1) = 1$, $\phi(2) = 1$, $\phi(3) = 2$, $\phi(5) = 4$, $\phi(6) = 2$, $\phi(10) = 4$, $\phi(15) = 8$, $\phi(30) = 8$ *and*

$$\phi(1) + \phi(2) + \phi(3) + \phi(5) + \phi(6) + \phi(10) + \phi(15) + \phi(30) = 30.$$

Now we can prove a companion theorem to Theorem 3.1 for reduced residue classes.

Theorem 3.3. *Suppose that* $(k, m) = 1$ *and that*

$$a_1, a_2, \ldots, a_{\phi(m)}$$

forms a set of reduced residue classes modulo m. *Then*

$$ka_1, ka_2, \ldots, ka_{\phi(m)}$$

also forms a reduced set of residues modulo m.

Proof. In view of the earlier theorem the residue classes ka_j are distinct, and since $(a_j, m) = 1$ we have $(ka_j, m) = 1$ so they give $\phi(m)$ distinct reduced residue classes, so they are all of them in some order. □

We can now begin to examine the structure of complete and reduced systems of residue classes.

Theorem 3.4. *Suppose that m, $n \in \mathbb{N}$ and $(m, n) = 1$ and consider the mn numbers*

$$xn + ym.$$

with $1 \leq x \leq m$ and $1 \leq y \leq n$. Then they form a complete set of residues modulu mn. If instead x and y are further restricted to $(x, m) = 1$ and $(y, n) = 1$, then they form a reduced set of residues modulo m.

Proof. In the unrestricted case we have mn objects. Moreover if $xn + ym \equiv x'n + y'm \pmod{mn}$ then we would have $xn \equiv x'n \pmod{m}$, so that $x \equiv x' \pmod{m}$ and thus $x = x'$, and likewise $y = y'$. Thus we have mn distinct residues modulo mn and so a complete set. In the restricted case the same argument shows that the $xn + ym$ are distinct modulo mn. Moreover $(xn + ym, m) = (xn, m) = (x, m) = 1$ and likewise $(xn + ym, n) = 1$, so $(xn + ym, mn) = 1$ and the $xn + ym$ all belong to reduced residue classes. Now let z be an arbitrary reduced residue modulo mn. Choose x' and y' so that $nx' + my' = 1$ and choose $x \in \overline{x'z}$ modulo m and $y \in \overline{y'z}$ modulo n. Then one can check that $xn + yn \equiv x'zn + y'zm = z \pmod{mn}$ and hence every reduced residue class modulo mn is in some $xn + ym$. □

Example 3.3. *Here is a table of $xn + ym \pmod{mn}$ when $m = 5$, $n = 6$.*

x / y	1	2	3	4	5
1	11	17	23	29	5
2	16	22	28	4	10
3	21	27	3	9	15
4	26	2	8	14	20
5	1	7	13	19	25
6	6	12	18	24	30

The 30 numbers 1 through 30 appear exactly once each. The 8 reduced residue classes occur precisely in rows 1 and 5.

Immediate from Theorem 3.4 we have

Corollary 3.5. *If $(m, n) = 1$, then $\phi(mn) = \phi(m)\phi(n)$.*

Definition 3.6. *If an arithmetical function f which is not identically 0 satisfies*

$$f(mn) = f(m)f(n)$$

whenever $(m, n) = 1$ we say that f is multiplicative.

Corollary 3.6. *Euler's function is multiplicative.*

This enables a full evaluation of $\phi(n)$. If $n = p^k$, then the number of reduced residue classes modulo p^k is simply the number of x with $1 \leq x \leq p^k$ and $p \nmid x$. This is $p^k - N$ where N is the number of x with $1 \leq x \leq p^k$ and $p|x$, and $N = p^{k-1}$. Thus $\phi(p^k) = p^k - p^{k-1} = p^k(1 - 1/p)$. Putting this all together gives

Theorem 3.7. *Let $n \in \mathbb{N}$. Then*

$$\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$$

where, when $n = 1$ we interpret the product as an "empty" produce 1.

Example 3.4. *We have $\phi(9) = 6$, $\phi(5) = 4$, $\phi(45) = 24$. Note that $\phi(3) = 2$ and $\phi(9) \neq \phi(3)^2$.*

Here is a beautiful and as we shall see, useful, theorem.

Theorem 3.8 (Euler)**.** *Suppose that $m \in \mathbb{N}$ and $a \in \mathbb{Z}$ with $(a, m) = 1$. Then*

$$a^{\phi(m)} \equiv 1 \pmod{m}.$$

Proof. Let

$$a_1, a_2, \ldots, a_{\phi(m)}$$

be a reduced set of residues modulo m. Then

$$aa_1, aa_2, \ldots, aa_{\phi(m)}$$

is another. Hence

$$\begin{aligned} a_1 a_2 \ldots a_{\phi(m)} &\equiv aa_1 aa_2 \ldots aa_{\phi(m)} \pmod{m} \\ &\equiv a_1 a_2 \ldots a_{\phi(m)} a^{\phi(m)} \pmod{m}. \end{aligned}$$

Since $(a_1 a_2 \ldots a_{\phi(m)}, m) = 1$ we may cancel the

$$a_1 a_2 \ldots a_{\phi(m)}.$$

□

Corollary 3.9 (Fermat)**.** *Let p be a prime number and a an integer. Then*

$$a^p \equiv a \pmod{p}.$$

3.1.1 Exercises

Euler's function, congruences

1. Prove that if m, $n \in \mathbb{N}$ and $(m, n) = 1$, then $m^{\phi(n)} + n^{\phi(m)} \equiv 1 \pmod{mn}$.

2. For which values of $n \in \mathbb{N}$ is $\phi(n)$ odd?

3. Given that n is a product of two primes p and q with $p \leq q$, prove that

$$p = \frac{n + 1 - \phi(n) - \sqrt{(n + 1 - \phi(n))^2 - 4n}}{2}.$$

If you have a good calculator use this to factorise n where $n = 19749361535894833$ and $\phi(n) = 19749361232517120$.

3. Find all n such that $\phi(n) = 12$.

4. Show that if $f(x)$ is a polynomial with integer coefficients and if $f(a) \equiv k \pmod{m}$, then $f(a + tm) \equiv k \pmod{m}$ for every integer t.

5. Prove that for any integer n
   (i) $n^7 - n$ is divisible by 42,
   (ii) $n^{13} - n$ is divisible by 2730.

6. Prove that if m is an odd positive integer, then the sum of any complete set of residues modulo m is $0 \pmod{m}$. If m is any integer with $m > 2$, then prove the analogous result for any reduced system of residues modulo m.

7. The numbers $F_n = 2^{2^n} + 1$ are called Fermat numbers. F_0 through F_4 are prime. Fermat had conjectured that F_n is always prime.
   (i) Show that $641 | F_5$.

We now know that $F_5, \ldots, F_{19}$ are composite and it is now conjectured that there are no further Fermat primes!

Suppose that p is a prime with $p | F_n$ and let e denote the smallest positive integer such that $2^e \equiv 1 \pmod{p}$.
   (ii) Show that e exists and $e | 2^{n+1}$.
   (iii) Show that $e \nmid 2^n$.
   (iv) Show that $p \equiv 1 \pmod{2^{n+1}}$.

8. Prove that (i) if $(a, m) = (a - 1, m) = 1$, then

$$1 + a + a^2 + \cdots + a^{\phi(m)-1} \equiv 0 \pmod{m},$$

and
   (ii) prove that every prime other than 2 or 5 divides infinitely many of the integers 1, 11, 111, 1111, ...

9. Prove that if p is prime, and $a, b \in \mathbb{Z}$, then

$$(a + b)^p \equiv a^p + b^p \pmod{p}.$$

3.2 Linear congruences

Just as linear equations are the easiest to solve, so one might expect that linear congruences

$$ax \equiv b \pmod{m}$$

are the easiest to solve. In fact we have already solved this in principle since it is equivalent to the linear diophantine equation

$$ax + my = b.$$

Theorem 3.10. *The congruence*

$$ax \equiv b \pmod{m}$$

is soluble if and only if $(a, m)|b$, and then the general solution is given by the members of a residue class x_0 modulo $m/(a, m)$. The residue class x_0 can be found by applying Euclid's algorithm to solve $ax_0 + my_0 = b$.

Proof. The congruence is equivalent to the equation $ax + my = b$ and there can be no solution if $(a, m) \nmid b$. We know from Euclid's algorithm that if $(a, m)|b$, then

$$\frac{a}{(a, m)}x + \frac{m}{(a, m)}y = \frac{b}{(a, m)}$$

is soluble. Let x_0, y_0 be such a solution. Obviously every member of the residue class x_0 modulo $m/(a, m)$ gives a solution. Let x, y be another solution. Then

$$\frac{a}{(a, m)}(x - x_0) \equiv 0 \pmod{\frac{m}{(a, m)}}$$

and since

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

it follows that x is in the residue class x_0 modulo $m/(a, m)$. □

What about simultaneous linear congruences?

$$\begin{cases} a_1x & \equiv b_1 \pmod{q_1}, \\ \dots & \dots \\ a_rx & \equiv b_r \pmod{q_r}. \end{cases} \tag{3.1}$$

There can only be a solution when each individual equation is soluble, so we require $(a_j, q_j)|b_j$ for every j. Then we know that each individual equation is soluble for all the

members of some residue class modulo $q_j/(a_j, q_j)$. Thus the above system reduces to a collection of simultaneous congruences

$$\begin{cases} x & \equiv c_1 \pmod{m_1}, \\ \dots & \dots \\ x & \equiv c_r \pmod{m_r} \end{cases} \tag{3.2}$$

for some values of c_j and m_j. Now suppose that for some i and $j \neq i$ we have $(m_i, m_j) = d > 1$. Then x has to satisfy $c_i \equiv x \equiv c_j \pmod{d}$. This imposes further conditions on c_j which can get very complicated. Thus it is convenient to suppose that $(m_i, m_j) = 1$ when $i \neq j$, and in fact every system can, with some work, be reduced to this case.

Theorem 3.11. *Suppose that $(m_i, m_j) = 1$ for every $i \neq j$. Then the system (3.2) has as its complete solution precisely the members of a unique residue class modulo $m_1 m_2 \dots m_r$.*

Proof. We first show that there is a solution. Let $M = m_1 m_2 \dots m_r$ and $M_j = M/m_j$, so that $(M_j, m_j) = 1$. We know that there is an N_j so that $M_j N_j \equiv c_j \pmod{m_j}$ (solve $yM_j \equiv c_j \pmod{m_j}$ in y). Let x be any member of the residue class

$$N_1 M_1 + \dots + N_r M_r \pmod{M}.$$

Then for every j, since $m_j | M_i$ when $i \neq j$ we have

$$\begin{aligned} x &\equiv N_j M_j \pmod{m_j} \\ &\equiv c_j \pmod{m_j} \end{aligned}$$

so the residue class $x \pmod{M}$ gives a solution.

Now we have to show that this is unique. Suppose y is also a solution of the system. Then for every j we have

$$\begin{aligned} y &\equiv c_j \pmod{m_j} \\ &\equiv x \pmod{m_j} \end{aligned}$$

and so $m_j | y - x$. Since the m_j are pairwise co-prime we have $M | y - x$, so y is in the residue class x modulo M. □

Example 3.5. *Consider the system of congruences*

$$\begin{aligned} x &\equiv 3 \pmod{4}, \\ x &\equiv 5 \pmod{21}, \\ x &\equiv 7 \pmod{25}. \end{aligned}$$

We have $m_1 = 4$, $m_2 = 21$, $m_3 = 25$, $M = 2100$, $M_1 = 525$, $M_2 = 100$, $M_3 = 84$. First we have to solve

$$\begin{aligned} 525 N_1 &\equiv 3 \pmod{4}, \\ 100 N_2 &\equiv 5 \pmod{21}, \\ 84 N_3 &\equiv 7 \pmod{25}. \end{aligned}$$

Reducing the constants gives

$$\begin{aligned} N_1 &\equiv 3 \pmod 4, \\ (-5)N_2 &\equiv 5 \pmod{21}, \\ 9N_3 &\equiv 7 \pmod{25}. \end{aligned}$$

Thus we can take $N_1 = 3$, $N_2 = 20$, $7 \equiv -18 \pmod{25}$ *so* $N_3 \equiv -2 \equiv 23 \pmod{25}$. *Thus the complete solution is given by*

$$\begin{aligned} x &\equiv N_1M_1 + N_2M_2 + N_3M_3 \\ &= 3 \times 525 + 20 \times 100 + 23 \times 84 \\ &= 5507 \\ &\equiv 1307 \pmod{2100}. \end{aligned}$$

3.2.1 Exercises

1. Solve where possible.
   (i) $91x \equiv 84 \pmod{143}$
   (ii) $91x \equiv 84 \pmod{147}$

2. Suppose that $m_1, m_2 \in \mathbb{N}$, $(m_1, m_2) = 1$, $a, b \in \mathbb{Z}$. Prove that $a \equiv b \pmod{m_1}$ and $a \equiv b \pmod{m_2}$ if and only if $a \equiv b \pmod{m_1m_2}$.

3. Solve $11x \equiv 21 \pmod{105}$.

4. Prove that when a natural number is written in the usual decimal notation, (i) it is divisible by 3 if and only if the sum if its digits is divisible by 3 and (ii) it is divisible by 9 if and only if the sum if its digits is divisible by 9.

5. Show that the last decimal digit of a perfect square cannot be $2, 3, 7$ or 8.

6. Prove that, for any integer a, $6|a(a+1)(2a+1)$.

7. Solve the simultaneous congruences

$$\begin{aligned} x &\equiv 4 \pmod{19} \\ x &\equiv 5 \pmod{31} \end{aligned}$$

8. Solve the simultaneous congruences

$$\begin{aligned} x &\equiv 6 \pmod{17} \\ x &\equiv 7 \pmod{23} \end{aligned}$$

9. Solve the simultaneous congruences

$$\begin{aligned} x &\equiv 3 \pmod 6 \\ x &\equiv 5 \pmod{35} \\ x &\equiv 7 \pmod{143} \\ x &\equiv 11 \pmod{323} \end{aligned}$$

10. Eggs in basket problem (India 7c.). Find the smallest number of eggs such that when egs are removed 2, 3, 4, 5 or 6 at a time 1 remains, but when eggs are removed 7 at a time none remain.

11. The numbers $F_n = 2^{2^n} + 1$ are called Fermat numbers. F_0 through F_4 are prime. Fermat had conjectured that F_n is always prime. Show that $641|F_5$. We now know that $F_5, \ldots, F_{19}$ are composite and it is now conjectured that there are no further Fermat primes!

Suppose that p is a prime with $p|F_n$. Let e denote the smallest positive integer such that $2^e \equiv 1 \pmod p$.

(i) Show that e exists and $e|2^{n+1}$.

(ii) Show that $e \nmid 2^n$.

(iii) Show that $p \equiv 1 \pmod{2^{n+1}}$.

12. Show that every integer satisfies at least one of the following congruences; $x \equiv 0 \pmod 2$, $x \equiv 0 \pmod 3$, $x \equiv 1 \pmod 4$, $x \equiv 1 \pmod 6$, $x \equiv 11 \pmod{12}$. Such a collection of congruences (with the moduli all different) is known as a covering class. Paul Erdős asked whether there are covering classes with all the moduli arbitrarily large. For a long time it was an open question. Eventually Bob Hough showed that there are none.

13. Prove that any fourth power must have one of 0, 1, 5, 6 for its unit digit.

14. Show that $61! + 1 \equiv 63! + 1 \equiv 0 \pmod{71}$.

15. Let $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$ be a sequence of n integers (not necessarily distinct). Show that some non-empty subsequence of of $\mathcal{A}$ has a sum which is divisible by n.

16. Let a, b, and x_0 be positive integers and define x_n iteratively for $n \geq 1$ by $x_n = ax_{n-1} + b$. Prove that not all the x_n are prime.

3.3 Non-linear polynomial congruences

The solution of a general polynomial congruence can be quite tricky, even for a polynomial with a single variable

$$f(x) := a_0 + a_1 x + \cdots + a_j x^j + \cdots a_J x^J \equiv 0 \pmod m \tag{3.3}$$

where the a_j are integers. The largest k such that $a_k \not\equiv 0 \pmod{m}$ is the degree of f modulo m. If $a_j \equiv 0 \pmod{p}$ for every j, then the degree of f modulo m is not defined, and so does not exist.

We have already seen that

$$x^2 \equiv 1 \pmod{8}$$

is solved by any odd x, so that it has four solutions modulo 8, $x \equiv 1, 3, 5, 7 \pmod{8}$. That is, more than the degree 2. However, when the modulus is prime we have the more familiar conclusion.

When we have a solution x to a polynomial congruence such as (3.3) we may sometimes refer to such values as a *root* of the polynomial modulo m.

Theorem 3.12 (Lagrange)**.** *Suppose that p is prime, and $f(x) = a_0 + a_1x + \cdots + a_jx^j + \cdots$ is a polynomial with integer coefficients a_j and it has degree k modulo p. Then the number of incongruent solutions of*

$$f(x) \equiv 0 \pmod{p}$$

is at most k.

Proof. The case of degree 0 is obvious. Thus we can suppose $k \geq 1$. We use induction on the degree k. If a polynomial f has degree 1 modulo p, so that $f(x) = a_0 + a_1x$ with $p \nmid a_1$, then the congruence becomes

$$a_1x \equiv -a_0 \pmod{p}$$

and since $a_1 \not\equiv 0 \pmod{p}$ (because f has degree 1) we know that this is soluble by precisely the members of a unique residue class modulo p.

Now suppose that the conclusion holds for all polynomials of a given degree k and suppose that f has degree $k+1$. If

$$f(x) \equiv 0 \pmod{p}$$

has no solutions, then we are done. Hence we may suppose it has (at least) one, say $x \equiv x_0 \pmod{p}$. By the division algorithm for polynomials we have

$$f(x) = (x - x_0)q(x) + f(x_0)$$

where $q(x)$ is a polynomial of degree k with integer coefficients. [To see this observe first that $x^j - x_0^j = (x - x_0)(x^{j-1} + x^{j-2}x_0 + \cdots x_0^{j-1})$ and so collecting together the terms we get $f(x) - f(x_0) = (x - x_0)q(x)$. Moreover the leading coefficient of $q(x)$ is $a_k \not\equiv 0 \pmod{p}$]. But $f(x_0) \equiv 0 \pmod{p}$, so that

$$f(x) \equiv (x - x_0)q(x) \pmod{p}$$

If $f(x_1) \equiv 0 \pmod{p}$, with $x \not\equiv x_0 \pmod{p}$, then $p \nmid x_1 - x_0$ so that $p | q(x_1)$. [Note that if the modulus is not prime we cannot make this deduction; $m_1m_2 | ab$ could hold because $m_1 | a$ and $m_2 | b$]. By the inductive hypothesis there are at most k possibilities for x_1, so at most $k+1$ in all. □

A curious, but sometimes useful, application of the above is the following

Theorem 3.13 (Wilson)**.** *Let p be a prime number, then $(p-1)! \equiv -1 \pmod{p}$.*

Proof. The case $p = 2$ is $(2-1)! = 1 \equiv -1 \pmod{2}$. Thus we may suppose that $p \geq 3$. Consider the polynomial

$$x(x-1)(x-2)\dots(x-p+1) - x^p + x.$$

We show that all the coefficients of p are divisible by p, for then the coefficient of x is $(-1)^{p-1}(p-1)! + 1$ and the result follows when $p \geq 3$. Suppose on the contrary that the degree of f modulo p exists. Obviously $x(x-1)(x-2)\dots(x-p+1) \equiv 0 \pmod{p}$ for every x, and by Fermat's little theorem $x^p - x \equiv 0 \pmod{p}$ for every x. Hence $f(x) \equiv 0 \pmod{p}$ for every x. But when one multiplies out the product the leading term is x^p which is cancelled out by the $-x^p$. Thus f has degree at most $p-1$ but has p roots modulo p, which contradicts Lagrange's theorem. □

It is useful at this stage to consider generally the number of solutions of a polynomial congruence.

Definition 3.7. *Suppose that f is a polynomial with integer coefficients. Given a modulus $m \in \mathbb{N}$, we define the $N_f(m)$ to be the number of different residue classes x modulo m such that $f(x) \equiv 0 \pmod{m}$.*

For example when $f(x) = x^2 - 1$ we have $N_f(8) = 4$, and for an odd prime p, $N_f(p) = 2$, but $N_f(2) = 1$. If $g(x) = x^2 + 5$, then $N_g(2) = 1$, $N_g(3) = 2$, $N_g(5) = 1$, $N_g(7) = 2$, $N_g(11) = 0$, $N_g(21) = 4$. Is there a general formula here? The answer is yes, but we don't yet have the tools to decide this. To get the last example you could compute all 21 values modulo 21, but it is easier to use the following.

Theorem 3.14. *Suppose that f is a polynomial with integer coefficients. The $N_f(m)$ is a multiplicative function of m.*

Note that in the first case above $N_f(8) \neq N_f(2)^3$.

Proof. Suppose that $(m_1, m_2) = 1$. Choose n_j so that $n_2 m_2 \equiv 1 \pmod{m_1}$ and $n_1 m_1 \equiv 1 \pmod{m_2}$. Suppose that x_1, x_2 are such that $f(x_j) \equiv 0 \pmod{m_j}$. Let

$$x \equiv x_1 n_2 m_2 + x_2 n_1 m_1 \pmod{m_1 m_2}.$$

Then

$$x \equiv x_1 n_2 m_2 \equiv x_1 \pmod{m_1}$$

and

$$f(x) \equiv f(x_1) \equiv 0 \pmod{m_1}.$$

Likewise $f(x) \equiv 0 \pmod{m_2}$. Hence $f(x) \equiv 0 \pmod{m_1 m_2}$. Moreover the x are distinct modulo $m_1 m_2$. Thus we have constructed $N_f(m_1)N_f(m_2)$ solutions to the latter congruence, so that $N_f(m_1)N_f(m_2) \leq N_f(m_1 m_2)$.

On the other hand, if we have $f(x) \equiv 0 \pmod{m_1 m_2}$, then we can choose x_1, x_2 uniquely modulo m_1 and m_2 respectively so that $x_1 n_2 m_2 \equiv x \pmod{m_1}$ and $x_2 n_1 m_1 \equiv x \pmod{m_2}$, and then $x \equiv x_1 n_2 m_2 + x_2 n_1 m_1 \pmod{m_1 m_2}$. Hence

$$f(x_1) \equiv f(x_1 n_2 m_2 + x_2 n_1 m_1) \equiv 0 \pmod{m_1}$$

and likewise $f(x_2) \equiv 0 \pmod{m_2}$. Thus $N_f(m_1 m_2) \leq N_f(m_1) n_f(m_2)$. □

In view of the multiplicative of the structure of the roots of a polynomial congruence it suffices to concentrate on the case when m is a prime power. It turns out that the really hard case is when the modulus is prime. If we can deal with that, then the case of higher powers of primes becomes more amenable. Incredibly we can imitate Newton's method from calculus. This gives a possible method of lifting from solutions modulo p to solutions modulo higher powers of p. Note that if we have a solution to

$$f(x) \equiv 0 \pmod{p^{t+1}}, \tag{3.4}$$

then it must also be a solution to

$$f(x) \equiv 0 \pmod{p^t}. \tag{3.5}$$

Theorem 3.15 (Hensel's Lemma)**.** *Suppose that f is a polynomial with integer coefficients and there is an x_1 such that $f(x_1) \equiv 0 \pmod{p^t}$. There are three cases.*

(i) If $p | f'(x_1)$ but $p^{t+1} \nmid f(x_1)$, then there is no solution x to (3.4) with $x \equiv x_1 \pmod{p^t}$.

(ii) If $p | f'(x_1)$ and $p^{t+1} | f(x_1)$, then there are p solutions x_2 to (3.4) with $x_2 \equiv x_1 \pmod{p^t}$, given by taking all possible such x_2.

(iii) If $p \nmid f'(x_1)$, then there is a unique solution x_2 to (3.4) with $x_2 \equiv x_1 \pmod{p^t}$ given by

$$x_2 \equiv x_1 + p^t j \pmod{p^{t+1}}, \quad j f'(x_1) \equiv -f(x_1) p^{-t} \pmod{p}$$

Proof. We use the Taylor expansion of f about x_1. We have

$$f(x_1 + h) = f(x_1) + h f'(x_1) + h^2 \frac{f''(x_1)}{2} + \cdots + h^j \frac{f^{(j)}(x_1)}{j!} + \cdots$$

Since f is a polynomial there are only a finite number of terms and each of the coefficients $\frac{f^{(j)}(x_1)}{j!}$ is an integer. Now put $h = p^t j$ where j is at our disposal. All the terms except the first two are divisible by p^{2t} and $2t \geq t + 1$. Thus

$$f(x_1 + p^t j) \equiv f(x_1) + p^t j f'(x_1) \pmod{p^{t+1}}.$$

The first case is clear; when $p|f'(x_1)$ but $p^{t+1} \nmid f(x_1)$, then there can be no solution. Also in the second case, $p|f'(x_1)$ and p^{t+1} then there is a solution for every choice of j, so for every x_2 modulo p^{t+1} with $x_2 \equiv x_1 \pmod{p^t}$. Finally in the third case there is exactly one solution j modulo p so that

$$jf'(x_1) \equiv -f(x_1)p^{-t} \pmod{p}$$

and so there is a unique $x_2 \equiv x_1 + p^t j \pmod{p^{t+1}}$ with $f(x_2) \equiv 0 \pmod{p^{t+1}}$.

If we think of this as saying

$$x_1 + p^t j \text{“} = \text{”} x_1 - \frac{f(x_1)}{f'(j)}$$

then we can see this exactly imitates Newton's method for finding roots. □

Example 3.6. *Find all roots of* $x^2 - 2 \equiv 0 \pmod{7^r}$ *with* $1 \leq r \leq 3$.

3 *and* 4 *are solutions modulo* 7.

(i) $x_1 = 3$, $f(x) = x^2 - 2$, $f'(x) = 2x$, $f(3) = 7$, $f'(3) = 6 \not\equiv 0 \pmod{7}$, *so* 3 *lifts to a unique solution modulo* 7^2. $6j = jf'(3) \equiv -f(3)/7 \equiv -1 \pmod{7}$, $j = 1$, $x_1 + 7j = 3 + 7 = 10$, *so* $x_2 \equiv 10 \pmod{7^2}$. *(ii)* $f(10) = 98 = 2 \times 7^2$, $f'(10) = 20 \not\equiv 0 \pmod{7}$, *so* 10 *lifts to a unique solution modulo* 7^3. $20j = jf'(10) \equiv -f(10)/(7^2) = -2 \pmod{7}$, $j \equiv 2 \pmod{7}$, $x_3 = 10 + 2 \times 7^2 = 108$. $f(108) = 11662 \equiv 0 \pmod{7^3}$. *(iii)* $x_1 = 4$, $f(4) = 14$, $f'(4) = 8 \not\equiv 0 \pmod{7}$. $8j = jf'(4) \equiv -f(4)/7 = -2 \pmod{7}$, $j = 5$, $x_2 = x_1 + 7j = 39 \pmod{7^2}$, $f(39) \equiv 0 \pmod{7^2}$. *(iv)* $x_2 = 39$, $f(39) = 1519$, $f'(39) = 78 \equiv 1 \pmod{7}$, $j \equiv jf'(39) \equiv -f(39)/(7^2) = -31 \equiv 4 \pmod{7}$. $x_3 = x_2 + 7^2 j = 39 + 196 = 235 \pmod{7^3}$. $f(235) = 55223 = 161 \times 7^3$.

Example 3.7. *Find all solutions of* $x^3 - 2 \pmod{3^r}$. *By trial, the only solution modulo* 3 *is* $x_1 = 2$. $f(x) = x^3 - 2$, $f'(x) = 3x^2$. *Thus* $f'(2) \equiv 0 \pmod{3}$ *and* $f(2) = 6$. *But* $3^2 \nmid f(2)$ *so we are in case (i) so there is no solution modulo* 3^2 *and hence none modulo* 3^r *with* $r \geq 2$.

3.3.1 Exercises

1. Let p denote a prime number.

(ii) Let

$$f(x) = \prod_{i=1}^{p-1}(x - i) = x^{p-1} + \sum_{i=0}^{p-2} a_i x^i.$$

Show that if $i = 1, 2, \ldots, p-2$, then $p|a_i$.

(iii) Suppose that $p > 3$. When $(a, p) = 1$, a^* denotes a solution of $ax \equiv 1 \pmod{p^2}$. Show that $1^* + 2^* + \cdots + (p-1)^* \equiv 0 \pmod{p^2}$ (Wolstenholme's congruence).

3. Prove that $3n^2 - 1$ can never be a perfect square.

4. (i) Prove that if $x \in \mathbb{Z}$, then $x^2 \equiv 0$ or $1 \pmod{4}$.

(ii) Prove that $5y^2 + 2 = z^2$ has no solutions with $y, z \in \mathbb{Z}$.

5. (i) Prove that if $x \in \mathbb{Z}$, then $x^3 \equiv 0$ or $\pm 1 \pmod 7$.
(ii) Prove that $y^3 - z^3 = 3$ has no solutions with $y, z \in \mathbb{Z}$.

6. Let $f(x)$ denote a polynomial of degree at least 1 with integer coefficients and positive leading coefficient.
(i) Show that if $f(x_0) = m > 0$, then $f(x) \equiv 0 \pmod m$ whenever $x \equiv x_0 \pmod m$.
(ii) Show that there are infinitely many $x \in \mathbb{N}$ such that $f(x)$ is not prime.

7. (i) Suppose that p is an odd prime and x is an integer with $p|x^2 + 1$. Prove that x has order 4 and $p \equiv 1 \pmod 4$.
(ii) Prove that there are infinitely many primes $p \equiv 1 \pmod 4$.

8. Find all solutions (if there are any) to each of the following congruences
(i) $x^2 \equiv -1 \pmod 7$, (ii) $x^2 \equiv -1 \pmod{13}$, (iii) $x^5 + 4x \equiv 0 \pmod 5$.

9. Prove that the Carmichael function $\lambda(n)$ satisfies $\lambda(n)|\phi(n)$.

10. (i) Solve $x^2 + x + 23 \equiv 0 \pmod 5$.
(ii) Use the Hensel-Newton method to find all solutions to

$$x^2 + x + 23 \equiv 0 \pmod{5^2}.$$

11. (i) Solve $x^3 + 2x^2 + 4 \equiv 0 \pmod 5$.
(ii) Use the Hensel-Newton method to find all solutions to

$$x^3 + 2x^2 + 4 \equiv 0 \pmod{5^r}$$

when $r = 2$, 3.

12. Solve $x^2 + x + 47 \equiv 0 \pmod{7^r}$ when $r = 1$, 2 and 3.

13. Find all solutions to the congruence $x^3 \equiv 27 \pmod{3^r}$ when $r = 4$.

14. (i) Prove that if p is an odd prime and $0 < k < p$, then (assuming $0! = 1$) $(p-k)!(k-1)! \equiv (-1)^k \pmod p$.
(ii) Prove that if $p \equiv 1 \pmod 4$, then the congruence $x^2 + 1 \equiv 0 \pmod p$ is soluble.

15. Let $f(x) = x^3 + x^2 - 5$. Show that for $j = 1, 2, 3, \ldots$ there is a unique $x_j \pmod{7^j}$ such that $f(x_j) \equiv 0 \pmod{7^j}$.

16. For $k = 1$, 2, 3, solve where possible.
(i) $x^3 - 2x + 3 \equiv 0 \pmod{3^k}$.
(ii) $x^3 - 5x^2 + 3 \equiv 0 \pmod{3^k}$.
(iii) $x^3 - 2x + 4 \equiv 0 \pmod{5^k}$.

17. (i) Let $m \in \mathbb{N}$. Prove that

$$(y-1)(y^{m-1} + y^{m-2} + \cdots + y + 1) = y^m - 1.$$

(ii) Let $n \in \mathbb{N}$. Prove that

$$(x^2+1)(x^2-1)(x^{4n-4}+x^{4n-8}+\cdots+x^4+1) = x^{4n}-1.$$

(iii) Let p be a prime number with $p \equiv 1 \pmod 4$. Prove that $x^2 \equiv -1 \pmod p$ has exactly two solutions.

18. Prove that if a has order 3 modulo a prime p, then $1+a+a^2 \equiv 0 \pmod p$, and $1+a$ belongs to the exponent 6.

19. Suppose that $(10a, q) = 1$, and that k is the order of 10 modulo q. Show that the decimal expansion of the rational number a/q is periodic with least period k.

20. Let $n \in \mathbb{Z}$. Prove that if $p|n^2+n+1$ and $p > 3$, then $p \equiv 1 \pmod 6$. Deduce that there are infinitely many primes $p \equiv 1 \pmod 6$.

3.4 Notes

§1 The concept of residue classes and the idea that the residue classes modulo n partition the integers was introduced by Euler about 1750. The notation $\equiv$ was introduced by Gauss in 1801. For a modern translation see C. F. Gauss, Disquisitiones Arithmeticæ, Yale University Press, 1965. Euler introduced the eponymous function in 1763.

The first complete solution of the Chinese Remainder Theorem in the general case occurs in the treatise of Ch'in Chiu-shao of 1247.

§3 Wilson's thereom was first stated by Ibn al-Haytham about 1000AD. The first proof was given by Lagrange in 1771. Hensel proved his lemma in 1897. The proof in the non-singular case is motivated by Newton's method in numerical analysis. The function $\operatorname{ord} am$ has its roots in work of Lagrange. Carmichael introduced his function in R. D. Carmichael (1910), "Note on a new number theory function". Bulletin of the American Mathematical Society. 16 (5), 232–238.

Chapter 4

Primitive Roots

4.1 Primitive Roots

We have seen that on the residue classes modulo m we can perform many of the standard operations of arithmetic. Such an object is called ring. In this case it is usually denoted by $\mathbb{Z}/m\mathbb{Z}$ or $\mathbb{Z}_m$. In this chapter we will look at its multiplicative structure. In particular we will consider the reduced residue classes modulo m. An obvious question is what happens if we take powers of a fixed residue a?

Definition 4.1. *Given $m \in \mathbb{N}$, $a \in \mathbb{Z}$, $(a, m) = 1$ we define the order $\mathrm{ord}_m(a)$ of a modulo m to be the smallest positive integer t such that*

$$a^t \equiv 1 \pmod{m}.$$

Note that by Euler's theorem, $a^{\phi(m)} \equiv 1 \pmod{m}$, so that $\mathrm{ord}_m(a)$ exists. The Carmichael function $\lambda(m)$ is the smallest positive number such that $\mathrm{ord}_a(m)|\lambda(m)$ whenever $(a, m) = 1$. It follows that $\lambda(m)|\phi(m)$.

Theorem 4.1. *Suppose that $m \in \mathbb{N}$, $(a, m) = 1$ and $n \in \mathbb{N}$ is such that $a^n \equiv 1 \pmod{m}$. Then $\mathrm{ord}_m(a)|n$. In particular $\mathrm{ord}_m(a)|\phi(m)$.*

Proof. For concision let $t = \mathrm{ord}_m(a)$. Since t is minimal we have $t \leq n$. Thus by the division algorithm there are q and r with $0 \leq r < t$ such that $n = tq + r$. Hence

$$a^r \equiv (a^t)^q a^r = a^{qt+r} = a^n \equiv 1 \pmod{m}.$$

But $0 \leq r < t$. If we would have $r > 0$, then we would contradict the minimality of t. Hence $r = 0$. □

Here is an application we will make use of later.

Theorem 4.2. *Suppose that $d|p-1$. Then the congruence $x^d \equiv 1 \pmod{p}$ has exactly d solutions.*

Proof. We have

$$x^{p-1} - 1 = (x^d - 1)(x^{p-1-d} + x^{d-p-2d} + \cdots + x^d + 1).$$

To see this just multiply out the right hand side and observe that the terms telescope. We know from Euler's theorem that there are exactly $p-1$ incongruent roots to the left hand side modulo p. On the other hand the second factor has at most $p-1-d$ such roots, so the first factor must account for at least d. On the other hand it has at most d. □

We have already seen that, when $(a,m) = 1$, a has order modulo m which divides $\phi(m)$. One question one can ask is, given any $d|\phi(m)$, are there elements of order d? In the special case $d = \phi(m)$ this would mean that $a, a^2, \ldots, a^{\phi(m)}$ are distinct modulo m, because otherwise we would have $a^u \equiv a^v \pmod{m}$ with $1 \leq u < v \leq \phi(m)$ and then $a^{v-u} \equiv 1 \pmod{m}$ and $1 \leq v - u < \phi(m)$ contradicting the assumption that $\mathrm{ord}_m(a) = \phi(m)$.

Example 4.1. $m = 7$.
$a = 1$, $\mathrm{ord}_7(1) = 1$.
$a = 2$, $2^2 = 4$, $2^3 = 8 \equiv 1$. $\mathrm{ord}_7(2) = 3$.
$a = 3$, $3^2 = 9 \equiv 2$, $3^3 = 27 \equiv 6$, $3^4 \equiv 18 \equiv 4$,
$3^5 \equiv 12 \equiv 5$, $3^6 \equiv 1$, $\mathrm{ord}_7(3) = 6$.
$a = 4$, $4^2 \equiv 2$, $4^3 \equiv 2^6 \equiv 1$, $\mathrm{ord}_7(4) = 3$.
$a = 5$, $5^2 = 25 \equiv 4$, $5^3 \equiv 20 \equiv 6$, $5^4 \equiv 30 \equiv 2$,
$5^5 \equiv 10 \equiv 3$, $5^6 \equiv 1$, $\mathrm{ord}_7(5) = 6$.
$a = 6$, $6^2 = 36 \equiv 1$, $\mathrm{ord}_7(6) = 2$.
Thus there is one element of order 1*, one element of order* 2*, two of order* 3 *and two of order* 6*.*

Is it a fluke that for each $d|6 = \phi(7)$ *the number of elements of order* d *is* $\phi(d)$*?*

Definition 4.2. *Suppose that* $m \in \mathbb{N}$ *and* $(a,m) = 1$*. If* $\mathrm{ord}_m(a) = \phi(m)$ *then we say that* a *is a primitive root modulo* m*.*

We know that we do not always have primitive roots. For example, any number a with $(a,8) = 1$ is odd and so $a^2 \equiv 1 \mod 8$, whereas $\phi(8) = 4$. There are primitive roots to some moduli. For example, modulo 7 the powers of 3 are successively $3, 2, 6, 4, 5, 1$.

Gauss determined precisely which moduli possess primitive roots. The first step is the case of prime modulus.

Theorem 4.3 (Gauss)**.** *Suppose that* p *is a prime number. Let* $d|p-1$ *then there are* $\phi(d)$ *residue classes* a *with* $\mathrm{dlog}_p(a) = d$*. In particular there are* $\phi(p-1) = \phi(\phi(p))$ *primitive roots modulo* p*.*

Proof. We have already seen that the order of every reduced residue class modulo p divides $p-1$. For a given $d|p-1$ let $\psi(d)$ denote the number of reduced residues of order d modulo p. We know that the congruence $x^d \equiv 1 \pmod{p}$ has exactly d solutions. Thus every solution has order dividing d. Moreover every reduced residue which has order dividing d must be a solution. Thus for each $d|p-1$ we have

$$\sum_{r|d} \psi(r) = d.$$

Thus is reminiscent of an earlier formula

$$\sum_{r|d} \phi(r) = d.$$

Let $1 = d_1 < d_2 < \ldots < d_k = p-1$ be the divisors of $p-1$ in order. We have a relationship

$$\sum_{r|d_j} \psi(r) = d_j$$

for each $j = 1, 2, \ldots$ and, of course, the sum is over a subset of the divisors of $p-1$. I claim that this determines $\psi(d_j)$ uniquely. We could prove this by observing that if N is the number of positive divisors of $p-1$, then we have N linear equations in the N unknowns $\psi(r)$ and we can we can write this in matrix notation $\boldsymbol{\psi}\mathcal{U} = \mathbf{d}$. Moreover $\mathcal{U}$ is an upper triangular matrix with non-zero entries on the diagonal and so is invertible. Hence the $\psi(d_j)$ are uniquely determined. But we already know a solution, namely $\psi = \phi$. If we wish to avoid the linear algebra we can prove this by induction on j. For the base case we have $\psi(1) = 1$. Suppose that $\psi(d_1), \ldots, \psi(d_j)$ are determined. Then we have

$$\sum_{r|d_{j+1}} \psi(r) = d_{j+1}.$$

Hence

$$\psi(d_{j+1}) = d_{j+1} - \sum_{\substack{r|d_{j+1} \\ r<d_{j+1}}} \psi(r)$$

and every term on the right hand side is already determined. Thus we can conclude there is only one solution to our system of equations. But we already know one solution, namely $\psi(r) = \phi(r)$. □

Example 4.2. *Here is the proof when $p = 13$, so we are concerned with the divisors of 12.*

$$\big(\psi(1), \psi(2), \psi(3), \psi(4), \psi(6), \psi(12)\big) \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} = (1, 2, 3, 4, 6, 12)$$

How about higher powers of odd primes? We can use the idea of "lifting" which we already saw in connection with solutions of congruences.

Theorem 4.4 (Gauss)**.** *Suppose that p is an odd prime and $d|\phi(p^k) = p^{k-1}(p-1)$. Then there are $\phi(d)$ residue classes modulo p^k which have order d.*

Proof. We first prove the existence of a primitive root modulo p^k when $k > 1$. Let g be a primitive root modulo p. It is clear that a primitive root modulo p^k will also be one modulo p, so it makes sense to examine $g + jp$. We show that there is a j so that

$$(g + jp)^{p-1} = 1 + h_1 p$$

with $p \nmid h_1$. Observe that $g^{p-1} = 1 + lp$ for some l. Then, by the binomial expansion, for every j

$$\begin{aligned}(g + jp)^{p-1} &\equiv g^{p-1} + (p-1)g^{p-2}jp \pmod{p^2} \\ &\equiv 1 + (l - g^{p-2}j)p \pmod{p^2}\end{aligned}$$

and we may choose j so that $p \nmid l - g^{p-2}p$.

Now we show that with this j, for every t there is an h_t such that

$$(g + jp)^{p^{t-1}(p-1)} = 1 + h_t p^t \quad (p \nmid h_t). \tag{4.1}$$

We do this by induction on t. We have already established the base case. Suppose we have already established the result for some t. Then

$$\begin{aligned}(g + jp)^{p^t(p-1)} &= (1 + h_t p^t)^p \\ &\equiv 1 + h_t p^{t+1} + \frac{p(p-1)}{2} h_t^2 p^{2t} \pmod{p^{3t}}.\end{aligned}$$

We have both $2t + 1 \geq t + 2$ and $3t \geq t + 2$. Hence we have

$$(g + jp)^{p^t(p-1)} \equiv 1 + h_t p^{t+1} \pmod{p^{t+2}}$$

and since $p \nmid h_t$ this gives the desired conclusion.

Now consider the number $g+jp$. We show that this is a primitive root modulo p^k, and we may suppose that $k \geq 2$. Let $d = \mathrm{ord}_p^k(g + jp)$. Then $d|\phi(p^k) = p^{k-1}(p-1)$. Hence $d = p^t v$ for some t and v with $0 \leq t \leq k-1$ and $v|p-1$. We have $p^t = (p - 1 + 1)^t \equiv 1 \pmod{p-1}$. Hence

$$\begin{aligned}1 &\equiv (g + jp)^d \equiv (g + jp)^{p^t v} \pmod{p^k} \\ &\equiv (g + jp)^v \pmod{p} \\ &\equiv g^v \pmod{p}\end{aligned}$$

and since g is a primitive root modulo p we have $v = p-1$. Now repeating the argument we have

$$\begin{aligned} 1 &\equiv (g+jp)^d \pmod{p^k} \\ &\equiv (g+jp)^{p^t(p-1)} \pmod{p^k} \\ &= 1 + h_{t+1}p^{t+1} \end{aligned}$$

by (4.1). Since $p \nmid h_{t+1}$ this can only be $\equiv 1 \pmod{p^k}$ if $t = k-1$.

Now suppose that $d|\phi(p^k)$ and g is a primitive root modulo p^k and consider the $\phi(d)$ residue classes

$$g^{b\phi(p^k)/d},$$

modulo p^k with $(b,d) = 1$ and $1 \le b \le d$. Since

$$\left(g^{b\phi(p^k)/d}\right)^d \equiv 1 \pmod{p^k}$$

they have order r dividing d. Moreover g would have order

$$\big(b\phi(p^k)r/d, \phi(p^k)\big) = (br,d)\phi(p^k)/d = \phi(p^k)r/d,$$

and so $r = d$. □

It is easy to see that 1 is a primitive root modulo 2 and 3 is a primitive root modulo 4, and we have already seen that there are no primitive roots modulo 8, and hence there are none modulo higher powers of 2. Thus we are half-way to proving the following theorem.

Theorem 4.5 (Gauss)**.** *We have primitive roots modulo m when $m = 2$, $m = 4$, $m = p^k$ and $m = 2p^k$ with p and odd prime and in no other cases.*

Proof. The one positive case left to settle is $m = 2p^k$. We have $\phi(2p^k) = \phi(p^k)$. Let g be a primitive root modulo p^k and let $G = g$ if g is odd and $G = g + p^k$ if g is even. Then G is odd and a primitive root modulo p^k. Hence, given x with $1 \le x \le 2p^k$ and $(x, 2p^k) = 1$ there is a y so that $G^y \equiv x \pmod{p^k}$ and (regardless of the value of y) $G^y \equiv x \pmod 2$. Hence $G^y \equiv x \pmod{2p^k}$.

It remains to show that for all other m there are no residue classes of order $\phi(m)$. We have already dealt with $m = 2^k$ with $k \ge 3$. Write $m = 2^k p_1^{k_1} \dots p_r^{k_r}$. we can suppose that (i) $k = 0$ or 1 and $r \ge 2$ or (ii) $k \ge 2$ and $r \ge 1$. The key to the proof is that given a with $(a,m) = 1$ the orders of a modulo 2^k, $p_j^{k_j}$ divides $\phi(2^k)$ and $\phi(p_j^{k_j})$ respectively. Thus the order of a modulo m divides the least common multiple of $\phi(2^k), \phi(p_1^{k_1}), \dots \phi(p_r^{k_r})$. That is

$$\operatorname{dlog}_m(a) | [2^{k-1}, p_1^{k_1-1}(p_1-1), \dots, p_r^{k_r-1}(p_r-1)]$$

and this LCM is strictly smaller than $\phi(m)$ because 2 divides at least two terms. Thus in case (ii) $[p_1^{k_1-1}(p_1-1), p_2^{k_2-1}(p_2-1)] = p_1^{k_1-1}p_2^{k_2-1}[p_1-1, p_2-1] \le \frac{1}{2}\phi(p_1^{k_1}p_2^{k_2})$. Likewise in case (i) we have $[2^{k-1}, p_1^{k_1-1}(p_1-1)] = 2^{k-2}p_1^{k_1-1}[2, p_1-1] = 2^{k-2}p_1^{k_1-1}(p_1-1) < \phi(2^k p_1^{k_1})$. □

Example 4.3. *Primitive roots modulo* 7 *and* 7^2.

(i) Modulo 7. *Try* 2. *Divisors of* $\phi(7) = 6$ *are* $1, 2, 3, 6$ *and the order of* 2 *must beone of these.* $2^1 = 2 \not\equiv 1$, $2^2 = 4 \not\equiv 1$, $2^3 = 8 \equiv 1$ *so* 2 *not a primitive root.*

Try 3. $3^1 = 3 \not\equiv 1$, $3^2 = 9 \equiv 2 \not\equiv 1$, $3^3 = 27 \equiv 6 \not\equiv 1$. *Hence* 3 *has order* 6 *and so is a primitive root modulo* 7. *One can now find all primitive roots modulo* 7 *by considering* 3^x *with* $1 \leq x \leq 6$ *and* $(x, 6) = 1$. *The only choices for* x *are* 1 *and* 5, *so the only other primitive root modulo* 7 *is* $3^5 = 243 \equiv 5 \pmod 7$. *Thus* $3, 5$ *are the primitive roots modulo* 7.

By the way, this trial and error method is the best general method that we have. It is believed that in general one does not have to search very far, but we cannot prove it.

(ii) Modulo 7^2. *We know that a primitive root modulo* 7^2 *has to be one modulo* 7, *so we can start with* 3. *The divisors of* $\phi(7^2) = 6.7$ *are* $1, 2, 3, 6, 7, 14, 21, 42$. *We know that* $3^x \not\equiv 1 \pmod 7$ *when* $x = 1, 2, 3$ *and so* $3^x \not\equiv 1 \pmod{7^2}$ *in those cases. Also since* $3^7 \equiv 3 \pmod 7$, $3^{14} \equiv 3^2 \equiv 2 \pmod 7$ *and* $3^{21} \equiv 3^3 \equiv 6 \pmod 7$ *so* $3^x \not\equiv 1 \pmod{7^2}$ *in those cases either. Thus we only need check* $3^6 = 729 \equiv 43 \not\equiv 1 \pmod{7^2}$. *Thus* 3 *is also a primitive root modulo* 7^2.

We know from the Chinese Remainder Theorem that we can reduce a polynomial congruence modulo m when m is composite to its prime power constituents. However we were not able to treat the case $m = 2^k$ in general because when $k \geq 3$ primitive roots do not exist. Nevertheless we can usually apply the following theorem.

Theorem 4.6 (Gauss)**.** *Suppose that* $k \geq 3$. *Then the numbers* $(-1)^u 5^v$ *with* $u = 0, 1$ *and* $0 \leq v < 2^{k-2}$ *form a set of reduced residues modulo* 2^k

Proof. We first prove that if $r \geq 3$, then

$$5^{2^{r-2}} = 1 + 2^r j_r \tag{4.2}$$

with $2 \nmid j_r$. We prove this by induction on r. It is clear when $r = 3$, since $5^2 = 25 = 1 + 2^3.3$. If (4.2) holds, then

$$5^{2^{r-1}} = 1 + 2^{r+1} j_r + 2^{2r} j_r^2$$

and $2 \nmid j_r + 2^{r-1} j_r^2$. We also know that $\mathrm{dlog}_5(2^k) | \phi(2^k) = 2^{k-1}$, so $\mathrm{dlog}_5(2^k) = 2^r$ for some $0 \leq r \leq k-1$. The relationship (4.2) shows that $r = k-2$. Hence the numbers

$$1, 5, 5^2, 5^3, \ldots, 5^{2^{k-2}-1}$$

are distinct modulo 2^k. Likewise the numbers

$$-1, -5, -5^2, -5^3, \ldots, -5^{2^{k-2}-1}$$

are distinct modulo 2^k. Moreover the numbers in the fist list are $\equiv 1 \pmod 4$ and those in the second one are $\equiv -1 \pmod 4$. Thus the members of the first list are all different modulo 2^k to those in the second. Thus the two lists together give a complete cover of the 2^{k-1} reduced residues modulo 2^k. □

In terms of group theory this says that the reduced residues modulo 2^k with $k \geq 3$, under multiplication form a direct product of a cyclic group of order 2 and one of order 2^{k-2}.

4.1.1 Exercises

1. Find all the primitive roots of 7, 14, 49.

2. First find a primitive root modulo 19 and then find all primitive roots modulo 19.

3. Prove that $1^k + 2^k + \cdots + (p-1)^k \equiv 0 \pmod{p}$ when $p - 1 \nmid k$ and is $\equiv -1 \pmod{p}$ when $p - 1 | k$.

4. Let g be a primitive root modulo p. Prove that no k exists satisfying $g^{k+2} \equiv g^{k+1} + 1 \equiv g^k + 2 \pmod{p}$.

5. Suppose that $p = 2^m + 1$ is a prime, $p \nmid a$ and a is a quadratic non-residue (i.e., $x^2 \equiv a \pmod{p}$ is insoluble) modulo p. Show that a is a primitive root modulo p.

6. [Gauss] Prove that for any prime number $p \neq 3$ the product of its primitive roots is 1 (mod p).

4.2 Binomial Congruences

As an application of this theory we can say something about the solution of congruences of the form

$$x^k \equiv a \pmod{p}$$

when p is odd. The case $a = 0$ is easy. The only solution is $x \equiv 0 \pmod{p}$. Suppose $a \not\equiv 0 \pmod{p}$. Then we can pick a primitive root g modulo p and then there will be a c so that $g^c \equiv a \pmod{p}$. Also, since any solution x will have $p \nmid x$ we can define y so that $g^y \equiv x \pmod{p}$. Thus our congruence becomes

$$g^{ky} \equiv g^c \pmod{p}.$$

Hence it follows that

$$ky \equiv c \pmod{p-1}.$$

We have turned a polynomial congruence into a linear one. This is a bit like using logarithms on real numbers. Sometimes the exponents c and y are referred to as the discrete logarithms modulu p to the base g. Computing them numerically is hard and there is a protocol (Diffie-Hellman) which uses them to exchange secure keys and passwords. Our new congruence is soluble if and only if $(k, p-1)|c$, and when this holds the y which satisfy it lie in a residue class modulo $\frac{p-1}{(k,p-1)}$, i.e. $(k, p-1)$ different residue classes modulo $p - 1$. Thus, when $a \not\equiv 0 \pmod{p}$ the original congruence is either insoluble or has $(k, p-1)$ solutions. Thus we just proved the following theorem.

Theorem 4.7. *Suppose p is an odd prime. When $p \nmid a$ the congruence $x^k \equiv a \pmod{p}$ has 0 or $(k, p-1)$ solutions, and the number of reduced residues a modulo p for which it is soluble is $\frac{p-1}{(k,p-1)}$.*

4.2.1 Discrete Logarithms

The above theorem suggests that we can use primitive roots to create the residue class equivalent of logarithms.

Definition 4.3. *Given a primitive root g and a reduced residue class a modulo m we define the discrete logarithm $\operatorname{dlog}_g(a)$ to be that unique residue class l modulo $\phi(m)$ such that $g^l \equiv a \pmod{m}$*

Example 4.4. *Find a primitive root modulo 11 and construct a table of discrete logarithms. First we check 2. The divisors of $11 - 1 = 10$ are 1, 2, 5, 10 and $2^1 = 2 \not\equiv 1 \pmod{11}$, $2^2 = 4 \not\equiv 1 \pmod{11}$, $2^5 = 32 \equiv 10 \not\equiv 1 \pmod{11}$, so 2 is a primitive root modulo 11.*

Now we construct a table of powers of 2 modulo 11

y	1	2	3	4	5	6	7	8	9	10
$x \equiv 2^y$	2	4	8	5	10	9	7	3	6	1

Now we construct the "inverse" table

x	1	2	3	4	5	6	7	8	9	10
$y = \operatorname{dlog}_2(x)$	10	1	8	2	4	9	7	3	6	5

Note that while x is a residue modulo p (here $p = 11$), the y are residues modulo $p-1$ (here 10). The number y is the order, or exponent, to which 2 has to be raised to give x modulo p. Sometimes the notation $\operatorname{ind}_g(x)$ is used, where g is the given primitive root, but we will use $\operatorname{dlog}_g(x)$. In other words $x \equiv g^{\operatorname{dlog}_g(x)} \pmod{p}$.

Example 4.5. *We can use this to solve, if possible, the congruences,*

$$x^3 \equiv 6 \pmod{11},$$
$$x^5 \equiv 9 \pmod{11},$$
$$x^{65} \equiv 10 \pmod{11}$$

Consider the first one, $x^3 \equiv 6 \pmod{11}$. We can write $x \equiv 2^y \pmod{11}$, so that $x^3 = 2^{3y}$ and we see from the second table that $6 \equiv 2^9 \pmod{11}$. Thus what we need is that $3y$ and 9 match. This means that we need

$$3y \equiv 9 \pmod{10}.$$

Recall that the modulus here is $p - 1 = 10$ since $2^{10} \equiv 1 \pmod{11}$. This has the unique solution

$$y \equiv 3 \pmod{10}.$$

Going to the first table we find that $x \equiv 8 \pmod{11}$.

For the second congruence we find that $5y \equiv 6 \pmod{10}$ *and now we see that this has no solutions because* $(5, 10) = 5 \nmid 6$.

In the third case we have $65y \equiv 5 \pmod{10}$ *and this is equivalent to* $13y \equiv 1 \pmod{2}$ *and this has one solution modulo* $y \equiv 1 \pmod{2}$*, and so* 5 *solutions modulo* 10 *given by* $y \equiv 1$, 3, 5, 7 *or* 9 *modulo* 10. *Hence the original congruence has five solutions given by*

$$x \equiv 2,\ 8,\ 10,\ 7,\ 6 \pmod{11}$$

4.2.2 Exercises

1. Show that 3 is a primitive root modulo 17 and draw up a table of discrete logarithms to this base modulo 17. Hence, or otherwise, find all solutions to the following congruences.
(i) $x^{12} \equiv 16 \pmod{17}$,
(ii) $x^{48} \equiv 9 \pmod{17}$,
(iii) $x^{20} \equiv 13 \pmod{17}$,
(iv) $x^{11} \equiv 9 \pmod{17}$.

2. (i) Find the discrete logarithms of 2, 3 and 5 modulo 23.
(ii) Find a primitive root modulo 23, construct a table of discrete logarithms, and solve the congruence $x^6 \equiv 4 \pmod{23}$.

3. Show that 2 is a primitive root modulo 13 and draw up a table of discrete logarithms to this base. Hence, or otherwise, find all solutions to the following congruences.
(i) $x^{16} \equiv 3 \pmod{13}$,
(ii) $x^{21} \equiv 3 \pmod{13}$,
(iii) $x^{31} \equiv 7 \pmod{13}$.

4. Show that 2 is a primitive root modulo 11 and draw up a table of discrete logarithms to this base modulo 11. Hence, or otherwise, find all solutions to the following congruences.
(i) $x^6 \equiv 7 \pmod{11}$,
(ii) $x^{48} \equiv 9 \pmod{11}$,
(iii) $x^7 \equiv 8 \pmod{11}$.

4.3 Notes

Euler invented the term *primitive root*, and Gauss (1801) was the first to prove that they exist modulo p for every prime p.

Chapter 5

Quadratic Residues

5.1 Quadratic Congruences

We can now apply the theory we have developed to study quadratic congruences, and especially

$$x^2 \equiv c \pmod{m}.$$

The structure here is especially rich and was thus subject to much work in the eighteenth century, culminating in a famous theorem of Gauss.

From the various theories we have developed we know that the first, or base, case we need to understand is that when the modulus is a prime p, and since the case $p = 2$ is rather easy we can suppose that $p > 2$. Then we are interested in

$$x^2 \equiv c \pmod{p}. \tag{5.1}$$

By the way, the apparently more general congruence $ax^2 + bx + c \equiv 0 \pmod{p}$ (with $p \nmid a$ of course) can be reduced by "completion of the square" via $4a(ax^2 + bx + c) \equiv 0 \pmod{p}$ to $(2ax + b)^2 \equiv b^2 - 4ac \pmod{p}$ and since $2ax + b$ ranges over a complete set of residues as x does this is equivalent to solving $x^2 \equiv b^2 - 4ac \pmod{p}$. Thus it suffices to know about the solubility of the congruence (5.1).

We know that (5.1) has at most two solutions, and that sometimes it is soluble and sometimes not

Example 5.1. $x^2 \equiv 6 \mod 7$ *has no solution (check* $x \equiv 0, 1, 2, 3 \pmod{7}$*), but*

$$x^2 \equiv 5 \pmod{11}$$

has the solutions

$$x \equiv 4, 7 \pmod{11}.$$

If $c \equiv 0 \pmod{p}$, then the only solution to (5.1) is $x \equiv 0 \pmod{p}$ (note that $p|x^2$ implies that $p|x$). If $c \not\equiv 0 \pmod{p}$ and the congruence has one solution, say $x \equiv x_0 \pmod{p}$, then $x \equiv p - x_0 \pmod{p}$ gives another.

The fundamental question here is can we characterise or classify those c for which the congruence (5.1) is soluble? Better still can we quickly determine, given c, whether (5.1) is soluble?

Definition 5.1. *If $c \not\equiv 0 \pmod{p}$, and (5.1) has a solution, then we call c a quadratic residue which we abbreviate to QR. If it does not have a solution, then we call c a quadratic non-residue or QNR.*

Some authors also call 0 a quadratic residue. Others leave it undefined. We will follow the latter course. Zero does behave differently. Now we can prove the following simple, but surprisingly useful, theorem.

Theorem 5.1. *Let p be an odd prime number. The numbers*

$$1, 2^2, 3^2, \ldots, \left(\frac{p-1}{2}\right)^2$$

are distinct modulo p and give a complete set of (non-zero) quadratic residues modulo p. There are exactly $\frac{1}{2}(p-1)$ quadratic residues modulo p and exactly $\frac{1}{2}(p-1)$ quadratic non-residues.

Proof. Suppose that $1 \leq x < y \leq \frac{1}{2}(p-1)$. If $p|y^2 - x^2 = (y-x)(y+x)$, then $p|y-x$ or $p|y+x$. But $0 < y - x < y + x < 2y \leq p - 1 < p$. Thus the numbers $1, 2^2, 3^2, \ldots, \left(\frac{p-1}{2}\right)^2$ are distinct modulo p.

Now suppose that c is a quadratic residue modulo p. Then there is an x with $1 \leq x \leq p-1$ such that $x^2 \equiv c \pmod{p}$. If $x \leq \frac{1}{2}(p-1)$, then x^2 is in our list and represents c. If $\frac{1}{2}(p-1) < x \leq p-1$, then $(p-x)^2 \equiv x^2 \equiv c \pmod{p}$, $(p-x)^2$ represents c, and $1 \leq p - x \leq \frac{1}{2}(p-1)$. Moreover $(p-x)^2$ is in our list. Thus every QR is in our list and every member of our list is distinct and a QR. Hence there are exactly $\frac{1}{2}(p-1)$ QR. Moreover then the remaining $p - 1 - \frac{1}{2}(p-1) = \frac{1}{2}(p-1)$ non-zero residues have to be QNR. □

We can use this in various ways.

Example 5.2. *Find a complete set of quadratic residues r modulo 19 with $1 \leq r \leq 18$.*

We can solve this by first observing that $1^2 = 1, 2^2 = 4, 3^2 = 9, 4^2 = 16, 5^2 = 25, 6^2 = 36, 7^2 = 49, 8^2 = 64, 9^2 = 81$ is a complete set of quadratic residues and then reduce them modulo 19 to give

$$1, 4, 9, 16, 6, 17, 11, 7, 5$$

which we can rearrange as

$$1, 4, 5, 6, 7, 9, 11, 16, 17.$$

To help us understand quadratic residues we make the following definition.

Definition 5.2. *Given an odd prime number p and an integer c we define the Legendre symbol*

$$\left(\frac{c}{p}\right)_L = \begin{cases} 0 & c \equiv 0 \pmod{p}, \\ 1 & c \text{ a } QR \pmod{p}, \\ -1 & c \text{ a } QNR \pmod{p}, \end{cases} \tag{5.2}$$

The Legendre symbol is a remarkable function with lots of interesting properties.

Example 5.3. *One very important property is that it has the same value if one replaces c by $c + kp$ regardless of the value of k. Thus given p it is periodic in c with period p.*

Example 5.4. *Suppose that p is an odd prime and $a \not\equiv 0 \pmod{p}$. Then*

$$\sum_{x=1}^{p} \left(\frac{ax+b}{p}\right)_L = 0. \tag{5.3}$$

The proof of this is rather easy. The expression $ax + b$ runs through a complete set of residues as x does and so one of the terms is 0, half the rest are $+1$, and the remainder are -1.

Example 5.5. *The number of solutions of the congruence*

$$x^2 \equiv c \pmod{p}$$

is

$$1 + \left(\frac{c}{p}\right)_L.$$

We already know that the number of solutions is 1 when $p|c$, 2 when c is a QR, and 0 when c is a QNR and this matches the above exactly.

We can use this to count the solutions of more complicated congruences.

Example 5.6. *How many solutions does*

$$x^2 + y^2 \equiv c \pmod{p}$$

have in x and y? Denote the number by $N(p; c)$. We can rewrite the congruence as $z + w \equiv c \pmod{p}$, and then for each solution z, w ask for the number of solutions of $x^2 \equiv z \pmod{p}$ and $y^2 \equiv w \pmod{p}$. From above this is

$$\left(1 + \left(\frac{z}{p}\right)_L\right)\left(1 + \left(\frac{w}{p}\right)_L\right).$$

Also $w \equiv c - z \pmod{p}$*, thus the total number of solutions is*

$$N(p;c) = \sum_{z=1}^{p} \left(1 + \left(\frac{z}{p}\right)_L\right)\left(1 + \left(\frac{c-z}{p}\right)_L\right).$$

If we multiply this out we get

$$p + \sum_{z=1}^{p} \left(\frac{z}{p}\right)_L + \sum_{z=1}^{p} \left(\frac{c-z}{p}\right)_L + \sum_{z=1}^{p} \left(\frac{z}{p}\right)_L \left(\frac{c-z}{p}\right)_L.$$

By (5.3) the first and second sums are 0*, so that*

$$N(p;c) = p + \sum_{z=1}^{p} \left(\frac{z}{p}\right)_L \left(\frac{c-z}{p}\right)_L.$$

It is possible also to evaluate the sum here, but we need to know a little more about the Legendre symbol.

The Legendre symbol is a prototype for an important class of number theoretic functions called Dirichlet characters. A simple example would be to take an odd prime p and a primitive root modulo g modulo p, and then for a fixed h we can define $\chi(g^k) = e^{2\pi ihk/(p-1)}$ and $\chi(n) = 0$ if $p|n$. The Legendre symbol is the special case $h = \frac{p-1}{2}$. Dirichlet used them to prove that if $(a, m) = 1$, then there are infinitely many primes in the residue class a modulo m.

We can combine the definition of the Legendre symbol with a criterion first enunciated by Euler.

Theorem 5.2 (Euler's Criterion)**.** *Suppose that* p *is an odd prime number. Then*

$$\left(\frac{c}{p}\right)_L \equiv c^{\frac{p-1}{2}} \pmod{p}$$

and the Legendre symbol, as a function of c*, is totally multiplicative.*

Remark 5.1. *By multiplicative we mean a function* f *which satisfies*

$$f(n_1 n_2) = f(n_1) f(n_2)$$

whenever $(n_1, n_2) = 1$*. Totally multiplicative means that the condition* (n_1, n_2) *can be dropped.*

Remark 5.2. *The totally multiplicative property means that if* x *and* y *are both QR, or both QNR, then their product is a QR, and their product can only be a QNR if one is a QR and the other is a QNR.*

Proof. If c is a quadratic residue, then there is an $x \not\equiv 0 \pmod{p}$ such that $x^2 \equiv c \pmod{p}$. Hence $c^{\frac{p-1}{2}} \equiv x^{p-1} \equiv 1 = \left(\frac{c}{p}\right)_L \pmod{p}$. We know that the congruence

$$c^{\frac{p-1}{2}} \equiv 1 \pmod{p}$$

has at most $\frac{p-1}{2}$ solutions and so we have just shown that is has exactly that many solutions. We also have

$$\left(c^{\frac{p-1}{2}} - 1\right)\left(c^{\frac{p-1}{2}} + 1\right) = c^{p-1} - 1$$

and we know that this has exactly $p-1$ roots modulo p. In particular every QNR is a solution, but cannot be a root of $c^{\frac{p-1}{2}} - 1$. Hence if c is a QNR, then $c^{\frac{p-1}{2}} \equiv -1 = \left(\frac{c}{p}\right)_L \pmod{p}$. This proves the first part of the theorem.

To prove the second part, we have to show that for any integers c_1, c_2 we have

$$\left(\frac{c_1 c_2}{p}\right)_L = \left(\frac{c_1}{p}\right)_L \left(\frac{c_2}{p}\right)_L.$$

$c_1 \equiv 0 \pmod{p}$ or $c_2 \equiv 0 \pmod{p}$ then both sides are 0, so we can suppose that $c_1 c_2 \not\equiv 0 \pmod{p}$. Now

$$\begin{aligned}\left(\frac{c_1 c_2}{p}\right)_L &\equiv (c_1 c_2)^{\frac{p-1}{2}} \\ &\equiv c_1^{\frac{p-1}{2}} c_2^{\frac{p-1}{2}} \\ &\equiv \left(\frac{c_1}{p}\right)_L \left(\frac{c_2}{p}\right)_L \pmod{p}.\end{aligned}$$

Thus p divides

$$\left(\frac{c_1 c_2}{p}\right)_L - \left(\frac{c_1}{p}\right)_L \left(\frac{c_2}{p}\right)_L.$$

But this is $-2, 0$ or 2 and so has to be 0 since $p > 2$ □

We can use this to evaluate the Legendre symbol in special cases.

Example 5.7. *Suppose that p is an odd prime. Then*

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

Observe that by Euler's criterion

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

Now the difference between the left and right hand sides is $-2, 0$ *or* 2 *and the same argument as above gives equality.*

This example has some interesting consequences.

1. Every odd prime divisor p *of the polynomial* $x^2 + 1$ *satisfies* $p \equiv 1 \pmod 4$.

2. There are infinitely many primes of the form $4k + 1$.

To see 1. one only has to observe that for any such prime factor -1 *has to be a quadratic residue, so its Legendre symbol is* 1. *To deduce 2., follow Euclid's argument by supposing there are only finitely many such, say* $p_1, \dots, p_r$, *and take* x *to be* $2p_1 \dots p_r$.

A famous question, first asked by I. M. Vinogradov in 1919, concerns the size $n_2(p)$ of the *least* positive QNR modulo p. One thing one can see straight away is that $n_2(p)$ has to be prime, since it must have a prime factor which is a QNR. He conjectured that for any fixed positive number $\varepsilon > 0$ we should have $n_2(p) < C(\varepsilon)p^{\varepsilon}$ and then proceeded to prove this at least when $\varepsilon > \frac{1}{2\sqrt{e}}$ where e is the base of the natural logarithm! In 1959 David Burgess, in his PhD thesis (!!) reduced this to any $\varepsilon > \frac{1}{4\sqrt{e}}$. Where on earth does the $\sqrt{e}$ come from? This was one of the things that got me interested in number theory when I was a student. Here is an easier result.

Theorem 5.3. *Suppose that* p *is an odd prime. Then*

$$n_2(p) \leq \frac{1}{2} + \sqrt{p - \frac{3}{4}}.$$

Proof. Let h be the smallest positive integer such that $h \equiv -p \pmod{n_2(p)}$. Obviously $n_2(p)|p + h$, $1 \leq h \leq n_2(p)$, and we cannot have $h = n_2(p)$ since then $n_2(p)|p$ which is not possible since $n_2(p) < p$. Thus $1 \leq h \leq n_2(p) - 1$. Since $h < n_2(p)$ it follows that h is a QR, and hence $p + h$ is a QR. But as $n_2(p)|p + h$ and $n_2(p)$ is a QNR it follows that $(p + h)/n_2(p)$ is a QNR, and so $p + h \geq n_2(p)^2$. Inserting the upper bound for h gives $p - 1 + n_2(p) \geq n_2(p)^2$. This can be rearranged as $n_2(p)^2 - n_2(p) \leq p - 1$, so $(n_2(p) - \frac{1}{2})^2 \leq p - \frac{3}{4}$. The theorem follows by taking the square root. □

The multiplicative property of the Legendre symbol tells us that it suffices to understand

$$\left(\frac{q}{p}\right)_L$$

when p is an odd prime and q is prime. When q is also odd, Euler found a remarkable relationship between this Legendre symbol and

$$\left(\frac{p}{q}\right)_L$$

but no one in the eighteenth century was able to prove it. Gauss proved it when he was 19! The relationship enables one to imitate the Euclid algorithm and so rapidly evaluate the Legendre symbol.

5.1.1 Exercises

1. Find a complete set of quadratic residues r modulo 13 in the range $1 \leq r \leq 12$.

2. Find a complete set of quadratic residues r modulo 17 in the range $1 \leq r \leq 16$.

3. Find a complete set of quadratic residues r modulo 23 in the range $1 \leq r \leq 22$.

4. Suppose that p is an odd prime and g is a primitive root modulo p. Prove that g is a quadratic non-residue modulo p.

5. Prove that $7n^3 - 1$ can never be a perfect square.

6. Prove that if p is an odd prime, then

$$\sum_{x=1}^{p} \sum_{y=1}^{p} \left(\frac{xy+1}{p} \right)_L = p.$$

7. (i) Recall that for every reduced residue class r modulo p there is a unique reduced residue class s_r modulo p such that $1 \equiv rs_r \pmod{p}$, and that for every reduced residue class s modulo p there is a unique r such that $s_r \equiv s \pmod{p}$. Hence prove that if p is an odd prime, then

$$\sum_{r=1}^{p-1} \left(\frac{r(r+1)}{p} \right)_L = \sum_{s=1}^{p-1} \left(\frac{1+s}{p} \right)_L = -1.$$

(ii) Prove that if p is an odd prime, then the number of residues r modulo p for which both r and $r+1$ are quadratic residues is

$$\frac{p - (-1)^{\frac{p-1}{2}}}{4} - 1.$$

8. Let $N(p;c)$ be as in Example 5.6 so that

$$N(p;c) = p + \sum_{z=1}^{p} \left(\frac{z(c-z)}{p} \right)_L.$$

(i) Prove that if $c \equiv 0 \pmod{p}$, then

$$N(p;0) = p + (-1)^{\frac{p-1}{2}}(p-1).$$

(ii) Prove that if $c \not\equiv 0 \pmod{p}$, then

$$\sum_{z=1}^{p} \left(\frac{z(c-z)}{p} \right)_L = \sum_{z=1}^{p-1} \left(\frac{z^2(cs_z - 1)}{p} \right)_L = \sum_{s=1}^{p-1} \left(\frac{cs-1}{p} \right)_L = -(-1)^{\frac{p-1}{2}}.$$

(iii) Deduce that if $c \not\equiv 0 \pmod{p}$, then

$$N(p;c) = p - (-1)^{\frac{p-1}{2}}.$$

9. Let g be a primitive root modulo p. Prove that the quadratic residues are precisely the residue classes g^{2k} with $0 \le k < \frac{1}{2}(p-1)$. Show that the sum of the quadratic residues modulo p is the 0 residue.

10. Prove that every quadratic non-residue modulo p is a primitive root modulo p if and only if $p = 2^{2^n} + 1$ for some non-negative integer n.

11. Suppose that $p \nmid a$. Show that the number of solutions to $ax^2 + bx + c \equiv 0 \pmod{p}$ is $1 + \left(\frac{b^2-4ac}{p}\right)_L$.

12. Prove that $\sum_{x=1}^{p} \left(\frac{x}{p}\right)_L = 0$ and that if $p \nmid a$, then $\sum_{x=1}^{p} \left(\frac{ax+b}{p}\right)_L = 0$.

13. Let $S(p,a,b,c) = \sum_{x=1}^{p} \left(\frac{ax^2+bx+c}{p}\right)_L$.

(i) Show that $S(p,1,b,0) = \sum_{y=1}^{p-1} \left(\frac{1+by}{p}\right)_L$. (Hint: For each x with $1 \le x \le p-1$ let y denote the unique solution to $xy \equiv 1 \pmod{p}$, so that $x(x+b) \equiv x^2(1+by)$.) Deduce that $S(p,1,b) = p-1$ when $p|b$ and is -1 when $p \nmid b$.

(ii) Show that $S(p,1,0,c) = \sum_{y=0}^{p} \left(\frac{y+b}{p}\right)_L \left(1 + \left(\frac{y}{p}\right)_L\right)$. (Hint: Note that for each y with $1 \le y \le p$ the number of solutions in x to $x^2 \equiv y \pmod{p}$ is $1 + \left(\frac{y}{p}\right)_L$) Deduce that $S(p,1,0,c) = S(p,1,c,0) = p-1$ when $p|c$ and is -1 when $p \nmid c$.

(iii) Show that if $p \nmid a$, then $S(p,a,b,c) = \left(\frac{4a}{p}\right)_L S(p,1,0,4ac-b^2)$. Deduce that $S(p,a,b,c) = p\left(\frac{c}{p}\right)_l$ when $p|a$ and $p|b$, is 0 when $p|a$ and $p \nmid b$, and satisfies

$$S(p,a,b,c) = \begin{cases} \left(\frac{a}{p}\right)_L (p-1) & \text{when } p \nmid a \text{ and } p|b^2-4ac, \\ -\left(\frac{a}{p}\right)_L & \text{when } p \nmid a(b^2-4ac). \end{cases} \tag{5.4}$$

5.2 Quadratic Reciprocity

What Euler spotted was a very curious relationship between the values of

$$\left(\frac{q}{p}\right)_L$$

when p and q are different odd primes, which only depended on their residue classes modulo 4. Of course, this was before the Legendre symbol was invented and he described the phenomenon in terms of quadratic residues and non-residues.

Example 5.8. *Here is a short table of values for primes out to* 29.

$p\backslash q$	3	5	7	11	13	17	19	23	29
3	0	−1	1	−1	1	−1	1	−1	−1
5	−1	0	−1	1	−1	−1	1	−1	1
7	−1	−1	0	1	−1	−1	−1	1	1
11	1	1	−1	0	−1	−1	−1	1	−1
13	1	−1	−1	−1	0	1	−1	1	1
17	−1	−1	−1	−1	1	0	−1	−1	−1
19	−1	1	1	1	−1	1	0	1	−1
23	1	−1	−1	−1	1	−1	−1	0	1
29	−1	1	1	−1	1	−1	−1	1	0

Table of $\left(\frac{q}{p}\right)_L$ for odd primes $p, q \leq 23$.

Apparently if $p \equiv 1 \pmod 4$ or $q \equiv 1 \pmod 4$, then $\left(\frac{q}{p}\right)_L = \left(\frac{p}{q}\right)_L$, but if $p \equiv q \equiv 3 \pmod 4$, then $\left(\frac{q}{p}\right)_L \neq \left(\frac{p}{q}\right)_L$.

Gauss was fascinated by this and eventually found seven (!) different proofs. The first step in many of them is Gauss' Lemma. In it's statement the we use the following function.

Definition 5.3. *For real numbers α we define the* **floor function** $\lfloor \alpha \rfloor$ *to be the largest integer not exceeding α.*

Example 5.9. *Thus $\lfloor \frac{5}{2} \rfloor = 2$ and $\lfloor -\sqrt{2} \rfloor = -2$.*

The only property we will use here is that for any real number α and integer k we have $\lfloor \alpha - k \rfloor = \lfloor \alpha \rfloor - k$, which is easy to check, and otherwise it is just a useful shorthand. We will investigate its properties in more detail in Chapter 8.

Theorem 5.4 (Gauss' Lemma)**.** *Suppose that p is an odd prime and $(a, p) = 1$. Apply the division algorithm to write each of the $\frac{1}{2}(p-1)$ numbers ax with $1 \leq x < \frac{1}{2}p$ as $ax = q_x p + r_x$ with $0 \leq r_x < p$. Let m be the number of r_x with $\frac{1}{2}p < r_x < p$. Then we have*

$$\left(\frac{a}{p}\right)_L = (-1)^m$$

and

$$m \equiv \sum_{1 \leq x < p/2} \left\lfloor \frac{2ax}{p} \right\rfloor \pmod 2.$$

This theorem enables us to evaluate quite a number of cases directly with some ease.

Example 5.10. *Take $a = 2$. Then we begin by considering the numbers $2x$ with $1 \leq x < \frac{1}{2}p$. These numbers satisfy $2 \leq 2x < p$. In view of the latter inequality, they are their own remainder, i.e. $r_x = 2x$, so we need to count the number of x with $\frac{1}{2}p < 2x < p$, that is $\frac{1}{4}p < x < \frac{1}{2}p$. Hence the number of such x is*

$$m = \left\lfloor \frac{p}{2} \right\rfloor - \left\lfloor \frac{p}{4} \right\rfloor.$$

Now suppose that $p = 8k + 1$. Then $m = 4k - 2k$ is even. Likewise when $p = 8k + 7$ we have $m = 2k + 2$ is also even. It can be checked similarly that if $p \equiv 3$ or $5 \pmod 8$, then m is odd. Thus

$$\left(\frac{2}{p}\right)_L = \begin{cases} 1 & (p \equiv \pm 1 \pmod 8), \\ -1 & (p \equiv \pm 3 \pmod 8). \end{cases} \tag{5.5}$$

One can check that another way of writing this is

$$\left(\frac{2}{p}\right)_L = (-1)^{\frac{p^2-1}{8}}.$$

It is relatively easy to deal with the case $a = 3$ in a similar way.

Proof of Gauss' Lemma. The proof is combinatorial - a kind of counting argument. We consider the product

$$a^{\frac{p-1}{2}} \prod_{1 \leq x < p/2} x = \prod_{1 \leq x < p/2} ax.$$

This is

$$\equiv \prod_{1 \leq x < p/2} r_x \pmod p.$$

Let $\mathcal{A}$ be the set of x with $p/2 < r_x < p$ and $\mathcal{B}$ the x with $1 \leq r_x < p/2$. Then $\operatorname{card} \mathcal{A} = m$ and we can rearrange the product to give

$$a^{\frac{p-1}{2}} \prod_{1 \leq x < p/2} x \equiv \left(\prod_{x \in \mathcal{A}} r_x \right) \prod_{x \in \mathcal{B}} r_x \equiv (-1)^m \left(\prod_{x \in \mathcal{A}} (p - r_x) \right) \prod_{x \in \mathcal{B}} r_x \pmod p. \tag{5.6}$$

Since $|r_x - r_y| < p$ and $r_x - r_y \equiv a(x - y) \pmod p$ we have $r_x \neq r_y$ when $x \neq y$. Thus the r_x are distinct. Also since $p \nmid a$ and $1 \leq x, y < p/2$ we have $p - r_x - r_y \equiv -a(x + y) \not\equiv 0 \pmod p$. Thus the $p - r_x$ with $x \in \mathcal{A}$ are distinct from the r_y with $y \in \mathcal{B}$. Thus in the expression on the right in (5.6) the $\frac{1}{2}(p-1)$ numbers $p - r_x$ and r_x are just a permutation of the numbers z with $1 \leq z \leq \frac{1}{2}(p-1)$. Thus (5.6) becomes

$$a^{\frac{p-1}{2}} \prod_{1 \leq x < p/2} x \equiv (-1)^m \prod_{1 \leq x < p/2} x \pmod p$$

and so, by Euler's Criterion,

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

Now we can complete the proof of the first formula in the theorem by our usual observation that the difference between the two sides is -2, 0 or 2.

For the final formula we note that

$$r_x = ax - p\left\lfloor \frac{ax}{p} \right\rfloor \tag{5.7}$$

so that $0 \leq r_x < p$. Now $0 < 2r_x/p < 2$ and so $\lfloor 2r_x/p \rfloor = 0$ or 1 and is 1 precisely when $p/2 < r_x < p$. Thus

$$m = \sum_{1 \leq x < p/2} \lfloor 2r_x/p \rfloor.$$

Moreover, by (5.7)

$$\begin{aligned} \lfloor 2r_x/p \rfloor &= \left\lfloor \frac{2ax}{p} - 2\left\lfloor \frac{ax}{p} \right\rfloor \right\rfloor \\ &= \left\lfloor \frac{2ax}{p} \right\rfloor - 2\left\lfloor \frac{ax}{p} \right\rfloor \\ &\equiv \left\lfloor \frac{2ax}{p} \right\rfloor \pmod{2} \end{aligned}$$

and the final formula follows. □

If we restrict our attention to odd a there is a useful variant of this.

Theorem 5.5. *Suppose that p is an odd prime and $(a, 2p) = 1$. Then*

$$\left(\frac{a}{p}\right)_L = (-1)^n$$

where

$$n = \sum_{1 \leq x < p/2} \left\lfloor \frac{ax}{p} \right\rfloor.$$

We also have

$$\left(\frac{2}{p}\right)_L = (-1)^{\frac{p^2-1}{8}}.$$

Proof. We have

$$\begin{aligned}\left(\frac{2}{p}\right)_L\left(\frac{a}{p}\right)_L &= \left(\frac{2}{p}\right)_L\left(\frac{a+p}{p}\right)_L \\ &= \left(\frac{4}{p}\right)_L\left(\frac{(a+p)/2}{p}\right)_L \\ &= \left(\frac{(a+p)/2}{p}\right)_L \\ &= (-1)^l\end{aligned}$$

where

$$\begin{aligned}l &= \sum_{x=1}^{(p-1)/2}\left\lfloor\frac{(a+p)x}{p}\right\rfloor \\ &= \sum_{x=1}^{(p-1)/2}\left\lfloor\frac{ax}{p}+x\right\rfloor \\ &= \sum_{x=1}^{(p-1)/2}\left(\left\lfloor\frac{ax}{p}\right\rfloor+x\right) \\ &= n+\frac{p^2-1}{8}.\end{aligned}$$

If we take $a = 1$, then we have recovered the stated formula for

$$\left(\frac{2}{p}\right)_L.$$

Then factoring out the formula for this give the result for

$$\left(\frac{a}{p}\right)_L$$

□

Now we come to the big one. This is the Law of Quadratic Reciprocity. Gauss called it "Theorema Aureum", the Golden Theorem.

Theorem 5.6 (The Law of Quadratic Reciprocity)**.** *Suppose that p and q are odd prime numbers. Then*

$$\left(\frac{q}{p}\right)_L\left(\frac{p}{q}\right)_L = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}},$$

or equivalently

$$\left(\frac{q}{p}\right)_L = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}\left(\frac{p}{q}\right)_L,$$

We can use this to compute rapidly Legendre symbols.

Example 5.11. *Is* $x^2 \equiv 951 \pmod{2017}$ *soluble?* 2017 *is prime but* $951 = 3 \times 317$. *Thus*

$$\left(\frac{951}{2017}\right)_L = \left(\frac{3}{2017}\right)_L \left(\frac{317}{2017}\right)_L.$$

Now by the law, since $2017 \equiv 1 \pmod{4}$,

$$\left(\frac{3}{2017}\right)_L = \left(\frac{2017}{3}\right)_L = \left(\frac{1}{3}\right)_L = 1$$

and

$$\left(\frac{317}{2017}\right)_L = \left(\frac{2017}{317}\right)_L = \left(\frac{115}{317}\right)_L = \left(\frac{5}{317}\right)_L \left(\frac{23}{317}\right)_L.$$

Again applying the law, we have

$$\left(\frac{5}{317}\right)_L = \left(\frac{317}{5}\right)_L = \left(\frac{2}{5}\right)_L = -1$$

and

$$\left(\frac{23}{317}\right)_L = \left(\frac{317}{23}\right)_L = \left(\frac{18}{23}\right)_L = \left(\frac{2}{23}\right)_L = 1$$

so that

$$\left(\frac{317}{2017}\right)_L = -1$$

and thus

$$\left(\frac{951}{2017}\right)_L = -1.$$

Thus the congruence is insoluble.

We can also use the law to obtain general rules, like that for $2 \pmod{p}$.

Example 5.12. *Let* $p > 3$ *be an odd prime. Then*

$$\left(\frac{3}{p}\right)_L = (-1)^{\frac{p-1}{2}} \left(\frac{p}{3}\right)_L.$$

Now p *is a QR modulo 3 iff* $p \equiv 1 \pmod{3}$. *Thus*

$$\left(\frac{3}{p}\right)_L = \begin{cases} (-1)^{\frac{p-1}{2}} & (p \equiv 1 \pmod{3}) \\ -(-1)^{\frac{p-1}{2}} & (p \equiv 2 \pmod{3}). \end{cases}$$

We can also combine this with the formula in the case of $-1 \pmod{p}$ *which follows from the Euler Criterion. Thus*

$$\left(\frac{-3}{p}\right)_L = \begin{cases} 1 & (p \equiv 1 \pmod{3}) \\ -1 & (p \equiv 2 \pmod{3}). \end{cases}$$

We now turn to the proof of the law.

Proof of the Law of Quadratic Reciprocity. We start from two applications of the previous theorem. Thus

$$\left(\frac{q}{p}\right)_L \left(\frac{p}{q}\right)_L = (-1)^{u+v}$$

where

$$u = \sum_{1 \le x < p/2} \left\lfloor \frac{qx}{p} \right\rfloor$$

and

$$v = \sum_{1 \le y < q/2} \left\lfloor \frac{py}{q} \right\rfloor .$$

Observe that $\left\lfloor \frac{qx}{p} \right\rfloor$ is the number of positive integers y with $1 \le y \le qx/p$. Since $p \nmid qx$ this is the same as the number of positive integers with $1 \le y < qx/p$. Thus the sum is the number of ordered pairs x, y with $1 \le x < p/2$ and $1 \le y < qx/p$. Likewise $\sum_{1 \le y < q/2} \left\lfloor \frac{py}{q} \right\rfloor$ is the number of ordered pairs x, y with $1 \le y < q/2$ and $1 \le x < py/q$, that is with $1 \le x < p/2$ and $xq/p < y < q/2$. Hence $u + v$ is the number of ordered pairs x, y with $1 \le x < p/2$ and $1 \le y < q/2$. This is

$$\frac{p-1}{2} \cdot \frac{q-1}{2}$$

and completes the proof. This argument is due to Eisenstein. □

5.2.1 Exercises

1. Evaluate the following Legendre symbols.
   (i) $\left(\frac{2}{127}\right)_L$,
   (ii) $\left(\frac{-1}{127}\right)_L$,
   (iii) $\left(\frac{5}{127}\right)_L$,
   (iv) $\left(\frac{11}{127}\right)_L$.

2. (i) Prove that 3 is a QR modulo p when $p \equiv \pm 1 \pmod{12}$ and is a QNR when $p \equiv \pm 5 \pmod{12}$.

   (ii) Prove that -3 is a QR modulo p for primes p with $p \equiv 1 \pmod 6$ and is a QNR for primes $p \equiv -1 \pmod 6$.

   (iii) By considering $4x^2 + 3$ show that there are infinitely many primes in the residue class $1 \pmod 6$.

3. Show that for every prime p the congruence

$$x^6 - 11x^4 + 36x^2 - 36 \equiv 0 \pmod{p}$$

is always soluble.

4. Find the number of solutions of the congruence (i) $x^2 \equiv 226 \pmod{563}$, (ii) $x^2 \equiv 429 \pmod{563}$.

5. Decide whether $x^2 \equiv 150 \pmod{1009}$ is soluble or not.

6. Find all primes p such that $x^2 \equiv 13 \pmod{p}$ has a solution.

7. Show that $(x^2 - 2)/(2y^2 + 3)$ is never an integer when x and y are integers.

5.3 The Jacobi symbol

In Example 5.13, there were several occasions when we needed to factorise the a in $\left(\frac{a}{p}\right)_L$. Jacobi introduced an extension of the Legendre symbol which avoids this.

Definition 5.4. *Suppose that m is an odd positive integer and a is an integer. Let $m = p_1^{r_1} \dots p_s^{r_s}$ be the canonical decomposition of m. Then we define the Jacobi symbol by*

$$\left(\frac{a}{m}\right)_J = \prod_{j=1}^{s} \left(\frac{a}{p_j}\right)_L^{r_j}.$$

Note that interpreting 1 as being an "empty product of primes" means that

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

Remarkably the Jacobi symbol has exactly the same properties as the Legendre symbol, except for one. That is, for a general odd modulus m it does not tell us about the solubility of $x^2 \equiv a \pmod{m}$.

Example 5.13. *We have*

$$\left(\frac{2}{15}\right)_J = \left(\frac{2}{3}\right)_L \left(\frac{2}{5}\right)_L = (-1)^2 = 1,$$

but $x^2 \equiv 2 \pmod{15}$ is insoluble because any solution would also be a solution of $x^2 \equiv 2 \pmod{3}$ which we know is insoluble.

Properties of the Jacobi symbol

1. Suppose that m is odd. Then
$$\left(\frac{a_1 a_2}{m}\right)_J = \left(\frac{a_1}{m}\right)_J \left(\frac{a_2}{m}\right)_J.$$
2. Suppose that the m_j are odd. Then
$$\left(\frac{a}{m_1 m_2}\right)_J = \left(\frac{a}{m_1}\right)_J \left(\frac{a}{m_2}\right)_J.$$
3. Suppose that m is odd and $a_1 \equiv a_2 \pmod{m}$. Then
$$\left(\frac{a_1}{m}\right)_J = \left(\frac{a_2}{m}\right)_J.$$
4. Suppose that m is odd. Then
$$\left(\frac{-1}{m}\right)_J = (-1)^{\frac{m-1}{2}}.$$
5. Suppose that m is odd. Then
$$\left(\frac{2}{m}\right)_J = (-1)^{\frac{m^2-1}{8}}.$$
6. Suppose that m and n are odd and $(m,n) = 1$. Then
$$\left(\frac{n}{m}\right)_J \left(\frac{m}{n}\right)_J = (-1)^{\frac{m-1}{2}\cdot\frac{n-1}{2}}.$$

The first three of these follow easily from the definition. The rest depend on algebraic identities combined with inductions on the number of prime factors, but proving them is tiresome. For 4. we need to know that
$$\frac{m_1-1}{2} + \frac{m_2-1}{2} \equiv \frac{m_1 m_2 - 1}{2} \pmod{2},$$
5. depends on
$$\frac{m_1^2-1}{8} + \frac{m_2^2-1}{8} \equiv \frac{m_1^2 m_2^2 - 1}{8} \pmod{2}.$$
6. Finally here one needs
$$\frac{l-1}{2}\cdot\frac{m-1}{2} + \frac{n-1}{2}\cdot\frac{m-1}{2} \equiv \frac{ln-1}{2}\cdot\frac{m-1}{2} \pmod{2}.$$

Example 5.14. *Return to Example 5.11, where we evaluated* $\left(\frac{951}{2017}\right)_L$. *Now we don't have to factor* 951. *By the Jacobi version of the law*
$$\begin{aligned}\left(\frac{951}{2017}\right)_L &= \left(\frac{2017}{951}\right)_J = \left(\frac{115}{951}\right)_J = -\left(\frac{951}{115}\right)_J \\ &= -\left(\frac{31}{115}\right)_J = \left(\frac{115}{31}\right)_J = \left(\frac{22}{31}\right)_J \\ &= -\left(\frac{31}{11}\right)_J = -\left(\frac{9}{11}\right)_J = -1.\end{aligned}$$

5.3.1 Exercises

1. Let $n \in \mathbb{Z}$ and let $n = (-1)^u 2^v p_1^{v_1} \dots p_r^{v_r}$ be the canonical decomposition of n with $u = 0$ or 1, $v \geq 0$, and each $v_j > 0$ when $r \geq 1$.

(i) If v is odd, then let $n_0 = |n|2^{-v}$ and choose $m \in \mathbb{N}$ so that $m \equiv 5 \pmod 8$ and $m \equiv 1 \pmod{n_0}$. Prove that

$$\left(\frac{n}{m}\right)_J = -1.$$

(ii) If v is even, but there is a j for which v_j is odd, let $n_j = |n|2^{-v} p_j^{-v_j}$ and choose $m \in \mathbb{N}$ so that $m \equiv 1 \pmod{(4n_j)}$ and m is a QNR modulo p_j. Prove that

$$\left(\frac{n}{m}\right)_J = -1.$$

(iii) If v is even, v_j is even for every j and $u = 1$, choose $m \in \mathbb{N}$ so that $m \equiv 3 \pmod 4$. Prove that

$$\left(\frac{n}{m}\right)_J = -1.$$

(iv) Prove that if n is not a perfect square, then there is an odd prime number p such that

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

(v) Prove that if n is a QR for every odd prime number p not dividing n, then n is a perfect square.

This is an example of the "local to global" principle.

5.4 Other questions

There are many interesting problems associated with quadratic residues and the Legendre and Jacobi symbols.

1. How many consecutive quadratic residues are there, that is how many x with $1 \leq x \leq p-2$ have the property that x and $x+1$ are both quadratic residues modulo p? This number is

$$\sum_{x=1}^{p-2} \frac{1}{4}\left(1 + \left(\frac{x}{p}\right)_L\right)\left(1 + \left(\frac{x+1}{p}\right)_L\right).$$

The method of exercise 5.1.1.13 is useful here. How about the number of triples $x, x+1, x+2$, or how about a fixed sequence of QR and QNR?

2. Given an N with $0 \leq N \leq p$, how small can you make M, regardless of the value of N, and ensure that the interval $(N, N+M]$ contains a quadratic non-residue?

3. Let m be an odd positive integer, and for brevity write $\chi(x)$ for the Jacobi symbol $\left(\frac{x}{m}\right)_J$. For a complex number z define

$$L(z;\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^z}.$$

This converges for $\Re z > 0$. There is a Riemann hypothesis for this function but we cannot prove it. Also $L(1,\chi)$ has some interesting values. For example if $m = 3$, then

$$L(1,\chi) = \frac{\pi}{3\sqrt{3}}.$$

4. The Gauss sum

$$\tau_p = \sum_{x=1}^{p} \left(\frac{x}{p}\right)_L e^{2\pi i x/p}$$

was studied by Gauss in connection with several of his proofs of the law of quadratic reciprocity. He showed that

$$\tau_p = \begin{cases} \sqrt{p} & (p \equiv 1 \pmod 4) \\ i\sqrt{p} & (p \equiv 3 \pmod 4). \end{cases}$$

5.4.1 Exercises

1. (i) Prove that if $\chi_1(n) = (-1)^{(n-1)/2}$ when n is odd and $\chi_1(n) = 0$ when n is even, then $L(1,\chi_1) = \frac{\pi}{4}$

(ii) Prove that if $\chi(n) = \left(\frac{n}{3}\right)_L$, then $L(1,\chi) = \frac{\pi}{3\sqrt{3}}$

(iii) Prove that if $\chi(n) = \left(\frac{n}{5}\right)_L$, then $L(1,\chi) = \frac{1}{\sqrt{5}} \log \frac{3+\sqrt{5}}{2}$

2. Let $c_n \in \mathbb{C}$ $(n = 1, 2, \ldots, p)$. Prove that

$$\sum_{a=1}^{p} \left| \sum_{n=1}^{p} c_n e^{2\pi i a n/p} \right|^2 = p \sum_{n=1}^{p} |c_n|^2.$$

3. For an odd prime p define

$$S(p,a) = \sum_{y=1}^{p} e^{2\pi i a y^2/p}$$

(i) Prove that if $p \nmid a$, then

$$\begin{aligned} S(p,a) &= \sum_{x=1}^{p} \left(1 + \left(\frac{x}{p}\right)_L\right) e^{2\pi i a x/p} \\ &= \sum_{x=1}^{p} \left(\frac{x}{p}\right)_L e^{2\pi i a x/p} \\ &= \left(\frac{a}{p}\right)_L \tau_p. \end{aligned}$$

(ii) Prove that

$$\sum_{a=1}^{p} |S(p,a)|^2 = p(2p-1).$$

(iii) Prove that

$$\begin{aligned}(p-1)|\tau_p|^2 &= \sum_{a=1}^{p-1}\left|\left(\frac{a}{p}\right)_L \tau_p\right|^2 \\ &= \sum_{a=1}^{p-1}|S(p,a)|^2 \\ &= p(p-1),\end{aligned}$$

whence $|\tau_p| = \sqrt{p}$.

5.5 Notes

§1. Fermat and Euler had studied questions which in modern terminology can be described in terms of the solubility of quadratic congruences. A. M. Legendre's eponymous symbol was introduced by him in "Essai sur la théorie des nombres", Paris, 1798, p. 186. I. M. Vinogradov made his conjecture on the least quadratic non-residue in "On the distribution of quadratic residues and non-residues", Zh. Fiz,-Mt. Obshch. Univ. Perm 2, 1-16, 1919. The estimate of D. A. Burgess's result is in "The distribution of quadratic residues and non-residues", Mathematika 4(1957), 106-112.

§2. Euler (1983) had formulated a conjecture that if we take the primes p in the residue class r modulo $4m$, then the residue class m modulo p is always a QR modulo p or always a QNR modulo p and moreover $4m-r$ is the same. That is, when $p \nmid 4m$,

$$\left(\frac{m}{p}\right)_L = \left(\frac{m}{p}\right)_L$$

depends only on the residue class in which p lies modulo $4m$, and is the same for primes in the residue class $4m-r$. This follows at once from the LQR in our modern formulation. The first correct proof is due to Gauss (1796). This was before Legendre invented his symbol and Gauss used the much clumsier notation aRp and aNp to indicate whether a was a quadratic residue modulo p or a quadratic non-residue.

§3. Jacobi defined his symbol in C. G. J. Jacobi (1837), "Über die Kreisteilung und ihre Anwendung auf die Zahlentheorie", Bericht Ak. Wiss. Berlin, 127–136.

§4. The investigation of the distribution of patterns of k consecutive QR and QNR is intimately connected with questions concerning the zeros of the zeta function of curves $y^2 = f(x)$ over finite fields. See the article on "Quadratic residue patterns modulo a prime" by Keith Conrad at https://kconrad.math.uconn.edu/blurbs/

Exercise 5.5.3 shows that the sum

$$S(p,a) = \sum_{x=1}^{p} e(ax^2/p)$$

is closely related to τ_p. Gauss showed that $\tau_p = \sqrt{p}$ when $p \equiv 1 \pmod 4$ and $\tau_p = i\sqrt{p}$ when $p \equiv 3 \pmod 4$ and used this as the basis of one of his proofs of LQR.

We know less about the sums

$$S_k(a,p) = \sum_{x=1}^{p} e(ax^k/p).$$

We do know that if $p \nmid a$, then

$$|S_k(p,a)| \leq \big((k,p-1)-1\big)\sqrt{p}.$$

but in general we do not know how

$$p^{-1/2}S_k(p,a)$$

is distributed. In a few cases, especially the cubic case when $p \equiv 1 \pmod 3$ it is known that the argument is "uniformly distributed". See D. R. Heath-Brown, "Kummer's conjecture for cubic Gauss sums", Israeli. J. Math. 120(2000), 97–124 and the reference to the earlier paper of Heath-Brown and Patterson.

Chapter 6

Sums of Squares

6.1 Some Evidence

The basic results on sums of squares depend on the theory of quadratic residues, so this chapter is a natural continuation of the previous one.

Example 6.1.

1	0^2+1^2	$0^2+0^2+0^2+1^2$	13	2^2+3^2	$0^2+0^2+3^2+3^2$
2	1^2+1^2	$0^2+0^2+1^2+1^2$	17	1^2+4^2	$0^2+0^2+1^2+4^2$
3		$0^2+1^2+1^2+1^2$	19		$1^2+1^2+1^2+4^2$
4	0^2+2^2	$0^2+0^2+0^2+2^2$	23		$1^2+2^2+3^2+3^2$
5	1^2+2^2	$0^2+0^2+1^2+2^2$	29	2^2+5^2	$0^2+0^2+2^2+5^2$
6		$0^2+1^2+1^2+2^2$	31		$1^2+1^2+2^2+5^2$
7		$1^2+1^2+1^2+2^2$	37	1^2+6^2	$1^2+1^2+1^2+2^2$
8	2^2+2^2	$0^2+0^2+2^2+2^2$	41	4^2+5^2	$0^2+0^2+4^2+5^2$
9	0^2+3^2	$0^2+1^2+2^2+2^2$	43		$1^2+1^2+4^2+5^2$
10	1^2+3^2	$0^2+0^2+1^2+3^2$	47		$1^2+1^2+3^2+6^2$
11		$0^2+1^2+1^2+3^2$	53	2^2+7^2	$0^2+0^2+2^2+7^2$
12		$1^2+1^2+1^2+3^2$	59		$0^2+1^2+3^2+7^2$

So it looks like every number is the sum of four squares and it seems that the primes $p \equiv 1 \pmod 4$ always have a representation, but those $\equiv 3 \pmod 4$ never have one. But what about general n? Fermat found a rule which tells us precisely which numbers are the sum of two squares. Euler tried very hard unsuccessfully to prove the four square theorem, which I find very surprising because one can adapt Fermat's work on two squares. Eventually Lagrange proved the four square theorem.

Example 6.2. *By the way, although we won't use it, one can see the glimmerings of algebraic number theory. If $p = x^2 + y^2$, then we can write it as*

$$p = (x + iy)(x - iy)$$

where $i = \sqrt{-1}$. Also you might guess that quaternions are relevant to sums of four squares.

6.2 Sums of Two Squares

Let us start by considering the solubility of $p = x^2 + y^2$ where p is an odd prime and x and y are integers.

If we had $p|y$, then we would have to have $p|x$, but then the right hand side would be divisible by p^2, which is obvious nonsense. Thus we may assume that $p \nmid y$. If we rewrite the equation as $x^2 = p - y^2$, then we have $x^2 \equiv -y^2 \pmod{p}$. Thus $-y^2$ has to be a QR modulo p. Hence

$$1 = \left(\frac{-y^2}{p}\right)_L = \left(\frac{-1}{p}\right)_L = (-1)^{\frac{p-1}{2}}$$

by Euler's criterion. Thus $p \equiv 1 \pmod{4}$, and we have proved one half of the following theorem.

Theorem 6.1 (Fermat)**.** *An odd prime p is the sum of two squares if and only if $p \equiv 1 \pmod{4}$.*

Proof. It remains to prove that if $p \equiv 1 \pmod{4}$, then p is the sum of two squares. We give a rather slick proof due to Axel Thue, a Norwegian mathematician. We know that -1 is a QR. Choose Z so that $Z^2 \equiv -1 \pmod{p}$. Consider the

$$(1 + \lfloor\sqrt{p}\rfloor)^2 > (\sqrt{p})^2 = p$$

numbers $xZ + y$ with $0 \le x < \sqrt{p}$ and $0 \le y < \sqrt{p}$ (note that since a prime is not a perfect square you cannot have equality). Here $\lfloor * \rfloor$ is defined in Definition 5.3. Since there are more than p of them, there must be a residue class modulo p which contains at least two of them (the Dirichlet Box Principle (sometimes called the *pigeon hole principle*, or *Schubfachprinzip*). That is, we have $x_1Z + y_1 \equiv x_2Z + y_2 \pmod{p}$, and since the pairs x_1, y_1 and x_2, y_2 are different we have $xZ + y \equiv 0 \pmod{p}$ with $|x| = |x_1 - x_2| < \sqrt{p}$, $|y| = |y_1 - y_2| < \sqrt{p}$ and x and y not both 0. Now $x^2 + y^2 \equiv x^2 + (-xZ)^2 = x^2(Z^2 + 1) \equiv 0 \pmod{p}$. Moreover $0 < x^2 + y^2 < p + p = 2p$. Hence $x^2 + y^2 = p$. □

Example 6.3. *Sums of two squares have a remarkable multiplicative property. Consider the following table.*

2	$1^2 + 1^2$	26	$1^2 + 5^2$	68	$2^2 + 8^2$	100	$6^2 + 8^2$
4	$0^2 + 2^2$	29	$2^2 + 5^2$	72	$6^2 + 6^2$	104	$2^2 + 10^2$
5	$1^2 + 2^2$	34	$3^2 + 5^2$	74	$5^2 + 7^2$	106	$5^2 + 9^2$
8	$2^2 + 2^2$	40	$2^2 + 6^2$	80	$4^2 + 8^2$	116	$4^2 + 10^2$
9	$0^2 + 3^2$	45	$3^2 + 6^2$	81	$0^2 + 9^2$	117	$6^2 + 9^2$
10	$1^2 + 3^2$	50	$5^2 + 5^2$	82	$1^2 + 9^2$	122	$1^2 + 11^2$
13	$2^2 + 3^2$	52	$4^2 + 6^2$	85	$2^2 + 9^2$	125	$5^2 + 10^2$
20	$2^2 + 4^2$	58	$3^2 + 7^2$	90	$3^2 + 9^2$	128	$8^2 + 8^2$
25	$0^2 + 5^2$	65	$1^2 + 8^2$	98	$7^2 + 7^2$	130	$3^2 + 11^2$

This looks as though, if a number n has a factorisation ab with both a and b being sums of two squares, then n is also the sum of two squares. For example $117 = 9 \times 13$ and both 9 and 13 are sums of two squares.

Example 6.4. *It turns out that there is a neat identity which proves this. Given x, y, X and Y we have*

$$(x^2 + y^2)(X^2 + Y^2) = (xX - yY)^2 + (xY + yX)^2. \tag{6.1}$$

The simplest proof is to multiply out both sides

$$x^2X^2 + x^2Y^2 + y^2X^2 + y^2Y^2,$$

$$x^2X^2 - 2xXyY + y^2Y^2 + x^2Y^2 + 2xYyX + y^2X^2$$

and observe that the cross product terms on the right cancel and then the two sides are equal. Another way of seeing this identity is to write it as

$$\begin{aligned}(x^2 + y^2)(X^2 + Y^2) = |x + iy|^2|X + iY|^2 = |(x + iy)(X + iY)|^2 \\ = |xX - yY + i(xY + yX)|^2 = (xX - yY)^2 + (xY + yX)^2.\end{aligned}$$

Now we can prove Fermat's theorem

Theorem 6.2 (Fermat)**.** *Let n have the canonical decomposition*

$$n = p_1^{a_1} \dots p_r^{a_2} q_1^{b_1} \dots q_s^{b_s}$$

where the q_j are the primes in the factorisation with $q_j \equiv 3 \pmod 4$ and the p_j are the prime 2 (if n is even) and the primes $p_j \equiv 1 \pmod 4$. Then n is the sum of two squares if and only if all the exponents b_j are even.

Proof. If n satisfies the necessary condition, then the conclusion follows from repeated use of the identity and the special cases $p = 2$, $p \equiv 1 \pmod 4$ and q^2 which we already know.

To prove the converse observe that if q is a prime with $q \equiv 3 \pmod 4$ and $n = x^2 + y^2 \equiv 0 \pmod q$, then we must have $q|x$ and $q|y$, for if not, then

$$1 = \left(\frac{-y^2}{q}\right)_L = \left(\frac{-1}{q}\right)_L = -1$$

which is absurd. Hence n/q^2 is a sum of two squares and we can use an inductive argument (Fermat used a "descent argument" - the modern equivalent is the "well ordering principle"). □

6.2.1 Exercises

1. Suppose that p is an odd prime and define

$$S(a) = \sum_{x=1}^{p} \left(\frac{x^3 + ax}{p} \right)_L .$$

(i) Show that if $p \nmid r$ and $a \equiv r^2 b \pmod{p}$, then $S(a) = \left(\frac{r}{p}\right)_l S(b)$.

(ii) Show that for any quadratic non-residue n modulo p we have

$$\sum_{a=1}^{p} |S(a)|^2 = \frac{p-1}{2}|S(1)|^2 + \frac{p-1}{2}|S(n)|^2.$$

(iii) Show that

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

Formula (5.4) is useful here.

(iv) Show that for any a, $S(a)$ is an even integer.

(v) Show that if $p \equiv 1 \pmod{4}$, then for any quadratic non-residue n modulo p,

$$|S(1)/2|^2 + |S(n)/2|^2 = p,$$

giving an explicit representation of p as the sum of two squares, and show that if $p \equiv 3 \pmod{4}$, then, for any integer a, $S(a) = 0$.

6.3 Binary Quadratic Forms

It is also possible to show a similar result for numbers of the form

$$x^2 + 2y^2$$

and likewise for

$$x^2 + 3y^2.$$

The general rule here is that if -2 (or -3 in the second case) is a QR modulo p, then p can be represented and there is an identity $(x^2 + \lambda y^2)(X^2 + \lambda Y^2) = (xX - \lambda yY)^2 + \lambda(xY + yX)^2$ which works in both cases.

In the case of $x^2 + 2y^2$ Thue's argument shows that if $p \equiv 1$ or $3 \pmod{8}$, then there are x and y such that $x^2 + 2y^2 = mp$ with $m = 1$ or 2. But if $m = 2$, then $2|x$ and the equation reduces to $2(x/2)^2 + y^2 = p$.

For the form $x^2 + 3y^2$, when $p \equiv 1 \pmod{3}$, Thue reduces to $x^2 + 3y^2 = mp$ with $m = 1, 2$ or 3. $m = 3$ can be dealt with as before. $m = 2$ cannot happen because when

$p > 2$ one cannot have $2|xy$, so the left hand side is $\equiv 1 + 3 \equiv 4 \pmod 8$ and 4 does not divide $2p$.

This phenomenon does not occur for more general binary quadratic forms

$$ax^2 + bxy + cy^2$$

because it is possible in most cases that $D = b^2 - 4ac$ is a QR modulo p, but the form does not represent p. It turns out there is a different form with the same value of discriminant D which represents p.

Example 6.5. *In the case when $D = -20$, there are basically two forms, (everything else with that discriminant can be reduced to them) $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$. The discriminant -20 is a QR for $p = 7$ and $p = 29$, but only the second form represents 7 and only the first one represents 29. This is related to the "class number problem", and the fact that the quadratic number field $\mathbb{Q}(\sqrt{-5})$ fails to have uniqueness of factorisation. This phenomenon was extensively studied by Gauss in Disquisitiones Arithmeticae in 1798 (he was 21). It is a very elegant theory. First, in modern notation, one can write*

$$ax_1^2 + bx_1x_2 + cx_2^2 = \mathbf{x}A\mathbf{x}^T$$

where $\mathbf{x}$ denotes the vector (x_1, x_2), $\mathbf{x}^T$ its transpose and A is the matrix

$$A = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}.$$

If the 2×2 matrix U has integer entries and determinant $\det U = \pm 1$, then it is invertible and the inverse matrix has integer entries. Thus

$$\mathbf{x}UAU^T\mathbf{x}^T$$

will represent the same integers as $\mathbf{x}A\mathbf{x}^T$. Hence one can divide the forms $ax_1^2 + bx_1x_2 + cx_2^2$, i.e. matrices A, with a given discriminant $D = -4 \det A$, into "equivalence classes". The number of different equivalence classes is called the class number $h(D)$. There is a canonical or "reduced" form - in which the coefficients satisfy a certain minimality condition - which is normally taken to be the representative of the class.

Example 6.6. *When $D = -20$ the class number $h(-20) = 2$ and the two reduced forms are $x_1^2 + 5y_2^2$ and $2x_1^2 + 2x_1x_2 + 3x_2^2$.*

In the modern era the subject of binary quadratic forms is subsumed in the study of quadratic number fields $\mathbb{Q}(\sqrt{D})$.

6.3.1 Exercises

1. Let $\mathcal{P}$ denote the set of odd primes p for which -7 is a quadratic residue modulo p.

(i) Prove that $p \in \mathcal{P}$ if and only if p is odd and $p \equiv 1$, 2 or 4 (mod 7). Note that if $p \in \mathcal{P}$, then $p > 7$.

(ii) Prove that if $p \in \mathcal{P}$, then there are x, y, m such that

$$x^2 + 7y^2 = mp$$

and $1 \leq m \leq 7$.

(iii) Prove that if $m = 7$, then $7|x$, and $y^2 + 7(x/7)^2 = p$.

(iv) Prove that if x and y are both odd, then $x^2 + 7y^2 \equiv 0 \pmod 8$. Deduce that in (ii) x and y cannot both be odd.

(v) Prove that if x and y are both even, then $x^2 + 7y^2 \equiv 0 \pmod 4$ and so in (ii) m would be 4. Deduce that $(x/2)^2 + 7(y/2)^2 = p$.

(vi) That leaves $m = 1$, 3 or 5 in (ii). Prove that if $m = 3$ or 5 in (ii), then $(m, xy) = 1$.

(viii) Prove that if $(3, xy) = 1$, then $x^2 + 7y^2 \not\equiv 0 \pmod 3$ and hence that in (ii) $m \neq 3$.

(viii) Prove that if $(5, xy) = 1$, then $x^2 + 7y^2 \not\equiv 0 \pmod 5$ and hence that in (ii) $m \neq 5$.

(xi) Prove that $p \in \mathcal{P}$ if and only if $p \neq 7$ and there are x and y such that $x^2 + 7y^2 = p$.

(x) Prove that there are infinitely many primes in $\mathcal{P}$.

This is a curious example. The discriminant of the form $x^2 + 7y^2$ is $-4.1.7 = -28$ and there is another "reduced" form $2x^2 + 2xy + 4y^2$ with the same discriminant. However this form represents only even numbers, so the only prime it can represent is 2, which in some sense is one of the "missing" primes from $\mathcal{P}$. It also "imprimitive" in the sense that the coefficients have a common factor greater than 1. The other missing prime is 7, which *is* represented by $x^2 + 7y^2$.

2. (i) Prove that if $p \neq 2$, then $\left(\frac{-5}{p}\right)_L = 1$ iff $p \equiv 1, 3, 7$ or 9 (mod 20).

(ii) List those $n \leq 25$ for which $x^2 + 5y^2 = n$ is soluble in integers x and y. Are there any primes of the form $p \equiv 1, 3, 7$ or 9 (mod 20) for which $x^2 + 5y^2 = p$ is insoluble in x and y? Which of them are represented by $2x^2 + 2xy + 3y^2$?

3. Find all solutions to the diophantine equation $x^2 + y^2 = 3z^2 + 3t^2$.

4. Find all solutions to the diophantine equation

$$2x^2 + 3y^2 = 26z^2 + 39t^2.$$

6.4 Sums of Four Squares

The proof of Lagrange's four square theorem has a similar structure. As for two squares there is an identity, discovered by Euler, which just as the two square example is related to complex numbers the four square example is related to quaternions.

Theorem 6.3 (Euler's four squares identity)**.** *For any numbers*

$$a, b, c, d, w, x, y, z,$$

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

Proof. I am not sure how Euler discovered this. Of course, as Littlewood said, "all identities are trivial". One could check it by multiplying both sides out. Here is an alternative. Think of it as a polynomial in the variable x. The coefficient of x^2 on both sides is $a^2+b^2+c^2+d^2$. The coefficient if x on the left is obviously 0, and a little checking shows that the x-terms on the right cancel. That leaves the "constant" term. To check that put $x=0$ and repeat the argument with y. And then z, and finally w. □

Now we can prove Lagrange's theorem.

Theorem 6.4 (Lagrange)**.** *Every natural number is the sum of four squares.*

Proof. In view of Euler's identity and $1^2+1^2=2$, it suffices to prove that every odd prime is such a sum.

Lemma 6.5. *If n is even and is a sum of four squares, then so is $\frac{n}{2}$.*

Proof of Lemma 6.5. When $n=a^2+b^2+c^2+d^2$ is even, an even number of the squares will be odd. and so the a, b, c, d can be rearranged so that a, b have the same parity and so do c, d. Thus $\frac{n}{2}=\left(\frac{a+b}{2}\right)^2+\left(\frac{a-b}{2}\right)^2+\left(\frac{c+d}{2}\right)^2+\left(\frac{c-d}{2}\right)^2$. □

Lemma 6.6. *If p is an odd prime, then there are integers a, b, c, d and an m so that $0<a^2+b^2+c^2+d^2=mp<\frac{p^2}{2}$.*

Proof of Lemma 6.6. The $\frac{p+1}{2}$ numbers

$$0^2, 1^2, \ldots, \left(\frac{p-1}{2}\right)^2$$

are pairwise incongruent modulo p. Thus the $\frac{p+1}{2}$ numbers u^2 with $0\le u\le\frac{p-1}{2}$ will lie in separate residue classes modulo p and the $\frac{p+1}{2}$ numbers $-v^2-1$ with $0\le v\le\frac{p-1}{2}$ will lie in separate residue classes modulo p. Since $\frac{p+1}{2}+\frac{p+1}{2}=p+1>p$ there will be atleast one residue class which contains one of each. Hence there are u, v such that $u^2\equiv -v^2-1 \pmod p$. They cannot both be 0 and so $0<u^2+v^2+1\le\frac{p^2-2p+3}{2}<\frac{p^2}{2}$. □

We could have proved this lemma by recalling the result we proved in chapter 5 on the number $N(p,c)$ of solutions of $x^2+y^2 \equiv c \pmod{p}$.

By Lemma 6.6 there is an integer m with $0 < m < p$ so that for some a,b,c,d we have

$$a^2+b^2+c^2+d^2 = mp$$

and we may suppose that m is chosen minimally. Moreover, by Lemma 6.5 we may suppose that m is odd. If $m=1$, then we are done. Suppose $m>1$. If m were to divide each of a,b,c,d, then we would have $m|p$ contradicting $m<p$. Choose x,y,z,w so that $x \equiv a \pmod{m}$, $|x| \le \frac{m-1}{2}$, $y \equiv -b \pmod{m}$, $|y| \le \frac{m-1}{2}$, $z \equiv -c \pmod{m}$, $|z| \le \frac{m-1}{2}$, $w \equiv -d \pmod{m}$, $|w| \le \frac{m-1}{2}$, and then not all of x,y,z,w can be 0. Moreover $x^2+y^2+z^2+w^2 \equiv 0 \pmod{m}$ and so $0 < x^2+y^2+z^2+w^2 = mn \le 4\left(\frac{m-1}{2}\right)^2 = (m-1)^2$. Thus $0<n<m$. Now $ax-by-cz-dw \equiv a^2+b^2+c^2+d^2 \equiv 0 \pmod{m}$, $ay+bx+cw-dz \equiv -ab+ab-cd+dc \equiv 0 \pmod{m}$, $az+cx+dy-bw \equiv -ac+ac-db+db \equiv 0 \pmod{m}$, $aw+dx+bz-cy \equiv -ad+ad-bc+bc \equiv 0 \pmod{m}$. By Euler's identity m^2np is the sum of four squares and each of the squares is divisible by m^2. Hence np is the sum of four squares. But $n<m$ contradicting the minimality of m. □

6.4.1 Exercises

1. Throughout $R(n)$ denotes the number of solutions of the equation

$$x_1^2+x_2^2+x_3^2+x_4^2 = n$$

in integers x_1, x_2, x_3, x_4.

(i) Prove that if $k \ge 3$, then $R(2^k) = R(2^{k-2})$.

(ii) Prove that $R(2^k) = 24$ when k is odd and 32 when k is even.

(iii) In this part it is necessary to have some familiarity with the notation of §7.3 and ideas of Theorem 7.11 below. Prove that if $X \ge 1$, then

$$\sum_{n \le X} R(n) = \frac{\pi^2}{2}X^2 + O(X^{3/2}).$$

We remark that although the last part shows that the average value of $R(n)$ is about $\pi^2 n$ and so is unbounded the earlier parts show that infinitely often $R(n)$ is bounded by 24. Is it, therefore, perhaps a fluke that $R(n) > 0$ for all positive n?

6.5 Three Squares?

Ok, so many numbers are not the sum of two squares and every number is the sum of four, so what about sums of three squares? This is quite hard and was first solved by Legendre in 1798. In the case of two squares we saw that the primes $p \equiv 3 \pmod{4}$, when they occur to an odd power, were excluded by a simple congruence argument, $x^2+y^2 \equiv 0 \pmod{p}$ with $p \nmid xy$ which requires -1 to be a quadratic residue modulo p.

Example 6.7. *We know that* $x^2 \equiv 0,\ 1$ *or* $4 \pmod 8$. *Thus one can check that*

$$x_1^2 + x_2^2 + x_3^2 \equiv 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6, \pmod 8$$

but

$$x_1^2 + x_2^2 + x_3^2 \not\equiv 7 \pmod 8.$$

Thus if $x_1^2+x_2^2+x_3^2 = n$, *then we have to have* $n \equiv 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6, \pmod 8$. *Moreover if*

$$x_1^2 + x_2^2 + x_3^2 \equiv 0 \pmod 4,$$

then the variables x_j *have to be all even and we can factor out a* 4 *on both sides and reduce to*

$$(x_1/2)^2 + (x_2/2)^2 + (x_3/2)^2 = n/4.$$

Then we have just proved

Theorem 6.7. *If* $n = 4^h(8k + 7)$ *for some non-negative integers* h *and* k, *then* n *is not the sum of three squares.*

Legendre proved that all other n *are* the sum of three squares. The proof is quite complicated and I do not plan to give it here.

6.6 Other Questions

Given a positive integer s, how many ways are there of writing n as the sum of two squares of integers? We count $(-x)^2$ separately from x^2 when $x \neq 0$. Suppose $z \in \mathbb{C}$ and $|z| < 1$. Consider the series

$$f(z) = \sum_{n=-\infty}^{\infty} z^{n^2} = 1 + 2\sum_{n=1}^{\infty} z^{n^2}.$$

Then formally

$$f(z)^s = \sum_{n_1} \cdots \sum_{n_s} z^{n_1^2+\cdots+n_s^2} = \sum_{n=0}^{\infty} r_s(n) z^n$$

where $r_s(n)$ is the number of ways of writing n as the sum of s squares. The function $f(z)$ has lots of structure and this can be used to find formulas for $r_s(n)$, and was exploited extensively by Jacobi.

In 1770 Edward Waring stated without proof that "every positive integer is the sum of at most four squares, nine cubes, nineteen biquadrates, and so on". What we think he meant was that if we define $g(k)$ to be the smallest number s such that every positive integer is the sum of at most s k-th powers, then $g(2) = 4$, $g(3) = 9$, $g(4) = 19$. Many

mathematicians have worked on Waring's Problem, including Hilbert, Landau, Hardy, Littlewood, Davenport, ... What we believe is that

$$g(k) = 2^k + \left\lfloor \left(\frac{3}{2}\right)^k \right\rfloor - 2$$

and we know that this is true for all but at most a finite number of exceptions, and there are no exceptions with $k \le 471,600,000$.

The value of $g(k)$ depends on the peculiarities of a few small numbers, and probably

$$\begin{aligned} n &= 2^k \left\lfloor \left(\frac{3}{2}\right)^k \right\rfloor - 1 \\ &= \left(\left\lfloor \left(\frac{3}{2}\right)^k \right\rfloor - 1 \right) \times 2^k + \left(2^k - 1\right) \times 1^k \end{aligned}$$

is extremal because one can only use 2^k and 1^k.

A harder problem which avoids the peculiarities of small numbers, is to take $G(k)$ to be the smallest s such that every *sufficiently large* integer is the sum of at most s k-th powers. This has only been solved in two cases, $G(2) = 4$ (Lagrange) and $G(4) = 16$ (Davenport). For example we only know that $4 \le G(3) \le 7$ (Linnik) and $6 \le G(5) \le 17$ (RCV and Wooley).

6.6.1 Exercises

1. Prove that the number

$$n = 2^k \left\lfloor \left(\frac{3}{2}\right)^k \right\rfloor - 1$$

cannot be represented by the sum of fewer than

$$2^k + \left\lfloor \left(\frac{3}{2}\right)^k \right\rfloor - 2$$

positive k-th powers.

2. (i) Prove that for any x we have $x^4 \equiv 0$ or $1 \pmod{16}$.

(ii) Prove that $G(4) \ge 16$.

6.7 Notes

§§1 and 2. The provenance of Theorems 6.1 and 6.2 is clouded by time. Sums of two squares were investigated by Diophantus in the 3rd century AD. Possibly the first person

to articulate the assertion of these theorems is Albert Girard (1595-1632), but the identity of Example 6.4 that connects the two was known to Brahmagupta and Fibonacci. Fermat gave a more detailed version of Theorem 6.2 in a letter to Marin Mersenne dated 25th December 1640, but he wrote down no proof. Nevertheless it would have been within his compass. The first published proof is by Euler using the "method of descent". This is a version of the well ordering principle and is equivalent to proof by induction. In over all structure it would have been similar to the proof given, except that the value of m which starts the argument could be larger. The proof was discovered in the late 1740s, but was not published for some years.

Binary quadratic forms were first studied systematically by Lagrange in 1775, and Legendre added to this theory. However Gauss radically developed the subject in §5 of his Disquisitiones Arithmeticae (1801). Although entirely classical in style it can be considered the starting point for modern algebraic number theory and abstract algebra. For modern accounts of binary quadratic forms see D. A. Buell (1989), "Binary Quadratic Forms", Springer, New York, or J. Buchmann & U. Vollmer (2007), "Binary Quadratic Forms", Springer, Berlin.

§3. Lagrange's theorem is sometimes known as Bachet's conjecture, but there are examples in Diophantus' "Arithmetica" which indicate that the assertion was known in antiquity.

§4. Legendre had also made an assertion about the number of solutions for n which are the sum of three squares, but his proof of that turned out to be faulty and the first correct proof was given, of course, by Gauss in *Disquisitiones Mathematicae*.

§5. The function f was used by Jacobi to obtain formulæ for $r_s(n)$ for various values of s. Thus

$$r_2(n) = 4\sum_{m|n} \chi_1(m)$$

where

$$\chi_1(n) = \begin{cases} (-1)^{\frac{n-1}{2}} & (2 \nmid n), \\ 0 & (2|n) \end{cases}$$

and

$$r_4(n) = 8\sum_{\substack{m|n \\ 4\nmid m}} m = \begin{cases} 8\sigma(n) & (4 \nmid n) \\ 8\sigma(n) - 32\sigma(n/4) & (4|n) \end{cases}$$

where

$$\sigma(l) = \sum_{m|l} m.$$

For more about χ_1 and σ see Definition 7.7 *et seq.* and Exercise 7.2.1.1.

Chapter 7

Arithmetical Functions

7.1 Introduction

It is convenient to make the following definition.

Definition 7.1. *Let $\mathcal{A}$ denote the set of arithmetical functions, that is the functions defined by*

$$\mathcal{A} = \{f : \mathbb{N} \to \mathbb{C}\}.$$

Of course the range of any particular function might well be a subset of $\mathbb{C}$, such as $\mathbb{R}$ or $\mathbb{Z}$. There are quite a number of important arithmetical functions. Some examples are

Definition 7.2 (**The divisor function**). *The number of positive divisors of n.*

$$d(n) = \sum_{m|n} 1.$$

Definition 7.3 (**The Möbius function**). *This is a more peculiar function. It is defined by*

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct primes,} \\ 0 & \text{if there is a prime } p \text{ such that } p^2|n. \end{cases}$$

It is also convenient to introduce three very boring functions.

Definition 7.4 (**The Unit**).

$$e(n) = \begin{cases} 1 & (n = 1), \\ 0 & (n > 1). \end{cases}$$

Definition 7.5 (**The One**).

$$\mathbf{1}(n) = 1 \text{ for every } n.$$

Definition 7.6 (**The Identity).**

$$N(n) = n.$$

Two other functions which have interesting structures but which we will say less about at this stage are

Definition 7.7 (**The primitive character modulo** 4**).** *We define*

$$\chi_1(n) = \begin{cases} (-1)^{\frac{n-1}{2}} & 2 \nmid n, \\ 0 & 2|n. \end{cases}$$

Similar functions we have already met are Euler's function ϕ, the Legendre symbol and its generalization the Jacobi symbol

$$\left(\frac{n}{m}\right)_J.$$

Here we think of it as a function of n, keeping m fixed, but we could also think of it as a function of m keeping n fixed.

Definition 7.8 (**Sums of two squares).** *We define $r(n)$ to be the number of ways of writing n as the sum of two squares of integers.*

Example 7.1. *For example, $1 = 0^2 + (\pm 1)^2 = (\pm 1)^2 + 0^2$, so $r(1) = 4$, $r(3) = r(6) = r(7) = 0$, $r(9) = 4$, $65 = (\pm 1)^2 + (\pm 8)^2 = (\pm 4)^2 + (\pm 7)^2$ so $r(65) = 16$. This is the function $r_2(n)$ of the previous chapter.*

The functions d, ϕ, e, $\mathbf{1}$, N, χ_1, $\left(\frac{\cdot}{m}\right)_J$ have an important property. That is that they are multiplicative. We already discussed this in connection with Euler's function and the Legendre and Jacobi symbols. Here is a reminder.

Definition 7.9. *An arithmetical function f which is not identically 0 is* **multiplicative** *when it satisfies*

$$f(mn) = f(m)f(n) \tag{7.1}$$

whenever $(m, n) = 1$. Let $\mathcal{M}$ denote the set of multiplicative functions. If (7.1) holds for **all** *m and n, then we say that f is* **totally multiplicative**.

The function $r(n)$ is not multiplicative, since $r(65) = 16$ but $r(5) = r(13) = 8$. Indeed the fact that $r(1) \neq 1$ would contradict the next theorem. However it is true that $r(n)/4$ is multiplicative, but this is a little trickier to prove.

Theorem 7.1. *Suppose that $f \in \mathcal{M}$. Then $f(1) = 1$.*

Proof. Since f is not identically 0 there is an n such that $f(n) \neq 0$. Hence $f(n) = f(n \times 1) = f(n)f(1)$, and the conclusion follows. □

It is pretty obvious that e, $\mathbf{1}$ and N are in $\mathcal{M}$, and it is actually quite easy to show

Theorem 7.2. *We have $\mu \in \mathcal{M}$.*

Proof. Suppose that $(m,n)=1$. If $p^2|mn$, then $p^2|m$ or $p^2|n$, so $\mu(mn)=0=\mu(m)\mu(n)$. If

$$m = p_1 \dots p_k, \quad n = p_1' \dots p_l'$$

with the p_i, p_j' distinct, then

$$\mu(mn) = (-1)^{k+l} = (-1)^k(-1)^l = \mu(m)\mu(n).$$

□

The following is very useful.

Theorem 7.3. *Suppose the $f \in \mathcal{M}$, $g \in \mathcal{M}$ and h is defined for each n by*

$$h(n) = \sum_{m|n} f(m)g(n/m).$$

Then $h \in \mathcal{M}$.

Proof. Suppose $(n_1,n_2)=1$. Then a typical divisor m of n_1n_2 is uniquely of the form m_1m_2 with $m_1|n_1$ and $m_2|n_2$. Hence

$$\begin{aligned} h(n_1n_2) &= \sum_{m_1|n_1} \sum_{m_2|n_2} f(m_1m_2)g(n_1n_2/(m_1m_2)) \\ &= \sum_{m_1|n_1} f(m_1)g(n_1/m_1) \sum_{m_2|n_2} f(m_2)g(n_2/m_2). \end{aligned}$$

□

This enables us to establish an interesting property of the Möbius function.

Theorem 7.4. *We have*

$$\sum_{m|n} \mu(m) = e(n).$$

Proof. By the definition of $\mathbf{1}$ the sum here is

$$\sum_{m|n} \mu(m)\mathbf{1}(n/m)$$

and so by the previous theorem it is in $\mathcal{M}$. Moreover if $k \geq 1$, then

$$\sum_{m|p^k} \mu(m) = \mu(1) + \mu(p) = 1 - 1 = 0$$

□

7.1.1 Exercises

1. Show that

$$\left(\sum_{m|n} d(m)\right)^2 = \sum_{m|n} d(m)^3.$$

2. (i) Show that

$$\sum_{l|(m,n)} \mu(l)$$

is 1 when $(m,n)=1$ and is 0 otherwise.

(ii) Prove that

$$\sum_{\substack{m=1 \\ (m,n)=1}}^{n} m = \frac{1}{2} n\phi(n) \quad \text{when} \quad n>1.$$

3. A *squarefree* number is one which has no square other than 1 dividing it. Let $s(n)$ denote the characteristic function of the squarefree numbers.

(i) Prove that

$$s(n) = \sum_{m^2|n} \mu(m).$$

(ii) Prove that $s(n)$ is multiplicative.

4. A positive integer n is perfect when $\sigma(n) = 2n$.

(i) (Euclid) Prove that if $2^{(l+1)}-1$ is prime, then $2^l(2^{l+1}-1)$ is perfect.

(ii)(Euler) Suppose that $n = 2^l m$, m odd, is an even perfect number. Prove that $\sigma(m) = m + \frac{m}{2^{l+1}-1}$. Prove that m has exactly two positive divisors and so is prime, and that $m = 2^{l+1}-1$.

(iii) Prove that there is no squarefree perfect number apart from 6.

5. Show that the only totally multiplicative function f for which $\sum_{m|n} f(m)$ is totally multiplicative is the unit e.

6. Prove that for every positive integer n,

$$\sum_{m|n} \mu(m)d(m) = (-1)^{\omega(n)},$$

where $\omega(n)$ is the number of different prime factors of n, as defined in §7.5.

7. Show that the sum of all the primitive roots modulo p lies in the residue class $\mu(p-1)$ modulo p.

8. Let $k \in \mathbb{N}$. Prove that there are infinitely many n such that $\mu(n+1) = \mu(n+2) = \cdots = \mu(n+k)$.

9. Determine the arithmetic function f such that for every natural number n we have $\mu(n) = \sum_{m|n} f(m)$, i.e. is it multiplicative, and what are its values on the prime powers?

10. Show that every odd number n can be written as the difference of two squares, $n = x^2 - y^2$. How many different choices for the integers x and y are there?

11. Suppose that $n \geq 2$ and n has the distinct prime factors $p_1, p_2, \ldots, p_r$. Show that

$$\sum_{\substack{m=1 \\ (m,n)=1}}^{n} m = \frac{1}{2}\phi(n)n$$

and

$$\sum_{\substack{m=1 \\ (m,n)=1}}^{n} m^2 = \frac{1}{3}\phi(n)n^2 + \frac{1}{6}(-1)^r\phi(n)p_1p_2\ldots p_r.$$

12. Show that if n is a natural number, then

$$\prod_{m|n} m = n^{\frac{1}{2}d(n)}.$$

13. Suppose that $f : \mathbb{N} \to \mathbb{Z}$ is a totally multiplicative function with $f(n) = 0$ or ± 1. Prove that

$$\sum_{m|n} f(m) \geq 0$$

and

$$\sum_{m|n^2} f(m) \geq 1.$$

14. (a) Prove that if $x \geq 1$, then

$$\sum_{n \leq x} \mu(n) \left\lfloor \frac{x}{n} \right\rfloor = 1.$$

Here $\lfloor * \rfloor$ is defined in Definition 5.3.
(b) Prove that

$$-1 + 1/x \leq \sum_{n \leq x} \frac{\mu(n)}{n} \leq 1 + 1/x.$$

In fact we know that

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n} = 0,$$

but this is equivalent to the prime number theorem in the sense that if follows from the prime number theorem and there is a relatively simple proof that it implies the prime number theorem.

15.[Schneider] Suppose that $|x| < 1$. (i) Prove that

$$-\sum_{k=1}^{\infty} \frac{\phi(k)}{k} \log(1-x^k) = \frac{x}{1-x}.$$

(ii) Prove that

$$-\sum_{k=1}^{\infty} \frac{\mu(k)}{k} \log(1-x^k) = x.$$

(iii) Prove that if $\omega = \frac{\sqrt{5}-1}{2}$, so that $1/\omega$ is the *golden ratio*, then

$$\sum_{k=1}^{\infty} \frac{\mu(k)-\phi(k)}{k} \log(1-\omega^k) = 1.$$

7.2 Dirichlet Convolution

Theorem 7.3 suggests a general way of defining new functions.

Definition 7.10. *Given two arithmetical functions f and g we define the* **Dirichlet convolution** *$f * g$ to be the function defined by*

$$(f*g)(n) = \sum_{m|n} f(m)g(n/m).$$

Note that this operation is commutative because

$$f*g(n) = \sum_{m|n} f(m)g(n/m) = \sum_{m|n} g(n/m)f(m)$$

and the mapping $m \leftrightarrow n/m$ is a bijection.

It is also quite easy to see that the relation is associative

$$(f*g)*h = f*(g*h).$$

To see this write the left hand side as

$$\sum_{m|n} \left(\sum_{l|m} f(l)g(m/l) \right) h(n/m)$$

and interchange the order of summation and replace m by kl, so that $kl|n$, i.e $l|n$ and $k|n/l$. Thus the above is

$$\sum_{l|n} f(l) \sum_{k|n/l} g(k)h\big((n/l)/k\big) = f*(g*h)(n).$$

Dirichlet convolution has some interesting properties.
1. $f * e = e * f = f$ for any $f \in \mathcal{A}$, so e is really acting as a unit.
2. $\mu * \mathbf{1} = \mathbf{1} * \mu = e$, so μ is the inverse of $\mathbf{1}$, and *vice versa*.
3. Theorem 7.3 tells us that if $f \in \mathcal{M}$ and $g \in \mathcal{M}$, then $f * g \in \mathcal{M}$.
4. Theorem 3.2 says that $\phi * \mathbf{1} = N$.
5. $d = \mathbf{1} * \mathbf{1}$, so $d \in \mathcal{M}$. Hence
6. $d(p^k) = k + 1$ and $d(p_1^{k_1} \dots p_r^{k_r}) = (k_1 + 1) \dots (k_r + 1)$.

Theorem 7.5 (Möbius inversion I)**.** *Suppose that $f \in \mathcal{A}$ and $g = f * \mathbf{1}$. Then $f = g * \mu$.*

Proof. We have

$$g * \mu = (f * \mathbf{1}) * \mu = f * (\mathbf{1} * \mu) = f * e = f.$$

□

Theorem 7.6 (Möbius inversion II)**.** *Suppose that $g \in \mathcal{A}$ and $f = g * \mu$, then $g = f * \mathbf{1}$.*

The proof is similar.

Theorem 7.7. *We have $\phi = \mu * N$ and $\phi \in \mathcal{M}$. Moreover*

$$\phi(n) = n \sum_{m|n} \frac{\mu(m)}{m} = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$$

This gives new proofs of Corollary 3.6 and Theorem 3.7.

Proof. By property 4. and Theorem 7.5 we have

$$\phi = N * \mu = \mu * N.$$

Therefore, by property 3 and Theorem 7.2, $\phi \in \mathcal{M}$. Moreover $\phi(p^k) = p^k - p^{k-1}$ and we are done. □

Theorem 7.8. *Let $\mathcal{D} = \{f \in \mathcal{A} : f(1) \neq 0\}$. Then $\langle \mathcal{D}, * \rangle$ is an abelian group.*

Proof. Of course e is the unit, and closure is obvious. We already checked commutativity and associativity. It remains, given $f \in \mathcal{D}$, to construct an inverse. Define g iteratively by

$$\begin{aligned} g(1) &= 1/f(1) \\ g(n) &= -\sum_{\substack{m|n \\ m>1}} f(m) g(n/m)/f(1) \end{aligned}$$

and it is clear that $f * g = e$. □

7.2.1 Exercises

1. We define $\sigma(n)$ for $n \in \mathbb{N}$ to be the sum of the divisors of n,

$$\sigma(n) = \sum_{m|n} m.$$

(i) Prove that σ is a multiplicative function.
(ii) Evaluate $\sigma(1050)$.
(iii) Prove that

$$\sum_{m|n} \phi(m)\sigma(n/m) = nd(n).$$

(iii) Show that if $\sigma(n)$ is odd, then n is a square or twice a square.
(iv) Prove that

$$\sum_{m|n} \mu(m)\sigma(n/m) = n.$$

(v) Prove that

$$\sum_{m|n} \mu(n/m) \sum_{l|m} \mu(l)\sigma(m/l) = \phi(n).$$

2. (cf Hille (1937)) Suppose that $f(x)$ and $F(x)$ are complex-valued functions defined on $[1, \infty)$. Prove that

$$F(x) = \sum_{n \le x} f(x/n)$$

for all x if and only if

$$f(x) = \sum_{n \le x} \mu(n) F(x/n)$$

for all x.

3. Show for each positive integer k that there is a unique arithmetic function ϕ_k such that $\sum_{m|n} \phi_k(m) = n^k$. Obtain a formula for $\phi_k(n)$ and show that $\phi_k(n)$ is multiplicative.

4. Evaluate $h(n) = \sum_{m|n} (-1)^m \mu(n/m)$.

5. Suppose that the arithmetical function $\eta(n)$ satisfies $\sum_{m|n} \eta(m) = \phi(n)$. Show that $\eta(n)$ is multiplicative and evaluate $\eta(p^k)$.

6. Let $g(n)$ denote the number of ordered k-tuples of integers $x_1, x_2, \ldots, x_k$ such that $1 \le x_j \le n$ $(j = 1, 2 \ldots, k)$ and

$$(x_1, x_2, \ldots, x_k, n) = 1,$$

and let $G(n) = \sum_{m|n} g(m)$. Prove that $G(n) = n^k$ and that

$$g(n) = n^k \prod_{p|n} \left(1 - p^{-k}\right).$$

7. This question investigates whether there exists an arithmetic function θ such that $\theta * \theta = \mu$ and $\theta(1) \geq 0$.

(i) Prove that θ exists and is uniquely determined.

(ii) Prove that

$$\theta(p^k) = (-1)^k \binom{\frac{1}{2}}{k}.$$

This is the coefficient of z^k in the Taylor expansion of $(1-z)^{1/2}$ centred at 0. It is easily checked that

$$\theta(p^k) = -\frac{(2k)!}{2^{2k}(k!)^2} = -\frac{1}{2^{2k}}\binom{2k}{k}.$$

(iii) By considering the function $\theta_1(n) = \prod_{p^k \| n} \theta(p^k)$, or otherwise, show that $\theta \in \mathcal{M}$.

8. Let $s \in \mathbb{N}$. Generalise the results of question 7 to the situation $\theta * \theta * \cdots * \theta = \mu$ where on the left one has the s-fold product.

7.3 Averages of Arithmetical Functions

One of the most powerful techniques we have is to take an average.

Example 7.2. *Suppose we have an arithmetical function f and we would like to know that is it often non-zero. If we could show, for example, that for each large X we have*

$$\sum_{n \leq X} f(n)^2 > C_1 X^{5/3}$$

and

$$|f(n)| < C_2 X^{1/3} \quad (n \leq X),$$

where C_1 and C_2 are positive constants, then it follows that

$$C_1 X^{5/3} < \sum_{n \leq X} f(n)^2 \leq (C_2 X^{1/3})^2 \operatorname{card}\{n \leq X : f(n) \neq 0\}$$

and so

$$\operatorname{card}\{n \leq X : f(n) \neq 0\} > C_1 C_2^{-2} X.$$

A more sophisticated version of this would be that if one could show that

$$\sum_{X < n \leq 2X} \left(f(n) - C_3 n^{1/3}\right)^2 < C_4 X^{4/3},$$

then it would follow that for most n the function $f(n)$ is about $n^{1/3}$.

This technique has been used to show that "almost all" even numbers are the sum of two primes.

We are going to need some notation which avoids the continual use of $C_1, C_2, \ldots$, etc., to denote unspecified constants.

Given functions f and g defined on some domain $\mathcal{X}$ with $g(x) \geq 0$ for all $x \in \mathcal{X}$ we write

$$f(x) = O(g(x)) \tag{7.2}$$

to mean that there is some constant C such that

$$|f(x)| \leq Cg(x)$$

for every $x \in \mathcal{X}$. We also use

$$f(x) = o(g(x)$$

to mean that if there is some limiting operation, such as $x \to \infty$, then

$$\frac{f(x)}{g(x)} \to 0$$

and

$$f(x) \sim g(x)$$

to mean

$$\frac{f(x)}{g(x)} \to 1.$$

The symbol O was introduced by Bachmann in 1894, and the symbol o by Landau in 1909. The O-symbol can be a bit clumsy for complicated expressions and we will often instead use the Vinogradov symbols, which I. M. Vinogradov introduced about 1934. Thus we will use

$$f \ll g \tag{7.3}$$

to mean (7.2). This has the advantage that we can write strings of inequalities in the form

$$f_1 \ll f_2 \ll f_3 \ll \ldots.$$

Also if f is also non-negative we may use

$$g \gg f$$

to mean (7.3).

Our first theorem on averages concerns the function $r(n)$ and is due to Gauss. The proof illustrates a rather general principle.

Theorem 7.9 (Gauss)**.** *Let $X \geq 1$ and $G(X)$ denote the number of lattice points in the disc centre 0 of radius $\sqrt{X}$, i.e. the number of ordered pairs of integers x, y with $x^2 + y^2 \leq X$. Then*

$$G(X) = \sum_{n \leq X} r(n)$$

and

$$G(X) = \pi X + O(X^{1/2}).$$

Let

$$E(X) = G(X) - \pi X.$$

The question of the actual size of $E(X)$ is one of the classic problems of analytic number theory.

Proof. The first part of this is immediate from the definition of $r(n)$.

To prove the second part we associate with each lattice point (x, y) the unit square $S(x, y) = [x, x + 1) \times [y, y + 1)$ and this gives a partition of the plane. The squares with $x^2 + y^2 \leq X$ are contained in the disc centred at 0 of radius $\sqrt{X} + \sqrt{2}$ (apply the triangle inequality). One the other hand their union contains the disc centered at 0 of radius $\sqrt{X} - \sqrt{2}$. Moreover their area is $G(X)$ and it lies between the areas of the two discs, so

$$\pi(\sqrt{X} - \sqrt{2})^2 \leq G(X) \leq \pi(\sqrt{X} + \sqrt{2})^2, x$$

i.e.

$$\pi X - \pi 2\sqrt{2}\sqrt{X} + 2 < G(X) \leq \pi X + \pi 2\sqrt{2}\sqrt{X} + 2\pi,$$

Hence $|G(X) - \pi X| \leq \pi 2\sqrt{2}\sqrt{X} + 3\pi \ll \sqrt{X}$. □

The general principle involved in the above proof is that if one has some finite convex region in the plane and one expands it homothetically, then the number of lattice points in the region is approximately the area of the region with an error of order the length of the boundary. Thus in the theorem above the unit disc centered at the origin has its linear dimensions blown up by a factor of $\sqrt{X}$ (its radius) and the number of lattice points is approximately its area, πX with an error of order the length of the boundary $2\pi\sqrt{X}$.

Before proceeding to look further at some of the arithmetical functions we have defined above, consider the important sum

$$S(X) = \sum_{n \leq X} \frac{1}{n}$$

where $X \geq 1$. This crops up in many places. One thing you might guess straight away is that if there were only a finite number of primes, then this sum would converge as $X \to \infty$, and one could see this more or less immediately by multiplying out

$$\prod_p \left(1 - \frac{1}{p}\right)^{-1} = \prod_p \left(1 + \frac{1}{p} + \cdots + \frac{1}{p^k} + \cdots\right).$$

But, of course, the sum $S(X)$ behaves a bit like the integral so is a bit like $\log X$ which tends to infinity with X. In fact there is something more precise which one can say, which was discovered by Euler, and we will look at this in Chapter 8. Recall that $\lfloor * \rfloor$ is defined in Definition 5.3.

Theorem 7.10 (Euler)**.** *When $X \geq 1$ the sum $S(X)$ satisfies*

$$S(X) = \log X + C_0 + O\left(\frac{1}{X}\right)$$

where $C_0 = 0.577\ldots$ is Euler's constant

$$C_0 = 1 - \int_1^\infty \frac{t - \lfloor t \rfloor}{t^2} dt.$$

Proof. We have

$$\begin{aligned} S(X) &= \sum_{n \leq X} \left(\frac{1}{X} + \int_n^X \frac{dt}{t^2} \right) = \frac{\lfloor X \rfloor}{X} + \int_1^X \frac{\lfloor t \rfloor}{t^2} dt \\ &= \int_1^X \frac{dt}{t} + 1 - \int_1^X \frac{t - \lfloor t \rfloor}{t^2} dt - \frac{X - \lfloor X \rfloor}{X} \\ &= \log X + C_0 + \int_X^\infty \frac{t - \lfloor t \rfloor}{t^2} dt - \frac{X - \lfloor X \rfloor}{X}. \end{aligned}$$

□

Euler computed C_0 to 19 decimal places (by hand of course). Actually that is not so hard.

One of the more famous theorems concerning averages of arithmetical functions is

Theorem 7.11 (Dirichlet)**.** *Suppose that $X \in \mathbb{R}$ and $X \geq 2$. Then*

$$\sum_{n \leq X} d(n) = X \log X + (2C_0 - 1)X + O(X^{1/2}).$$

Let

$$\Delta(X) = \sum_{n \leq X} d(n) - X \log X - (2C_0 - 1)X.$$

As with the similar question for the Gauss lattice point problem one can ask "how does $\Delta(X)$ really behave?"

Proof. The divisor function $d(n)$ can be thought of as the number of ordered pairs of positive integers m, l such that $ml = n$. Thus when we sum over $n \leq X$ we are just counting the number of ordered pairs m, l such that $ml \leq X$. In other words we are counting the number of *lattice points* m, l under the rectangular hyperbola

$$xy = X.$$

The method that Gauss employed for his lattice point problem fails here, because the area under the rectangular hyperbola is infinite, and so is the boundary. Nevertheless the number of lattice points under the curve is finite.

We follow Dirichlet's ingenious proof method, which has become known as the *method of the hyperbola*. We *could* just crudely count, given $m \leq X$, the number of choices for l, namely

$$\left\lfloor \frac{X}{m} \right\rfloor$$

and obtain

$$\sum_{m \leq X} \frac{X}{m} + O(X)$$

and then apply Euler's estimate for $S(X)$, but this gives a much weaker error term.

Dirichlet's idea is to divide the region under the hyperbola into two parts. That with

$$m \leq \sqrt{X},\ l \leq \frac{X}{m}$$

and that with

$$l \leq \sqrt{X},\ m \leq \frac{X}{l}.$$

Clearly each region has the same number of lattice points. However the points m, l with $m \leq \sqrt{X}$ and $l \leq \sqrt{X}$ are counted in both regions. Thus we obtain

$$\begin{aligned}\sum_{n \leq X} d(n) &= 2 \sum_{m \leq \sqrt{X}} \left\lfloor \frac{X}{m} \right\rfloor - \lfloor \sqrt{X} \rfloor^2 \\ &= 2 \sum_{m \leq \sqrt{X}} \frac{X}{m} - X + O(X^{1/2}) \\ &= 2X\big(\log(\sqrt{X}) + C_0\big) - X + O(X^{1/2}).\end{aligned}$$

where in the last line we used Euler's estimate. □

One can also compute an average for Euler's function

Theorem 7.12. *Suppose that $x \in \mathbb{R}$ and $x \geq 2$. Then*

$$\sum_{n \leq x} \phi(n) = \frac{x^2}{2} \sum_{m=1}^{\infty} \frac{\mu(m)}{m^2} + O(x \log x).$$

We remark that the infinite series here is "well known" to be $\frac{6}{\pi^2}$.

Proof. We have $\phi = \mu * N$. Thus

$$\sum_{n \leq x} \phi(n) = \sum_{n \leq x} n \sum_{m \mid n} \frac{\mu(m)}{m} = \sum_{m \leq x} \mu(m) \sum_{l \leq x/m} l.$$

We want a good approximation to the inner sum. This is just the sum of an arithmetic progression of $\lfloor x/m \rfloor$ terms with first term 1 and last term $\lfloor x/m \rfloor$. Thus the sum is

$$\frac{1}{2}\left\lfloor \frac{x}{m} \right\rfloor \left(1 + \left\lfloor \frac{x}{m} \right\rfloor\right) = \frac{1}{2}\left(\frac{x}{m}\right)^2 + O\left(\frac{x}{m}\right).$$

Inserting this in the formula above gives

$$\sum_{n \le x} \phi(n) = \frac{x^2}{2} \sum_{m \le x} \frac{\mu(m)}{m^2} + O\left(\sum_{m \le x} \frac{x}{m}\right).$$

The error term is $\ll x \log x$ by Euler's bound applied to the sum. The main term is

$$\frac{x^2}{2} \sum_{m=1}^{\infty} \frac{\mu(n)}{m^2} + O\left(\sum_{m > x} \frac{x^2}{m^2}\right)$$

The error term here, by the monotonicity of the general term is

$$\ll x^2 \int_x^{\infty} \frac{dy}{y^2} \ll x.$$

Collecting together our bounds gives the theorem. □

7.3.1 Exercises

1. Prove that for any positive fixed real numbers C and ε we have $(\log n)^C \ll n^{\varepsilon}$.
2. Suppose that $f(x)$ is differentiable on $[1, X]$ with a continuous derivative on $[1, X]$.
   (i) Prove that

$$\sum_{n \le X} f(n) = \lfloor X \rfloor f(X) - \int_1^X \lfloor t \rfloor f'(t) dt$$

$$= \int_1^X f(t)dt + f(1) - (X - \lfloor X \rfloor) f(X) + \int_1^X (t - \lfloor t \rfloor) f'(t) dt.$$

   (ii) Suppose further that f is differentiable on $[1, \infty)$ with a continuous derivative on $[1, \infty)$ and that

$$\int_0^{\infty} |f'(t)| dt$$

converges. Prove that

$$\sum_{n \le X} f(n) = \int_1^X f(t)dt + C - (X - \lfloor X \rfloor) f(X) - \int_X^{\infty} (t - \lfloor t \rfloor) f'(t) dt$$

where

$$C = f(1) + \int_1^{\infty} (t - \lfloor t \rfloor) f'(t) dt.$$

3. Prove that $\sum_{n\le x}\frac{\sigma(n)}{n}=\frac{\pi^2}{6}x+O(\log x)$ for $x\ge 2$.

4. Let $D(x)=\sum_{n\le x}d(n)$.
(i) Prove that
$$\sum_{n\le x}\frac{d(n)}{n}=\frac{D(x)}{x}+\int_1^x\frac{D(u)}{u^2}du.$$
(ii) Prove that
$$\sum_{n\le x}\frac{d(n)}{n}=\frac{1}{2}(\log x)^2+O(\log x).$$

5. A number $n\in\mathbb{N}$ is *squarefree* when it has no repeated prime factors. For $X\in\mathbb{R}$, $X\ge 1$ let $Q(X)$ denote the number of squarefree numbers not exceeding X.
(i) Prove that
$$Q(X)=\frac{6}{\pi^2}X+O\left(\sqrt{X}\right).$$
(ii) Prove that if $n\in\mathbb{N}$, then
$$Q(n)\ge n-\sum_p\left[\frac{n}{p^2}\right].$$
(iii) Prove that
$$\sum_p\frac{1}{p^2}<\frac{1}{4}+\sum_{k=1}^{\infty}\frac{1}{(2k+1)^2}<\frac{1}{4}+\sum_{k=1}^{\infty}\frac{1}{4k(k+1)}=\frac{1}{2}.$$
(iv) Prove that $Q(n)>n/2$ for all $n\in\mathbb{N}$.
(v) Prove that every integer $n>1$ is a sum of two squarefree numbers.

6. Let $f(n)$ denote the number of solutions of $x^3+y^3=n$ in natural numbers x,y. Show that
$$\sum_{n\le X}f(n)=AX^{2/3}+O\left(X^{1/3}\right)\quad\text{where}\quad A=\int_0^1(1-\alpha^3)^{1/3}d\alpha.$$
Note that $A=\frac{1}{3}B(4/3,1/3)=\frac{\Gamma(4/3)^2}{\Gamma(5/3)}=\frac{1}{\pi}3^{3/2}\Gamma(4/3)^3$. Here $B(\alpha,\beta)$ is the Beta function.

7. Show that the number $N(X)$ of different natural numbers of the form 2^r3^s with $r\in\mathbb{N}$, $s\in\mathbb{N}$ and $2^r3^s\le X$ satisfies
$$N(X)=\frac{(\log X)^2}{2(\log 2)(\log 3)}+O(\log X)$$
as $X\to\infty$. Hint: Note that the condition $2^r3^s\le X$ is equivalent to $r\log 2+s\log 3\le\log X$.

8. Let $M(X)$ denote the number of ordered pairs (m, n) with $m \neq n$, $m \leq X$ and $n \leq X$ such that $\gcd(m, n) = 1$. Prove that

$$M(X) = 2 \sum_{2 \leq n \leq X} \phi(n) = \frac{6}{\pi^2} X^2 + O(X \log X),$$

that is, the probability that two different integers chosen at random from $[1, X]$ are coprime is $\frac{6}{\pi^2}$.

9. Let Let

$$d_k(n) = \sum_{\substack{m_1, m_2, \ldots, m_k \\ m_1 m_2 \ldots m_k = n}} 1.$$

Prove that

$$\sum_{n \leq X} d_k(n) \sim X \frac{(\log X)^{k-1}}{(k-1)!} \quad \text{as} \quad X \to \infty.$$

10. (i) Prove that $d(mn) \leq d(m)d(n)$

(ii) Prove that

$$\sum_{n \leq x} d(n)^2 \ll x (\log x)^3.$$

(iii) Let k be a fixed positive integer. Prove that

$$\sum_{n \leq x} d(n)^k \ll x (\log x)^{2^k - 1}.$$

7.4 Orders of Magnitude of Arithmetical Functions.

It is sometimes useful to know something about the way that an arithmetical function grows. Multiplicative functions tend to oscillate quite a bit in size. For example $d(p) = 2$ but if we take n to be the product of the first k primes where k is large, then

$$d(n) = 2^k.$$

The function $d(n)$ also arises in comparisons, for example in deciding the convergence of certain important series. Thus it is useful to have a simple universal upper bound.

Theorem 7.13. *Let $\varepsilon > 0$. Then there is a positive number C which depends at most on ε such that for every $n \in \mathbb{N}$ we have*

$$d(n) < C n^{\varepsilon}.$$

Note, such a statement is often written as

$$d(n) = O_{\varepsilon}(n^{\varepsilon})$$

or

$$d(n) \ll_{\varepsilon} n^{\varepsilon}.$$

Proof. It suffices to prove the theorem when

$$\varepsilon \leq \frac{1}{\log 2}.$$

Write $n = p_1^{k_1} \dots p_r^{k_r}$ where the p_j are distinct. Recall that

$$d(n) = (k_1 + 1) \dots (k_r + 1).$$

Thus

$$\frac{d(n)}{n^\varepsilon} = \prod_{j=1}^{r} \frac{k_j + 1}{p_j^{\varepsilon k_j}}.$$

Since we are only interested in an upper bound the terms for which $p_j^\varepsilon > 2$ can be thrown away since $2^k \geq k + 1$. However there are only $\leq 2^{1/\varepsilon}$ primes p_j for which

$$p_j^\varepsilon \leq 2.$$

Morever for any such prime we have

$$\begin{aligned} p_j^{\varepsilon k_j} &\geq 2^{\varepsilon k_j} \\ &= \exp(\varepsilon k_j \log 2) \\ &\geq 1 + \varepsilon k_j \log 2 \\ &\geq (k_j + 1)\varepsilon \log 2. \end{aligned}$$

Thus

$$\frac{d(n)}{n^\varepsilon} \leq \left(\frac{1}{\varepsilon \log 2}\right)^{2^{1/\varepsilon}}. \tag{7.4}$$

□

The above can be refined.

Theorem 7.14. *Let $\varepsilon > 0$. Then for every $n \in \mathbb{N}$ we have*

$$d(n) \ll \exp\left(\frac{(\log 2 + \varepsilon)\log n}{\log\log n}\right)$$

In Theorem 8.9 we will show that this is essentially best possible.

Proof. We may suppose that n is larger than some function of ε. In (7.4) replace the ε of that inequality by

$$\frac{\log 2 + \frac{\varepsilon}{2}}{\log\log n}.$$

The n^ε becomes

$$\exp\left(\frac{\left(\log 2 + \frac{\varepsilon}{2}\right)\log n}{\log\log n},\right)$$

and the right hand side becomes

$$\exp\left(2^{\frac{\log\log n}{\log 2+\varepsilon/2}}\log\frac{\log\log n}{(\log 2+\varepsilon/2)\log 2}\right)=\exp\left((\log n)^{1-\frac{\varepsilon/2}{\log 2+\varepsilon/2}}\log\frac{\log\log n}{(\log 2+\varepsilon/2)\log 2}\right)$$
$$\ll\exp\left(\frac{\varepsilon\log n}{2\log\log n}\right).$$

□

The product

$$\prod_{p|n}\left(1-\frac{1}{p}\right),$$

or similar such objects, can arise in many contexts. Crudely,

$$(1-1/p)^{-1}\le 2=d(p)\le d(p^k).$$

Thus

$$\prod_{p|n}\left(1-\frac{1}{p}\right)\ge\frac{1}{d(n)}\gg n^{-\varepsilon}.$$

Thus

$$n\exp\left(-(\log 2+\varepsilon)\frac{\log n}{\log\log n}\right)\le\phi(n)<n.$$

In Chapter 8 we will do much better than this.

7.4.1 Exercises

1. Let

$$d_k(n)=\sum_{\substack{m_1,m_2,\dots,m_k\\ m_1m_2\dots m_k=n}}1.$$

(i) Prove that $d_k\in\mathcal{M}$.
(ii) Prove that for any fixed $\varepsilon>0$ we have

$$d_k(n)\ll n^{\varepsilon}.$$

7.5 Notes

§1. Möbius discovered Mobius inversion in 1832. The exercise 7.2.1.11 is in E. Hille (1937). *The inversion problem of Möbius*, Duke Math. J. **3**, 549–568.

§3. As in the remark after Gauss' Theorem 7.9 let $E(X)=G(X)-\pi X$. The best bound we have for $E(X)$ is in Huxley 2002, "Integer points, exponential sums and

the Riemann zeta function", Number theory for the millennium, II (Urbana, IL, 2000) pp.275–290, pub. A K Peters, where it is shown that

$$E(X) = O(X^\theta)$$

for any $\theta > \frac{131}{416}$. We also know (Hardy and Landau, independently [1915]) that one cannot take $\theta < \frac{1}{4}$.

Euler investigated $S(X)$ and C_0 in 1735. Sometimes γ is used to denote C_0 (Mascheroni 1790).

Theorem 7.11 occurs in J.P.G.L. Dirichlet (1849) "Über die Bestimmung der mittleren Werte in der Zahlentheorie," Abh. Akad. Wiss. Berlin,2, 49–66. A huge amount of work has gone into bounding $\Delta(X)$. Suppose that θ is such that

$$\Delta(X) \ll X^\theta$$

for every $X \geq 1$. Then the current world record is that this holds for any $\theta > 131/416 = 0.31490\cdots$ and is in M. N. Huxley (2003), "Exponential sums and lattice points III", Proc. London Math. Soc. 87 (3), 591–609. In the other direction Hardy [1916] proved that one cannot take $\theta < \frac{1}{4}$.

Theorem 7.12, or rather the exercise 7.3.1.8 is sometimes known as the primitive lattice point problem. The error term is connected with the Riemann Hypothesis.

Apropos Exercise 7.3.1.10, Ramanujan (1916) "Some formulæ in the analytic theory of numbers", Messenger of Mathematics, 45, 81-84, formula (3), states that

$$\sum_{n\leq x} d(n)^2 = \frac{1}{\pi^2} x(\log x)^3 + Bx(\log x)^2 + Cx\log x + Dx + O(x^\theta)$$

holds for certain constants B, C and D and for any $\theta > 3/5$.

Chapter 8

The Distribution of Primes

8.1 Euler and Primes

There is a function which we have already seen in Definition 5.3, but we have only used so far as a form of shorthand. This is the floor function. It is not an arithmetical function - it is defined on $\mathbb{R}$, not $\mathbb{Z}$. For convenience we repeat the definition here.

Definition 8.1. *For real numbers α we define the* **floor function** $\lfloor \alpha \rfloor$ *to be the largest integer not exceeding α.*

Occasionally it is also useful to define the **ceiling function** $\lceil x \rceil$ *as the smallest integer u such that $x \leq u$. The difference $x - \lfloor x \rfloor$ is often called* **the fractional part** *of x and is sometimes denoted by $\{x\}$.*

Example 8.1. $\lfloor \pi \rfloor = 3$, $\lceil \pi \rceil = 4$, $\lfloor \sqrt{2} \rfloor = 1$, $\lfloor -\sqrt{2} \rfloor = -2$, $\lceil -\sqrt{2} \rceil = -1$.

Another related function which is very useful in some parts of number theory, although we will not use it here is $\|x\|$, *the distance of x from a nearest integer*,

$$\|x\| = \min_{n \in \mathbb{Z}} |x - n| = \min(x - \lfloor x \rfloor, \lceil x \rceil - x).$$

The floor function has some useful properties.

Theorem 8.1 (Properties of the floor function)**.** *(i) For any $x \in \mathbb{R}$ we have $0 \leq x - \lfloor x \rfloor < 1$.*
(ii) For any $x \in \mathbb{R}$ and $k \in \mathbb{Z}$ we have $\lfloor x + k \rfloor = \lfloor x \rfloor + k$.
(iii) For any $x \in \mathbb{R}$ and any $n \in \mathbb{N}$ we have $\lfloor x/n \rfloor = \lfloor \lfloor x \rfloor / n \rfloor$.
(iv) For any $x, y \in \mathbb{R}$ we have $\lfloor x \rfloor + \lfloor y \rfloor \leq \lfloor x + y \rfloor \leq \lfloor x \rfloor + \lfloor y \rfloor + 1$.
(v) For $x \in \mathbb{R}$ define $b(x) = \lfloor x \rfloor - 2\lfloor x/2 \rfloor$. Then $b(x)$ is periodic with period 2 and $b(x) = 0$ when $0 \leq x < 1$ and 1 when $1 \leq x < 2$.

Proof. (i) For any $x \in \mathbb{R}$ we have $0 \leq x - \lfloor x \rfloor < 1$. This is pretty obvious. If $x - \lfloor x \rfloor < 0$, then $x < \lfloor x \rfloor$ contradicting the definition. If $1 \leq x - \lfloor x \rfloor$, then $1 + \lfloor x \rfloor \leq x$ also contradicting the definition. This also shows that $\lfloor x \rfloor$ is unique.

(ii) For any $x \in \mathbb{R}$ and $k \in \mathbb{Z}$ we have $\lfloor x+k \rfloor = \lfloor x \rfloor + k$. One way to see this is to observe that by (i) we have $x = \lfloor x \rfloor + \theta$ for some θ with $0 \le \theta < 1$. Then $x+k-\lfloor x \rfloor - k = \theta$ and since there is only one integer l with $0 \le x+k-l < 1$, and this l is $\lfloor x+k \rfloor$ we must have $\lfloor x+k \rfloor = \lfloor x \rfloor + k$.

(iii) For any $x \in \mathbb{R}$ and any $n \in \mathbb{N}$ we have $\lfloor x/n \rfloor = \lfloor \lfloor x \rfloor / n \rfloor$. We know by (i) that $\theta = x/n - \lfloor x/n \rfloor$ satisfies $0 \le \theta < 1$. Now $x = n\lfloor x/n \rfloor + n\theta$ and so by (ii) $\lfloor x \rfloor = n\lfloor x/n \rfloor + \lfloor n\theta \rfloor$. Hence $\lfloor x \rfloor / n = \lfloor x/n \rfloor + \lfloor n\theta \rfloor / n$ and so $\lfloor x/n \rfloor \le \lfloor x \rfloor / n < \lfloor x/n \rfloor + 1$ and so $\lfloor x/n \rfloor = \lfloor \lfloor x \rfloor / n \rfloor$.

(iv) For any $x, y \in \mathbb{R}$ we have $\lfloor x \rfloor + \lfloor y \rfloor \le \lfloor x+y \rfloor \le \lfloor x \rfloor + \lfloor y \rfloor + 1$. Put $x = \lfloor x \rfloor + \theta$ and $y = \lfloor y \rfloor + \phi$ where $0 \le \theta, \phi < 1$. Then $\lfloor x+y \rfloor = \lfloor \theta + \phi \rfloor + \lfloor x \rfloor + \lfloor y \rfloor$ and $0 \le \theta + \phi < 2$.

(v) For $x \in \mathbb{R}$ define $b(x) = \lfloor x \rfloor - 2\lfloor x/2 \rfloor$. Then $b(x)$ is periodic with period 2 and $b(x) = 0$ when $0 \le x < 1$ and 1 when $1 \le x < 2$.

The periodicity is easy, since for any $k \in \mathbb{Z}$ we have

$$\begin{aligned} b(x+2k) &= \lfloor x \rfloor + 2k - 2\lfloor (x/2) + k \rfloor \\ &= \lfloor x \rfloor + 2k - 2\lfloor (x/2) \rfloor - 2k \\ &= b(x). \end{aligned}$$

Hence we only have to evaluate it when $0 \le x < 2$. It is pretty clear that $b(x) = 0$ when $0 \le x < 1$ and $= 1$ when $1 \le x < 2$. □

Here is a proof of the infinitude of primes which is essentially due to Euler, and is analytic in nature and quite different from Euclid's. It is the beginning of the modern approach. Return to

$$S(x) = \sum_{n \le x} \frac{1}{n}.$$

Less precise than Euler's result is the observation that

$$S(x) \ge \sum_{n \le x} \int_n^{n+1} \frac{dt}{t} \ge \int_1^x \frac{dt}{t} = \log x.$$

Now consider

$$P(x) = \prod_{p \le x} (1 - 1/p)^{-1}$$

where the product is over the primes not exceeding x. Then

$$P(x) = \prod_{p \le x} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots \right) \ge \sum_{n \le x} \frac{1}{n} \ge \log x.$$

Note that when one multiplies out the left hand side every fraction $\frac{1}{n}$ with $n \le x$ occurs. Since $\log x \to \infty$ as $x \to \infty$, there have to be infinitely many primes. Actually one can get something a bit more precise. Take logs on both sides. Thus

$$-\sum_{p \le x} \log(1 - 1/p) \ge \log \log x.$$

Moreover the expression on the left is

$$-\sum_{p\le x}\log(1-1/p)=\sum_{p\le x}\sum_{k=1}^{\infty}\frac{1}{kp^k}.$$

Here the terms with $k\ge 2$ contribute at most

$$\sum_{p\le x}\frac{1}{2}\sum_{k=2}^{\infty}\frac{1}{p^k}\le\frac{1}{2}\sum_{n=2}^{\infty}\frac{1}{n(n-1)}=\frac{1}{2}.$$

Hence we have just proved that

$$\sum_{p\le x}\frac{1}{p}\ge\log\log x-\frac{1}{2}.$$

Euler's result on primes is often quoted as follows.

Theorem 8.2 (Euler)**.** *The sum*

$$\sum_{p}\frac{1}{p}$$

diverges.

The above is quite close to the truth, and we will show in a while that there is a constant C_1 such that

$$\sum_{p\le x}\frac{1}{p}=\log\log x+C_1+o(1).$$

Since

$$\int_2^x\frac{dt}{t\log t}=\log\log x-\log\log 2$$

it suggests that about $1/\log n$ of the numbers near n are prime, or in other words the "probability" that n is prime is $1/\log n$. Hence one might guess that $\pi(x)$ is indeed about

$$\mathrm{li}(x)=\int_0^x\frac{dt}{\log t}$$

and the following table indicates that this is indeed true for x out to about 10^{27}. Note that the function $\mathrm{li}(x)$ is often confused with

$$\mathrm{Li}(x)=\int_2^{\infty}\frac{dt}{\log t}$$

and the two differ by about $\mathrm{li}(2)=1.045163\cdots$. Of course in $\mathrm{li}(x)$ one has to take the symmetric limit

$$\lim_{\varepsilon\to 0}\left(\int_0^{1-\varepsilon}+\int_{1+\varepsilon}^{x}\right)\frac{dt}{\log t}$$

at $t=1$ and Li avoids this. On the other hand $\mathrm{Li}(x)<\pi(x)$ for $2\le x<3$.

x	$\pi(x)$	$\mathrm{li}(x)$	$\mathrm{li}(x) - \pi(x)$
2	1	1.04	0.04
10	4	5.12	1.12
10^2	25	29.08	4.08
10^3	168	176.56	8.56
10^4	1229	1245.09	16.09
10^5	9592	9628.76	36.76
10^6	78498	78626.50	128.50
10^7	664579	664917.36	338.36
10^8	5761455	5762208.33	753.33
10^9	50847534	50849233.90	1699.90
10^{10}	455052511	455055613.54	3102.54
10^{11}	4118054813	4118066399.58	11586.58
10^{12}	37607912018	37607950279.76	38261.76
10^{13}	346065536839	346065458090.05	108969.92
10^{14}	3204941750802	3204942065690.91	314888.91
10^{15}	29844570422669	29844571475286.54	1052617.54
10^{16}	279238341033925	279238344248555.75	3214630.75
10^{17}	2623557157654233	2623557165610820.07	7956587.07
10^{18}	24739954287740860	24739954309690413.98	21949553.98
10^{19}	234057667276344607	234057667376222382.22	99877775.22
10^{20}	2220819602560918840	2220819602783663483.55	222744643.55
10^{21}	21127269486018731928	21127269486616126182.33	597394254.33
10^{22}	201467286689315906290	201467286691248261498.15	1932355208.15
10^{23}	1925320391606803968923		7250186216.00
10^{24}	18435599767349200867866		17146907278.00
10^{25}	176846309399143769411680		55160980939.00
10^{26}	1699246750872437141327603		155891678121.00
10^{27}	16352460426841680446427399		508666658006.00

8.1.1 Exercises

1. Prove that if n is a natural number and α is a real number, then

$$\sum_{k=0}^{n-1} \left\lfloor \alpha + \frac{k}{n} \right\rfloor = \lfloor n\alpha \rfloor.$$

2. Let $n \in \mathbb{N}$ and p be a prime number, show that the largest t such that $p^t | n$ satisfies

$$t = \sum_{h=1}^{\infty} \left\lfloor \frac{n}{p^h} \right\rfloor.$$

3. Let $P(Y) = \prod_{p \leq Y} p$. Prove that if $X \geq 1$, then

$$\pi(X) = \pi(\sqrt{X}) - 1 + \sum_{m | P(\sqrt{X})} \mu(m) \left\lfloor \frac{X}{m} \right\rfloor.$$

4. When $X \geq 1$ let

$$T(X) = \sum_{m \leq X} \frac{\mu(m)}{m}.$$

(i) Prove that

$$\sum_{m \leq X} \mu(m) \left\lfloor \frac{X}{m} \right\rfloor = 1.$$

(ii) Prove that

$$-1 + \frac{1}{X} \leq T(X) \leq \frac{1}{X} + 1.$$

Actually $T(X) \to 0$ as $X \to \infty$, but this is non-trivial, and con be proved by the same methods as those used to prove the prime number theorem.

8.2 Elementary Prime number theory

The strongest results we know about the distribution of primes use complex analytic methods. However there are some very useful and basic results that can be established elementarily. Many expositions of the results we are going to describe use nothing more than properties of binomial coefficients, but it is good to start to get the flavour of more sophisticated interpretations. We start by introducing

Definition 8.2 (**The von Mangoldt function).** *This is defined by*

$$\Lambda(n) = \begin{cases} 0 & \text{if } n = 1, \\ 0 & \text{if } p_1 p_2 | n \text{ with } p_1 \neq p_2, \\ \log p & \text{if } n = p^k. \end{cases}$$

The support of Λ is the prime powers. The higher powers are quite rare, at most $O(\sqrt{x})$ of them not exceeding x, and so the function is mostly concentrated on the primes themselves. This function is definitely not multiplicative, since $\Lambda(1) = 0$, but nevertheless it has an interesting and useful relationship with a familiar function as a consequence of the extension to prime powers.

Lemma 8.3. *Let $n \in \mathbb{N}$. Then*

$$\sum_{m | n} \Lambda(m) = \log n,$$

Proof. Write $n = p_1^{k_1} \dots p_r^{k_r}$ with the p_j distinct. Then for a non-zero contribution to the sum we have $m = p_s^{j_s}$ for some s with $1 \leq s \leq r$ and j_s with $1 \leq j_s \leq k_s$. Thus the sum is

$$\sum_{s=1}^{r} \sum_{j_s=1}^{k_s} \log p_s = \log n.$$

□

We need to know something about the average of $\log n$.

Lemma 8.4 (Stirling)**.** *Suppose that $X \in \mathbb{R}$ and $X \geq 2$. Then*

$$\sum_{n \leq X} \log n = X(\log X - 1) + O(\log X).$$

This can be thought of as the logarithm of Stirling's formula for $\lfloor X \rfloor !$.

Proof. We have

$$\begin{aligned} \sum_{n \leq X} \log n &= \sum_{n \leq X} \left(\log X - \int_n^X \frac{dt}{t} \right) \\ &= \lfloor X \rfloor \log X - \int_1^X \frac{\lfloor t \rfloor}{t} dt \\ &= X(\log X - 1) + \int_1^X \frac{t - \lfloor t \rfloor}{t} dt + O(\log X). \end{aligned}$$

□

Now we can say something about averages of the von Mangoldt function.

Theorem 8.5. *Suppose that $X \in \mathbb{R}$ and $X \geq 2$. Then*

$$\sum_{m \leq X} \Lambda(m) \left\lfloor \frac{X}{m} \right\rfloor = X(\log X - 1) + O(\log X).$$

Proof. The sum in question is

$$= \sum_{m \leq X} \Lambda(m) \sum_{k \leq X/m} 1.$$

Collecting together the ordered pairs $mk = n$ for a given n and rearranging gives

$$\sum_{n \leq X} \sum_{\substack{k,m \\ km=n}} \Lambda(m)$$

and this is

$$\sum_{n\leq X}\sum_{m|n}\Lambda(m).$$

By the first lemma this is

$$\sum_{n\leq X}\log n$$

and by the second it is

$$X(\log X-1)+O(\log X).$$

□

At this stage it is necessary to introduce some of the fundamental counting functions of prime number theory. For $X\geq 0$ we define

$$\psi(X)=\sum_{n\leq X}\Lambda(n),$$
$$\vartheta(X)=\sum_{p\leq X}\log p,$$
$$\pi(X)=\sum_{p\leq X}1.$$

The following theorem shows the close relationship between these three functions.

Theorem 8.6. *Suppose that $X\geq 2$. Then*

$$\psi(X)=\sum_{k}\vartheta(X^{1/k}),$$
$$\vartheta(X)=\sum_{k}\mu(k)\psi(X^{1/k}),$$
$$\pi(X)=\frac{\vartheta(X)}{\log X}+\int_2^X\frac{\vartheta(t)}{t\log^2 t}dt,$$
$$\vartheta(X)=\pi(X)\log X-\int_2^X\frac{\pi(t)}{t}dt.$$

Note that each of these functions are 0 when $X<2$, so the sums are all finite.

Proof. By the definition of Λ we have

$$\psi(X)=\sum_{k}\sum_{p\leq X^{1/k}}\log p=\sum_{k}\vartheta(X^{1/k}).$$

Hence we have

$$\sum_{k}\mu(k)\psi(X^{1/k})=\sum_{k}\mu(k)\sum_{l}\vartheta(X^{1/(kl)}).$$

Collecting together the terms for which $kl = m$ for a given m this becomes

$$\sum_m \vartheta(X^{1/m}) \sum_{k|m} \mu(k) = \vartheta(X).$$

We also have

$$\begin{aligned}\pi(X) &= \sum_{p \le X} (\log p) \left(\frac{1}{\log X} + \int_p^X \frac{dt}{t \log^2 t} \right) \\ &= \frac{\vartheta(X)}{\log X} + \int_2^X \frac{\vartheta(t)}{t \log^2 t} dt.\end{aligned}$$

The final identity is similar.

$$\vartheta(X) = \sum_{p \le X} \log X - \sum_{p \le X} \int_p^X \frac{dt}{t}$$

etcetera. □

Now we come to a series of theorems which are still used frequently.

Theorem 8.7 (Chebyshev)**.** *There are positive constants C_1 and C_2 such that for each $X \in \mathbb{R}$ with $X \ge 2$ we have*

$$C_1 X < \psi(X) < C_2 X.$$

Proof. Recall the function

$$b(x) = \lfloor x \rfloor - 2 \left\lfloor \frac{x}{2} \right\rfloor$$

defined in Theorem 8.1 for $x \in \mathbb{R}$. There we showed that b is periodic with period 2 and

$$b(x) = \begin{cases} 0 & (0 \le x < 1), \\ 1 & (1 \le x < 2). \end{cases}$$

Hence

$$\begin{aligned}\psi(X) &\ge \sum_{n \le X} \Lambda(n) b(X/n) \\ &= \sum_{n \le X} \Lambda(n) \left\lfloor \frac{X}{n} \right\rfloor - 2 \sum_{n \le X/2} \Lambda(n) \left\lfloor \frac{X/2}{n} \right\rfloor.\end{aligned}$$

Here we used the fact that there is no contribution to the second sum when $X/2 < n \le X$. Now we apply Theorem 8.5 and obtain for $x \ge 4$

$$X(\log X - 1) - 2\frac{X}{2} \left(\log \frac{X}{2} - 1) \right) + O(\log X) = X \log 2 + O(\log X).$$

This establishes the first inequality of the theorem for all $X > C$ for some positive constant C. Since $\psi(X) \geq \log 2$ for all $X \geq 2$ the conclusion follows if C_1 is small enough.

We also have, for $X \geq 4$,

$$\psi(X) - \psi(X/2) \leq \sum_{n \leq X} \Lambda(n) f(X/n)$$

and we have already seen that this is

$$X \log 2 + O(\log X).$$

Hence for some positive constant C we have, for all $X > 0$,

$$\psi(X) - \psi(X/2) \leq CX.$$

Hence, for any $k \geq 0$,

$$\psi(X2^{-k}) - \psi(X2^{-k-1}) < CX2^{-k}.$$

Summing over all k gives the desired upper bound. □

We can now obtain the following.

Corollary 8.8 (Chebyshev)**.** *There are positive constants C_3, C_4, C_5, C_6 such that for every $X \geq 2$ we have*

$$\begin{aligned} C_3 X <& \vartheta(X) < C_4 X, \\ \frac{C_5 X}{\log X} <& \pi(X) < \frac{C_6 X}{\log X}. \end{aligned}$$

Proof. The second result of Theorem 8.6 states that

$$\vartheta(X) = \sum_{k=1}^{\infty} \mu(k) \psi(X^{1/k}).$$

Remember that the series is really finite because the terms are all 0 when $X^{1/k} < 2$, i.e $k > (\log X)/(\log 2)$. Thus by the previous theorem

$$\left| \sum_{k=2}^{\infty} \mu(k) \psi(X^{1/k}) \right| \leq C_2 X^{1/2} + C_2 X^{1/3} \frac{\log X}{\log 2} < CX^{1/2}$$

for some constant C. Thus

$$|\vartheta(X) - \psi(X)| < CX^{1/2}$$

and so by the previous theorem again

$$C_1 X - CX^{1/2} < \vartheta(X) < C_2 + CX^{1/2} < C_4 X$$

with, say $C_4 = C_2 + C$. If we take $0 < C' < C_1$, then

$$C'X < C_1X - CX^{1/2}$$

provided that $X > X_0 = \left(\frac{C}{C_1 - C'}\right)^2$. Since $\vartheta(X_) \geq \log 2$ whenever $X \geq 2$ we can take C_3 to be the minimum of C' and

$$\min_{2 \leq X \leq X_0} \left(\frac{\vartheta(X)}{X}\right).$$

Now turn to $\pi(X)$. By the third formula in Theorem 8.6 we have

$$\pi(X) = \frac{\vartheta(X)}{\log X} + \int_2^X \frac{\vartheta(t)}{t \log^2 t} dt.$$

Thus, at once

$$\pi(X) \geq \frac{\vartheta(X)}{\log X} \geq \frac{C_3 X}{\log X}.$$

The upper bound is more annoying. We have

$$\pi(X) \leq \frac{C_4 X}{\log X} + \int_2^X \frac{C_4 dt}{\log^2 t}.$$

The integral here is bounded by

$$\int_2^{\sqrt{X}} \frac{C_4 dt}{(\log 2)^2} + \int_{\sqrt{X}}^X \frac{C_4 dt}{(\log \sqrt{X})^2} < \frac{C_4 \sqrt{X}}{(\log 2)^2} + \frac{4C_4 X}{(\log X)^2} < \frac{C'X}{\log X}.$$

□

Chebychev's theorem can be used to establish a companion to Theorem 7.14.

Theorem 8.9. *For every $\varepsilon > 0$ there are infinitely many n such that*

$$d(n) > \exp\left(\frac{(\log 2 - \varepsilon) \log n}{\log \log n}\right).$$

Proof. Let $n = \prod_{p \leq X} p$ so that

$$\log n = \vartheta(X).$$

Then, by Chebyshev

$$X \ll \log n \ll X$$

and so

$$\log X \sim \log \log n.$$

Moreover

$$d(n) = 2^{\pi(X)},$$

whence

$$\begin{aligned}\log d(n) &= (\log 2)\pi(X)\\ &\geq (\log 2)\frac{\vartheta(X)}{\log X}\\ &\sim (\log 2)\frac{\log n}{\log\log n}.\end{aligned}$$

□

It is also possible to establish a more precise version of Euler's result on the primes.

Theorem 8.10 (Mertens)**.** *There is a constant B and a positive constant c such that whenever $X \geq 2$ we have*

$$\sum_{n\leq X}\frac{\Lambda(n)}{n} = \log X + O(1), \tag{8.1}$$

$$\sum_{p\leq X}\frac{\log p}{p} = \log X + O(1), \tag{8.2}$$

$$\sum_{p\leq X}\frac{1}{p} = \log\log X + B + O\left(\frac{1}{\log X}\right), \tag{8.3}$$

$$\prod_{p\leq X}\left(1-\frac{1}{p}\right) = \frac{c}{\log X} + O\left(\frac{1}{(\log X)^2}\right). \tag{8.4}$$

Proof. By Theorem 8.5 we have

$$\sum_{m\leq X}\Lambda(m)\left\lfloor\frac{X}{m}\right\rfloor = X(\log X - 1) + O(\log X).$$

The left hand side is

$$X\sum_{m\leq X}\frac{\Lambda(m)}{m} + O(\psi(X)).$$

Hence by Cheyshev's theorem we have

$$X\sum_{m\leq X}\frac{\Lambda(m)}{m} = X\log X + O(X).$$

Dividing by X gives the first result.

We also have

$$\sum_{m\leq X}\frac{\Lambda(m)}{m} = \sum_{k}\sum_{p^k\leq X}\frac{\log p}{p^k}.$$

The terms with $k \geq 2$ contribute

$$\leq \sum_p \sum_{k \geq 2} \frac{\log p}{p^k} \leq \sum_{n=2}^{\infty} \frac{\log n}{n(n-1)}$$

which is convergent, and this gives the second expression.

Finally we can see that

$$\begin{aligned} \sum_{p \leq X} \frac{1}{p} &= \sum_{p \leq X} \frac{\log p}{p} \left(\frac{1}{\log X} + \int_p^X \frac{dt}{t \log^2 t} \right) \\ &= \frac{1}{\log X} \sum_{p \leq X} \frac{\log p}{p} + \int_2^X \sum_{p \leq t} \frac{\log p}{p} \frac{dt}{t \log^2 t}. \end{aligned}$$

Let

$$E(t) = \sum_{p \leq t} \frac{\log p}{p} - \log t$$

so that by the second part of the theorem we have $E(t) \ll 1$. Then the above is

$$\begin{aligned} &= \frac{\log X + E(X)}{\log X} + \int_2^X \frac{\log t + E(t)}{t \log^2 t} dt \\ &= \log \log X + 1 - \log \log 2 + \int_2^{\infty} \frac{E(t)}{t \log^2 t} dt \\ &\quad + \frac{E(X)}{\log X} - \int_X^{\infty} \frac{E(t)}{t \log^2 t} dt. \end{aligned}$$

The first integral here converges and the last two terms are

$$\ll \frac{1}{\log X}.$$

For the final assertion of the theorem observe that

$$-\log \left(1 - \frac{1}{p} \right) = \sum_{k=1}^{\infty} \frac{1}{kp^k}$$

and so

$$-\log \prod_{p \leq X} \left(1 - \frac{1}{p} \right) = \sum_{p \leq X} \frac{1}{p} + B_1 - \sum_{p > X} \sum_{k=2}^{\infty} \frac{1}{kp^k}$$

where

$$B_1 = \sum_p \sum_{k=2}^{\infty} \frac{1}{kp^k}$$

which converges absolutely since

$$\sum_{k=2}^{\infty} \frac{1}{kp^k} \leq \sum_{k=2}^{\infty} \frac{1}{p^k} = \frac{1}{p(p-1)}.$$

The other series is bounded by

$$\sum_{p>X} \frac{1}{p(p-1)} \ll X^{-1}.$$

Hence, by the third part of the theorem,

$$-\log \prod_{p \leq X} \left(1 - \frac{1}{p}\right) = \log\log X + B_2 + O\left(\frac{1}{\log X}\right)$$

for some real constant B_2. Exponentiating both sides gives the desired conclusion. □

There are several interesting applications of the above which lead to some important developments.

Theorem 8.11. *Suppose that $n \geq 3$. Let c be the constant of Theorem 8.10. Then*

$$\prod_{p|n} \left(1 - \frac{1}{p}\right) \geq \frac{c}{\log\log n} + O\left(\frac{1}{(\log\log n)^2}\right)$$

and

$$\frac{cn}{\log\log n} + O\left(\frac{n}{(\log\log n)^2}\right) \leq \phi(n) < n.$$

Proof. Suppose that n has k different prime factors and p_j denotes the j-th prime in order of magnitude. Then

$$\prod_{p|n} \left(1 - \frac{1}{p}\right) \geq \prod_{j=1}^{k} \left(1 - \frac{1}{p_j}\right) = \prod_{p \leq p_k} \left(1 - \frac{1}{p}\right).$$

By Theorem 8.10 this is

$$\frac{c}{\log p_k} + O\left(\frac{1}{(\log p_k)^2}\right).$$

Moreover

$$n \geq \prod_{j \leq k} p_j = \exp\big(\vartheta(p_k)\big).$$

Hence $\log n \geq \vartheta(p_k)$ and so by Chebyshev's theorem $p_k \ll \log n$. Hence $\log p_k \leq \log\log n + O(1)$ and the conclusions follow.

□

8.2.1 Exercises

1. Let $A(x) = \lfloor x \rfloor - \lfloor x/2 \rfloor - \lfloor x/3 \rfloor - \lfloor x/6 \rfloor$.
(i) Prove that $A(x)$ is periodic with period 6 and

$$A(x) = \begin{cases} 0 & x \in [0,1), \\ 1 & x \in [1,5), \\ 2 & x \in [5,6). \end{cases}$$

(ii) Let

$$S(x) = \sum_{m \le x} \Lambda(m) A(x/m).$$

Prove that if $x \ge 6$, then $S(x) = cx + O(\log x)$ where

$$c = \frac{1}{2}\log 2 + \frac{1}{3}\log 3 + \frac{1}{6}\log 6 = 1.01140\ldots.$$

(iii) Prove that if $x \ge 0$, then

$$\psi(x) + \psi(x/5) - 2\psi(x/6) \le S(x) \le \psi(x) + \psi(x/5).$$

(iv) Prove that if $x \ge 2$, then

$$\psi(x) \le \frac{6c}{5}x + O(\log^2 x).$$

2. For $x \ge 0$ define $B(x) = \lfloor x \rfloor - \lfloor x/2 \rfloor - \lfloor x/3 \rfloor - \lfloor x/5 \rfloor + \lfloor x/30 \rfloor$.
(i) Prove that $B(x)$ is periodic with period 30,

$$B(x) = \begin{cases} 0 & x \in [0,1), \\ 1 & x \in [1,6), \\ 0 & x \in [6,7), \\ 1 & x \in [7,10), \\ 0 & x \in [10,11), \\ 1 & x \in [11,12), \\ 0 & x \in [12,13), \\ 1 & x \in [13,15) \end{cases}$$

and that if $0 \le x < 15$, then $B(x+15) = B(x) + \lfloor x/2 \rfloor - \lfloor (x+1)/2 \rfloor$. Deduce that $0 \le B(x) \le 1$ for all x.
(ii) Let $T(x) = \sum_{m \le x} \Lambda(m) B(x/m)$. Prove that $B(x) = c'x + O(\log x)$ where $c' = \frac{1}{2}\log 2 + \frac{1}{3}\log 3 + \frac{1}{5}\log 5 - \frac{1}{30}\log 30 = 0.9212\ldots$.

(iii) Prove that $\psi(x) - \psi(x/6) \le T(x) \le \psi(x)$.
(iv) Prove that if $x \ge 2$, then

$$c'x + O(\log x) \le \psi(x) \le \frac{6c'}{5}x + O(\log^2 x).$$

Remark: $6c'/5 = 1.1054\ldots$.

3. Suppose that $a \in \mathbb{N}$, $a \ge 2$, $k \in \mathbb{N}$.
(i) Prove that if $l \in \mathbb{N}$, then

$$\log\left(a^l - 1\right) = \sum_{m | a^l - 1} \Lambda(m) = \sum_{\operatorname{ord} m a | l} \Lambda(m)$$

where Λ is the function of §7.4.
(ii) Show that

$$\sum_{l|k} \mu(l) \log\left(a^{k/l} - 1\right) = \sum_{\operatorname{ord} ma = k} \Lambda(m).$$

(iii) Show that

$$\sum_{l|k} \mu(l) \log\left(a^{k/l} - 1\right) = \phi(k) \log a + O\left(a^{-1}\right).$$

(iv) Prove that if $r \in \mathbb{N}$ and $e_{p^r}(a) = k$, then $p \nmid a$.
(v) Show that if $p \nmid k$, $r \in \mathbb{N}$ and $e_{p^r}(a) = k$, then $k | p - 1$.
(vi) By taking a to be an appropriate large multiple of k deduce that there are infinitely many primes in the residue class 1 modulo k.

Remarks: The expression on the left of (ii) and (iii) is $\log\left(\Phi_k(a)\right)$. This is the ultimate version of Euclid's proof that there are infinitely many primes. For each divisor m of 24 there are polynomials which establish that there are infinitely many primes in each reduced residue class modulo m, but it is not known whether for an arbitrary reduced residue class to an arbitrary modulus there are polynomials in one (or more) variables which work .

4. (i) Prove that if $x \ge 1$, then

$$\int_1^x \frac{\psi(u)}{u^2} du = \log x + O(1).$$

(ii) Prove that $\limsup_{x\to\infty} \frac{\psi(x)}{x} \ge 1$ and $\liminf_{x\to\infty} \frac{\psi(x)}{x} \le 1$.
(iii) Prove that if there is a constant c such that $\psi(x) \sim cx$ as $x \to \infty$, then $c = 1$.
(iv) Prove that if there is a constant c such that $\pi(x) \sim c\frac{x}{\log x}$ as $x \to \infty$, then $c = 1$.

6. (i) Let $d_n = \operatorname{lcm}[1, 2, \ldots, n]$. Show that $d_n = e^{\psi(n)}$.
(ii) Let $P \in \mathbb{Z}[x]$, $\deg P \le n$. Put $I = I(P) = \int_0^1 P(x)\, dx$. Show that $I d_{n+1} \in \mathbb{Z}$, and hence that $d_{n+1} \ge 1/|I|$ if $I \ne 0$.

(iii) Show that there is a polynomial P as above so that $Id_{n+1} = 1$.
(iv) Verify that $\max_{0\le x\le 1} |x^2(1-x)^2(2x-1)| = 5^{-5/2}$.
(v) For $P(x) = \big(x^2(1-x)^2(2x-1)\big)^{2n}$, verify that $0 < I < 5^{-5n}$.
(vi) Show that $\psi(10n+1) \ge (\frac{1}{2}\log 5)\cdot 10n$.

8.3 The Normal Number of Prime Factors

As a companion to the definition of a multiplicative function we have

Definition 8.3. *An $f \in \mathcal{A}$ is* **additive** *when it satisfies $f(mn) = f(m) + f(n)$ whenever $(m,n) = 1$.*

Now we introduce two further functions.

Definition 8.4. *We define $\omega(n)$ to be the number of different prime factors of n and $\Omega(n)$ to be the total number of prime factors of n.*

Example 8.2. *We have $360 = 2^3 3^2 5$ so that $\omega(360) = 3$ and $\Omega(360) = 6$. Generally, when the p_j are distinct, $\omega(p_1^{k_1}\dots p_r^{k_r}) = r$ and $\Omega(p_1^{k_1}\dots p_r^{k_r}) = k_1 + \cdots k_r$.*

One might expect that most of the time Ω is appreciably bigger than ω, but in fact this is not so. By the way, there is some connection with the divisor function. It is not hard to show that

$$2^{\omega(n)} \le d(n) \le 2^{\Omega(n)}.$$

In fact this is a simple consequence of the chain of inequalities

$$2 \le k+1 \le 2^k.$$

Theorem 8.12. *Suppose that $X \ge 2$. Then*

$$\sum_{n\le X} \omega(n) = X\log\log X + BX + O\left(\frac{X}{\log X}\right)$$

where B is the constant of Theorem 8.10, and

$$\sum_{n\le X} \Omega(n) = X\log\log X + \left(B + \sum_p \frac{1}{p(p-1)}\right) X + O\left(\frac{X}{\log X}\right).$$

Proof. We have

$$\begin{aligned}\sum_{n\le X}\omega(n) &= \sum_{n\le X}\sum_{p\mid n} 1 = \sum_{p\le X}\left\lfloor \frac{X}{p}\right\rfloor \\ &= X\sum_{p\le X}\frac{1}{p} + O\big(\pi(X)\big)\end{aligned}$$

and the result follows by combining Corollary 8.8 and (8.3) of Theorem 8.10.

The case of Ω is similar. We have

$$\sum_{n\le X}\Omega(n) = X\sum_{\substack{p,k\\ p^k\le X}}\frac{1}{p^k} + O\left(\sum_{k\le(\log X)/(\log 2)}\pi(X^{1/k})\right).$$

When $k\ge 2$ the terms in the error are $\ll X^{1/2}$ and so the total contribution from the $k\ge 2$ is $\ll X^{1/2}\log X$. In the main term, when $k\ge 2$ it remains to understand the behaviour of

$$\sum_{k\ge 2}\sum_{p>X^{1/k}}\frac{1}{p^k} \le \sum_{p>X^{1/2}}\frac{1}{p^2} + \sum_{k\ge 3}\frac{1}{(X^{1/k})^{k/2}}\sum_p\frac{1}{p^{k/2}}.$$

The first sum is $\ll X^{-1/2}$ and the second is

$$\ll X^{-1/2}\sum_p\frac{1}{p(p^{1/2}-1)} \ll X^{-1/2}.$$

□

Hardy and Ramanujan made the remarkable discovery that $\log\log n$ is not just the average of $\omega(n)$, but is its normal order. Later Turán found a simple proof of this.

Theorem 8.13 (Hardy & Ramanujan)**.** *Suppose that $X\ge 2$. Then*

$$\sum_{n\le X}\left(\omega(n)-\sum_{p\le X}\frac{1}{p}\right)^2 \ll X\sum_{p\le X}\frac{1}{p},$$

$$\sum_{n\le X}(\omega(n)-\log\log X)^2 \ll X\log\log X$$

and

$$\sum_{2\le n\le X}(\omega(n)-\log\log n)^2 \ll X\log\log X.$$

This theorem says that the normal number of prime factors of n is $\log\log n$.

Proof. (Turán). By (8.3), we have

$$\sum_{n\le X}\left(\sum_{p\le X}\frac{1}{p}-\log\log X\right)^2 \ll X$$

and, since for $\sqrt{X}<n\le X$, we have

$$\begin{aligned}0\le \log\log X-\log\log n &< \log\log X-\log\log\sqrt{X}\\ &= \log X-\log\frac{1}{2}\log X\\ &= \log 2,\end{aligned}$$

it follows that

$$\sum_{2\le n\le X}(\log\log X-\log\log n)^2 \ll \sqrt{X}(\log\log X)^2+\sum_{\sqrt{X}n\le X}1$$
$$\ll X.$$

Thus it suffices to prove the second statement in the theorem. We have

$$\begin{aligned}\sum_{n\le X}\omega(n)^2 &= \sum_{n\le X}\sum_{p_1|n}\sum_{p_2|n}1\\ &= \sum_{p_1\le X}\sum_{\substack{p_2\le X\\ p_2\ne p_1}}\left\lfloor\frac{X}{p_1p_2}\right\rfloor+\sum_{p\le X}\left\lfloor\frac{X}{p}\right\rfloor\\ &\le \sum_{p_1\le X}\sum_{\substack{p_2\le X\\ p_2\ne p_1}}\frac{X}{p_1p_2}+\sum_{p\le X}\frac{X}{p}\\ &\le X(\log\log X)^2+O(X\log\log X)\end{aligned}$$

by (8.3). Hence, by 8.12

$$\sum_{n\le X}(\omega(n)-\log\log X)^2\le 2X(\log\log X)^2-2(\log\log X)\sum_{n\le X}\omega(n)+O(X\log\log X)$$

and this is $\ll X\log\log X$. □

One way of interpreting this theorem is to think of it probabilistically. It is saying that the events $p|n$ are approximately independent and occur with probability $\frac{1}{p}$. Thus we can think of $\omega(n)$ as being a sum of independent random variables, and so the cental limit theorem should apply. That is, one might guess that the distribution is normal. This indeed is true and was established by Erdős and Kac in 1940. Let

$$\Phi(a,b)=\lim_{x\to\infty}\frac{1}{x}\text{card}\{n\le x : a<\frac{\omega(n)-\log\log n}{\sqrt{\log\log n}}\le b\}.$$

Then

$$\Phi(a,b)=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}dt.$$

This lead to a whole new subject, *Probabilistic Number Theory.*

8.3.1 Exercises

1. Let $\lambda(n)=(-1)^{\Omega(n)}$ (Liouville's function). Prove that

$$\lambda(n)=\sum_{m^2|n}\mu\left(n/m^2\right).$$

2. Prove that $\Omega(n) \leq \frac{\log n}{\log 2}$.

3. Let y be any real number with $y > 1$.

(i) By considering the prime divisors p of n with $p > y$, or otherwise, prove that $y^{\omega(n)-y} \leq n$, i.e.

$$\omega(n) \leq y + \frac{\log n}{\log y}.$$

(ii) Prove that $f(x) = 2x^{\frac{1}{2}} - \log x$ is an increasing function of x for $x \geq 1$. Deduce that if $n \geq 3$, then

$$(\log n)^{\frac{1}{2}} < \frac{2\log n}{\log\log n}.$$

(iii) Prove that if $n \geq 3$, then $\omega(n) \leq \frac{4\log n}{\log\log n}$.

4. Suppose that $X \geq 2$. Prove that

$$\sum_{n\leq X}\left(\Omega(n) - \sum_{p\leq X}\frac{1}{p}\right)^2 \ll X\sum_{p\leq X}\frac{1}{p},$$

$$\sum_{n\leq X}\left(\Omega(n) - \log\log X\right)^2 \ll X\log\log X$$

and

$$\sum_{2\leq n\leq X}\left(\Omega(n) - \log\log n\right)^2.$$

5. Let $\varepsilon > 0$. Prove that the set $E(X)$ of $n \leq X$ for which

$$(\log n)^{\log 2-\varepsilon} < d(n) < (\log n)^{\log 2+\varepsilon}$$

does not hold satisfies $\operatorname{card} E(X) \ll \frac{X}{\log\log X}$.

This reveals the curious fact that whereas the average value of $d(n)$ is $\log n$, $d(n)$ is normally smaller, about $(\log n)^{\log 2}$. The reason is that the average is dominated by the exceptionally large values of $d(n)$.

8.4 Primes in arithmetic progressions

We finish the chapter by developing the ultimate version of Euclid's proof that there are infinitely many primes. Let $k \in \mathbb{N}$ and let $\Phi_k(z)$ denote the k-th cyclotomic polynomial.

$$\Phi_k(z) = \prod_{\substack{l=1\\(k,k)=1}}^{k}(z - \varpi^l)$$

where

$$\varpi = e^{2\pi i/k}.$$

Thus Φ_k is the monic polynomial whose roots are the primitive k-th roots of unity and its degree is Euler's function $\phi(k)$. Note that $\Phi_k(z)$ is a (polynomial) factor of $z^k - 1$.

We can use the Möbius function to remove the condition that $(l, k) = 1$. Thus

$$\begin{aligned}\Phi_k(z) &= \prod_{l=1}^{k}(z - \varpi^l)^{\sum_{m|(l,k)}\mu(m)} \\ &= \prod_{l=1}^{k}\prod_{m|(l,k)}(z - \varpi^l)^{\mu(m)} \\ &= \prod_{m|k}\left(\prod_{n=1}^{k/m}(z - \varpi^{nm})\right)^{\mu(m)}.\end{aligned}$$

Therefore

$$\Phi_k(z) = \prod_{m|k}(z^{k/m} - 1)^{\mu(m)}. \tag{8.5}$$

Example 8.3. *The cases $k = 4$ and 6 are*

$$\Phi_4(z) = (z - i)(z + i) = z^2 + 1 = \frac{z^4 - 1}{z^2 - 1}$$

and

$$\Phi_6(z) = (z - \varpi)(z - \varpi^5) = z^2 - z + 1 = \frac{(z^6 - 1)(z - 1)}{(z^3 - 1)(z^2 - 1)}.$$

For any prime p

$$\Phi_p(z) = z^{p-1} + z^{p-2} + \cdots + z + 1.$$

We can use (8.5) to prove that the cyclotomic polynomials have integer coefficients.

Theorem 8.14. *The k-th cyclotomic polynomial has integer coefficients.*

Proof. By the formula (8.5), when $|z| < 1$, we have

$$\begin{aligned}z^{\phi(k)}\Phi_k(1/z) &= \prod_{m|k}(1 - z^{k/m})^{\mu(m)} \\ &= \prod_{\substack{m|k \\ \mu(m)=1}}(1 - z^{k/m}) \prod_{\substack{m|k \\ \mu(m)=-1}}(1 + z^{k/m} + z^{2k/m} + \cdots).\end{aligned}$$

□

We have a finite product of absolutely convergent series with integer coefficients whose product is a polynomial. Collecting together terms shows that $\Phi_k(z)$ has integer coefficients.

The constant term of $\Phi_k(z)$ is

$$\prod_{\substack{l=1\\(l,k)=1}}^{k} (-\varpi^l)$$

which has modulus 1. Thus it is ± 1.

We can use these polynomials to show that given any $k \in \mathbb{N}$ there are infinitely many primes of the form $kx+1$.

Theorem 8.15. *Suppose that $k \in \mathbb{N}$. Then there are infinitely many primes of the form $kx+1$.*

Proof. Suppose that $r \in \mathbb{N}$, $r > 1$ and p is a prime with $p \nmid k$ and $p|\Phi_k(r)$. Then $p|r^k - 1$ and $p \nmid r$. Thus $e = \operatorname{ord}_p r|k$, and if $m|k$ and $p|r^m - 1$, then $e|m$. Write $r^e = 1 + up^v$ for some positive integers u and v with $p \nmid u$. Then

$$r^{el} - 1 = (1 + up^v)^l - 1 \equiv lup^v \pmod{p^{2v}}.$$

Thus if $l|k$, so that $p \nmid l$, p^v is the exact power of p dividing $r^{el} - 1$. Thus the exact power of p dividing $\Phi_k(r)$ is

$$\prod_{\substack{m|k\\e|m}} (p^v)^{\mu(m)} = p^{v\sum_{l|k} /e\mu((k/e)/l)}.$$

and the exponent is 0 unless $e = k$. Thus we have shown that if $p \nmid k$ and $p|\Phi_k(r)$, then r has order k modulo p. Thus $k = \operatorname{ord} p(r)|p-1$.

Now suppose there are only a finite number of primes $p_1, \ldots, p_j$ in the residue class 1 modulo k and let $r = kyp_1 \ldots p_j$ where y is chosen to ensure that $\Phi_k(r) > 1$. Then there is at least one prime with $p|\Phi_k(r)$ and from above $p \equiv 1 \pmod{k}$. Thus $p|r$ also. Hence p divides the constant term of $\Phi_k(z) = \pm 1$ which is absurd. □

8.4.1 exercises

1. Prove that if p is a prime, then

$$\Phi_{pk}(z) = \begin{cases} \frac{\Phi_k(z^p)}{\Phi_k(z)} & (p \nmid k), \\ \Phi_k(z^p) & (p|k). \end{cases}$$

2. Prove that if $2 \nmid k$, $j \geq 1$ and $k > 1$, then

$$\Phi_{2^j k} = \Phi_k\big(-z^{2^{j-1}}\big).$$

3. Prove that if $k > 1$, then $\Phi_k(0) = 1$.
4. (i) Prove that if k is the product of at most two distinct primes, then its coefficients are ± 1 or 0.

   (ii) Prove that the coefficient of z^7 in $\Phi_{105}(z)$ is -2.
5. Prove that $\Phi_k(1) = e^{\Lambda(k)}$, where Λ is the von Mangoldt function.

8.5 Notes

§1. The seminal paper of B. Riemann (1860) stating a connection between π and the zeros of the Riemann zeta function is "Über die Anzahl der Primzahlen unter einer gegebenen Grösse", Monatsberichte der Königlichen Preussichen Akademie der Wissenschaften zu Berlin aus dem Jahre 1859, 671-680. The first proofs of the prime number theorem are by J. Hadamard (1896), "Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques", Bull. Soc. Math. France 24, 199-220 and Charles-Jean Étienne Gustave Nicolas, baron de la Vallée Poussin (1896), "Recherches analytiques sur la théorie des nombres premiers", I–III, Ann. Soc. Sci. Bruxelles 20, 183–256,281–362, 363–397. The strongest form we currently know of the prime number theorem which does not assume any unproven hypothesis is in N. M. Korobov (1958), "Weyl's estimates of sums and the distribution of primes", Dokl. Akad. Nauk SSSR 123, 28–31 and "Estimates of trigonometric sums and their applications", Uspehi Mat. Nauk, 13(4 (82)), 185–192, and I. M. Vinogradov (1958), "A new evaluation of $\zeta(1+it)$", Izv. Akad. Nauk SSSR 22, 161-164, again independently (Vinogradov is a little hand-wavy and, presumably mistakenly, omits the $\log\log$ factor). The result is

$$\pi(x) - \mathrm{li}(x) \ll x \exp\left(-\frac{C(\log x)^{3/5}}{(\log\log x)^{1/5}}\right)$$

for some positive constant C.

§2. Chebyshev established Theorems 8.7 and 8.8 in P. L. Chebyshev (1848, 1850), "Sur la fonction qui détermine la totalité des nombres premiers inférieurs à une limite donné", Mem. Acad. Sci. St. Petersburg 6, 1-19 and "Mémoire sur nombres premiers", Mem. Acad. Sci. St. Petersburg 7, 17-33. The various parts of Theorem 8.10 appeared in F. Mertens (both 1874), "Über einige asymptotische Gesetze der Zahlentheorie", J. Reine Angew. Math. 77, 289-338 and "Ein Beitrag zur analytischen Zahlentheorie, J. Reine Angew. Math. 78, 46-62.

§3. Theorem 8.12 is in G. H. Hardy & S. Ramanujan (1920) "The normal order of prime factors of a number n", Quart. J. Math. 48, 76-92 and the proof we give is in P. Turán (1934) "On a theorem of Hardy and Ramanujan", J. London Math. Soc. 9, 274-276. The Erdős-Kac theorem is in P. Erdős & M. Kac (1940). "The Gaussian Law of Errors in the Theory of Additive Number Theoretic Functions", American Journal of Mathematics. 62 (1/4), 738–742.

§4. Theorem 8.15 was first proved by Legendre in 1830. Curiously there seems to be no way of developing these ideas further to establish that a general reduced residue class contains infinitely many primes. Dirichlet' s proof of this instead is essentially analytic and can be considered the ultimate version of Euler's proof. However there are connections between Dirchlet's proof and algebraic number theory, especially the zeta function associated with a ring of integers.

Exercise 4 was first noticed by A. Migotti, "Aur Theorie der Kreisteilungsgleichung", Z. B. der Math.-Naturwiss, Classe der Kaiserlichen Akademie der Wissenschaften, Wien,

87, 7-14 (1883). In spite of initial appearances to the contrary the coefficients can get surprisingly large. Let $A(k)$ denote the absolute value of the largest coefficient of $\Phi_k(z)$. Schur in a letter to Landau in 1935 showed that the sequence $A(k)$ is unbounded, and following work of P. Erdős, "On the coefficients of the cyclotomic polynomials", Bull. Amer. Math. Soc., 52, 179-181, (1946) and "On the coefficients of the cyclotomic polynomials", Portugal. Math. 8, 63-71 (1949), it was shown in R. C. Vaughan, "Bounds for the coefficients of cyclotomic polynomials", Michigan Math. J. 21, 289-295 (1975) that there are arbitrarily large n such that

$$A(n) > \exp\left(\exp\left((\log 2)\frac{\log n}{\log\log n}\right)\right)$$

and that this is essentially best possible.

Chapter 9

Diophantine Equations and Approximation

9.1 Introduction

The subject of diophantine equations is typically the study of the integral solutions of polynomial equations with integral coefficients. This book is littered with typical examples, such as

$$x^2 + y^2 = 585$$

or

$$\begin{cases} 2x \equiv 91 \pmod{73}, \\ 3x \equiv 17 \pmod{101}. \end{cases}$$

There are often close connections with questions of diophantine approximation, that is the study of rational approximations to real numbers. For example, consider Pell's equation $x^2 - dy^2 = 1$ in the special case $d = 2$,

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

Suppose that x and y are both positive. Then $(x, y) = 1$ and this can be rewritten as

$$\frac{x}{y} - \sqrt{2} = \frac{1}{y(x + \sqrt{2}y)}$$

so it gives a solution to

$$\left| \sqrt{2} - \frac{a}{q} \right| < \frac{1}{\sqrt{2}q^2}$$

with $(a, q) = 1$. On the other hand if we have

$$\left| \sqrt{2} - \frac{a}{q} \right| < \frac{1}{cq^2}$$

for some positive constant c, then it follows, since $a^2 - 2q^2$ is a non-zero integer, that

$$1 \le |a^2 - 2q^2| = q(a + q\sqrt{2})|\sqrt{2} - a/q| < q^2|\sqrt{2} - a/q|^2 + 2\sqrt{2}q^2|\sqrt{2} - a/q| < \frac{1 + 2\sqrt{2}}{c}.$$

But this is impossible if $c > 1 + 2\sqrt{2}$, so there is a limitation on how good rational approximations to $\sqrt{2}$ can be.

9.2 Dirichlet's Theorem

A property of the integers which we frequently use, and already did so above, is that if $|h| < 1$, then $h = 0$, alternatively that if $h \neq 0$, then $|h| \ge 1$. The rational numbers are dense in $\mathbb{R}$ but two rationals with small denominators cannot be too close together. Thus when a/q and b/r are two different rational numbers we have

$$\frac{a}{q} - \frac{b}{r} = \frac{ar - bq}{qr}$$

and since they are unequal the numerator is non-zero. Thus

$$\left|\frac{a}{q} - \frac{b}{r}\right| \ge \frac{1}{qr} \tag{9.1}$$

There is a very simple, and useful, theorem due to Dirichlet which tells us how well a real number can be approximated by a rational number a/q in terms of the denominator q.

Theorem 9.1 (Dirichlet)**.** *For any real number α and any integer $Q \ge 1$ there exist integers a and q with $1 \le q \le Q$ such that*

$$\left|\alpha - \frac{a}{q}\right| \le \frac{1}{q(Q+1)}.$$

As an immediate consequence of casting out all common factors of a and q in a/q we have

Corollary 9.2. *The conclusion holds with the additional condition $(a, q) = 1$.*

Proof. Let I_n denote the interval $\left[\frac{n-1}{Q+1}, \frac{n}{Q+1}\right)$ and consider the Q numbers

$$\{\alpha\}, \{2\alpha\}, \ldots, \{Q\alpha\}.$$

(Here we use $\{*\} = * - \lfloor * \rfloor$ to denote the "fractional" part). If one of these numbers, say $\{q\alpha\}$, lies in I_1, then we are done. We take $a = \lfloor q\alpha \rfloor$ and then $0 \le \alpha - a < \frac{1}{Q+1}$. Similarly when one of the numbers lies in I_{Q+1}, then $1 - \frac{1}{Q+1} \le q\alpha - \lfloor q\alpha \rfloor < 1$, whence

$-\frac{1}{Q+1} \le q\alpha - (\lfloor q\alpha \rfloor + 1) < 0$ and we can take $a = \lfloor q\alpha \rfloor + 1$. When neither of these situations occurs the Q numbers must lie in the $Q-1$ intervals $I_2, \dots, I_Q$, so there must be at least one interval which contains at least two of the the numbers (the *pigeon hole principle*, or *box argument*, or *Schubfachprinzip*). Thus there are q_1, q_2 with $q_1 < q_2$ such that $|(\alpha q_2 - \lfloor \alpha q_2 \rfloor) - (\alpha q_1 - \lfloor \alpha q_1 \rfloor)| < \frac{1}{Q+1}$. We put $q = (q_2 - q_1)$, $a = (\lfloor \alpha q_2 \rfloor - \lfloor \alpha q_1 \rfloor)$. □

This turns out to be a very powerful theorem and in many applications it is all that one needs to know about the approximation of reals by rational numbers. It is obviously best possible. Take $b = 1$, $r = Q + 1$ in (9.1) above.

Theorem 9.3. *Suppose that α is irrational. Then there exist infinitely many rational numbers a/q with $(a, q) = 1$ such that $|\alpha - a/q| < q^{-2}$. In particular there are arbitrarily large q for which this inequality holds.*

Proof. Choose Q_1 to be an integer > 1 and choose a_1, q_1 in accordance with Corollary 9.2. Then $|\alpha - a_1/q_1| \le \frac{1}{q_1(Q_1+1)} < q_1^{-2}$. Now, given $a_1/q_1, \dots, a_n/q_n$ with $(a_m, q_m) = 1$ and $|\alpha - a_m/q_m| < q_m^{-2}$ we obtain a_{n+1}, q_{n+1} as follows. Since α is irrational we have $\alpha \ne a_m/q_m \, (m = 1, \dots, n)$. Choose

$$Q_{n+1} > \max\left\{|\alpha - a_1/q_1|^{-1}, \dots, |\alpha - a_n/q_n|^{-1}\right\}$$

and then choose a_{n+1}, q_{n+1} in accordance with Corrollary 1. Obviously

$$|\alpha - a_{n+1}/q_{n+1}| \le \frac{1}{q_{n+1}(Q_{n+1}+1)} < q_{n+1}^{-2}$$

and

$$|\alpha - a_{n+1}/q_{n+1}| < \min\left\{|\alpha - a_1/q_1|, \dots, |\alpha - a_n/q_n|\right\}$$

so we must have a_{n+1}/q_{n+1} distinct from any of $a_1/q_1, \dots, a_n/q_n$. Moreover it is clear that for any q_m the a_m is uniquely defined by the inequality

$$|\alpha - a_m/q_m| \le \frac{1}{q_m(Q_m+1)}.$$

Thus the q_m are distinct and so there are arbitrarily large q_m. □

When α is rational, say $\alpha = a_0/q_0$, the inequality $|\alpha - a/q| < q^{-2}$ has only a finite number of solutions in a, q with $(a, q) = 1$ since, by (1), we have $|\alpha - a/q| \ge \frac{1}{q_0 q}$ whenever $a/q \ne a_0/q_0$. Indeed the inequality $|\alpha - a/q| < \frac{1}{q_0 q}$ has the unique solution $a/q = a_0/q_0$.

Theorem 9.4. *The real number α is irrational if and only if for every $\varepsilon > 0$ there are $a \in \mathbb{Z}$, $q \in \mathbb{N}$ such that $0 < |q\alpha - a| < \varepsilon$*

Proof. If $\alpha \in \mathbb{R}\backslash\mathbb{Q}$, then choose $Q = \lfloor 1/\varepsilon \rfloor$. Then by Theorem 1, there are a, q such that $|q\alpha - a| \leq \frac{1}{Q+1} < \varepsilon$. Moreover, $q\alpha \neq a$. If $\alpha \in \mathbb{Q}$, then there are $b \in \mathbb{Z}$ and $r \in \mathbb{N}$ such that $(b, r) = 1$ and $\alpha = b/r$. Choose $\varepsilon = \frac{1}{2r}$ and suppose that there are $a \in \mathbb{Z}$, $q \in \mathbb{N}$ such that $|q\alpha - a| < \varepsilon$. Then $|\alpha - a/q| < \frac{1}{2rq}$ and $\alpha - a/q = b/r - a/q = \frac{bq-ar}{rq}$. Thus $|bq - ar| < \frac{1}{2}$. Hence $bq = ar$, whence $q\alpha - a = 0$. □

Example 9.1. *$e = \sum_0^\infty \frac{1}{k!}$ is irrational. To prove this let $q = K!$, $a = K! \sum_{k=0}^K \frac{1}{k!}$. Then $0 < \sum_{k=K+1}^\infty \frac{K!}{k!} = qe - a$ and*

$$\sum_{k=K+1}^{\infty} \frac{K!}{k!} = \frac{1}{K+1} \sum_{k=K+1}^{\infty} \frac{1}{(k-K-1)!\binom{k}{K+1}} \leq \frac{e}{K+1} < \varepsilon$$

if K is large enough.

We have already seen examples of quadratic surds which cannot be very well approximated. There is a far reaching generalisation of this.

Theorem 9.5 (Liouville)**.** *Suppose that α is an algebraic number of degree n (≥ 1). Then there is a positive constant $c = c(\alpha)$ such that*

$$\left|\alpha - \frac{a}{q}\right| > cq^{-n}$$

whenever $a \in \mathbb{Z}$, $q \in \mathbb{N}$ and $a/q \neq \alpha$ (this latter condition can be omitted when $n \geq 2$).

Proof. By "algebraic of degree n" we mean that α is a root of a non-constant polynomial with integer coefficients and the degree n corresponds to the minimal degree amongst all such polynomials. It is not hard to see that we may suppose that there is a unique polynomial

$$P(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_n$$

such that

(i) $a_j \in \mathbb{Z}$ for $0 \leq j \leq n$,
(ii) $a_0 > 0$,
(iii) $(a_0, a_1, \ldots, a_n) = 1$,
(iv) $P(\alpha) = 0$,
(v) n minimal.

Firstly a polynomial satisfying (i) and (iv) must exist by definition of α. Taking one of minimal degree ensures (v). By multiplying through by ± 1 we can ensure (ii) and by taking out common factors we can ensure (iii). Moreover if there were two distinct such polynomials P and P^*, then by (ii) and (iii) the one cannot be a multiple of the other so we could obtain, by considering $a_0^* P(\lambda) - a_0 P(\lambda)$, one of lower degree satisfying (i) and (iv) and then repeat the above process to obtain (ii) and (iii) and so contradict (v).

It suffices to show that there is a $c(\alpha)$ such that if $|\alpha - a/q| \leq 1$, then $|\alpha - a/q| > c(\alpha)q^{-n}$ for then we can replace $c(\alpha)$ by $\min\big(1, c(\alpha)\big)$ and so the conclusion follows also when $|\alpha - a/q| \geq 1$.

Since the a_j are integers we have

$$q^n P\left(\frac{a}{q}\right) \in \mathbb{Z}.$$

Moreover $P(a/q) \neq 0$, for otherwise we could factor out $\lambda - a/q$ and obtain a polynomial of lower degree $Q(\lambda) = P(\lambda)/(\lambda - a/q)$ which satisfies $Q(\alpha) = 0$. Although in the first instance this could be guaranteed only to have rational coefficients by multiplying through by a suitably integer we could recover a polynomial Q^* of degree $n-1$ with integer coefficients and satisfying $Q^*(\alpha) = 0$. Hence

$$q^n \left| P\left(\frac{a}{q}\right) \right| \geq 1.$$

On the other hand, by the mean value theorem of the differential calculus

$$-P(a/q) = P(\alpha) - P(a/q) = (\alpha - a/q)P'(\beta)$$

where β lies between α and a/q. Since we are supposing that $|\alpha - a/q| \leq 1$ it follows that

$$|P'(\beta)| \leq \max\{|P'(\lambda) : \lambda \in \lfloor \alpha - 1, \alpha + 1 \rfloor\} = c(\alpha).$$

Hence

$$1 \leq q^n |P(a/q)| \leq |\alpha - a/q| c(\alpha).$$

□

Example 9.2. *The number*

$$\theta = \sum_{k=0}^{\infty} \frac{1}{2^{k!}}$$

is transcendental, i.e. is not algebraic. To see this suppose on the contrary that it is algebraic and let n be its degree. Let $q = q_K = 2^{K!}$, $a = a_K = \sum_{k=0}^{K} 2^{K!-k!}$. Then $0 < \theta - a/q = \sum_{k=K+1}^{\infty} \frac{1}{2^{k!}} \leq \frac{1}{2^{(K+1)!}} \sum_{l=0}^{\infty} \frac{1}{2^l} = \frac{2}{q^{K+1}}$, and so if K is sufficiently large we have

$$0 < |\theta - a_K/q_K| < \frac{c(\theta)}{q_K^n}$$

which contradicts Liouville's theorem.

9.2.1 Exercises

1. Show that if $\alpha_1, \ldots, \alpha_n$ are real numbers and $R \geq 2$ is an integer, then there are $a_1, \ldots, a_n$ and q with $1 \leq q \leq R^n - 2^n + 1$ such that

$$|\alpha_1 - a_1/q| \leq q^{-1}R^{-1}, \ldots, |\alpha_n - a_n/q| \leq q^{-1}R^{-1}.$$

2. Show that if $\alpha_1, \ldots, \alpha_n$ are real numbers and $Q_1, \ldots, Q_n$ are positive integers, then there are $q_1, \ldots, q_n$ not all zero and a with $|q_1| \leq Q_1, \ldots, |q_n| \leq Q_n$ such that

$$|\alpha_1 q_1 + \cdots + \alpha_n q_n - a| \leq ((Q_1 + 1) \ldots (Q_n + 1))^{-1}.$$

Note this conclusion is not very useful unless $\alpha_1, \ldots, \alpha_n, 1$ are linearly independent over $\mathbb{Q}$. In the contrary case it is trivial provided that the Q_j are large enough.
3. Let p denote a prime number with $p \equiv 1 \pmod 4$. Then we know that there is an x with $0 < x < p$ such that $x^2 + 1 \equiv 0 \pmod p$. By Dirichlet's Theorem, or otherwise, show that there are integers a, q with $1 \leq q < \sqrt{p}$ such that $s = xq - pa$ satisfies $|s| < \sqrt{p}$. Prove that $s^2 + q^2 = p$.
4. (R. Sherman Lehman, 1974.) Suppose that n has a divisor d with $n^{\frac{1}{3}} < d \leq n^{\frac{1}{2}}$. Show that there is a t with $1 \leq t \leq n^{\frac{1}{3}} + 1$, y with $4tn \leq y^2 \leq 4tn + n^{\frac{2}{3}}$ and an x so that $4tn = y^2 - x^2$. Deduce that this gives a simple method in $O(n^{\frac{1}{3}})$ steps for finding non-trivial factors of composite numbers and of proving primality for prime numbers. This can be used as a basis for an algorithm which is both practical on small pocket calculators and appreciably faster than trial division.

Hint: Use Dirichlet's theorem to find a and q with $x = |\frac{n}{d}q - ad|$ suitably small and put $t = aq$.

9.3 Pell's equation

There is a nice application of Dirichlet's theorem on diophantine approximation to Pell's equation, $x^2 - dy^2 = 1$. When d is a perfect square the solubility of the equation is boringly trivial. By factorising the left hand side and equating each factor to ± 1 we see that the only solutions are $x = \pm 1$, $y = 0$ in that case. When d is not a perfect square things get much more interesting.

Example 9.3. *Let $d = 2$. Then we have $x_1 = 3$, $y_1 = 2$. Now, by the binomial theorem*

$$(x_1 + y_1\sqrt{2})^2 = 9 + 12\sqrt{2} + 8 = 17 + 12\sqrt{2},$$

$$17^2 - 2.12^2 = 289 - 288 = 1,$$

and

$$(x_1 + y_1\sqrt{2})^3 = 27 + 3.3^2.2\sqrt{2} + 3.3.8 + 16\sqrt{3} = 99 + 70\sqrt{2},$$

$$99^2 - 2.70^2 = 9801 - 9800 = 1.$$

Example 9.4. *Let $d = 5$. Then we have $x_1 = 9$, $y_1 = 4$. Now, by the binomial theorem*

$$(x_1 + y_1\sqrt{5})^2 = 81 + 72\sqrt{5} + 80 = 161 + 72\sqrt{5},$$

$$(x_1 + y_1\sqrt{5})^3 = 9^3 + 3.9^2.4\sqrt{5} + 3.9.4^2.5 + 4^3.5\sqrt{5} = 2889 + 1292\sqrt{5}$$

and one can check that $2889^2 - 5.1292^2 = 1$

Therefore we henceforward suppose that d is not a perfect square. In particular $\sqrt{d}$ is irrational.

Let $\alpha = \sqrt{d}$ in Dirichlet's theorem. Since $\sqrt{d}$ is irrational, by the repeated application of Theorem 9.1 we can obtain an infinite sequence of triples of integers a_1, q_1, Q_1; a_2, q_2, Q_2; a_3, q_3, Q_3;... with

$$\left|\sqrt{d} - \frac{a_n}{q_n}\right| < \frac{1}{q_n(Q_n + 1)}, \quad Q_{n+1} > \left|\sqrt{d} - \frac{a_n}{q_n}\right|^{-1}.$$

Thus

$$\begin{aligned} |a_n^2 - dq_n^2| &= \left|a_n - q_n\sqrt{d}\right| \left|a_n + q_n\sqrt{d}\right| \\ &\leq \frac{1}{Q_n + 1} \left|a_n - q_n\sqrt{d} + 2q_n\sqrt{d}\right| \\ &\leq \frac{1}{Q_n + 1} \left(\frac{1}{Q_n + 1} + 2Q_n\sqrt{d}\right) \\ &< 2\sqrt{d}. \end{aligned}$$

Thus we have found infinitely many solutions to the inequality

$$|x^2 - dy^2| < 2\sqrt{d}.$$

Hence, by the box principle, there exists an integer t with $0 < |t| < 2\sqrt{d}$ such that there are infinitely many pairs x, y with

$$x^2 - dy^2 = t. \tag{9.2}$$

Again by the box principle, there are infinitely many pairs x and y so that not only (9.2) holds but x is in a fixed residue class modulo $|t|$ and y is in a fixed residue class modulo $|t|$.

Let x_0, y_0 be a given such pair and let x, y be another with x and y large (obviously if one is, then so is the other). Then

$$x \sim y\sqrt{d}.$$

Choose

$$u = \frac{|xx_0 - dyy_0|}{|t|},\ v = \frac{|yx_0 - xy_0|}{|t|}.$$

Then $v \sim y|x_o - y_0\sqrt{d}||t|^{-1} \to \infty$ with y since $\sqrt{d}$ is irrational. Moreover

$$\begin{aligned} u^2 - dv^2 &= t^{-2}\left((xx_0 - dyy_0)^2 - d(yx_0 - xy_0)^2\right) \\ &= t^{-2}(x^2x_0^2 - dy^2x_0^2 - dx^2y_0^2 + d^2y^2y_0^2) \\ &= t^{-2}(x^2 - dy^2)(x_0^2 - dy_0^2) \\ &= 1. \end{aligned}$$

Thus we have produced infinitely many solutions to Pell's equation. It is, at least theoretically, possible to calculate solutions, for a given d, by this method, but this is very inefficient and there is a much faster way *via* the theory of continued fractions. However, it is now possible to obtain the structure of the complete solution set to Pell's equation. Let x_1, y_1 be the solution with $x_1 > 0$, $y_1 > 0$, $x_1 + y_1\sqrt{d}$ minimal. Then, by the binomial theorem there are $x_k > 0$, $y_k > 0$ such that

$$x_k + y_k\sqrt{d} = (x_1 + y_1\sqrt{d})^k$$

and it is easily verified that

$$x_k^2 - dy_k^2 = 1.$$

Suppose that there is another solution

$$X^2 - dY^2 = 1$$

with $X > 0$, $Y > 0$ and not in this list. Then for some $k \geq 1$

$$x_k + y_k\sqrt{d} < X + Y\sqrt{d} < x_{k+1} + y_{k+1}\sqrt{d}.$$

Hence

$$1 < (X + Y\sqrt{d})(x_1 - y_1\sqrt{d})^k < x_1 + y_1\sqrt{d}$$

and again by the binomial theorem for some non-zero integers X', Y' we have

$$1 < X' + Y'\sqrt{d} = (X + Y\sqrt{d})(x_1 - y_1\sqrt{d})^k$$

and $X'^2 - dY'^2 = 1$. Clearly X' and Y' cannot both be negative and if X' is positive and Y' is negative, then we would have

$$X' + Y'\sqrt{d} = 1/(X' - Y'\sqrt{d}) < 1$$

and if X' is negative and Y' is positive, then the above formula shows that $X' + Y'\sqrt{d}$ is negative. Hence both X' and Y' are positive which would contradict the minimality of $x_1 + y_1\sqrt{d}$. Therefore, we have established the following theorem.

Theorem 9.6. *Suppose that d is positive but not a perfect square. Then the equation*

$$x^2 - dy^2 = 1$$

has infinitely many solutions in non-zero integers x, y, and if x_1, y_1 is the solution with $x_1 > 0$, $y_1 > 0$ and $x_1 + y_1\sqrt{d}$ minimal, then every solution is given by $x = \pm x_k$, $y = \pm y_k$ and x_k and y_k are determined by

$$x_k + y_k\sqrt{d} = (x_1 + y_1\sqrt{d})^k.$$

Also the trivial solution $x = \pm 1$, $y = 0$ corresponds to $k = 0$.

We remark that although we have established the structure of the solution set to Pell's equation the above argument gives no easy way of finding x_1, y_1.

The above is just the tip of an iceburg. We have just described the structure of units in the quadratic number field $\mathbb{Q}(\sqrt{d})$ when d is a positive integer, not a square. Thus one can be lead to the study of such algebraic structures in the setting of algebraic number theory. See also section 6.7.

More generally one can ask, given integers $a_0, \ldots, a_k$, about the solubility of

$$f(x, y) = a_0x^k + a_1x^{k-1}y + \cdots + a_ky^k = n$$

in integers x and y. When k = 2 the form f is just a binary quadratic form of the kind mentioned in §6.3, and generally the number of solutions will be bounded if the discriminant is negative. If the discriminant is positive, the theory of Pell's equation can be brought to bear on the question. When $k \geq 3$, the equation is often called the Thue equation, since he showed in that case that if $f(1, y)$ is irreducible over $\mathbb{Q}$, then there are only a finite number of solutions. Thue's result is closely connected with our understanding of how well we can approximate algebraic numbers by rational numbers and lead to further important work on diophantine approximation by others, including Siegel, Dyson, Roth and Baker.

9.3.1 Exercises

1. Find all solutions in integers to

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

2. Find all solutions in integers to

$$x^2 - 3y^2 = 1.$$

3. Find all solutions in integers to

$$x^2 - 13y^2 = 1.$$

4. Find all solutions in integers to

$$x^2 - 19y^2 = 1.$$

9.4 Notes

§2. Numbers like θ in Example 9.2, or

$$\sum_{k=0}^{\infty} \frac{1}{10^{k!}},$$

were the first numbers to be proved to be transcendental, by Joseph Liouville (1851), "Sur des classes très étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques", Journal für die reine und angewandte Mathematik, 16, 133-–142.

§3. Pell's name became attached to the eponymous equation, apparently, because Euler mistook Pell for Lord Brouncker who had worked on the equation! However it was first studied, at least in the special case $x^2 - 2y^2 = 1$ about 400 BC in Greece and India, and it seems that Archimedes knew how to solve it. Later it was studied by Fermat and Lagrange. The proof we give of Theorem 9.6 is essentially Dirichlet's. Generally the quickest way to find the fundamental solution is by the theory of continued fractions. See `https://en.wikipedia.org/wiki/Pell's_equation`

Thue's theorem on the eponymous equation is in A. Thue (1909), "Über Annäherungswerte algebraischer Zahlen", Journal für die reine und angewandte Mathematik 1909(135), 284—305. For some of the later work on diophantine approximation see C. L. Siegel, "Approximation algebraischer Zahlen", Mathematische Zeitschrift, 10(1921), 173-–213, F. J. Dyson, "The approximation to algebraic numbers by rationals", Acta Mathematica, 79(1947), 225—240, K. F. Roth, "Rational approximations to algebraic numbers", Mathematika, 2(1955), 1—20, 168, A. Baker, Transcendental Number Theory, Cambridge University Press, 1975, ISBN 0-521-20461-5.