Lectures on Number Theory

Lars-Åke Lindahl

2002

Contents

Preface

The present lecture notes contain material for a 5 credit points course in Elementary Number Theory. The formal prerequisites for the material are minimal; in particular no previous course in abstract algebra is required. High school mathematics, familiarity with proofs by mathematical induction and with the basic properties of limits of sequences of real numbers (in particular the fact that a bounded monotone sequence of real numbers is convergent) are all that is needed. (The discussion of the prime number counting function $\pi(x)$ in section 2 requires more calculus skills, but this part could be skipped without any loss of continuity.)

A preliminary version of these notes has been carefully reviewed by Joakim Elgh, and I would like to thank him for some very useful suggestions and improvements.

Uppsala, 2002
Lars-Åke Lindahl

1 Divisibility

Definition 1.1 An integer b is *divisible* by an integer a, written $a \mid b$, if there is an integer x such that $b = ax$. We also say that b is a *multiple* of a, and that a is a *divisor* of b.

Any integer a has ± 1 and $\pm a$ as divisors. These divisors are called *trivial.*

The proof of the following simple properties are left to the reader.

Proposition 1.2 *Let a, b and c be integers.*

(i) If $a \mid b$ and $b \neq 0$, then $|a| \leq |b|$.

(ii) If $a \mid b$, then $a \mid bc$.

(iii) If $a \mid b$ and $b \mid c$, then $a \mid c$.

(iv) If $c \mid a$ and $c \mid b$, then $c \mid (ax + by)$ for all integers x and y.

(v) If $a \mid b$ and $b \mid a$, then $a = \pm b$.

(vi) Assume $c \neq 0$. Then $a \mid b$ if and only if $ac \mid bc$.

Definition 1.3 Every nonzero integer a has finitely many divisors. Consequently, any two integers a and b, not both $= 0$, have finitely many common divisors. The greatest of these is called the *greatest common divisor* and it is denoted by (a, b).

In order not to have to avoid the special case $a = b = 0$, we also *define* $(0, 0)$ as the number 0. (One good reason for this choice will appear in Theorem 1.9.)

By definition, if at least one of the numbers a and b is nonzero, then

$$d = (a, b) \Leftrightarrow d \mid a \wedge d \mid b \wedge (x \mid a \wedge x \mid b \Rightarrow x \leq d).$$

Obviously, $(b, a) = (a, b) = (-a, b) = (a, -b) = (-a, -b)$, so when calculating the greatest common divisor of two numbers we may replace them by their absolute values.

Example 1 The number 102 has the positive divisors 1, 2, 3, 6, 17, 34, 51, 102, and the number -170 has the positive divisors 1, 2, 5, 10, 17, 34, 85, and 170. The common positive divisors are 1, 2, 17, and 34. Hence $(102, -170) = 34$. To determine the greatest common divisor by finding all common divisors is obviously not a feasible method if the given numbers are large. □

Proposition 1.4 *For all integers n, $(a, b) = (a - nb, b)$.*

Proof. Write $r = a - nb$; then $a = r + nb$. Assuming $c \mid b$ we now see from Proposition 1.2 (iv) that $c \mid a$ if and only if $c \mid r$. Consequently, the pairs a, b and a, r have the same common divisors. In particular, they have the same greatest common divisor. □

We can extend the definition of greatest common divisor in a straightforward way. Given n integers $a_1, a_2, \ldots, a_n$ not all zero, we define their *greatest common divisor* $(a_1, a_2, \ldots, a_n)$ to be the greatest integer which divides all the given numbers. Finally, we define $(0, 0, \ldots, 0) = 0$.

If $(a, b) = 1$ we say that a and b are *relatively prime.* More generally, the integers $a_1, a_2, \ldots, a_n$ are called relatively prime if $(a_1, a_2, \ldots, a_n) = 1$, and they are called *pairwise relatively prime* if any two of them are relatively prime.

Example 2 The numbers 4, 6, and 9 are relatively prime but not pairwise relatively prime. □

Theorem 1.5 (The Division Algorithm) *Given integers a and b with $a > 0$ there exist two unique integers q and r such that $b = aq + r$ and $0 \le r < a$.*

The number q is called the *quotient* and r is called the *(principal) remainder.* Obviously, $q = [b/a]$ ($=$ the greatest integer $\le b/a$).

Proof. Consider the arithmetic progression

$$\dots, b-3a, b-2a, b-a, b, b+a, b+2a, b+3a, \dots$$

This sequence contains a smallest non-negative number r. By definition, $r = b - qa$ for some integer q, and clearly $0 \le r < a$. This proves the existence.

To prove uniqueness, suppose we also have $b = aq' + r'$ with $0 \le r' < a$. Then

$$r - r' = a(q' - q) \quad \text{and} \quad -a < r - r' < a.$$

Thus $a \mid (r - r')$, and it follows that $r - r' = 0$ or $|a| \le |r - r'|$. Since the latter case is excluded, we conclude that $r - r' = 0$, that is $r = r'$. Therefore $a(q - q') = 0$, which implies $q - q' = 0$, i.e. $q = q'$. □

More generally, we say that r' is a remainder when b is divided by a whenever there is an integer q' such that $b = aq' + r'$ without any further restriction on r'. If r' is an arbitrary remainder and r is the principal remainder then obviously $r' - r = na$ for some integer n, and conversely. For the principal remainder r we either have $0 \le r \le a/2$ or $a/2 < r < a$, and in the latter case the remainder $r' = r - a$ satisfies the inequality $-a/2 < r' < 0$. Hence, there is always a uniqe remainder r satisfying the inequality $-a/2 < r \le a/2$. This is the *remainder of least absolute value.* We thus have the following division algorithm, which for some purposes is more efficient than the ordinary one.

Theorem 1.5' (Modified Division Algorithm) *Given integers a and b with $a > 0$ there exist two unique integers q and r such that $b = aq + r$ and $-a/2 < r \le a/2$.*

Example 3 $37 = 2 \cdot 13 + 11 = 3 \cdot 13 - 2$. 11 is the principal remainder and -2 is the remainder of least absolute value. □

We now turn to an important class of subsets of $\mathbf{Z}$.

Definition 1.6 A non-empty set A of integers is called an *ideal* if it is closed under subtraction and under multiplication by arbitrary integers, that is if it has the following two properties:
(i) $x, y \in A \Rightarrow x - y \in A$
(ii) $x \in A$, $n \in \mathbf{Z} \Rightarrow nx \in A$.

Example 4 The sets $\{0\}$, $\mathbf{Z}$, and $\{0, \pm3, \pm6, \pm9, \dots\}$ are ideals. More generally, given any integer g, the set $A = \{ng \mid n \in \mathbf{Z}\}$ consisting of all multiples of g is an ideal. This ideal is said to be *generated* by the number g, and it will be denoted by $g\mathbf{Z}$. Thus, using this notation, $3\mathbf{Z} = \{0, \pm3, \pm6, \pm9, \dots\}$.

Note that the trivial ideal $\{0\}$ is generated by 0 and that the whole set $\mathbf{Z}$ is generated by 1. □

To show that a subset A of Z is an ideal it suffices to verify that (i) holds, because we have the following result.

Proposition 1.7 *A non-empty subset A of $\mathbf{Z}$ is an ideal if $x, y \in A \Rightarrow x-y \in A$.*

Proof. Suppose A is a non-empty subset with property (i) of Definition 1.6, and let x_0 be an element of A. Since $0 = x_0 - x_0$ we first note that $0 \in A$. Then we see that $x \in A \Rightarrow -x = 0 - x \in A$ and that

$$x, y \in A \Rightarrow x, -y \in A \Rightarrow x + y \in A,$$

i.e. the set A is closed under addition.

Next assume that the implication $x \in A \Rightarrow nx \in A$ holds for a certain nonnegative integer n (this is certainly true for $n = 0$). Then we also have $x \in A \Rightarrow (n + 1)x = nx + x \in A$. Hence, it follows by induction that the implication $x \in A \Rightarrow nx \in A$ holds for each nonnegative integer n. Finally, if $x \in A$ and n is a negative integer, then $-n$ is positive, so it follows first that $(-n)x \in A$ and then that $nx = -(-n)x \in A$. This shows that property (ii) of Definition 1.6 holds for A. □

Remark. The ideal concept is a *ring* concept. A ring is a set with two operations, addition and multiplication, satisfying certain natural axioms. The integers $\mathbf{Z}$ form a ring, and another important example is given by the set of polynomials with ordinary polynomial addition and multiplication as operations. For ideals in general rings, property (ii) does not follow from property (i). Thus the ring $\mathbf{Z}$ is special in that respect.

The ideals that are listed in Example 4 are all generated by a single number g. We next show that all ideals of $\mathbf{Z}$ have this property.

Theorem 1.8 *Every ideal A is generated by a unique nonnegative number g, that is $A = g\mathbf{Z} = \{ng \mid n \in \mathbf{Z}\}$. If A is not equal to the zero ideal $\{0\}$, then the generator g is the smallest positive integer belonging to A.*

Proof. The zero ideal is generated by 0, so assume that A contains some nonzero integer x_0. Since by (ii), A also contains the number $-x_0$ $(= (-1)x_0)$, A certainly contains a positive integer. Let g be the least positive integer belonging to A.

We will prove that A is generated by the number g. That ng belongs to A for every integer n follows immediately from (ii), so we only have to prove that there are no other numbers in A. Therefore, let $b \in A$ and divide b by g. By the division algorithm, there exist integers q and r with $0 \le r < g$ such that $b - qg = r$. Since $qg \in A$ it follows from (i) that $r \in A$, and since g is the least positive integer in A, we conclude that $r = 0$. Hence $b = qg$ as claimed. □

We will now use Theorem 1.8 to characterize the greatest common divisor. Let a and b be two integers and consider the set

$$A = \{ax + by \mid x, y \in \mathbf{Z}\}.$$

The set A is clearly closed under subtraction, i.e. A is an ideal, and by the previous theorem, A is generated by a unique nonnegative number g. This number has the following two properties:

(i) There exist integers x_0, y_0 such that $ax_0 + by_0 = g$

(ii) For all integers x and y there exists an integer n such that $ax + by = ng$.

Taking $x = 1$ and $y = 0$ in (ii) we see that $a = ng$ for some integer n and hence $g \mid a$. Similarly, $g \mid b$, so g is a common divisor of a and b. Using (i), we see that every common divisor of a and b is a divisor of g. In particular, the greatest common divisor $d = (a, b)$ divides g and hence $d \leq g$. It follows that g is the greatest common divisor, i.e. $g = (a, b)$.

This is also true in the trivial case $a = b = 0$, for then $g = 0$ and we have defined $(0, 0)$ to be the number 0.

Our discussion is summarized in the following theorem.

Theorem 1.9 *The ideal $\{ax+by \mid x, y \in \mathbf{Z}\}$ is generated by the greatest common divisor (a, b), i.e.*

(i) There exist integers x_0 and y_0 such that $ax_0 + by_0 = (a, b)$.

(ii) $ax + by$ is a multiple of (a, b) for all integers x and y.

The proof of Theorem 1.9 is easily extended to cover the case of n integers $a_1, a_2, \dots, a_n$ instead of two integers a and b. The general result reads as follows.

Theorem 1.9' *Let $a_1, a_2, \dots, a_n$ be any integers. The ideal*

$$\{a_1x_1 + a_2x_2 + \cdots + a_nx_n \mid x_1, x_2, \dots, x_n \in \mathbf{Z}\}$$

is generated by the greatest common divisor $d = (a_1, a_2, \dots, a_n)$, i.e.

(i) There exist integers $y_1, y_2, \dots, y_n$ such that $a_1y_1 + a_2y_2 + \cdots + a_ny_n = d$.

(ii) $a_1x_1 + a_2x_2 + \cdots + a_nx_n$ is a multiple of d for all integers $x_1, x_2, \dots, x_n$.

Corollary 1.10 *If $c \mid a$ and $c \mid b$, then $c \mid (a, b)$, i.e. every common divisor of a and b is a divisor of the greatest common divisor (a, b).*

Proof. By Theorem 1.9 (i) we have $ax_0 + by_0 = (a, b)$, and the conclusion of the corollary now follows from Proposition 1.2 (iv). □

Corollary 1.11 *(i) $(ca, cb) = c(a, b)$ for every nonnegative integer c.*

(ii) If $d = (a, b) \neq 0$, then $\left(\frac{a}{d}, \frac{b}{d}\right) = 1$.

Proof. (i) Write $d = (a, b)$. By Theorem 1.9, the ideal $\{ax + by \mid x, y \in \mathbf{Z}\}$ is generated by d. Now $cax + cby = c(ax + by)$, so it follows that the ideal $\{cax + cby \mid x, y \in \mathbf{Z}\}$ is generated by cd. But the latter ideal is according to Theorem 1.9 also generated by the number (ca, cb). Since the nonnegative generator is unique, we conclude that $(ca, cb) = cd$.

(ii) By (i), $d\left(\frac{a}{d}, \frac{b}{d}\right) = (a, b) = d$. The result now follows upon division by d. □

Theorem 1.12 *If $(a, b) = 1$ and $a \mid bc$, then $a \mid c$.*

Proof. Assume $(a, b) = 1$ and $a \mid bc$. Since clearly $a \mid ac$, it follows that a is a common divisor of ac and bc. By Corollary 1.11, $(ac, bc) = c(a, b) = c$, and the conclusion $a \mid c$ now follows from Corollary 1.10. □

Theorem 1.13 *If $a \mid c$, $b \mid c$ and $(a, b) = 1$, then $ab \mid c$.*

Proof. By assumption, $c = am$ for some integer m. Since $b \mid am$ and $(b, a) = 1$, we conclude from Theorem 1.12 that $b \mid m$, that is $m = bn$ for some integer n. Hence, $c = abn$, i.e. $ab \mid c$. □

Theorem 1.14 *If* $(a, b) = (a, c) = 1$*, then* $(a, bc) = 1$*.*

Proof. By Theorem 1.9 there are integers x, y and z, w such that $ax + by = 1$ and $az + cw = 1$. Then $by \cdot cw = (1 - ax)(1 - az) = 1 - an$, where $n = x + z - axz$ is an integer. Hence, $an + bcyw = 1$, and we conclude from Theorem 1.9 that $(a, bc) = 1$. □

We now turn to the problem of efficiently calculating the greatest common divisor of two integers a and b. We can of course assume that both are non-negative and that $a \geq b$.

If $b = 0$ then $(a, b) = (a, 0) = a$ and there is nothing more to do. Otherwise, we use Proposition 1.4 to see that $(a, b) = (a - nb, b)$ for all integers n. In particular, using the ordinary division algoritm $a = qb + r$ with $0 \leq r < b$ we obtain

$$(a, b) = (a - qb, b) = (r, b) = (b, r). \tag{1}$$

If $r = 0$, then we are finished, because $(a, b) = (b, 0) = b$. Otherwise, (1) allows us to replace the pair (a, b) with the smaller pair (b, r), where $r < b < a$, and we can repeat the whole procedure. Since at each step we get a new pair with smaller integers, we must finally reach a stage where one of the numbers is 0.

The whole procedure may be summarized as follows.

The Euclidean Algorithm

Let a *and* b *be integers with* $a \geq b \geq 0$*. Put* $a_0 = a$ *and* $b_0 = b$*.*

(i) If $b_0 = 0$*, then* $(a, b) = a_0$*.*

(ii) Otherwise, using the division algorithm calculate q *and* r *such that* $a_0 = qb_0 + r$ *with* $0 \leq r < b_0$*.*

(iii) Put $a_0 = b_0$ *and* $b_0 = r$ *and go to (i).*

The algorithm must terminate, because the successive b_0:s form a decreasing sequence of non-negative integers.

Instead of using the principal remainder, we could also use the remainder of least absolute value at each step. In general, this procedure will require fewer iterations. This modified algorithm runs as follows:

The Euclidean Algorithm with least absolute remainder

Let a *and* b *be integers with* $a \geq b \geq 0$*. Put* $a_0 = a$ *and* $b_0 = b$*.*

(i) If $b_0 = 0$*, then* $(a, b) = a_0$*.*

(ii) Otherwise, using the division algorithm calculate q *and* r *such that* $a_0 = qb_0 + r$ *with* $|r| \leq b_0/2$*.*

(iii) Put $a_0 = b_0$ *and* $b_0 = |r|$ *and go to (i).*

In (iii) we use the fact that $(a_0, b_0) = (a_0, -b_0)$ so it does not matter that we use $|r|$ in order to get a nonnegative number b_0. Again, the algorithm must terminate because at each step the new b_0 is at most half of the old one.

Example 5 Let us calculate $(247, 91)$. The ordinary division algorithm gives

$$\begin{aligned} 247 &= 2 \cdot 91 + 65 \\ 91 &= 1 \cdot 65 + 26 \\ 65 &= 2 \cdot 26 + 13 \\ 26 &= 2 \cdot 13. \end{aligned}$$

Hence $(247, 91) = (91, 65) = (65, 26) = (26, 13) = (13, 0) = 13$.

By instead using least absolute remainders, we obtain the following sequence as a result of the division algorithm:

$$\begin{aligned} 247 &= 3 \cdot 91 - 26 \\ 91 &= 3 \cdot 26 + 13 \\ 26 &= 2 \cdot 13. \end{aligned}$$

Hence $(247, 91) = (91, 26) = (26, 13) = (13, 0) = 13$. □

By Theorem 1.9, we know that the linear equation

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

has at least one integer solution x_0 and y_0. (We will see later that there are in fact infinitely many integer solutions.) As a by-product of the Euclidean Algorithm we have an algorithm for finding such a solution. Denoting the successive pairs (a_0, b_0) obtained during the process by (a_0, b_0), (a_1, b_1), (a_2, b_2), $\dots$, (a_n, b_n), with $b_n = 0$, we have

$$\begin{aligned} &a_0 = a, \quad b_0 = b \\ &a_i = b_{i-1}, \quad b_i = a_{i-1} - q_i b_{i-1} \quad \text{for suitable integers } q_i,\ i = 1, 2, \dots, n \\ &a_n = (a, b). \end{aligned}$$

It follows that each of the numbers a_i and b_i is a linear combination of the previous ones a_{i-1} and b_{i-1} and hence ultimately a linear combination of a and b, that is $a_i = x_i a + y_i b$ for suitable integers x_i, y_i, which can be found by calculating "backwards", and similarly for b_i. In particular, this holds for $(a, b) = a_n$.

Example 6 Going backwards in the calculations in Example 5, using the absolute remainder variant, we find that

$$13 = 91 - 3 \cdot 26 = 91 - 3 \cdot (3 \cdot 91 - 247) = 3 \cdot 247 - 8 \cdot 91.$$

Hence, the equation $247x + 91y = (247, 91)$ has $x = 3$, $y = -8$ as one of its integer solutions. □

The union $I \cup J$ of two ideals $I = a\mathbf{Z}$ and $J = b\mathbf{Z}$ in $\mathbf{Z}$ need not be an ideal. In fact, the union is an ideal if and only if one of the two ideals I and J is a subset of the other, i.e. if and only if one of the two generators a and b is divisible by the other. However, there is always a *smallest ideal* which contains the union $I \cup J$, namely the ideal $(a, b)\mathbf{Z} = \{ax + by \mid x, y \in \mathbf{Z}\}$. Thus, the greatest common divisor (a, b) is (uniquely determined as) the non-negative generator of the smallest ideal containing the union $a\mathbf{Z} \cup b\mathbf{Z}$.

On the other hand, it is completely obvious that the *intersection* $I \cap J$ of two ideals $I = a\mathbf{Z}$ and $J = b\mathbf{Z}$ is an ideal. (Indeed, the intersection of any number of ideals is an ideal.) By definition, an integer x belongs to this intersection if and only if $a|x$ and $b|x$, i.e. if and only if x is a *common multiple* of a and b.

Thus, the ideal $a\mathbf{Z} \cap b\mathbf{Z}$ coincides with the set of all common multiples of the numbers a and b. This observation leads us to the following concept, which is dual to the concept of greatest common divisor.

Definition 1.15 Let a and b be two integers. The nonnegative generator of the ideal $a\mathbf{Z} \cap b\mathbf{Z}$ is called the *least common multiple* of the two numbers, and it is denoted by $[a, b]$. More generally, given any sequence $a_1, a_2, \ldots, a_n$ of integers, we define their least common multiple $[a_1, a_2, \ldots, a_n]$ to be the uniquely determined nonnegative generator of the ideal $a_1\mathbf{Z} \cap a_2\mathbf{Z} \cap \cdots \cap a_n\mathbf{Z}$.

Note that $[a, b] = 0$ if $a = 0$ or $b = 0$, because the intersection $a\mathbf{Z} \cap b\mathbf{Z}$ is then equal to the trivial ideal $\{0\}$. If a and b are both nonzero, then $a\mathbf{Z} \cap b\mathbf{Z}$ is a nontrivial ideal since it certainly contains the number ab. Thus, nontrivial common multiples exist, and the least common multiple $[a, b]$ is a positive integer in that case.

EXAMPLE 7 $[30, 42]$=210, because in the sequence 30, 60, 90, 120, 150, 180, 210, ... of multiples of 30, the number 210 is the first one that is also a multiple of 42. □

Proposition 1.16 $[ca, cb] = c[a, b]$ *if c is a nonnegative number.*

Proof. $[ca, cb]\mathbf{Z} = ca\mathbf{Z} \cap cb\mathbf{Z} = c(a\mathbf{Z} \cap b\mathbf{Z}) = c[a, b]\mathbf{Z}$. □

Proposition 1.17 *Let a and b be nonnegative integers. Then* $[a, b] \cdot (a, b) = ab$.

Proof. If one of the two numbers equals zero, then $[a, b] = ab = 0$, wo we may assume that a and b are both positive. Let $d = (a, b)$. If $d = 1$, then any common multiple of a and b must also by a multiple of ab, by Theorem 1.13, and it follows that ab must be the least common multiple of a and b, i.e. $ab = [a, b] = [a, b] \cdot (a, b)$.

If $d > 1$, then $\left(\frac{a}{d}, \frac{b}{d}\right) = 1$. According to the case just proved, $\left[\frac{a}{d}, \frac{b}{d}\right] = \frac{a}{d} \cdot \frac{b}{d}$. Now multiply this equality by d^2 and apply Propostion 1.16 to obtain $ab = d^2\left[\frac{a}{d}, \frac{b}{d}\right] = d \cdot [a, b] = (a, b) \cdot [a, b]$. □

2 Prime Numbers

Definition 2.1 An integer > 1 is called a *prime number* or a *prime* if it has only trivial divisors. An integer > 1 which is not a prime is called *composite.*

Thus, $p > 1$ is a prime number if and only if $1 < x < p \Rightarrow x \not| p$.

Theorem 2.2 *Let p be a prime number. If $p \mid bc$, then $p \mid b$ or $p \mid c$.*

Proof. Assume that $p \mid bc$ but $p \not| b$. Since p has only trivial divisors, it follows that $(p, b) = 1$. Hence $p \mid c$ by Theorem 1.12. □

Theorem 2.2 is easily extended to

Theorem 2.2' *Let p be a prime number. If $p \mid b_1 b_2 \cdots b_n$, then $p \mid b_i$ for some i.*

Proof. By Theorem 2.2, $p \mid b_1 b_2 \cdots b_n \Rightarrow p \mid b_1 \vee p \mid b_2 \ldots b_n$. The result now follows by induction. □

Theorem 2.3 (The Fundamental Theorem of Arithmetic) *Every integer $n > 1$ can be expressed as a product of primes in a unique way apart from the order of the prime factors.*

Proof. The existence of such a factorization is proved by induction. Assume that every integer less than n can be written as a product of primes. If n is a prime, then we have a factorization of n consisting of one prime factor. If n is composite, than $n = n_1 n_2$ with $1 < n_1 < n$ and $1 < n_2 < n$, and it follows from the induction hypothesis that each of n_1 and n_2 is a product of primes. Therefore, n is also a product of primes.

Now suppose that there is an integer with to different factorizations. Then there is a least such number n. Let $n = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s$, where each p_i and q_j is a prime and where the two factorizations are different. Since p_1 divides the product $q_1 q_2 \cdots q_s$, it follows from Theorem 2.2$'$ that p_1 divides one of the prime numbers $q_1, \ldots, q_s$. Renumbering these numbers, we may assume that $p_1 | q_1$, which of course means that $p_1 = q_1$. Dividing n by p_1 we get a smaller number

$$\frac{n}{p_1} = p_2 p_3 \cdots p_r = q_2 q_3 \cdots q_s$$

with two different prime factorizations, but this would contradict the assumption that n is the smallest number with different factorizations. □

If the prime factorizations of two given numbers are known, then we can easily determine their greatest common divisor and least common multiple.

Proposition 2.4 *Let a and b be two positive integers and write*

$$a = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k} \quad \textit{and} \quad b = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k},$$

where $p_1, p_2, \ldots, p_k$ are different primes and m_1, m_2, ..., m_k and $n_1, n_2, \ldots, n_k$ are nonnegative integers. Put $d_j = \min(m_j, n_j)$ and $D_j = \max(m_j, n_j)$; then

$$(a, b) = p_1^{d_1} p_2^{d_2} \cdots p_k^{d_k} \quad \textit{and} \quad [a, b] = p_1^{D_1} p_2^{D_2} \cdots p_k^{D_k}.$$

Proof. Obvious. □

Theorem 2.5 *There exist infinitely many primes.*

Proof. We will show that given any finite collection of primes $p_1, p_2, \ldots, p_n$ there is a prime q which does not belong to the collection. Let $N = p_1 p_2 \cdots p_n + 1$. By Theorem 2.3, N has a prime factor q (which could be N itself). Since $(N, p_j) = (1, p_j) = 1$ for each j whereas $(N, q) = q$, it follows that $q \neq p_j$ for each j. □

On the other hand, there are arbitrarily large gaps in the sequence of primes:

Proposition 2.6 *For any natural number k there exist k consecutive composite numbers.*

Proof. Consider the numbers $(k+1)!+2$, $(k+1)!+3$, …, $(k+1)!+(k+1)$; they are composite, because they are divisible by 2, 3, …, $k+1$, respectively. □

Let $\pi(x)$ denote the number of primes that are less than or equal to the real number x. Thus

$$\pi(x) = \begin{cases} 0 & \text{if } x < 2 \\ 1 & \text{if } 2 \le x < 3 \\ 2 & \text{if } 3 \le x < 5 \\ \vdots & \\ n & \text{if } p_n \le x < p_{n+1} \end{cases}$$

where p_n denotes the nth prime number.

We will give a crude estimate for $\pi(x)$. To this end, we will need the following inequality.

Lemma 2.7 *Let x be a real number > 2. Then*

$$\sum_{p \le x} \frac{1}{p} > \ln\ln x - 1.$$

Here, the sum is over all primes p satisfying $p \le x$.

Since $\ln\ln x$ tends to ∞ with x it follows from the inequality above that the sum $\sum 1/p$ over all primes is infinite. This, of course, implies that there are infinitely many primes. Thus, by proving Lemma 2.7 we will obtain a new proof of Theorem 2.5.

Proof. Let $p_1, p_2, \ldots, p_n$ denote all primes $\le x$, and put

$$\mathcal{N} = \{p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} \mid k_1 \ge 0, k_2 \ge 0, \ldots, k_n \ge 0\},$$

i.e. $\mathcal{N}$ consists of 1 and all positive integers whose prime factorization only uses the primes $p_1, p_2, \ldots, p_n$.

Since the factorization of any number $\le x$ only uses primes that are $\le x$, the set $\mathcal{N}$ contains all of the numbers 1, 2, 3, …, $[x]$ (= the greatest integer $\le x$). Consequently,

$$\sum_{n \in \mathcal{N}} \frac{1}{n} \ge \sum_{n=1}^{[x]} \frac{1}{n} \ge \int_1^{[x]+1} \frac{dt}{t} = \ln([x]+1) > \ln x.$$

Now observe that

$$\prod_{p \le x} \left(1 - \frac{1}{p}\right)^{-1} = \prod_{p \le x} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^k} + \cdots\right) = \sum_{n \in \mathcal{N}} \frac{1}{n}.$$

Combining this with the previous inequality we obtain the inequality

$$\prod_{p\le x}\left(1-\frac{1}{p}\right)^{-1} > \ln x,$$

and, by taking the logarithm of both sides, the inequality

$$\sum_{p\le x}\ln\left(1-\frac{1}{p}\right)^{-1} > \ln\ln x. \tag{1}$$

Now use the Maclaurin expansion of $\ln(1+x)$ to get

$$-\ln(1-x) = x+\frac{x^2}{2}+\frac{x^3}{3}+\cdots \le x+\frac{x^2}{2}(1+x+x^2+\dots) = x+\frac{x^2}{2}\left(\frac{1}{1-x}\right)$$

for $0\le x<1$. Since $1/(1-x)\le 2$ when $x\le\frac{1}{2}$, we conclude that the inequality

$$\ln(1-x)^{-1} = -\ln(1-x) \le x+x^2$$

holds for $x\le\frac{1}{2}$. In particular, if p is a prime, then $\frac{1}{p}\le\frac{1}{2}$, and consequently,

$$\ln(1-\frac{1}{p})^{-1} \le \frac{1}{p}+\frac{1}{p^2}.$$

By summing these inequalities for all primes $p\le x$ and comparing with (1), we obtain

$$\sum_{p\le x}\frac{1}{p}+\sum_{p\le x}\frac{1}{p^2} > \ln\ln x. \tag{2}$$

Here the sum $\sum 1/p^2$ over all primes $\le x$ can be estimated as follows

$$\sum_{p\le x}\frac{1}{p^2} \le \sum_{n=2}^{\infty}\frac{1}{n^2} \le \sum_{n=2}^{\infty}\frac{1}{n(n-1)} = \sum_{n=2}^{\infty}\left(\frac{1}{n-1}-\frac{1}{n}\right) = 1,$$

and by combining this inequality with (2) we obtain the desired result

$$\sum_{p\le x}\frac{1}{p} > \ln\ln x - 1.$$

□

Lemma 2.8

$$\sum_{p\le x}\frac{1}{p} = \frac{\pi(x)}{x}+\int_2^x\frac{\pi(u)}{u^2}\,du.$$

Proof. Let $p_1<p_2<\cdots<p_n$ denote the primes $\le x$. Then

$$\begin{aligned}\int_2^x\frac{\pi(u)}{u^2}\,du &= \sum_{k=1}^{n-1}\int_{p_k}^{p_{k+1}}\frac{\pi(u)}{u^2}\,du+\int_{p_n}^{x}\frac{\pi(u)}{u^2}\,du\\ &= \sum_{k=1}^{n-1}\int_{p_k}^{p_{k+1}}\frac{k}{u^2}\,du+\int_{p_n}^{x}\frac{n}{u^2}\,du\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n-1} k\left(\frac{1}{p_k} - \frac{1}{p_{k+1}}\right) + n\left(\frac{1}{p_n} - \frac{1}{x}\right)\\
&= \sum_{k=1}^{n-1} \frac{k}{p_k} - \sum_{k=2}^{n} \frac{k-1}{p_k} + \frac{n}{p_n} - \frac{n}{x}\\
&= \sum_{k=1}^{n} \frac{1}{p_k} - \frac{\pi(x)}{x}. \qquad \square
\end{aligned}
$$

Theorem 2.9 *For any $\epsilon > 0$ and any real number ω, there exists a number $x > \omega$ such that*

$$\pi(x) > (1-\epsilon)\frac{x}{\ln x}.$$

Remark. For those who know the definition of lim sup we can state Theorem 2.9 as follows: $\limsup_{x\to\infty} \frac{\pi(x)}{x/\ln x} \geq 1$.

Proof. Assume the theorem to be false. Then there is an $\epsilon > 0$ and a real number ω such that $\pi(x) \leq (1-\epsilon)\frac{x}{\ln x}$ for all $x > \omega$. But then

$$
\begin{aligned}
\int_2^x \frac{\pi(u)}{u^2}\,du = \int_2^\omega \frac{\pi(u)}{u^2}\,du + \int_\omega^x \frac{\pi(u)}{u^2}\,du \leq C + (1-\epsilon)\int_\omega^x \frac{1}{u\ln u}\,du\\
= C + (1-\epsilon)(\ln\ln x - \ln\ln\omega) = D + (1-\epsilon)(\ln\ln x),
\end{aligned}
$$

where C and D are constants (depending on ω). Since obviously $\pi(x) < x$, it now follows from Lemma 2.8, that

$$\sum_{p\leq x} \frac{1}{p} \leq (1-\epsilon)\ln\ln x + \text{Constant}.$$

This contradicts Lemma 2.7. □

Theorem 2.9 can be sharpened considerably. The following result was conjectured by Gauss and proven by J. Hadamard and Ch. de la Vallée Poussin in 1896 using advanced methods from the theory of functions of a complex variable.

Theorem 2.10 (The Prime Number Theorem)

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\ln x} = 1.$$

The proof is too complicated to be given here.

We will now derive heuristically some conclusions from the Prime Number Theorem. Firstly, it follows that $\pi(x)/x < C/\ln x$ for some constant C, and hence the ratio $\pi(x)/x$ approaches 0 and the ratio $(x-\pi(x))/x$ approaches 1 as x tends to infinity. Since $n - \pi(n)$ is the number of composite numbers less than or equal to n, the ratio $(n-\pi(n))/n$ represents the proportion of composite numbers among the first n integers. That this ratio tends to 1 means in a certain sense that "almost all" positive integers are composite.

On the other hand, primes are not particularly scarce, because the logarithm function grows very slowly. By the Prime Number Theorem we can use $x/\ln x$

as an approximation of $\pi(x)$. If x is a large number and y is small compared to x then $\ln(x+y) \approx \ln x$, and hence

$$\pi(x+y) - \pi(x) \approx \frac{x+y}{\ln(x+y)} - \frac{x}{\ln x} \approx \frac{y}{\ln x}.$$

This means that in a relatively small interval of length y around the large number x there are approximately $y/\ln x$ primes, and we can expect to find a prime in the interval if the length is about $\ln x$. If the primes were randomly distributed the probability of a large number x being prime would be approximately $1/\ln x$. Taking for example $x = 10^{100}$ we have $\ln x \approx 230$. Thus, if we choose an integer N "at random" in the neigborhood of 10^{100} the probability that N is prime is roughly 1/230. Of course, we can raise this probability to 1/115 by avoiding the even numbers, and if we make sure that N is not divisible by 2, 3, or 5, the probability that N is prime grows to about 1/60. Thus, provided we use an efficient primality test, we can produce a very large prime by first choosing a number N at random not divisible by 2, 3, or 5 (and some other small primes) and testing it for primality. If N turns out to be a prime, then we are happy, otherwise we consider the next integer in the sequence $N+2$, $N+4$, $N+6$, ...that is not divisible by 3 and 5 (and the other selected small primes) and test this for primality. Because of the Prime Number Theorem we feel confident that we will find a prime after not too many tries.

3 The Linear Diophantine Equation ax+by=c

Let a, b and c be integers and consider the equation

$$ax + by = c. \tag{1}$$

We are interested in integer solutions x and y, only.

From section 1 we already know a lot about the equation. By Theorem 1.9, the set $\{ax + by \mid x, y \in \mathbf{Z}\}$ coincides with the set of all multiples $n(a,b)$ of the greatest common divisor of a and b. It follows that equation (1) is solvable if and only if $(a,b) \mid c$. Moreover, the Euclidean algorithm provides us with a method for finding a solution x_0, y_0 of the equation $ax + by = (a,b)$, and by multiplying this solution by $c/(a,b)$ we will get a solution of the original equation (1). What remains is to find the general solution given one particular solution. The complete story is summarized in the following theorem.

Theorem 3.1 *The equation $ax + by = c$ has integer solutions if and only if $(a,b) \mid c$. If x_0, y_0 is a solution, then all integer solutions are given by*

$$x = x_0 + \frac{b}{(a,b)}n, \quad y = y_0 - \frac{a}{(a,b)}n, \quad n \in \mathbf{Z}.$$

Proof. The numbers x and y defined above are integers, and one immediately verifies that they satisfy the equation. To see that these are all solutions, assume that x, y is an arbitrary integer solution. Then $ax + by = ax_0 + by_0$. It follows that $a(x - x_0) = b(y_0 - y)$, and that

$$\frac{a}{d}(x - x_0) = \frac{b}{d}(y_0 - y), \tag{2}$$

where we have written $d = (a, b)$ for short. Since $\left(\frac{a}{d}, \frac{b}{d}\right) = 1$, we conclude from Theorem 1.12 that $\frac{b}{d}$ is a divisor of $x - x_0$, i.e. there exists an integer n such that $x - x_0 = \frac{b}{d}\,n$. By inserting this into (2) and simplifying, we also obtain $y - y_0 = -\frac{a}{d}\,n$. □

The case $(a, b) = 1$ is so important that it is worth stating separately.

Corollary 3.2 *Suppose that $(a, b) = 1$. Then the linear equation $ax + by = c$ has integer solutions for all integers c. If x_0, y_0 is a solution, then all solutions are given by*

$$x = x_0 + bn, \quad y = y_0 - an, \quad n \in \mathbf{Z}.$$

According to Theorem 3.1, the distance between two consecutive x-solutions is b/d and the distance between two consecutive y-solutions is a/d, where $d = (a, b)$. It follows that, provided the equation is solvable, there is a solution (x, y) with $0 \le x \le b/d - 1$. We can find this solution by successively trying $x = 0$, $x = 1$, ..., solving the equation for y until an integer value for y is found. Of course, we can also solve the equation by looking for a solution y in the interval $0 \le y \le a/d - 1$. Hence, we can easily solve the equation $ax + by = c$ by trial and error whenever at least one of the numbers a/d and b/d is small.

Example 1 Solve the equation

$$247x + 91y = 39.$$

Solution 1: The equation is solvable, because $(247, 91) = 13$ and $13 \mid 39$. Since $\frac{91}{13} = 7$ the equation has an integer solution with $0 \le x \le 6$. Trying $x = 0$, 1, 2, we find that $x = 2$ gives the integer value $y = -5$. Therefore, the general solution of the equation is $x = 2 + 7n$, $y = -5 - 19n$.

Solution 2: In Example 6, section 1, we found that $x = 3$, $y = -8$ solves the equation $247x + 91y = 13$. By multiplying this solution by 3, we get the particular solution $x_0 = 9$, $y_0 = -24$ to our given equation, and the general solution is $x = 9 + 7n$, $y = -24 - 19n$. This parametrization of the solutions is different from that above, but the set of solutions is of course the same as in solution no. 1.

Solution 3: The solution above uses the Euclidean algorithm. We will now give another method, which is more or less equivalent to the Euclidean algorithm, but the presentation is different. To solve

$$247x + 91y = 39 \tag{3}$$

we start by writing $247 = 2 \cdot 91 + 65$, $247x = 91 \cdot 2x + 65x$ and $247x + 91y = 65x + 91(2x + y)$. Introducing new integer variables $x_1 = x$, $y_1 = 2x + y$, we now rewrite equation (3) as

$$65x_1 + 91y_1 = 39. \tag{4}$$

This equation has smaller coefficients. Note that if x_1 and y_1 are integers, then $x = x_1$ and $y = y_1 - 2x$ are integers, too. Hence, solving (4) for integer values is equivalent to solving (3) for integer values.

The same procedure can now be repeated. Write $91 = 65 + 26$ and $65x_1 + 91y_1 = 65(x_1 + y_1) + 26y_1$ in order to replace equation (4) with the equivalent equation

$$(5) \qquad 65x_2 + 26y_2 = 39, \qquad \text{with } x_2 = x_1 + y_1,\ y_2 = y_1.$$

We continue, noting that $65 = 2 \cdot 26 + 13$, and obtain

$$(6) \qquad 13x_3 + 26y_3 = 39, \qquad \text{with } x_3 = x_2,\ y_3 = 2x_2 + y_2.$$

Now $26 = 2 \cdot 13$, so

$$(7) \qquad 13x_4 + 0y_4 = 39, \qquad \text{with } x_4 = x_3 + 2y_3,\ y_4 = y_3.$$

From (7) we conclude that $x_4 = 39/13 = 3$ whereas y_4 is an arbitrary integer, n say. Going backwards, we find

$$\begin{aligned} y_3 &= y_4 = n, \quad x_3 = x_4 - 2y_3 = 3 - 2n \\ x_2 &= x_3 = 3 - 2n, \quad y_2 = y_3 - 2x_2 = n - 2(3 - 2n) = -6 + 5n \\ y_1 &= y_2 = -6 + 5n, \quad x_1 = x_2 - y_1 = 3 - 2n + 6 - 5n = 9 - 7n \\ x &= x_1 = 9 - 7n, \quad y = y_1 - 2x = -6 + 5n - 2(9 - 7n) = -24 + 19n. \end{aligned} \qquad \square$$

For linear equations with more than two variables we have the following result, which follows immediately from Theorem 1.9′.

Theorem 3.3 *The linear equation $a_1x_1 + a_2x_2 + \cdots + a_nx_n = c$ has integer solutions if and only if $(a_1, a_2, \ldots, a_n) \mid c$.*

The third solution method in Example 1 can easily be adopted to take care of equations with more than two variables.

Example 2 Solve the equation

$$6x + 10y + 15z = 5$$

for integer solutions.

Solution: The equation is solvable, because $(6, 10, 15) = 1$. Consider the least coefficient 6 and write $10 = 6 + 4$ and $15 = 2 \cdot 6 + 3$. Introducing new variables $x_1 = x + y + 2z$, $y_1 = y$, and $z_1 = z$ we can rewrite our linear equation as

$$6x_1 + 4y_1 + 3z_1 = 5.$$

Since $6 = 2 \cdot 3$ and $4 = 3 + 1$, we put $x_2 = x_1$, $y_2 = y_1$, and $z_2 = 2x_1 + y_1 + z_1$. This change of variables transforms our equation into

$$0x_2 + y_2 + 3z_2 = 5.$$

Now 1 is the least non-zero coefficient, and we put $x_3 = x_2$, $y_3 = y_2 + 3z_2$, and $z_3 = z_2$. Our equation now reads

$$0x_3 + y_3 + 0z_3 = 5$$

with the obvious solution $x_3 = m$, $y_3 = 5$, $z_3 = n$, m and n being arbitrary integers. Going backwards we get after some easy calculations:

$$x = 5 + 5m - 5n, \quad y = 5 - 3n, \quad z = -5 - 2m + 4n, \quad m, n \in \mathbf{Z}. \qquad \square$$

4 Congruences

Definition 4.1 Let m be a positive integer. If $m \mid (a-b)$ then we say that a is *congruent* to b *modulo* m and write $a \equiv b \pmod{m}$. If $m \nmid (a-b)$ then we say that a is not congruent to b modulo m and write $a \not\equiv b \pmod{m}$.

Obviously, $a \equiv b \pmod{m}$ is equivalent to $a = b + mq$ for some integer q.

We now list some useful properties, which follow easily from the definition.

Proposition 4.2 *Congruence modulo m is an equivalence relation, i.e.*

(i) $a \equiv a \pmod{m}$ *for all* a.

(ii) If $a \equiv b \pmod{m}$, *then* $b \equiv a \pmod{m}$.

(iii) If $a \equiv b \pmod{m}$ *and* $b \equiv c \pmod{m}$, *then* $a \equiv c \pmod{m}$.

Proof. We leave the simple proof to the reader. □

Our next proposition shows that congruences can be added, multiplied and raised to powers.

Proposition 4.3 *Let a, b, c and d be integers.*

(i) If $a \equiv b \pmod{m}$ *and* $c \equiv d \pmod{m}$, *then* $a + c \equiv b + d \pmod{m}$.

(ii) If $a \equiv b \pmod{m}$ *and* $c \equiv d \pmod{m}$, *then* $ac \equiv bd \pmod{m}$.

(iii) If $a \equiv b \pmod{m}$, *then* $a^k \equiv b^k \pmod{m}$ *for all non-negative integers* k.

(iv) Let $f(x)$ *be a polynomial with integral coefficients. If* $a \equiv b \pmod{m}$ *then* $f(a) \equiv f(b) \pmod{m}$.

Proof. (i) is left to the reader.

(ii) If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $a = b + mq$ and $c = d + mr$ for suitable integers q and r. It follows that $ac = bd + m(br + dq + mqr)$. Hence $ac \equiv bd \pmod{m}$.

(iii) Taking $c = a$ and $d = b$ in (ii) we see that $a \equiv b \pmod{m}$ implies $a^2 \equiv b^2 \pmod{m}$. Applying (ii) again, we get $a^3 \equiv b^3 \pmod{m}$, and the general case follows by induction.

(iv) Suppose $f(x) = \sum_{j=0}^{n} c_j x^j$. Using (iii) we first obtain $a^j \equiv b^j \pmod{m}$ for each j, and then $c_j a^j \equiv c_j b^j \pmod{m}$ by (ii). Finally, repeated application of (i) gives $f(a) = \sum_{j=0}^{n} c_j a^j \equiv \sum_{j=0}^{n} c_j b^j = f(b) \pmod{m}$. □

Remark on the computation of powers. In many applications we need to compute powers a^k modulo m. The naive approach would invoke $k-1$ multiplications. This is fine if k is small, but for large numbers k such as in the RSA-algorithm, to be discussed in section 7, this is prohibitively time consuming. Instead, one should compute a^k recursively using the formula

$$a^k = \begin{cases} (a^{k/2})^2 = (a^{[k/2]})^2 & \text{if } k \text{ is even,} \\ a \cdot (a^{(k-1)/2})^2 = a \cdot (a^{[k/2]})^2 & \text{if } k \text{ is odd.} \end{cases}$$

Thus, a^k is obtained from $a^{[k/2]}$ by using one multiplication (squaring) if k is even, and two multiplications (squaring followed by multiplication by a) if k is odd. Depending on the value of k, the innermost computation of the recursion will be a^2 or $a^3 = a \cdot a^2$.

The total number of multiplications required to compute a^k from a using recursion is of the order of magnitude $\log k$, which is small compared to k. Indeed, if k has the binary expansion $k = \alpha_r\alpha_{r-1}\dots\alpha_1\alpha_0 = \sum_{j=0}^{r}\alpha_j 2^j$, (with $\alpha_r = 1$), then $[k/2] = \alpha_r\alpha_{r-1}\dots\alpha_1$, and k is odd if $\alpha_0 = 1$ and even if $\alpha = 0$. It now easily follows that the number of squarings needed equals r, and that the number of extra multiplications by a equals the number of nonzero digits α_j minus 1. Thus, at most $2r$ multiplications are needed.

Example 1 The computation of $3^{1304} \pmod{121}$ by recursion can be summarized in the following table:

k	1304	652	326	163	162	81	80	40	20	10	5	4	2	1
$3^k \pmod{121}$	81	9	3	27	9	3	1	1	1	1	1	81	9	3

The numbers in the top row are computed from left to right. If a number is even, the next number is obtained by dividing it by 2, and if a number is odd the next one is obtained by subtracting 1. The numbers in the bottom row are computed from right to left. For instance, $3^4 = (3^2)^2 \equiv 9^2 \equiv 81$, $3^5 = 3\cdot 3^4 \equiv 3\cdot 81 \equiv 243 \equiv 1$, $3^{326} = (3^{163})^2 \equiv 27^2 \equiv 3$. □

We next investigate what happens when the modulus is multiplied or divided by a number. The simple proof of the following proposition is left to the reader.

Proposition 4.4 *Let c be an arbitrary positive integer, and let d be a positive divisor of m.*

(i) If $a \equiv b \pmod{m}$, then $ac \equiv bc \pmod{mc}$.

(ii) If $a \equiv b \pmod{m}$, then $a \equiv b \pmod{d}$.

In general, congruences may not be divided without changing the modulus. We have the following result.

Proposition 4.5 *Let c be a non-zero integer.*

(i) If $ca \equiv cb \pmod{m}$, then $a \equiv b \pmod{m/(c,m)}$

(ii) If $ca \equiv cb \pmod{m}$ and $(c,m) = 1$, then $a \equiv b \pmod{m}$.

Proof. (i) Let $d = (c,m)$. If $ca \equiv cb \pmod{m}$, then $m \mid c(a-b)$ and $\dfrac{m}{d} \,\Big|\, \dfrac{c}{d}(a-b)$. Since $\left(\dfrac{m}{d},\dfrac{c}{d}\right) = 1$, it follows that $\dfrac{m}{d} \,\Big|\, (a-b)$, i.e. $a \equiv b \pmod{m/d}$.

(ii) is a special case of (i). □

A system of congruences can be replaced by one congruence in the following way:

Proposition 4.6 *Let $m_1, m_2, \dots, m_r$ be positive integers. The following two statements are then equivalent:*

(i) $a \equiv b \pmod{m_i}$ for $i = 1, 2, \dots, r$.

(ii) $a \equiv b \pmod{[m_1, m_2, \dots, m_p]}$.

Proof. Suppose $a \equiv b \pmod{m_i}$ for all i. Then $(a-b)$ is a common multiple of all the m_is, and therefore $[m_1, m_2, \dots, m_p] \mid (a-b)$. This means that $a \equiv b \pmod{[m_1, m_2, \dots, m_r]}$.

Conversely, if $a \equiv b \pmod{[m_1, m_2, \dots, m_r]}$, then $a \equiv b \pmod{m_i}$ for each i, since $m_i \mid [m_1, m_2, \dots, m_r]$. □

For the rest of this section, we fix a positive integer m which we will use as modulus.

Definition 4.7 Let a be an integer. The set $\overline{a} = \{x \in \mathbf{Z} \mid x \equiv a \pmod{m}\}$ of all integers that are congruent modulo m to a is called a *residue class*, or *congruence class*, modulo m.

Since the congruence relation is an equivalence relation, it follows that all numbers belonging to the same residue class are mutually congruent, that numbers belonging to different residue classes are incongruent, that given two integers a and b either $\overline{a} = \overline{b}$ or $\overline{a} \cap \overline{b} = \emptyset$, and that $\overline{a} = \overline{b}$ if and only if $a \equiv b \pmod{m}$.

Proposition 4.8 *There are exactly m distinct residue classes modulo m, viz. $\overline{0}$, $\overline{1}$, $\overline{2}$, ..., $\overline{m-1}$.*

Proof. According to the division algorithm, there is for each integer a a unique integer r belonging to the interval $[0, m-1]$ such that $a \equiv r \pmod{m}$. Thus, each residue class $\overline{a}$ is identical with one of the residue classes $\overline{0}$, $\overline{1}$, $\overline{2}$, ..., $\overline{m-1}$, and these are different since $i \not\equiv j \pmod{m}$ if $0 \le i < j \le m-1$. □

Definition 4.9 Chose a number x_i from each residue class modulo m. The resulting set of numbers $x_1, x_2, \dots, x_m$ is called a *complete residue system* modulo m.

The set $\{0, 1, 2, \dots, m-1\}$ is an example of a complete residue system modulo m.

Example 2 $\{4, -7, 14, 7\}$ is a complete residue system modulo 4. □

Lemma 4.10 *If x and y belong to the same residue class modulo m, then $(x, m) = (y, m)$.*

Proof. If $x \equiv y \pmod{m}$, then $x = y + qm$ for some integer q, and it follows from Proposition 1.4 that $(x, m) = (y, m)$. □

Two numbers a and b give rise to the same residue class modulo m, i.e. $\overline{a} = \overline{b}$, if and only if $a \equiv b \pmod{m}$. The following definition is therefore consistent by virtue of Lemma 4.10.

Definition 4.11 A residue class $\overline{a}$ modulo m is said to be *relatively prime* to m if $(a, m) = 1$.

Definition 4.12 Let $\phi(m)$ denote the number of residue classes modulo m that are relatively prime to m. The function ϕ is called *Euler's ϕ-function.* Any set $\{r_1, r_2, \dots, r_{\phi(m)}\}$ of integers obtained by choosing one integer from each of the residue classes that are relatively prime to m, is called a *reduced residue system* modulo m.

The following two observations are immediate consequences of the definitions: The number $\phi(m)$ equals the number of integers in the interval $[0, m-1]$ that are relatively prime to m. $\{y_1, y_2, \dots, y_{\phi(m)}\}$ is a reduced residue system modulo m if and only if the numbers are pairwise incongruent modulo m and $(y_i, m) = 1$ for all i.

Example 3 The positive integers less than 8 that are relatively prime to 8 are 1, 3, 5, and 7. It follows that $\phi(8) = 4$ and that $\{1, 3, 5, 7\}$ is a reduced residue system modulo 8. □

Example 4 If p is a prime, then the numbers 1, 2, …, $p-1$ are all relatively prime to p. It follows that $\phi(p) = p-1$ and that $\{1, 2, \dots, p-1\}$ is a reduced residue system modulo p. □

Example 5 Let p^k be a prime power. An integer is relatively prime to p^k if and only if it is not divisible by p. Hence, in the interval $[0, p^k - 1]$ there are p^{k-1} integers that are not relatively prime to p, viz. the integers np, where $n = 0$, 1, 2, …, $p^{k-1} - 1$, whereas the remaining $p^k - p^{k-1}$ integers in the interval are relatively prime to p. Consequently,

$$\phi(p^k) = p^k - p^{k-1} = p^k\left(1 - \frac{1}{p}\right). \qquad \square$$

Theorem 4.13 *Let $(a, m) = 1$. Let $\{r_1, r_2, \dots, r_m\}$ be a complete residue system, and let $\{s_1, s_2, \dots, s_{\phi(m)}\}$ be a reduced residue system modulo m. Then $\{ar_1, ar_2, \dots, ar_m\}$ is a complete and $\{as_1, as_2, \dots, as_{\phi(m)}\}$ is a reduced residue system modulo m.*

Proof. In order to show that the set $\{ar_1, ar_2, \dots, ar_m\}$ is a complete residue system, we just have to check that the elements are chosen from distinct residue classes, i.e. that $i \neq j \Rightarrow ar_i \not\equiv ar_j \pmod{m}$. But by Proposition 4.5 (ii), $ar_i \equiv ar_j \pmod{m}$ implies $r_i \equiv r_j \pmod{m}$ and hence $i = j$.

Since $(s_i, m) = 1$ and $(a, m) = 1$, we have $(as_i, m) = 1$ for $i = 1, 2, \dots, \phi(m)$ by Theorem 1.14. Hence as_1, as_2, …, $as_{\phi(m)}$ are $\phi(m)$ numbers belonging to residue classes that are relatively prime to m, and by the same argument as above they are chosen from distinct residue classes. It follows that they form a reduced residue system. □

Theorem 4.14 (Euler's theorem) *If $(a, m) = 1$, then*

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

Proof. Let $\{s_1, s_2, \dots, s_{\phi(m)}\}$ be a reduced residue system modulo m. By Theorem 4.13, the set $\{as_1, as_2, \dots, as_{\phi(m)}\}$ is also a reduced residue system. Consequently, to each s_i there corresponds one and only one as_j such that $s_i \equiv as_j \pmod{m}$. By multiplying together and using Proposition 4.3 (ii), we thus get

$$\prod_{j=1}^{\phi(m)} (as_j) \equiv \prod_{i=1}^{\phi(m)} s_i \pmod{m},$$

and hence

$$a^{\phi(m)}\prod_{j=1}^{\phi(m)} s_j \equiv \prod_{i=1}^{\phi(m)} s_i \pmod{m}.$$

Since $(s_i, m) = 1$, we can use Proposition 4.5 (ii) repeatedly to cancel the s_i, and we obtain $a^{\phi(m)} \equiv 1 \pmod{m}$. □

The following theorem is an immediate corollary.

Theorem 4.15 (Fermat's theorem) *If p is a prime and $p \not| a$, then*

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

For every integer a, $a^p \equiv a \pmod{p}$.

Proof. If $p \not| a$, then (a, p)=1. Since $\phi(p) = p - 1$ by Example 4, the first part now follows immediately from Euler's theorem. By multiplying the congruence by a, we note that $a^p \equiv a \pmod{p}$, and this obvioulsy holds also in the case $a \equiv 0 \pmod{p}$. □

Example 6 Modulo 7 we get $3^1 \equiv 3$, $3^2 \equiv 2$, $3^3 \equiv 6$, $3^4 \equiv 4$, $3^5 \equiv 5$, and finally $3^6 \equiv 1$ in accordance with Fermat's theorem. Similarly, $2^1 \equiv 2$, $2^2 \equiv 4$, $2^3 \equiv 1$, and hence $2^6 \equiv 1$. □

5 Linear Congruences

The congruence

$$(1) \qquad ax \equiv b \pmod{m}$$

is equivalent to the equation

$$(2) \qquad ax - my = b$$

where we of course only consider integral solutions x and y. We know from Theorem 3.1 that this equation is solvable if and only if $d = (a, m)$ divides b, and if x_0, y_0 is a solution then the complete set of solution is given by

$$x = x_0 + \frac{m}{d}n, \quad y = y_0 + \frac{a}{d}n.$$

We get d pairwise incongruent x-values modulo m by taking $n = 0, 1, \ldots, d-1$, and any solution x is congruent to one of these. This proves the following theorem.

Theorem 5.1 *The congruence*

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

is solvable if and only if $(a, m) \mid b$. If the congruence is solvable, then it has exactly (a, m) pairwise incongruent solutions modulo m.

We have the following immediate corollaries.

Corollary 5.2 *The congruene* $ax \equiv 1 \pmod{m}$ *is solvable if and only if* $(a, m) = 1$*, and in this case any two solutions are congruent modulo* m*.*

Corollary 5.3 *If* $(a, m) = 1$*, then the congruence* $ax \equiv b \pmod{m}$ *is solvable for any* b *and any two solutions are congruent modulo* m*.*

Note that the existence of a solution in Corollories 5.2 and 5.3 can also be deduced from Euler's theorem. By taking $x_0 = a^{\phi(m)-1}$ and $x_1 = bx_0$ we obtain $ax_0 = a^{\phi(m)} \equiv 1 \pmod{m}$ and $ax_1 = bax_0 \equiv b \pmod{m}$.

However, in order to solve the congruence (1) it is usually more efficient to solve the equivalent equation (2) using the methods from section 3. Another possibility is to replace the congruence (1) by a congruence with a smaller modulus and then reduce the coefficients in the following way:

In (1) we can replace the numbers a and b with congruent numbers in the interval $[0, m-1]$, or still better in the interval $[-m/2, m/2]$. Assuming this done, we can now write equation (2) as

$$my \equiv -b \pmod{a} \tag{3}$$

with a module a that is less than the module m in (1). If $y = y_0$ solves (3), then

$$x = \frac{my_0 + b}{a}$$

is a solution to (1). Of course, the whole procedure can be iterated again and again until finally a congruence of the form $z \equiv c \pmod{n}$ is obtained.

Example 1 Solve the congruence

$$296x \equiv 176 \pmod{114}. \tag{4}$$

Solution: Since 2 divides the numbers 296, 176, and 114, we start by replacing (4) with the following equivalent congruence:

$$148x \equiv 88 \pmod{57}. \tag{5}$$

Now, reduce 148 and 88 modulo 57. Since $148 \equiv -23$ and $88 \equiv -26$, we can replace (5) with

$$23x \equiv 26 \pmod{57}. \tag{6}$$

Now we consider instead the congruence

$$57y \equiv -26 \pmod{23},$$

which of course is quivalent to

$$11y \equiv -3 \pmod{23}. \tag{7}$$

Again, replace this with the congruence

$$23z \equiv 3 \pmod{11}$$

which is at once reduced to

$$z \equiv 3 \pmod{11}.$$

Using this solution, we see that

$$y = \frac{23 \cdot 3 - 3}{11} = 6$$

is a solution to (7) and that all solutions have the form $y \equiv 6 \pmod{23}$. It now follows that

$$x = \frac{57 \cdot 6 + 26}{23} = 16$$

solves (6) and the equivalent congruence (4), and that all solutions are of the form $x \equiv 16 \pmod{57}$, which can of course also be written as $x \equiv 16, 73$ (mod 114). □

Concluding remarks. These remarks are intended for readers who are familiar with elementary group theory.

Let $\mathbf{Z}_m^*$ denote the set of all residue classes modulo m that are relatively prime to the module m. We can equip $\mathbf{Z}_m^*$ with a multiplication operation by defining the product of two residue classes as follows:

$$\overline{a} \cdot \overline{b} = \overline{ab}.$$

For this definition to be well behaved it is of course necessary that the residue class $\overline{ab}$ be dependent on the *residue classes* $\overline{a}$ and $\overline{b}$ only, and not on the particular numbers a and b chosen to represent them, and that $\overline{ab}$ belong to $\mathbf{Z}_m^*$. However, this follows from Proposition 4.3 (ii) and Theorem 1.14.

The multiplication on $\mathbf{Z}_m^*$ is obviously associative and commutative, and there is an identity element, namely the class $\overline{1}$. Moreover, it follows from Corollary 5.2 that the equation $\overline{a} \cdot \overline{x} = \overline{1}$ has a unique solution $\overline{x} \in \mathbf{Z}_m^*$ for each $\overline{a} \in \mathbf{Z}_m^*$. Thus, each element in $\mathbf{Z}_m^*$ has a unique multiplicative inverse.

This shows that $\mathbf{Z}_m^*$ is a finite abelian (commutative) group. The order of the group (i.e. the number of elements in the group) equals $\phi(m)$, by definition of the Euler ϕ-function.

One of the first theorems encountered when studying groups reads: If n is the order of a finite group with identity element e, then $a^n = e$ for every element a in the group. Applying this result to the group $\mathbf{Z}_m^*$, we recover Euler's theorem, since the statement

$$\overline{a}^{\,\phi(m)} = \overline{1}$$

is just another way of saying that

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

holds for every number a that is relatively prime to m.

6 The Chinese Remainder Theorem

Let us start by considering a system of two congruences

$$\begin{cases} x \equiv a_1 \pmod{m_1} \\ x \equiv a_2 \pmod{m_2} \end{cases}$$

where $(m_1, m_2) = 1$. The first congruence has the solutions $x = a_1 + m_1 y$, $y \in \mathbf{Z}$, and by substituting this into the second congruence, we obtain $a_1 + m_1 y \equiv a_2 \pmod{m_2}$, that is $m_1 y \equiv a_2 - a_1 \pmod{m_2}$. Now, since $(m_1, m_2) = 1$, this congruence has solutions of the form $y = y_0 + m_2 n$ and hence $x = a_1 + m_1 y_0 + m_1 m_2 n$. This shows that the system has a unique solution $x \equiv x_0 \pmod{m_1 m_2}$.

Consider now a system of three congruences

$$\begin{cases} x \equiv a_1 \pmod{m_1} \\ x \equiv a_2 \pmod{m_2} \\ x \equiv a_3 \pmod{m_3} \end{cases} \tag{1}$$

where the moduli m_1, m_2 and m_3 are pairwise relatively prime. As shown above, we can replace the first two congruences with an equivalent congruence of the form $x \equiv x_0 \pmod{m_1 m_2}$, and hence the whole system (1) is equivalent to a system of the form

$$\begin{cases} x \equiv x_0 \pmod{m_1 m_2} \\ x \equiv a_3 \pmod{m_3}. \end{cases} \tag{2}$$

Now, by assumption $(m_1 m_2, m_3) = 1$, and hence (2) has a unique solution $x \equiv x_1 \pmod{m_1 m_2 m_3}$.

By induction, it is now very easy to prove the following general result.

Theorem 6.1 (The Chinese Remainder Theorem) *The system*

$$\begin{cases} x \equiv a_1 \pmod{m_1} \\ x \equiv a_2 \pmod{m_2} \\ \quad \vdots \\ x \equiv a_r \pmod{m_r} \end{cases} \tag{3}$$

where $m_1, m_2, \dots, m_r$ are pairwise relatively prime, has a unique solution modulo $m_1 m_2 \cdots m_r$.

Proof. We will give a second proof of the theorem and also derive a formula for the solution.

Let for each $j = 1, 2, \dots, r$, δ_j be an integer satisfying

$$\delta_j \equiv \begin{cases} 1 \pmod{m_j} \\ 0 \pmod{m_i}, \quad \text{if } i \neq j. \end{cases}$$

Then obviously

$$x = \sum_{j=1}^{r} \delta_j a_j \tag{4}$$

satisfies the system (3).

It remains to prove that the numbers δ_j exist. Put $m = m_1 m_2 \cdots m_r$. By assumption $\left(\dfrac{m}{m_j}, m_j\right) = 1$ and hence, by Corollary 5.2, there is a number b_j such that

$$\frac{m}{m_j} b_j \equiv 1 \pmod{m_j}.$$

The numbers $\delta_j = \dfrac{m}{m_j} b_j$ will now clearly have the desired properties.

This proves the existence of a solution x to (3). To prove that the solution is unique modulo m, suppose x' is another solution. Then $x \equiv x' \pmod{m_j}$ holds for $j = 1, 2, \ldots, r$, and it follows from Proposition 4.6 that $x \equiv x' \pmod{m_1 m_2 \cdots m_r}$. □

Formula (4) is particularly useful when we are to solve several systems (3) with the same moduli but with different right hand members $a_1, a_2, \ldots, a_r$.

Example 1 Let us solve the system

$$\begin{cases} x \equiv 1 \pmod 3 \\ x \equiv 2 \pmod 4 \\ x \equiv 3 \pmod 5. \end{cases}$$

Solution 1: Using the method in our first proof of the Chinese Remainder Theorem, we replace the first congruence by $x = 1 + 3y$. Substituting this into the second congruence we obtain $3y + 1 \equiv 2 \pmod 4$ or $3y \equiv 1 \pmod 4$. This congruence has the solutions $y \equiv -1 \pmod 4$, i.e. $y = -1 + 4z$. Hence, $x = -2 + 12z$, and substituting this into the last congruence we end up in the congruence $12z - 2 \equiv 3 \pmod 5$ or $12z \equiv 5 \equiv 0 \pmod 5$. This congruence has the unique solution $z \equiv 0 \pmod 5$, that is $z = 5t$ and $x = -2 + 60t$. Hence, the system has the unique solution $x \equiv -2 \pmod{60}$.

Solution 2: Let us instead use the method of the second proof. Then we have first to find numbers b_1, b_2, and b_3 such that

$$20b_1 \equiv 1 \pmod 3, \qquad 15b_2 \equiv 1 \pmod 4, \qquad 12b_3 \equiv 1 \pmod 5.$$

One easily obtains $b_1 = 2$, $b_2 = 3$, and $b_3 = 3$. Next, we compute $\delta_1 = 20b_1 = 40$, $\delta_2 = 15b_2 = 45$, and $\delta_3 = 12b_3 = 36$. Finally,

$$x = \delta_1 + 2\delta_2 + 3\delta_3 = 40 + 90 + 108 = 238 \equiv 58 \pmod{60}.$$ □

The condition that the moduli $m_1, m_2, \ldots, m_r$ be pairwise relatively prime is absolutely essential for the conclusion of Theorem 6.1. Without that condition the system (3) is either unsolvable or there are more than one incongruent solution modulo $m_1 m_2 \cdots m_r$. Necessary and sufficient for the system to be solvable is that $(m_i, m_j) \mid (a_i - a_j)$ for all $i \neq j$. A given system can be solved or proved unsolvable by reasoning as in the first solution of Example 1.

We will now derive some important consequences of Theorem 6.1. Given a positive integer n we let $\mathcal{C}(n)$ denote a fixed complete residue system modulo n. The subset of all numbers in $\mathcal{C}(n)$ that are relatively prime to n forms a reduced residue system which we denote by $\mathcal{R}(n)$. The set $\mathcal{R}(n)$ contains $\phi(n)$ numbers. To be concrete, we could choose $\mathcal{C}(n) = \{0, 1, 2, \ldots, n-1\}$; then $\mathcal{R}(n) = \{j \mid 0 \leq j \leq n-1 \text{ and } (j, n) = 1\}$.

Let now m_1 and m_2 be two relatively prime numbers and put $m = m_1 m_2$. Then $\mathcal{C}(m)$ and the Cartesian product $\mathcal{C}(m_1) \times \mathcal{C}(m_2)$ contain the same number of elements, viz. m. We will construct a bijection τ between these two sets.

Given $x \in \mathcal{C}(m)$ and $j = 1$ or 2, we denote by x_j the unique number in $\mathcal{C}(m_j)$ that satisfies $x_j \equiv x \pmod{m_j}$. We then define $\tau(x) = (x_1, x_2)$.

A map between two sets with the same number of elements is a bijection if and only if it is surjective. But surjectivity of the map τ follows immediately from the Chinese Remainder Theorem, because given $(x_1, x_2) \in \mathcal{C}(m_1) \times \mathcal{C}(m_2)$, there is a (unique) $x \in \mathcal{C}(m)$ such that $x \equiv x_1 \pmod{m_1}$ and $x \equiv x_2 \pmod{m_2}$, which amounts to saying that $\tau(x) = (x_1, x_2)$.

We will next identify the image $\tau(\mathcal{R}(m))$ of the reduced residue system $\mathcal{R}(m)$ under the map τ. Since

$$(x, m) = 1 \Leftrightarrow (x, m_1) = (x, m_2) = 1$$

and

$$x \equiv x_j \pmod{m_j} \Rightarrow ((x, m_j) = 1 \Leftrightarrow (x_j, m_j) = 1)$$

it follows that $x \in \mathcal{R}(m) \Leftrightarrow \tau(x) \in \mathcal{R}(m_1) \times \mathcal{R}(m_2)$. Thus, τ maps the set $\mathcal{R}(m)$ bijectively onto the Cartesian product $\mathcal{R}(m_1) \times \mathcal{R}(m_2)$. The former set has $\phi(m)$ elements and the latter has $\phi(m_1)\phi(m_2)$ elements. Since the two sets must have the same number of elements, we have proved the following important theorem about Euler's ϕ-function.

Theorem 6.2 *If $m = m_1 m_2$, where the integers m_1 and m_2 are relatively prime, then*

$$\phi(m) = \phi(m_1)\phi(m_2).$$

Corollary 6.3 *If $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$, where $p_1, p_2, \dots, p_r$ are different primes, then*

$$\phi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_r}\right).$$

Proof. By repeated application of Theorem 6.2, we obtain

$$\phi(m_1 m_2 \cdots m_r) = \phi(m_1)\phi(m_2)\cdots\phi(m_r)$$

if the integers $m_1, m_2, \dots, m_r$ are pairwise relatively prime. In particular, this holds when the numbers m_i are powers of distinct primes. By Example 5 in section 4, $\phi(p^k) = p^{k-1}(p-1) = p^k(1 - 1/p)$ if p is prime. □

A polynomial $f(x) = \sum_{i=0}^n a_i x^i$ with coefficients $a_i \in \mathbf{Z}$ is called an *integral polynomial*, and the congruence

$$f(x) \equiv 0 \pmod{m},$$

is called a *polynomial congruence.* An integer a is called a solution or a *root* of the polynomial congruence if $f(a) \equiv 0 \pmod{m}$.

If a is a root of the polynomial congruence and if $b \equiv a \pmod{m}$, then b is also a root. Therefore, in order to solve the polynomial cogruence it is enough to find all roots that belong to a given complete residue system $\mathcal{C}(m)$ modulo m, e.g. to find all solutions among the numbers 0, 1, 2, …, $m-1$. By *the number of roots* of a polynomial congruence we will mean the number of such incongruent roots.

Next, consider a system

$$\begin{cases} f_1(x) \equiv 0 \pmod{m_1} \\ f_2(x) \equiv 0 \pmod{m_2} \\ \qquad \vdots \\ f_r(x) \equiv 0 \pmod{m_r} \end{cases}$$

of polynomial congruences, where the moduli $m_1, m_2, \dots, m_r$ are assumed to be pairwise relatively prime. By a solution of such a system we mean, of course, an integer which solves simultaneously all the congruences of the system. If a is a solution of the system, and if $b \equiv a \pmod{m_1 m_2 \cdots m_r}$, then b is also a solution of the system, since for each j we have $b \equiv a \pmod{m_j}$. Hence, to find all solutions of the system it suffices to consider solutions belonging to a complete residue system modulo $m_1 m_2 \cdots m_r$; by the number of solutions of the system we will mean the number of such incongruent solutions.

Theorem 6.4 *Let*

$$(5) \qquad \begin{cases} f_1(x) \equiv 0 \pmod{m_1} \\ f_2(x) \equiv 0 \pmod{m_2} \\ \qquad \vdots \\ f_r(x) \equiv 0 \pmod{m_r} \end{cases}$$

be a system of polynomial congruences, and assume that the the moduli m_1, m_2, $\dots$, m_r are pairwise relatively prime. Let X_j be a complete set of incongruent solutions modulo m_j of the jth congruence, and let n_j denote the number of solutions. The number of solutions of the system then equals $n_1 n_2 \cdots n_r$, and each solution of the system is obtained as the solution of the system

$$\begin{cases} x \equiv a_1 \pmod{m_1} \\ x \equiv a_2 \pmod{m_2} \\ \quad \vdots \\ x \equiv a_r \pmod{m_r} \end{cases}$$

with $(a_1, a_2, \dots, a_r)$ ranging over the set $X_1 \times X_2 \times \cdots \times X_r$.

Of course, a set X_j might be empty in which case $n_j = 0$

Proof. Write $m = m_1 m_2 \cdots m_r$, let $\mathcal{C}(m_j)$ be a complete residue system modulo m_j containing the solution set X_j $(j = 1, 2, \dots, r)$, and let $\mathcal{C}(m)$ be a complete residue system modulo m containing the solution set X of the system (5) of congruences. By the Chinese Remainder Theorem we obtain a bijection

$$\tau : \mathcal{C}(m) \to \mathcal{C}(m_1) \times \mathcal{C}(m_2) \times \cdots \times \mathcal{C}(m_r)$$

by defining

$$\tau(x) = (x_1, x_2, \dots, x_r),$$

where each $x_j \in \mathcal{C}(m_j)$ is a number satisfying the congruence $x_j \equiv x \pmod{m_j}$.

If $a \in X$, then a is a solution of each individual congruence in the system (5). Consequently, if $a_j \in \mathcal{C}(m_j)$ and $a_j \equiv a \pmod{m_j}$, then a_j is a solution of the j^{th} congruence of the system, i.e. a_j belongs to the solution set X_j. We conclude that $\tau(a) = (a_1, a_2, \dots, a_r)$ belongs to the set $X_1 \times X_2 \times \dots \times X_r$ for each $a \in X$, and the image $\tau(X)$ of X under τ is thus a subset of $X_1 \times X_2 \times \dots \times X_r$.

Conversely, if $\tau(a) = (a_1, a_2, \dots, a_r) \in X_1 \times X_2 \times \dots \times X_r$, then a solves each individual congruence and thus belongs to X. This follows from Proposition 4.3, because $a \equiv a_j \pmod{m_j}$ and $f_j(a_j) \equiv 0 \pmod{m_j}$ for each j. Hence, the bijection τ maps the subset X onto the subset $X_1 \times X_2 \times \dots \times X_r$, and we conclude that the number of elements in X equals $n_1 n_2 \cdots n_r$. □

Example 2 Consider the system

$$\begin{cases} x^2 + x + 1 \equiv 0 \pmod{7} \\ 2x - 4 \equiv 0 \pmod{6}. \end{cases}$$

By trying $x = 0$, ± 1, ± 2, ± 3, we find that $x \equiv 2 \pmod{7}$ and $x \equiv -3 \pmod{7}$ are the solutions of the first congruence. Similarly, we find that $x \equiv -1 \pmod{6}$ and $x \equiv 2 \pmod{6}$ solve the second congruence. We conclude that the system has 4 incongruent solutions modulo 42. To find these, we have to solve each of the following four systems:

$$\begin{cases} x \equiv 2 \pmod{7} \\ x \equiv -1 \pmod{6} \end{cases} \qquad \begin{cases} x \equiv 2 \pmod{7} \\ x \equiv 2 \pmod{6} \end{cases}$$

$$\begin{cases} x \equiv -3 \pmod{7} \\ x \equiv -1 \pmod{6} \end{cases} \qquad \begin{cases} x \equiv -3 \pmod{7} \\ x \equiv 2 \pmod{6}. \end{cases}$$

We use the solution formula (4) obtained in the proof of the Chinese Remainder Theorem. Thus, we determine b_1 and b_2 such that

$$\frac{42}{7} b_1 \equiv 1 \pmod{7} \quad \text{and} \quad \frac{42}{6} b_2 \equiv 1 \pmod{6}.$$

We easily find that $b_1 = -1$ and $b_2 = 1$ solve these congruences, and hence we can take $\delta_1 = -6$ and $\delta_2 = 7$. We conclude that four different solutions modulo 42 of our original system are

$$\begin{aligned} x_1 &= -6 \cdot 2 + 7 \cdot (-1) = -19 \equiv 23 \\ x_2 &= -6 \cdot 2 + 7 \cdot 2 = 2 \\ x_3 &= -6 \cdot (-3) + 7 \cdot (-1) = 11 \\ x_4 &= -6 \cdot (-3) + 7 \cdot 2 = 32. \end{aligned}$$ □

We now turn to an important special case of Theorem 6.4.

Theorem 6.5 *Let $f(x)$ be an integral polynomial. For each positive integer m, let $X(m)$ denote a complete set of roots modulo m of the polynomial congruence*

$$f(x) \equiv 0 \pmod{m},$$

and let $N(m)$ denote the number of roots.

Assume $m = m_1 m_2 \cdots m_r$, where the numbers $m_1, m_2, \ldots, m_r$ are pairwise relatively prime; then

$$N(m) = N(m_1)N(m_2)\cdots N(m_r).$$

Moreover, to each r-tuple $(a_1, a_2, \ldots, a_r) \in X(m_1) \times X(m_2) \times \cdots \times X(m_r)$ there corresponds a unique solution $a \in X(m)$ such that $a \equiv a_j \pmod{m_j}$ for each j.

Proof. By Proposition 4.6, the congruence $f(x) \equiv 0 \pmod{m}$ is equivalent to the system

$$\begin{cases} f(x) \equiv 0 \pmod{m_1} \\ f(x) \equiv 0 \pmod{m_2} \\ \quad\vdots \\ f(x) \equiv 0 \pmod{m_r}. \end{cases}$$

Hence, Theorem 6.4 applies. □

It follows that in order to solve a polynomial congruence modulo m it is sufficient to know how to solve congruences with prime power moduli.

Example 3 Let $f(x) = x^2 + x + 1$. Prove that the congruence $f(x) \equiv 0 \pmod{15}$ has no solutions.

Solution: By trying the values $x = 0, \pm 1, \pm 2$ we find that the congruence $f(x) \equiv 0 \pmod{5}$ has no solutions. Therefore, the given congruence modulo 15 $(= 5 \cdot 3)$ has no solutions. □

Example 4 Let $f(x) = x^2 + x + 9$. Find the roots of the congruence

$$f(x) \equiv 0 \pmod{63}.$$

Solution: Since $63 = 7 \cdot 9$, we start by solving the two congruences

$$f(x) \equiv 0 \pmod{7} \quad \text{and} \quad f(x) \equiv 0 \pmod{9}.$$

The first congruence has the sole root 3 (mod 7), and the second congruence has the roots 0 and -1 (mod 9). It follows that the given congruence has two roots modulo 63, and they are obtained by solving the congruences

$$\begin{cases} x \equiv 3 \pmod{7} \\ x \equiv 0 \pmod{9} \end{cases} \quad \text{and} \quad \begin{cases} x \equiv 3 \pmod{7} \\ x \equiv -1 \pmod{9}. \end{cases}$$

Using the Chinese remainder theorem, we find that the roots are 45 and 17 modulo 63. □

7 Public-Key Cryptography

In 1977 R.L. Rivest, A. Shamir and L.M. Adleman invented an asymmetric encryption scheme which has been called the *RSA algorithm* and which uses congruence arithmetic. The method uses two keys, one *public encryption key* and one *secret private decryption key.* The security of the algorithm depends on the hardness of factoring a large composite number and computing e^{th} roots modulo a composite number for a specified integer e.

The RSA algorithm is based on the following theorem.

Theorem 7.1 *Suppose that m is a positive square-free integer, i.e. that the canonical prime factorization $m = p_1p_2 \cdots p_r$ consists of distinct primes, and let e and d be positive integers such that $ed \equiv 1 \pmod{\phi(m)}$. Then $a^{ed} \equiv a \pmod{m}$ holds for each integer a.*

Proof. By Proposition 4.6 it suffices to prove that $a^{ed} \equiv a \pmod{p}$ holds for each prime $p = p_i$ dividing the modulus m. This is trivially true if $a \equiv 0 \pmod{p}$, because then $a^{ed} \equiv a \equiv 0 \pmod{p}$. Hence, we may assume that $a \not\equiv 0 \pmod{p}$.

By assumption, $ed = 1 + n\phi(m)$ for some nonnegative integer n, and

$$\phi(m) = \phi(p \cdot m/p) = \phi(p)\phi(m/p) = (p-1)\phi(m/p).$$

Hence, $ed = 1 + n(p-1)\phi(m/p) = 1 + (p-1)N$, where N is a nonnegative integer. Therefore, by Fermat's theorem

$$a^{ed} = a^{1+(p-1)N} = a \cdot (a^{p-1})^N \equiv a \cdot 1^N = a \pmod{p}.$$ □

An RSA public key consists of a pair (m, e) of integers. The number m will be used as the modulus, and e is the public exponent. The modulus m is the product of two distinct large primes p and q (the current recommendation is that each prime should have a size of at least 2^{512}). The exponent e must be relatively prime to $\phi(m)$, that is to $p-1$ and $q-1$, and it is normally chosen to be a small prime such as 3 $(= 2+1)$, 17 $(= 2^4 + 1)$, or 65537 $(= 2^{16} + 1)$, since powers $a^e \pmod{m}$ can be computed very fast for these particular choices of e.

In actual implementations of the RSA algorithm, the exponent e is first fixed. Then the primes p and q satisfying $(p-1, e) = (q-1, e) = 1$ are generated at random in such a way that each prime of the desired size (say 2^{512}) has the same probability of being chosen. Finally, $m = pq$.

The private key consists of the pair (m, d), where d is the unique positive number less than $\phi(m)$ satisfying $ed \equiv 1 \pmod{\phi(m)}$. The number d, as well as the primes p and q, and the number $\phi(m)$, are kept secrete by the owner of the private key.

Suppose now that somebody, say Alice, wants to send a secret message to Bob, the owner of the private key. The first thing to do is to convert the message into an integer a in the range $[0, m-1]$ in some standard way. (We could for example use the standard ASCII code. Since the ASCII code encodes "H" as 072, "e" as 101, "l" as 108, "o" as 111, and "!' as 033', the message "Hello!" would by concatenation become the number $a = 072101108108111033$.) If the message is too long, the message representative a will be out of range, but the message could then be divided into a number of blocks which could be encoded separately.

The sender Alice now uses the public encryption key to compute the unique number b satisfying $b \equiv a^e \pmod{m}$ and $0 \leq b \leq m-1$. This number b, the ciphertext representative of a, is transmitted to Bob.

When b is received, Bob uses his private exponent d to find the unique number c satisfying $0 \leq c < m$ and $c \equiv b^d \pmod{m}$. By Theorem 7.1, $c = a$ and hence the secret number a is recovered.

Suppose that some third party gains access to the number b. In order to recover the number a he has to extract the e^{th} root of b. There seems to be no other feasible method for this than to find the number d, and for that he need to know $\phi(m)$, and for that he has to factor the number m. But factorization of integers with 1000 binary digits seems to be beyond the reach of today's algorithms and fastest computers. Therefore, it is believed that the RSA method is very secure.

It is important that the message number a is not too small relative to m, because if $a^e < m$ then we can find a from the ciphertext representative $b = a^e$ by just extracting the ordinary e^{th} root of the integer b. Therefore, one has to use padding techniques which extend numbers with few non-zero digits in order to obtain a secure algorithm. A detailed description of how this is done is beyond the scope of this presentation.

8 Pseudoprimes

If a number n is composite, then it has a prime factor p which is less than or equal to $\sqrt{n}$. Hence, if n is not divisible by any (prime) number less than or equal to $\sqrt{n}$, then n is a prime. In the worst case, this means that we have to perform about $\sqrt{n}$ divisions in order to determine whether the number n is composite.

However, in most cases Fermat's theorem 4.15 can be used to show that a given number n is composite without having to find any factors, because if $(a, n) = 1$ and $a^{n-1} \not\equiv 1 \pmod{n}$, then necessarily n is composite. Our ability to evaluate powers $a^k \pmod{n}$ quickly makes this to a very efficient method. (The number of multiplications and divisions needed is proportional to $\log n$ which is considerably less than $\sqrt{n}$.)

Example 1 To show that the number 221 is composite withour having to factor it, we compute $2^{220} \pmod{221}$. The computation is summarized in the following table

k	220	110	55	54	27	26	13	12	6	3	2	1
$2^k \pmod{221}$	16	30	128	64	8	4	15	118	64	8	4	2

i.e. $2^{220} \equiv 16 \pmod{221}$. It follows that 221 must be composite. Indeed, $221 = 13 \cdot 17$. □

The converse of Fermat's theorem is not true, that is $a^{m-1} \equiv 1 \pmod{m}$ does not imply that m is a prime, so the above procedure is inconclusive in certain cases.

Example 2 The number 341 is composite ($= 11 \cdot 31$), but still $2^{340} \equiv 1 \pmod{341}$ as follows from the table

k	340	170	85	84	42	21	20	10	5	4	2	1
$2^k \pmod{341}$	1	1	32	16	4	2	1	1	32	16	4	2

Thus, the test does not detect that 341 is composite. But we can of course try another base than 2. Using 3 instead, we obtain

k	340	170	85	84	42	21	20	10	5	4	2	1
$3^k \pmod{341}$	56	67	254	312	163	201	67	56	243	81	9	3

Since $3^{340} \equiv 56 \not\equiv 1 \pmod{341}$, we conclude that the number 341 is composite. □

There is another, easier way to see that 341 is composite that relies on the following lemma.

Lemma 8.1 *Let p be a prime. Then $x^2 \equiv 1 \pmod{p}$ if and only if $x \equiv \pm 1 \pmod{p}$.*

Proof. The lemma is a special case of the more general Theorem 9.7, but since the proof is very simple we give a separate proof here. The congruence $x^2 \equiv 1 \pmod{p}$ may be expressed as $(x-1)(x+1) \equiv 0 \pmod{p}$, that is $p \mid (x-1)(x+1)$. Since p is prime, this is equivalent to $p \mid (x-1)$ or $p \mid (x+1)$, and we conclude that $x \equiv \pm 1 \pmod{p}$. □

Example 2 (revisited) Looking back to Example 2, we see that $(2^{85})^2 = 2^{170} \equiv 1 \pmod{341}$. Since $2^{85} \equiv 32 \not\equiv \pm 1 \pmod{341}$, we conclude from Lemma 8.1 that 341 must be a composite number. □

Definition 8.2 Let a and n be positive integers. If $(a, n) = 1$ and $a^{n-1} \equiv 1 \pmod{n}$, then n is called a *probable prime to the base a*. A composite probable prime is called a *pseudoprime.*

Of course, all primes are probable primes to every base, but there are probable primes that are composite, i.e. there are pseudoprimes. Example 2 shows that 341 is a pseudoprime to the base 2, but that it is not a probable prime to the base 3.

A natural question is whether there exists any number n that is a pseudoprime to every base a such that $(a, n) = 1$. In other words, is there a composite number n such that $a^{n-1} \equiv 1 \pmod{n}$ holds for all a with $(a, n) = 1$? The answer is yes. Such numbers n are called *Carmichael numbers*, and it is now known that there exist infinitely many Carmichael numbers, the smallest being 561.

Example 3 The number 561 is composite, because $561 = 3 \cdot 11 \cdot 17$. Apply Fermat's theorem for the primes 3, 11, and 17; for any number a which is relatively prime to 561, we obtain $a^2 \equiv 1 \pmod{3}$, $a^{10} \equiv 1 \pmod{11}$, and $a^{16} \equiv 1 \pmod{17}$. We next note that $560 = 2 \cdot 280 = 10 \cdot 56 = 16 \cdot 35$. Hence, by raising the first congruence to the power 280, the second to the power 56 and the third to the power 35, we get $a^{560} \equiv 1 \pmod{m}$ for $m = 3$, 11, and 17, and hence $a^{560} \equiv 1 \pmod{3 \cdot 11 \cdot 17}$. This shows that 561 is a probable prime for all bases that are relatively prime to 561, that is 561 is a Carmichael number. □

Example 2, revisited, indicates that there is a stronger way of testing whether a number is composite than just trying to show that it is not a probable prime to some base. This is the so called *strong pseudoprime test*, which will now be described.

Suppose we wish to show that an odd number n is composite. We start by dividing the even number $n-1$ by 2 repeatedly, in order to write $n - 1 = 2^k d$, with d odd. We then form the numbers

$$a^d, a^{2d}, a^{4d}, \ldots, a^{2^k d} \pmod{n}$$

by repeatedly squaring and reducing. If the last number is $\not\equiv 1 \pmod{n}$, then n is composite. If this last number is $\equiv 1 \pmod{n}$, then n is a probable prime to the base a. Assuming this is the case, let $a^{2^j d}$ be the first number in the above sequence that is $\equiv 1 \pmod{n}$. If $j \geq 1$ and the immediately preceding entry $a^{2^{j-1}d}$ is $\not\equiv -1 \pmod{n}$, then we may still conclude (Lemma 8.1) that n is composite. When this test is inconclusive, that is when either $a^d \equiv \pm 1 \pmod{n}$, or $j \geq 1$ is the least integer such that $a^{2^j d} \equiv 1 \pmod{n}$ and $a^{2^{j-1}d} \equiv -1 \pmod{n}$, then n is called a *strong probable prime to the base a.* An odd, composite, strong probable prime is called a *strong pseudoprime.*

It can be shown that there are no numbers that are strong pseudoprimes to every base.

The strong pseudoprimes are rare. In the interval $1 \leq n \leq 25 \cdot 10^9$, there are 1 091 987 405 primes, 2 163 Carmichael numbers, 4 842 strong pseudoprimes to the base 2, 184 numbers that are strong pseudoprimes to both the base 2 and the base 3, 13 that are strong pseudoprimes to the bases 2, 3, and 5, and only one number that is a strong pseudoprime to the bases 2, 3, 5, and 7.

The smallest strong pseudoprime to the base 2 is the number 2047.

EXAMPLE 4 To show that 2047 is a strong pseudoprime to the base 2, we write $2046 = 2 \cdot 1023$. Since $1023 = 2^{10} - 1 = \sum_{j=0}^{9} 2^j$, the binary expansion of 1023 will consist of ten 1s. By repeatedly squaring and reducing, we compute the powers 2^{2^j} modulo 2047 for $0 \leq j \leq 9$, and by multiplying them together we find that $2^{1023} \equiv 1 \pmod{2047}$. Hence, 2047 is a strong probable prime. Since $2047 = 23 \cdot 89$, it is a strong pseudoprime.

Instead of finding the factors, we can of course also try another base. Using base 3, we obtain $3^{1023} \equiv 1565 \pmod{2047}$ and $3^{2046} \equiv 1013 \pmod{2047}$. This proves that 2047 is not a probable prime to the base 3, that is 2047 is composite. □

9 Polynomial Congruences with Prime Moduli

In this and the following three sections we will study polynomial congruences. Theorem 6.5 reduces the study of congruences $f(x) \equiv 0 \pmod{m}$ with a general modulus m to the case when m is a prime power p^k. The case $k = 1$, that is congruences with prime moduli, is treated in this sections, while congruences with prime power moduli p^k and $k \geq 2$ will be discussed in section 10. Sections 11 and 12 are devoted to quadratic congruences. Finally, some additional results on the congruence $x^n \equiv a \pmod{m}$ appear in section 15.

First, however, we recall some general notions for polynomials and the division algorithm.

Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be an integral polynomial. The largest integer k such that $a_k \neq 0$ is called the *degree* of the polynomial, abbreviated $\deg f(x)$, and the corresponding coefficient a_k is called the *leading term* of the polynomial. This leaves the degree of $f(x)$ undefined when $f(x)$ is the zero polynomial, i.e. when all coefficients a_i are zero. To have the degree defined in that case, too, we define the degree of the zero polynomial to be the symbol $-\infty$, which we consider to be less than all integers.

Thus, a phrase like "$f(x)$ is a polynomial of degree $< n$" means that $f(x)$ is a non-zero polynomial of (ordinary) degree $< n$ or the zero polynomial.

If $f(x) = \sum_{i=0}^{n} a_i x^i$, $a_i \equiv b_i \pmod{m}$ and $g(x) = \sum_{i=0}^{n} b_i x^i$, then clearly $f(x) \equiv g(x) \pmod{m}$ for all x. Hence, in a congruence $f(x) \equiv 0 \pmod{m}$ we may reduce the coefficients modulo m, and in particular we may delete terms $a_i x^i$ with $a_i \equiv 0 \pmod{m}$ without changing the solution set.

Example 1 The congruence

$$20x^5 + 17x^4 + 12x^2 + 11 \equiv 0 \pmod{4}$$

is equivalent to the congruence

$$x^4 + 3 \equiv 0 \pmod{4},$$

and by trying -1, 0, 1, 2 we find the solutions $x \equiv \pm 1 \pmod{4}$. □

Since coefficients that are divisible by the modulus m can be treated as zero, it is sometimes very useful to consider the *degree modulo m* or *m-degree* of $f(x) = \sum_{i=0}^{n} a_i x^i$ which is defined to be the largest integer i such that $m \nmid a_i$. (If all coefficients are divisible by m, then the m-degree is defined to be $-\infty$.) Thus, the polynomial in Example 1 has 4-degree equal to 4. However, we will not need the notion of m-degree, and *henceforth, degree will mean the ordinary degree of a polynomial.*

When an integral polynomial $f(x)$ is divided by an integral polynomial $g(x)$, the quotient and remainder need not be integral polynomials. However, if the leading coefficient of $g(x)$ is 1, then the quotient and the remainder are integral.

Theorem 9.1 (The Division Algorithm for Integral Polynomials) *Let $f(x)$ and $g(x)$ be two integral polynomials, and assume the leading coefficient of $g(x)$ is equal to 1. Then there exist two unique integral polynomials $q(x)$ and $r(x)$ such that $f(x) = q(x)g(x) + r(x)$ and $\deg r(x) < \deg g(x)$.*

Proof. Let n be the degree of the polynomial $f(x)$, let ax^n be the leading term of $f(x)$, and let k be the degree of $g(x)$. If $k = 0$, then $g(x)$ is the constant 1 and there is nothing to prove, so assume $k \geq 1$. The proof of existence is by induction on n. If $n < k$, we take $q(x)$ to be the zero polynomial and $r(x) = f(x)$. Assume now that $n \geq k$ and that we have proved the existence of $q(x)$ and $r(x)$ for all polynomials $f(x)$ of degree less than n. Consider the polynomial polynomial $f(x) - ax^{n-k}g(x)$; it is a polynomial of degree $n_1 < n$, since the leading term of $f(x)$ is cancelled out, and by our induction hypothesis there exist polynomials $q_1(x)$ and $r(x)$, such that $f(x) - ax^{n-k}g(x) = q_1(x)g(x) + r(x)$ and $\deg r(x) < k$. Obviously, the polynomials $q(x) = ax^{n-k} + q_1(x)$ and $r(x)$ fulfill the requirements, and this completes the induction step. We leave to the reader to prove uniqueness. □

The following modulus version of the ordinary factor theorem for polynomials is an immediate consequence of the division algorithm.

Theorem 9.2 *Assume $f(x)$ is an integral polynomial. Then, the integer a is a root of the congruence $f(x) \equiv 0 \pmod{m}$ if and only if there exist an integral polynomial $q(x)$ and an integer b such that*

$$f(x) = (x - a)q(x) + mb.$$

Proof. Use the division algorithm to write $f(x) = (x-a)q(x) + c$, where the quotient $q(x)$ is an integral polynomial and the remainder c is a constant polynomial, i.e. an integer. Now, $f(a) = c$ and hence a is a root of the congruence if and only if $c \equiv 0 \pmod{m}$, i.e. if and only if $c = mb$ for some integer b. □

We now turn to polynomial congruences

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

where the modulus p is a prime number. If the degree of $f(x)$ is greater than or equal to p, we can reduce the degree in the following way: Divide the polynomial $f(x)$ by $x^p - x$; according to the division algorithm there are two integral polynomials $q(x)$ and $r(x)$ such that $f(x) = (x^p - x)q(x) + r(x)$ and $\deg r(x) < p$. By Fermat's theorem, $a^p - a \equiv 0 \pmod{p}$, and hence $f(a) \equiv r(a) \pmod{p}$ for all integers a. This proves the following result.

Theorem 9.3 *If p is a prime, then every polynomial congruence $f(x) \equiv 0 \pmod{p}$ is equivalent to a polynomial congruence $r(x) \equiv 0 \pmod{p}$, where $r(x)$ is a polynomial with degree less than p.*

Another way to obtain the polynomial $r(x)$ in Theorem 9.3 is to use the following lemma.

Lemma 9.4 *Assume $n \geq p$ and $n \equiv r \pmod{(p-1)}$, where $1 \leq r \leq p-1$. Then $x^n \equiv x^r \pmod{p}$ for all x.*

Proof. Write $n = q(p-1) + r$. By Fermat's theorem, $x^{p-1} \equiv 1 \pmod{p}$ if $x \not\equiv 0 \pmod{p}$, and hence $x^n = (x^{p-1})^q \cdot x^r \equiv 1^q \cdot x^r = x^r \pmod{p}$ holds for all $x \not\equiv 0 \pmod{p}$, and for $x \equiv 0 \pmod{p}$ the congruence is trivially true. □

Using Lemma 9.4 we can replace all terms of degree $\geq p$ in an integral polynomial $f(x)$ by equivalent terms of degree less than p, and this will lead to an integral polynomial $r(x)$ of degree less than p having the same roots modulo p as $f(x)$.

EXAMPLE 2 Consider the congruence $x^{11} + 2x^8 + x^5 + 3x^4 + 4x^3 + 1 \equiv 0 \pmod{5}$. Division by $x^5 - x$ yields

$$x^{11} + 2x^8 + x^5 + 3x^4 + 4x^3 + 1 = (x^6 + 2x^3 + x^2 + 1)(x^5 - x) + 5x^4 + 5x^3 + x + 1.$$

Hence, the given congruence is equivalent to the congruence $5x^4 + 5x^3 + x + 1 \equiv 0 \pmod{5}$, that is to $x+1 \equiv 0 \pmod{5}$, which has the sole solution $x \equiv 4 \pmod{5}$.

Instead, we could have used Lemma 9.4. Since $11 \equiv 3$, $8 \equiv 4$, and $5 \equiv 1$ modulo 4, we replace the terms x^{11}, $2x^8$, and x^5 by x^3, $2x^4$, and x, respectively. This results in the polynomial $x^3 + 2x^4 + x + 3x^4 + 4x^3 + 1 = 5x^4 + 5x^3 + x + 1 \equiv x + 1 \pmod{5}$. □

Theorem 9.5 *Let p be a prime. The non-congruent numbers $a_1, a_2, \dots, a_k$ are roots of the polynomial congruence $f(x) \equiv 0 \pmod{p}$ if and only if there exist two integral polynomials $q(x)$ and $r(x)$ such that*

$$f(x) = (x-a_1)(x-a_2)\cdots(x-a_k)q(x) + pr(x)$$

and $\deg r(x) < k$.

Proof. If such polynomials exist, then $f(a_j) = pr(a_j) \equiv 0 \pmod{p}$. The converse is proved by induction on the number k of roots. For $k = 1$, the existence of $q(x)$ and $r(x)$ was proved in Theorem 9.2. Assume the theorem is true for $k-1$ roots. Then there are two polynomials $q_1(x)$ and $r_1(x)$, with $\deg r_1(x) < k - 1$, such that

$$(1) \qquad f(x) = (x - a_1)(x - a_2) \cdots (x - a_{k-1})q_1(x) + pr_1(x).$$

From this, since $f(a_k) \equiv 0 \pmod{p}$, we obtain

$$(a_k - a_1)(a_k - a_2) \cdots (a_k - a_{k-1})q_1(a_k) \equiv 0 \pmod{p}.$$

Since $(a_k - a_j, p) = 1$ for $j = 1, 2, \dots, k-1$, we can cancel the factors $(a_k - a_j)$ in the above congruence to obtain $q_1(a_k) \equiv 0 \pmod{p}$. Hence by Theorem 9.2, there is a polynomial $q(x)$ and an integer b such that $q_1(x) = (x - a_k)q(x) + pb$, and by substituting this into (1) we find that the polynomials $q(x)$ and $r(x) = b(x - a_1)(x - a_2) \cdots (x - a_{k-1}) + r_1(x)$ satisfy all the requirements. □

As a corollary of the above theorem we get the following result.

Theorem 9.6 (Wilson's theorem) *If p is a prime, then $(p - 1)! \equiv -1 \pmod{p}$.*

Proof. By Fermat's theorem, the polynomial $x^{p-1} - 1$ has the roots 1,2, ..., $p - 1$ modulo p. Hence,

$$x^{p-1} - 1 = (x - 1)(x - 2) \cdots (x - (p - 1))q(x) + pr(x)$$

for suitable integral polynomials $q(x)$ and $r(x)$ with $\deg r(x) < p - 1$. By comparing degrees and leading coefficients, we see that $q(x) = 1$. Now inserting $x = 0$ we obtain $-1 \equiv (-1)^{p-1}(p - 1)! \pmod{p}$. If p is an odd prime, we conclude that $(p-1)! \equiv -1 \pmod{p}$, and for $p = 2$ we get the same result since $1 \equiv -1 \pmod{2}$. □

A polynomial congruence with a general modulus may have more roots than the degree of the polynomial. For example, the congruence $x^2 - 1 \equiv 0 \pmod{8}$ has the four roots 1, 3, 5, and 7. If however the modulus is prime, then the number of roots can not exceed the degree unless all coefficients of the polynomial are divisible by p. This follows as a corollary of Theorem 9.5.

Theorem 9.7 *Let p be a prime and let $f(x)$ be an integral polynomial of degree n not all of whose coefficients are divisible by p. Then the congruence $f(x) \equiv 0 \pmod{p}$ has at most n roots.*

Proof. Assume the congruence has k roots $a_1, a_2, \dots, a_k$, and use Theorem 9.5 to write $f(x) = (x - a_1)(x - a_2) \cdots (x - a_k)q_1(x) + pr_1(x)$. Here, the quotient $q_1(x)$ must be nonzero, because we have assumed that not all coefficients of $f(x)$ are divisible by p. Consequently, $n = \deg f(x) = k + \deg q_1(x) \geq k$. □

On the other hand, a polynomial congruence may well be without roots. The congruence $x^2 - 2 \equiv 0 \pmod{3}$ has no roots, and the congruence $x^p - x + 1 \equiv 0 \pmod{p}$ has no roots if p is prime, because of Fermat's theorem.

Here follows a criterion which guarantees that the number of roots equals the degree.

Theorem 9.8 *Let p be a prime, and suppose that the polynomial $f(x)$ has degree $n \le p$ and leading coefficient 1. Use the division algorithm to write $x^p - x = q(x)f(x) + r(x)$, where $\deg r(x) < \deg f(x)$. Then $f(x) \equiv 0 \pmod p$ has exactly n roots if and only if every coefficient of $r(x)$ is divisible by p.*

Remark. The assumption that the leading coefficient of $f(x)$ be 1 is really no restriction. If the leading coefficient is a, we may assume that $(a,p) = 1$. By choosing a' such that $a'a \equiv 1 \pmod p$ and replacing $f(x)$ by the polynomial $a'f(x) - (a'a-1)x^n$ we obtain a new polynomial with leading coefficient 1 and with the same roots modulo p as the original one.

Proof. Let m denote the degree of $q(x)$; then obviously $m + n = p$, and the leading coefficient of $q(x)$ is 1, too. If every coefficient of $r(x)$ is divisible by p, then by Fermats's theorem $q(a)f(a) \equiv a^p - a \equiv 0 \pmod p$ for each integer a. Since p is a prime, it follows that $q(a) \equiv 0 \pmod p$ or $f(a) \equiv 0 \pmod p$, i.e. every integer is a root of either $q(x) \equiv 0 \pmod p$ or $f(x) \equiv 0 \pmod p$. Now, by Theorem 9.7, the first congruence has at most m roots and the second has at most n roots, so together there are at most $m + n = p$ roots. Since there are p roots, we conclude that the congruence $f(x) \equiv 0 \pmod p$ must have precisely n roots.

Conversely, since $r(x) = x^p - x - q(x)f(x)$ it follows from Fermat's theorem that every root of $f(x)$ modulo p is a root of $r(x)$ modulo p. Hence, if $f(x)$ has n roots, then $r(x)$ has at least n roots. Since the degree of $r(x)$ is less than n, this is, however, impossible unless every coefficient of $r(x)$ is divisible by p. □

Corollary 9.9 *Assume p is a prime and that $d \mid (p-1)$. Then the congruence $x^d - 1 \equiv 0 \pmod p$ has exactly d roots.*

Proof. Write $p-1 = nd$. Use the identity $y^n - 1 = (y-1)(y^{n-1} + y^{n-2} + \cdots + y + 1)$ and replace y by x^d. We obtain $x^p - x = (x^{p-1} - 1)x = (x^d - 1)q(x)$, where $q(x) = x\sum_{j=0}^{n-1} x^{jd}$. Theorem 9.8 now applies. □

10 Polynomial Congruences with Prime Power Moduli

The general procedure for solving the polynomial congruence $f(x) \equiv 0 \pmod m$ when m is a prime power p^k, is to start with a root for the modulus p and use it to generate a root (or in some cases several roots) modulo p^2. Using the same technique, we produce roots modulo p^3, p^4, and so on, until we finally obtain roots for the original modulus p^k. The details will be given below.

Let us start by noting that if $f(x)$ is an integral polynomial and a is an integer, then there is an integral polynomial $g(t)$ such that

$$f(a+t) = f(a) + f'(a)t + t^2 g(t). \tag{1}$$

This is a special case of Taylor's formula. To prove it, we note that $f(a+t)$ is obviously a polynomial in t with integral coefficients, and hence $f(a+t) = A + Bt + t^2 g(t)$, where $g(t)$ is an integral polynomial. The coefficient A is obtained by putting $t = 0$, and to determine B we first differentiate and then take $t = 0$.

Let us now consider the congruence

$$f(x) \equiv 0 \pmod{p^2}, \tag{2}$$

where p is prime. Any solution a of this congruence must also be a solution of the congruence

$$f(x) \equiv 0 \pmod{p}. \tag{3}$$

Conversely, assume a is a solution of (3), and let us look for solutions b of (2) such that $b \equiv a \pmod{p}$, that is such that $b = a + pt$ for some integer t. By (1),

$$f(a+pt) = f(a) + f'(a)pt + p^2t^2g(pt) \equiv f(a) + pf'(a)t \pmod{p^2},$$

and hence $a + pt$ solves the congruence (2) if and only if $f(a) + pf'(a)t \equiv 0 \pmod{p^2}$, that is if and only if

$$f'(a)t \equiv -\frac{f(a)}{p} \pmod{p}. \tag{4}$$

If $(f'(a), p) = 1$, then (4) has a unique solution $t \equiv t_0 \pmod{p}$, and it follows that $x \equiv a + pt_0 \pmod{p^2}$ is a solution of the congruence (2) and that it is the only solution satisfying $x \equiv a \pmod{p}$.

If $p \mid f'(a)$, then (4) is solvable if and only if $p^2 \mid f(a)$, and in this case any number t solves (4). Hence $x \equiv a + pj \pmod{p^2}$ solves (2) for $j = 0, 1, \dots, p-1$. In this case, the congruence (2) has p roots that are congruent to a modulo p.

If $p \mid f'(a)$ and $p^2 \nmid f(a)$, then (2) has no solution that is congruent to a.

The step leading from p^k to p^{k+1} is analogous. Thus we have the following theorem.

Theorem 10.1 *Let p be a prime and let k be an arbitrary positive integer, and suppose that a is a solution of $f(x) \equiv 0 \pmod{p^k}$.*

(i) If $p \nmid f'(a)$, then there is precisely one solution b of $f(x) \equiv 0 \pmod{p^{k+1}}$ such that $b \equiv a \pmod{p^k}$. The solution is given by $b = a + p^k t$, where t is the unique solution of $f'(a)t \equiv -f(a)/p^k \pmod{p}$.

(ii) If $p \mid f'(a)$ and $p^{k+1} \mid f(a)$, then there are p solutions of the congruence $f(x) \equiv 0 \pmod{p^{k+1}}$ that are congruent to a modulo p^k; these solutions are $a + p^k j$ for $j = 0, 1, \dots, p-1$.

(iii) If $p \mid f'(a)$ and $p^{k+1} \nmid f(a)$, then there are no solutions of the congruence $f(x) \equiv 0 \pmod{p^{k+1}}$ that are congruent to a modulo p^k.

Proof. Let b be a solution of $f(x) \equiv 0 \pmod{p^{k+1}}$ that is congruent to a modulo p^k; then $b = a + p^k t$ for some integer t. By (1) we have

$$0 \equiv f(b) = f(a) + f'(a)p^kt + p^{2k}t^2g(p^kt) \equiv f(a) + f'(a)p^kt \pmod{p^{k+1}},$$

because $2k \ge k+1$. Since $f(a) \equiv 0 \pmod{p^k}$, it follows that $f(a)/p^k$ is integer, and we can divide the congruence above by p^k to obtain

$$f'(a)t \equiv -f(a)/p^k \pmod{p}.$$

The latter congruence has a unique solution if $(f'(a), p) = 1$, i.e. if $p \nmid f'(a)$. If $p \mid f'(a)$, then we must have $f(a)/p^k \equiv 0 \pmod{p}$, that is $p^{k+1} \mid f(a)$, in which case any value of t in a complete residue system will be a solution. Finally, if $p \mid f'(a)$ but $p^{k+1} \nmid f(a)$, there will be no t that solves the congruence. □

Corollary 10.2 *Let p be a prime and k an arbitrary positive integer. If a is a solution of $f(x) \equiv 0 \pmod{p}$ and $p \nmid f'(a)$, then there exists precisely one solution b of $f(x) \equiv 0 \pmod{p^k}$ such that $b \equiv a \pmod{p}$.*

Proof. By Theorem 10.1 (i) there exists a unique solution b_2 of $f(x) \equiv 0 \pmod{p^2}$ such that $b_2 \equiv a \pmod{p}$. It follows that $f'(b_2) \equiv f'(a) \pmod{p}$, and hence $p \nmid f'(b_2)$. Therefore, by the same theorem there exists a unique solution b_3 of $f(x) \equiv 0 \pmod{p^3}$ such that $b_3 \equiv b_2 \equiv a \pmod{p}$. Proceeding like this we will finally obtain the unique solution $b = b_k$ that is congruent to a modulo p of the congruence $f(x) \equiv 0 \pmod{p^k}$. □

Summary. The general procedure for finding all roots of $f(x) \equiv 0 \pmod{p^k}$ can be summarized as follows.

1. First find all solutions of the congruence $f(x) \equiv 0 \pmod{p}$.
2. Select one, say a_1; then there are either 0, 1 or p solutions of $f(x) \equiv 0 \pmod{p^2}$ congruent to a_1 modulo p; if solutions exist, they are found by solving the linear congruence $f'(a_1)t \equiv -f(a_1)/p \pmod{p}$. If there are no solutions, start again with a different a_1.
3. If there are solutions of $f(x) \equiv 0 \pmod{p^2}$, select one, say a_2, and find the corresponding roots of $f(x) \equiv 0 \pmod{p^3}$ by solving the congruence $f'(a_2)t \equiv -f(a_2)/p^2 \pmod{p}$. Do this for each root of $f(x) \equiv 0 \pmod{p^2}$. Note that since $a_2 \equiv a_1 \pmod{p}$, $f'(a_2) \equiv f'(a_1) \pmod{p}$, so we do not need to calculate $f'(a_2)$.
4. Proceeding in this fashion, we will eventually determine all solutions of $f(x) \equiv 0 \pmod{p^k}$.

It is worth emphasizing that if at any step in this procedure we obtain multiple solutions, then we must apply the above process to each solution.

Unfortunately, there is no general procedure for starting the above algorithm, that is for finding all solutions of $f(x) \equiv 0 \pmod{p}$. In the next section, we will discuss what can be said about the number of solutions, and in later sections we will treat some special cases.

Example 1 Solve the congruence $7x^6 + 4x + 12 \equiv 0 \pmod{135}$.

Solution: Since $135 = 3^3 \cdot 5$, the congruence is equivalent to the system

$$\begin{cases} 7x^6 + 4x + 12 \equiv 0 \pmod{5} \\ 7x^6 + 4x + 12 \equiv 0 \pmod{3^3}. \end{cases} \tag{5}$$

Write $f(x) = 7x^6 + 4x + 12$. Using Fermat's identity, we can replace the first congruence with $2x^2 + 4x + 2 \equiv 0 \pmod{5}$, which is equivalent to $(x+1)^2 \equiv 0 \pmod{5}$ and has the the sole root -1.

We now turn to the second congruence and start by solving the congruence $f(x) \equiv 0 \pmod{3}$, which is equivalent to $x^2 + x \equiv 0 \pmod{3}$, since $x^6 \equiv x^2 \pmod{3}$. Its solutions are $x \equiv 0 \pmod{3}$ and $x \equiv -1 \pmod{3}$. We have $f'(x) = 42x^5 + 4$. Since $f'(0) = 4 \equiv 1 \pmod{3}$ and $f'(-1) = -38 \equiv 1 \pmod{3}$, it follows from Corollary 10.2 that the congruence $f(x) \equiv 0 \pmod{3^3}$ has exactly two solutions.

To find these, we start from $x_1 = 0$ and solve the linear congruence $f'(0)t \equiv -f(0)/3 \pmod{3}$, that is $t \equiv -4 \equiv 2 \pmod{3}$. We conclude that $x_2 = 0 + 2 \cdot 3 = 6$ solves $f(x) \equiv 0 \pmod{3^2}$. Next, solve the congruence $f'(0)t \equiv f'(6)t \equiv$

$-f(6)/9 \pmod 3$, which yields $t \equiv 2 \pmod 3$. It follows that $x_3 = 6+2\cdot 9 = 24$ solves $f(x) \equiv 0 \pmod{3^3}$.

Starting instead from $y_1 = -1$ we first solve $f'(-1)t \equiv -f(-1)/3 \pmod 3$, which yields $t \equiv -5 \equiv 1 \pmod 3$. We conclude that $y_2 = -1 + 1 \cdot 3 = 2$ solves $f(x) \equiv 0 \pmod{3^2}$. Next, solve the congruence $f'(-1)t \equiv f'(2)t \equiv -f(2)/9 \pmod 3$, which yields $t \equiv 2 \pmod 3$. It follows that $y_3 = 2 + 2 \cdot 9 = 20$ solves $f(x) \equiv 0 \pmod{3^3}$.

To find the two solutions of our original congruence, we now use the Chinese Remainder Theorem to solve the two systems

$$\begin{cases} x \equiv -1 & \pmod 5 \\ x \equiv 24 & \pmod{3^3} \end{cases} \quad \text{and} \quad \begin{cases} x \equiv -1 & \pmod 5 \\ x \equiv 20 & \pmod{3^3}. \end{cases}$$

The solutions are $x \equiv 24 \pmod{135}$ and $x \equiv 74 \pmod{135}$. □

Example 2 Let us determine all solutions of $x^{10} \equiv 24 \pmod{125}$.

Solution: Since $125 = 5^3$ we start by solving the congruence $f(x) \equiv 0 \pmod 5$, where $f(x) = x^{10} - 24$. By Fermat's identity, $x^5 \equiv x \pmod 5$, and hence $f(x) \equiv x^2 - 24 \equiv x^2 - 4 \equiv (x-2)(x+2) \pmod 5$. We conclude that the congruence $f(x) \equiv 0 \pmod 5$ has two roots, $x \equiv \pm 2 \pmod 5$.

We next note that $f'(x) = 10x^9$ is divisible by 5 for any x and in particular for $x = \pm 2$. Hence, each of the two solutions ± 2 will give rise to 5 or 0 solutions modulo 25 of the congruence $f(x) \equiv 0 \pmod{25}$ depending on whether $25 \mid f(\pm 2)$ or not. Now $f(\pm 2) = 1000$ is divisible 25, and hence we get the ten incongruent solutions $\pm 2 + 5j$ modulo 25, $j = 0, 1, 2, 3, 4$. Of course, these can also be represented as ± 2, ± 7, ± 12, ± 17, and ± 22.

Let a_2 be one of these solutions; since $f'(a_2) \equiv f'(\pm 2) \equiv 0 \pmod 5$, a_2 will induce 5 or 0 solutions of the congruence $f(x) \equiv 0 \pmod{125}$ according as $125 \mid f(a_2)$ or not. Let us therefore compute $f(x)$ modulo 125 for each of the above solutions of the congruence $f(x) \equiv 0 \pmod{25}$. After some computations we obtain $f(\pm 2) \equiv 0$, $f(\pm 7) \equiv 100$, $f(\pm 12) \equiv 75$, $f(\pm 17) \equiv 50$, and $f(\pm 22) \equiv 25$. Consequently, we get 5 solutions from each of ± 2 and no solutions from the other roots of the congruence $f(x) \equiv 0 \pmod{25}$. The solutions are $\pm 2 + 25j$ modulo 125, $j = 0, 1, 2, 3, 4$, that is 2, 23, 27, 48, 52, 73, 77, 98, 102, and 123. □

11 The Congruence $\mathbf{x^2 \equiv a \pmod m}$

In this section we will study the congruence

$$(1) \qquad x^2 \equiv a \pmod m.$$

There are three main problems to consider. Firstly, when do solutions exist, secondly, how many solutions are there, and thirdly, how to find them.

We will first show that we can always reduce a congruence of the form (1) to a congruence of the same form with $(a, m) = 1$.

Assume therefore that $(a, m) > 1$, and let p be a prime dividing (a, m), that is $p \mid a$ and $p \mid m$. Suppose x is a solution of (1). Then $p \mid x^2$ and hence $p \mid x$. Write $x = py$; then (1) is equivalent to $p^2y^2 \equiv a \pmod m$. Divide by p to obtain

$$py^2 \equiv a/p \pmod{m/p}. \tag{2}$$

There are three separate cases to consider:

(i) If $p^2 \mid m$ and $p^2 \mid a$, then (2) is equivalent to the congruence $y^2 \equiv a/p^2 \pmod{m/p^2}$, and for each solution y_0 of this congruence (if there are any), there are p incongruent solutions modulo m of the original congruence (1). These are $x \equiv py_0 \pmod{m/p}$. If $(a/p^2, m/p^2) > 1$, we repeat the whole procedure.

(ii) If $p^2 \mid m$ but $p^2 \nmid a$, then (2) is a contradiction. Hence, (1) has no solutions in this case.

(iii) If $p^2 \nmid m$, then $(p, m/p) = 1$, and hence there is a number c such that $cp \equiv 1 \pmod{m/p}$. It follows that (2) is equivalent to the congruence $y^2 \equiv ca/p \pmod{m/p}$. Any solution y_0 of this congruence yields a unique solution $x \equiv py_0 \pmod{m}$ of (1). If $(ca/p, m/p) > 1$ we can repeat the whole procedure.

Note that if $p^2 \mid a$, then $ca/p = cp \cdot a/p^2 \equiv 1 \cdot a/p^2 \equiv a/p^2 \pmod{m/p}$, i.e. (2) is equivalent to the congruence $y^2 \equiv a/p^2 \pmod{m/p}$ in that case.

Example 1 Solve the four congruences:

$$\text{(i) } x^2 \equiv 36 \pmod{45}, \quad \text{(ii) } x^2 \equiv 15 \pmod{45},$$
$$\text{(iii) } x^2 \equiv 18 \pmod{21}, \quad \text{(iv) } x^2 \equiv 15 \pmod{21}.$$

Solution: (i) Here $(36, 45) = 9$ and writing $x = 3y$ we obtain the equivalent congruence $y^2 \equiv 4 \pmod 5$ with the solutions $y \equiv \pm 2 \pmod 5$. Hence $x \equiv \pm 6 \pmod{15}$, i.e. 6, 9, 21, 24, 36, and 39 are the solutions of (i).

(ii) Since $9 \mid 45$ but $9 \nmid 15$ there are no solutions of (ii).

(iii) Since $(18, 21) = 3$, we write $x = 3y$ and obtain the following sequence of equivalent congruences: $9y^2 \equiv 18 \pmod{21}$, $3y^2 \equiv 6 \pmod 7$, $y^2 \equiv 2 \pmod 7$ with the solutions $y \equiv \pm 3 \pmod 7$. Hence (iii) has the solutions $x \equiv \pm 9 \pmod{21}$.

(iv) Since $(15, 21) = 3$, we put $x = 3y$ and obtain $9y^2 \equiv 15 \pmod{21}$, that is $3y^2 \equiv 5 \pmod 7$. Since $5 \cdot 3 \equiv 1 \pmod 7$, we multiply the last congruence by 5, which yields $y^2 \equiv 4 \pmod 7$ with the solutions $y \equiv \pm 2 \pmod 7$. Consequently, $x \equiv \pm 6 \pmod{21}$ are the solutions of (iv). □

For the rest of this section, we will assume that $(a, m) = 1$.

Definition 11.1 Suppose that $(a, m) = 1$. Then a is called a *quadratic residue of* m if the congruence $x^2 \equiv a \pmod m$ has a solution. If there is no solution, then a is called a *quadratic nonresidue of* m.

By decomposing the modulus m into a product of primes and using Theorem 6.5, we reduce the study of the congruence (1) to a study of congruences of the form

$$x^2 \equiv a \pmod{p^k}$$

where the modulus is a prime power. Now, the techniques in section 10 apply. However, since the derivative of x^2 is $2x$, and $2x \equiv 0 \pmod 2$ we have to distinguish between the cases $p = 2$ and p odd prime.

Lemma 11.2 *If p is an odd prime, $(a,p)=1$ and a is a quadratic residue of p, then the congruence $x^2 \equiv a \pmod p$ has exactly two roots.*

Proof. By assumption, there is at least one root b. Obviously, $-b$ is a root, too, and $-b \not\equiv b \pmod p$, since $b \not\equiv 0$. By Theorem 9.7, there can not be more than two roots. □

Theorem 11.3 *If p is an odd prime and $(a,p)=1$, then $x^2 \equiv a \pmod{p^k}$ has exactly two solutions if a is a quadratic residue of p, and no solutions if a is a quadratic nonresidue of p.*

Proof. Let $f(x) = x^2 - a$; then $f'(x) = 2x$ is not divisible by p for any $x \not\equiv 0 \pmod p$. Hence, it follows from Corollary 10.2 and Lemma 11.2 that the equation $x^2 \equiv a \pmod{p^k}$ has exactly two roots for each k if a is a quadratic residue. Since every solution of the latter congruence also solves the congruence $x^2 \equiv a \pmod p$, there can be no solution if a is a quadratic nonresidue of p. □

The case $p = 2$ is different, and the complete story is given by the following theorem.

Theorem 11.4 *Suppose a is odd. Then*

(i) The congruence $x^2 \equiv a \pmod 2$ is always solvable and has exactly one solution;

(ii) The congruence $x^2 \equiv a \pmod 4$ is solvable if and only if $a \equiv 1 \pmod 4$, in which case there are precisely two solutions;

(iii) The congruence $x^2 \equiv a \pmod{2^k}$, with $k \geq 3$, is solvable if and only if $a \equiv 1 \pmod 8$, in which case there are exactly four solutions. If x_0 is a solution, then all solutions are given by $\pm x_0$ and $\pm x_0 + 2^{k-1}$.

Proof. (i) and (ii) are obvious.

(iii) Suppose $x^2 \equiv a \pmod{2^k}$ has a solution x_0. Then obviously $x_0^2 \equiv a \pmod 8$, and x_0 is odd since a is odd. But the square of an odd number is congruent to 1 modulo 8, and hence $a \equiv 1 \pmod 8$. This proves the necessity of the condition $a \equiv 1 \pmod 8$ for the existence of a solution. Moreover, $(-x_0)^2 = x_0^2 \equiv a \pmod{2^k}$ and $(\pm x_0 + 2^{k-1})^2 = x_0^2 \pm 2^k x_0 + 2^{2k-2} \equiv x_0^2 \equiv a \pmod{2^k}$, since $2k-2 \geq k$. It is easily verified that the four numbers $\pm x_0$ and $\pm x_0 + 2^{k-1}$ are incongruent modulo 2^k. Hence, the congruence has at least four solutions if there is any.

It remains to verify that the condition on a is sufficient and that there are at most four solutions. We show sufficiency by induction on k. The case $k = 3$ is clear, since $x^2 \equiv 1 \pmod 8$ has the solution $x \equiv 1$. Now assume that $x^2 \equiv a \pmod{2^k}$ is solvable with a solution x_0. Then we know that $\pm x_0$ and $\pm x_0 + 2^{k-1}$ also solve the congruence, and we will prove that one of them also solves the congruence

$$x^2 \equiv a \pmod{2^{k+1}}. \tag{3}$$

We know that $x_0^2 = a + 2^k n$ for some integer n. If n is even, then x_0 is obviously a solution of (3). If n is odd, then

$$(x_0 + 2^{k-1})^2 = x_0^2 + 2^k x_0 + 2^{2k-2} = a + 2^k(n + x_0) + 2^{2k-2} \equiv a \pmod{2^{k+1}},$$

because $(n+x_0)$ is even (since n and x_0 are both odd) and $2k-2 \geq k+1$ (since $k \geq 3$). This completes the induction step

Finally, in the interval $[1, 2^k]$ there are 2^{k-3} integers a that are congruent to 1 modulo 8. For each such number a we have already found 4 different solutions of the congruence $x^2 \equiv a \pmod{2^k}$ in the same interval, all of them odd. Taking all these solutions together we get $4 \cdot 2^{k-3} = 2^{k-1}$ solutions. But there are exactly 2^{k-1} odd numbers in the interval, so there is no room for any more solutions. Hence, each equation has exactly four solutions. □

If we combine the two theorems above with Theorem 6.5, we get the following complete answer to the question about the number of solutions of a congruence $x^2 \equiv a \pmod{m}$.

Theorem 11.5 *Let $m = 2^k p_1^{k_1} \cdots p_r^{k_r}$, where the p_i are distinct odd primes, and let a be a number which is relatively prime to m. Then the congruence $x^2 \equiv a \pmod{m}$ is solvable if and only if a is a quadratic residue of p_i for each i, and $a \equiv 1 \pmod 4$ in the case $k = 2$, and $a \equiv 1 \pmod 8$ in the case $k \geq 3$. If the congruence is solvable, then there are 2^r solutions if $k = 0$ or $k = 1$, 2^{r+1} solutions if $k = 2$, and 2^{r+2} solutions if $k \geq 3$.*

In order to apply Theorem 11.5 we need some criterion telling when a number is a quadratic residue of given prime p. First, note that there are as many quadratic residues as nonresidues of an odd prime.

Theorem 11.6 *Let p be an odd prime. Then there are exactly $(p-1)/2$ incongruent quadratic residues of p and exactly $(p-1)/2$ quadratic nonresidues of p.*

Proof. All quadratic residues can be found by squaring the elements of a reduced residue system. Since each solvable congruence $x^2 \equiv a \pmod{p}$ has exactly two solutions if $(a,p) = 1$, it follows that the number of quadratic residues equals half the number of elements in the reduced residue system, that is $(p-1)/2$. To get all quadratic residues one can for example take 1^2, 2^2, ..., $[(p-1)/2]^2$. □

Lemma 11.7 *Let p be an odd prime and suppose $a \not\equiv 0 \pmod{p}$. Then modulo p*

$$(p-1)! \equiv \begin{cases} a^{(p-1)/2} & \text{if } a \text{ is a quadratic nonresidue of } p \\ -a^{(p-1)/2} & \text{if } a \text{ is a quadratic residue of } p. \end{cases}$$

Proof. The congruence $mx \equiv a \pmod{p}$ is solvable for each integer m in the interval $1 \leq m \leq p-1$, i.e. for each m there is an integer n, $1 \leq n \leq p-1$ such that $mn \equiv a \pmod{p}$. If the congruence $x^2 \equiv a \pmod{p}$ has no solution, then $n \neq m$. If it has a solution, then it has exactly two solutions of the form $x \equiv m_0 \pmod{p}$ and $x \equiv p - m_0 \pmod{p}$, and it follows that $n \neq m$ for all but two values of m.

Now consider the product $(p-1)! = 1 \cdot 2 \cdot 3 \cdots (p-1)$. If the congruence $x^2 \equiv a \pmod{p}$ has no solution, then we can pair off the $p-1$ numbers into $(p-1)/2$ pairs such that the product of the two numbers in each pair is congruent to $a \pmod{p}$, and this means that $(p-1)!$ is congruent to $a^{(p-1)/2}$.

On the other hand, if the congruence has two solutions, m_0 and $p - m_0$, then we take away these two numbers and pair off the remaining $p-3$ numbers into $(p-3)/2$ pairs such that the product of the two numbers in each pair is congruent to $a \pmod p$. Since $m_0(p-m_0) \equiv -m_0^2 \equiv -a \pmod p$, it follows that $(p-1)! \equiv -a \cdot a^{(p-3)/2} \equiv -a^{(p-1)/2} \pmod p$. □

Now, let us recall Wilson's theorem, which we proved in the previous section as Theorem 9.6:

Wilson's theorem *If p is a prime then $(p-1)! \equiv -1 \pmod p$.*

First let us note that Wilson's theorem for $p > 2$ is a obtained as a special case of Lemma 11.7 by taking $a = 1$, which is obviously a quadratic residue of any prime p. Secondly, and more important, by combining Wilson's theorem with Lemma 11.7 we get the following solvability criterion due to Euler:

Theorem 11.8 (Euler's Criterion) *Let p be an odd prime and suppose $(a,p) = 1$. Then a is a quadratic residue or nonresidue of p according as $a^{(p-1)/2} \equiv 1 \pmod p$ or $a^{(p-1)/2} \equiv -1 \pmod p$.*

The following important result follows immediately from Euler's criterion.

Theorem 11.9 *Let p be a prime.Then -1 is a quadratic residue of p if and only if $p = 2$ or $p \equiv 1 \pmod 4$.*

Proof. -1 is a quadratic residue of 2 since $1^2 = 1 \equiv -1 \pmod 2$. For odd primes, we apply Euler's Criterion noting that $(-1)^{(p-1)/2} = 1$ if and only if $(p-1)/2$ is even, that is if and only if p is a prime of the form $4k+1$. □

Let us also note that Fermat's theorem is an easy consequence of Euler's criterion; by squaring we obtain

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

Let us finally address the question of finding a solution to the congruence $x^2 \equiv a \pmod p$ assuming that a is a quadratic residue of p. In the case $p \equiv 3 \pmod 4$ we have the following answer.

Theorem 11.10 *Let p be a prime and assume that $p \equiv 3 \pmod 4$. If a is a quadratic residue of p, then the congruence $x^2 \equiv a \pmod p$ has the two solutions $\pm a^{(p+1)/4}$.*

Proof. Since a is a quadratic residue, $a^{(p-1)/2} \equiv 1 \pmod p$. It follows that

$$\left(\pm a^{(p+1)/4}\right)^2 = a^{(p+1)/2} = a \cdot a^{(p-1)/2} \equiv a \pmod p. \qquad \square$$

Note that it is not necessary to verify in advance that $a^{(p-1)/2} \equiv 1 \pmod p$. It is enough to compute $x \equiv a^{(p+1)/4} \pmod p$. If $x^2 \equiv a \pmod p$, then $\pm x$ are the two solutions, otherwise $x^2 \equiv -a \pmod p$, and we can conclude that there are no solutions.

12 General Quadratic Congruences

A general quadratic congruence

$$(1) \qquad ax^2 + bx + c \equiv 0 \pmod{m},$$

can be reduced to a system consisting of a congruence of the form $y^2 \equiv d \pmod{m'}$ and a linear congruence by completing the square.

The simplest case occurs when $(4a, m) = 1$, because we may then multiply the congruence (1) by $4a$ without having to change the modulus m in order to get the following equivalent congruence

$$4a^2x^2 + 4abx + 4ac \equiv 0 \pmod{m},$$

that is,

$$(2ax + b)^2 \equiv b^2 - 4ac \pmod{m}.$$

Writing $y = 2ax + b$, we obtain the following result.

Theorem 12.1 *Assume that $(4a, m) = 1$. Then all solutions of the congruence*

$$ax^2 + bx + c \equiv 0 \pmod{m}$$

can be found by solving the following chain of congruences

$$y^2 \equiv b^2 - 4ac \pmod{m}, \quad 2ax \equiv y - b \pmod{m}.$$

Since $(2a, m) = 1$, the linear congruence has a unique solution modulo m for each root y.

Example 1 Let us solve the congruence $8x^2 + 5x + 1 \equiv 0 \pmod{23}$. Complete the square by multiplying by 32 to get $(16x+5)^2 \equiv 5^2 - 32 = -7 \equiv 16 \pmod{23}$. Thus $16x + 5 \equiv \pm 4$. Solving $16x \equiv -1 \pmod{23}$ gives $x \equiv 10$, and $16x \equiv -9 \pmod{23}$ yields $x \equiv 21$. Hence, 10 and 21 are the only solutions of the original congruence. □

When $(4a, m) \neq 1$, we start by factoring $4a = a_1a_2$ in such a way that $(a_2, m) = 1$. We may now multiply the congruence (1) by the number a_2 without having to change the modulus, but when we then multiply by a_1 we must change the modulus to a_1m in order to get the equivalent congruence $4a^2x^2 + 4abx + 4ac \equiv 0 \pmod{a_1m}$, which, of course, in turn is equivalent to the congruence $(2ax + b)^2 \equiv b^2 - 4ac \pmod{a_1m}$. This proves the following generalization of theorem 12.1.

Theorem 12.2 *Write $4a = a_1a_2$ with a_2 relatively prime to m. Then all solutions of the congruence*

$$ax^2 + bx + c \equiv 0 \pmod{m}$$

can be found by solving the following chain of congruences

$$y^2 \equiv b^2 - 4ac \pmod{a_1m}, \quad 2ax \equiv y - b \pmod{a_1m}.$$

Example 2 Let us solve the congruence $3x^2 + 3x + 2 \equiv 0 \pmod{10}$ using Theorem 12.2.. Since $(4 \cdot 3, 10) = 2 \neq 1$ but $(3, 10) = 1$, multiplication by $4 \cdot 3$ transforms the given congruence into the equivalent congruence

$$(6x + 3)^2 \equiv 3^2 - 4 \cdot 3 \cdot 2 = -15 \equiv 25 \pmod{40}.$$

The congruence $y^2 \equiv 25 \pmod{40}$ has four roots modulo 40, namely 5, 15, 25, and 35. For each root y we then solve the linear congruence $6x \equiv y - 3 \pmod{40}$. The solutions are in turn $x \equiv 7, 2, 17, 12 \pmod{20}$, which means that the solutions of our original congruence are $x \equiv 2$ and $x \equiv 7 \pmod{10}$. □

13 The Legendre Symbol and Gauss' Lemma

Definition 13.1 Let p be an odd prime.The *Legendre symbol* $\left(\dfrac{a}{p}\right)$ is defined as follows.

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{if } a \text{ is a quadratic residue of } p \\ -1, & \text{if } a \text{ is a quadratic nonresidue of } p \\ 0, & \text{if } p \mid a. \end{cases}$$

Theorem 13.2 *Let p be an odd prime. Then*

(i) $\left(\dfrac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}$,

(ii) $a \equiv b \pmod{p} \Rightarrow \left(\dfrac{a}{p}\right) = \left(\dfrac{b}{p}\right)$,

(iii) $\left(\dfrac{ab}{p}\right) = \left(\dfrac{a}{p}\right)\left(\dfrac{b}{p}\right)$,

(iv) *If* $(a, p) = 1$ *then* $\left(\dfrac{a^2}{p}\right) = 1$ *and* $\left(\dfrac{a^2 b}{p}\right) = \left(\dfrac{b}{p}\right)$,

(v) $\left(\dfrac{1}{p}\right) = 1$,

(vi) $\left(\dfrac{-1}{p}\right) = (-1)^{(p-1)/2} = \begin{cases} 1 & \text{if } p \equiv 1 \pmod{4}, \\ -1 & \text{if } p \equiv 3 \pmod{4}. \end{cases}$

Proof. If $p \mid a$ then (i) is obvious, and if $(p, a) = 1$ then (i) is just a reformulation of Euler's criterion (Theorem 11.8). The remaining parts are all simple consequences of (i). □

Because of Theorem 13.2 (iii) and (iv), in order to compute $\left(\dfrac{a}{p}\right)$ for an arbitrary integer a it is enough, given its prime factorization, to know $\left(\dfrac{-1}{p}\right)$, $\left(\dfrac{2}{p}\right)$ and $\left(\dfrac{q}{p}\right)$ for each odd prime q. We already know $\left(\dfrac{-1}{p}\right)$, and $\left(\dfrac{2}{p}\right)$ will be computed below. Finally, $\left(\dfrac{q}{p}\right)$ can be computed via the Reciprocity law, which will be discussed in the next section.

Theorem 13.3 (Gauss' Lemma) *Let p be an odd prime and suppose that the number a is relatively prime to p. Consider the least positive residues modulo p of the numbers a, $2a$, $3a$, $\dots$, $\frac{p-1}{2}a$. If N is the number of these residues that are greater than $p/2$, then* $\left(\dfrac{a}{p}\right) = (-1)^N$.

Proof. The numbers a, $2a$, $3a$, $\dots$, $\frac{p-1}{2}a$ are relatively prime to p and incongruent modulo p. Let $r_1, r_2, \dots, r_N$ represent the least positive residues that exceed $p/2$, and let $s_1, s_2, \dots, s_M$ denote the remaining residues, that is those that are less than $p/2$; then $N + M = (p-1)/2$.

The quotient q when ja is divided by p is $q = [ja/p]$. (Here $[x]$ denotes the greatest integer less than or equal to x.) It follows that

$$ja = [ja/p]\,p + \text{some } r_i \text{ or some } s_k. \tag{1}$$

The numbers $p - r_1$, $p - r_2$, $\dots$, $p - r_N$ are positive and less than $p/2$, relatively prime to p and incongruent in pairs modulo p. Also, no $p - r_i$ is an s_j. For suppose $p - r_i = s_j$, and let $r_i \equiv ma \pmod p$ and $s_j \equiv na \pmod p$, where m and n are distinct integers between 1 and $p/2$. Then $p = r_i + s_j \equiv (m+n)a \pmod p$, and since $(a,p) = 1$, we must have $p \mid (m+n)$, a contradiction since $0 < m + n < p$.

Thus, $p - r_1$, $p - r_2$, $\dots$, $p - r_N$, $s_1, s_2, \dots, s_M$ are all different integers in the intervall $[1, (p-1)/2]$, and since they are $M + N = (p-1)/2$ in number, they are equal in some order to the numbers 1, 2, $\dots$, $(p-1)/2$. Therefore,

$$(p - r_1)(p - r_2)\cdots(p - r_N)s_1s_2\cdots s_M = ((p-1)/2)!,$$

that is

$$(-1)^N r_1r_2\cdots r_N s_1s_2\cdots s_M \equiv ((p-1)/2)! \pmod p.$$

But the numbers $r_1, r_2, \dots, r_N$, $s_1, s_2, \dots, s_M$ are also congruent in some order to the numbers a, $2a$, $\dots$, $\frac{p-1}{2}a$, and hence

$$\left(\frac{p-1}{2}\right)! \equiv (-1)^N a \cdot 2a \cdot \dots \cdot \frac{p-1}{2}a = (-1)^N a^{(p-1)/2}\left(\frac{p-1}{2}\right)! \pmod p.$$

Since each factor in $((p-1)/2)!$ is relatively prime to p, we can divide each side of the last congruence by $((p-1)/2)!$ to obtain $a^{(p-1)/2} \equiv (-1)^N \pmod p$. The conclusion of the lemma now follows from part (i) of Theorem 13.2. □

As a simple application of Gauss' lemma, we now compute $\left(\dfrac{2}{p}\right)$.

Theorem 13.4 *Let p be an odd prime. Then 2 is a quadratic residue of p if $p \equiv \pm 1 \pmod 8$, and a quadratic nonresidue of p if $p \equiv \pm 3 \pmod 8$, that is*

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} 1, & \text{if } p \equiv \pm 1 \pmod 8, \\ -1, & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

Proof. Take $a = 2$ in Gauss' lemma; then N is the number of integers in the sequence 2, 4, $\dots$, $p - 1$ that are greater than $p/2$, that is N is the number of integers k such that $p/2 < 2k < p$, or equivalently $p/4 < k < p/2$. Consequently,

$N = [p/2] - [p/4]$. Taking $p = 4n + 1$ we get $N = 2n - n = n$, and $p = 4n - 1$ yields $N = (2n - 1) - (n - 1) = n$, too. Hence, N is even if n is even, i.e. if $p = 8m \pm 1$, and N is odd if n is so, i.e. if $p = 8m \pm 3$. □

EXAMPLE 1 The equation $x^2 \equiv 2 \pmod{17}$ is solvable since $17 \equiv 1 \pmod 8$. Indeed, $x \equiv \pm 6 \pmod{17}$ solves the congruence. □

Theorem 13.5 *If p is an odd prime and a is an odd number that is not divisible by p, then*

$$\left(\frac{a}{p}\right) = (-1)^n, \qquad \textit{where } n = \sum_{j=1}^{(p-1)/2} \left[\frac{ja}{p}\right].$$

Proof. We have to prove that n has the same parity as the number N in Gauss' lemma, i.e. that $n \equiv N \pmod 2$. We use the same notation as in the proof of the lemma. By summing over j in equation (1), we obtain

$$\sum_{j=1}^{(p-1)/2} ja = p \sum_{j=1}^{(p-1)/2} \left[\frac{ja}{p}\right] + \sum_{i=1}^{N} r_i + \sum_{k=1}^{M} s_k. \tag{2}$$

Since the numbers $(p-r_1)$, $(p-r_2)$, …, $(p-r_N)$, $s_1, s_2, \dots, s_M$ are the numbers 1, 2, …, $(p-1)/2$ in some order, we also have

$$\sum_{j=1}^{(p-1)/2} j = \sum_{i=1}^{N} (p - r_i) + \sum_{k=1}^{M} s_k.$$

Subtracting this from equation (2), we obtain

$$(a-1) \sum_{j=1}^{(p-1)/2} j = 2 \sum_{i=1}^{N} r_j + p\Big(\sum_{j=1}^{(p-1)/2} \left[\frac{ja}{p}\right] - N \Big) = 2 \sum_{i=1}^{N} r_j + p(n - N).$$

Since $a - 1$ is an even number, it follows that $p(n - N)$ is even, that is $n - N$ is even. □

EXAMPLE 2 Let us use Theorem 13.5 to compute $\left(\frac{3}{p}\right)$ for primes $p \geq 5$. Since

$$\left[\frac{3j}{p}\right] = \begin{cases} 0 & \text{if } 1 \leq j \leq [p/3], \\ 1 & \text{if } [p/3] + 1 \leq j \leq (p-1)/2. \end{cases}$$

it follows that $\left(\frac{3}{p}\right) = (-1)^n$, where $n = (p-1)/2 - [p/3]$. By considering the cases $p = 12k \pm 1$ and $p = 12k \pm 5$ separately, we see that n is even if and only if $p \equiv \pm 1 \pmod{12}$. Hence, $\left(\frac{3}{p}\right) = 1$ if and only if $p \equiv \pm 1 \pmod{12}$. □

Gauss' lemma and Theorem 13.5 are too cumbersome for numerical calculations of $\left(\frac{a}{p}\right)$. Instead, one uses Gauss' law of quadratic reciprocity, which will be the theme of next section.

14 Quadratic Reciprocity

Theorem 14.1 (The Gaussian Reciprocity Law) *Let p and q be two distinct odd primes. Then*

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

that is

$$\left(\frac{q}{p}\right) = \begin{cases} \left(\frac{p}{q}\right) & \text{if } p \equiv 1 \pmod 4 \text{ or } q \equiv 1 \pmod 4, \\ -\left(\frac{p}{q}\right) & \text{if } p \equiv 3 \pmod 4 \text{ and } q \equiv 3 \pmod 4. \end{cases}$$

Proof. By Theorem 13.5, $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^M(-1)^N = (-1)^{M+N}$, where

$$M = \sum_{k=1}^{(q-1)/2}\left[\frac{kp}{q}\right] \qquad \text{and} \qquad N = \sum_{j=1}^{(p-1)/2}\left[\frac{jq}{p}\right].$$

We will prove that $M + N = \{(p-1)/2\}\{(q-1)/2\}$.

To this end, consider the set

$$A = \{(j,k) \mid j = 1, 2, \dots, (p-1)/2 \text{ and } k = 1, 2, \dots, (q-1)/2\}.$$

We can represent A as a rectangular set of lattice points, i.e. points with integral coordinates, in a rectilinear coordinate system. Since $(p,q) = 1$, no number qj/p is an integer when $j = 1, 2, \dots, p-1$. Hence, there is no point from A on the line $y = \frac{q}{p}x$. Let B be all points in A that lie below this line, that is $(j,k) \in B$ if and only if $k < jq/p$. For a given j this condition is satisfied for $k = 1, 2, \dots, [jq/p]$. Hence there are exactly $[jq/p]$ points in B whose first coordinate equals j. Since this holds for $j = 1, 2, \dots, (p-1)/2$, we conclude that the total number of points in B equals N.

Similarly, M is the number of points in the set $C = \{(j,k) \in A \mid j < kp/q\} = \{(j,k) \in A \mid k > jq/p\}$, which can be represented as the set of all points in A above the line $y = \frac{q}{p}x$. Since A is the disjoint union of B and C, it follows that $M + N$ equals the number of points in A, which is $\{(p-1)/2\}\{(q-1)/2\}$.

This number is odd if and only if both $(p-1)/2$ and $(q-1)/2$ are odd, that is if and only if $p \equiv q \equiv 3 \pmod 4$. Hence $\left(\frac{p}{q}\right)$ and $\left(\frac{q}{p}\right)$ are of opposite sign if and only if $p \equiv q \equiv 3 \pmod 4$. □

Example 1 The number 991 is a prime, and we will compute the Legendre symbol $\left(\frac{402}{991}\right)$. We start by factoring 402; $402 = 2 \cdot 3 \cdot 67$. Next we compute

$$\begin{aligned}
\left(\frac{2}{991}\right) &= 1. && [991 \equiv -1 \pmod 8)] \\
\left(\frac{3}{991}\right) &= -\left(\frac{991}{3}\right) && [991 \equiv 3 \pmod 4)] \\
&= -\left(\frac{1}{3}\right) && [991 \equiv 1 \pmod 3)] \\
&= -1. \\
\left(\frac{67}{991}\right) &= -\left(\frac{991}{67}\right) && [991 \equiv 67 \equiv 3 \pmod 4)] \\
&= -\left(\frac{-14}{67}\right) && [991 \equiv -14 \pmod{67})] \\
&= -\left(\frac{-1}{67}\right)\left(\frac{2}{67}\right)\left(\frac{7}{67}\right) && [-14 = (-1)\cdot 2 \cdot 7] \\
&= -(-1)\cdot(-1)\cdot\left(-\left(\frac{67}{7}\right)\right) && [67 \equiv 3 \pmod 8) \text{ and } 7 \equiv 3 \pmod 4)] \\
&= \left(\frac{4}{7}\right) && [67 \equiv 4 \pmod 7)] \\
&= 1. && [4 \text{ is a square.}]
\end{aligned}$$

Finally, $\left(\frac{402}{991}\right) = \left(\frac{2}{991}\right)\left(\frac{3}{991}\right)\left(\frac{67}{991}\right) = 1 \cdot (-1) \cdot 1 = -1.$ □

Example 2 The number 2137 is a prime which is congruent to 1 modulo 8, and $666 = 2 \cdot 3^2 \cdot 37$. It follows that

$$\left(\frac{666}{2137}\right) = \left(\frac{2}{2137}\right)\left(\frac{37}{2137}\right) = 1 \cdot \left(\frac{2137}{37}\right).$$

Now $2137 \equiv -9 \pmod{37}$ and $37 \equiv 1 \pmod 4$. Hence

$$\left(\frac{2137}{37}\right) = \left(\frac{-9}{37}\right) = \left(\frac{-1}{37}\right) = 1, \quad \text{that is } \left(\frac{666}{2137}\right) = 1.$$ □

15 Primitive Roots

Let us start by computing the powers 3^i modulo 7 for $0 \le i < \phi(7) = 6$. We obtain $3^0 = 1$, $3^1 = 3$, $3^2 \equiv 2$, $3^3 \equiv 6$, $3^4 \equiv 4$, $3^5 \equiv 5$. Hence, the set $\{3^i \mid 0 \le i < \phi(7)\}$ is a reduced residue system modulo 7, that is every integer a not divisible by 7 is congruent to 3^i for a unique integer i modulo $\phi(7)$. This fact allows us to replace calculations using only multiplication and exponentiation modulo 7 by calculations using addition modulo $\phi(7)$ instead.

Example 1 1 Solve the equation $x^5 \equiv 6 \pmod 7$.

Solution: Let $x \equiv 3^y \pmod 7$. Since $6 \equiv 3^3 \pmod 7$, the given equation can now be written $3^{5y} \equiv 3^3 \pmod 7$, which is equivalent to the congruence $5y \equiv 3$

(mod 6). The latter congruence has the unique solution $y \equiv 3 \pmod 6$, and hence our original equation has the unique solution $x \equiv 6 \pmod 7$. □

Motivated by Example 1, we will investigate numbers m with the property that there exists a number g such that $\{g^i \mid 0 \le i < \phi(m)\}$ is a reduced residue system. That not all numbers m have this property follows from the following example.

Example 2 Since $1^2 \equiv 3^2 \equiv 5^2 \equiv 7^2 \equiv 1 \pmod 8$ and $\phi(8) = 4$, it follows that $\{a^i \mid 0 \le i < 4\}$ is never equal to a reduced residue system modulo 8. □

Proposition 15.1 *Let m be a positive integer and a any integer such that $(a, m) = 1$. Define*

$$A = \{k \in \mathbf{Z} \mid a^{|k|} \equiv 1 \pmod m\}.$$

Then A is an ideal in $\mathbf{Z}$.

Proof. We have to prove that the set A is closed under subtraction, i.e. that $j, k \in A \Rightarrow j - k \in A$. To prove this we may assume $j \ge k$, because $j - k$ belongs to A if and only if $k - j$ belongs to A.

Suppose $j, k \in A$. If $j \ge k \ge 0$, then $a^j \equiv a^k \equiv 1 \pmod m$, and hence $a^{j-k} \equiv a^{j-k}a^k = a^j \equiv 1 \pmod m$. If $j \ge 0 > k$, then $a^j \equiv a^{-k} \equiv 1 \pmod m$, and we obtain $a^{j-k} = a^j a^{-k} \equiv 1 \cdot 1 = 1 \pmod m$. Finally, if $0 > j \ge k$, then $a^{-j} \equiv a^{-k} \equiv 1 \pmod m$, and we conclude that $a^{j-k} \equiv a^{-j}a^{j-k} = a^{-k} \equiv 1 \pmod m$. Thus, in each case $j - k \in A$. □

Note that A contains nonzero integers, because $\phi(m)$ belongs to A by Euler's theorem. By Theorem 1.8, the ideal A is generated by a unique positive integer h, which is the smallest positive integer belonging to A, that is $a^h \equiv 1 \pmod m$ while $a^j \not\equiv 1 \pmod m$ for $1 \le j < h$.

Definition 15.2 The positive generator h of A, i.e. the smallest positive integer such that $a^h \equiv 1 \pmod m$, is called the *order of a modulo m* and is denoted by $\operatorname{ord} a$.

The order $\operatorname{ord} a$ of course depends on the modulus m, but since the modulus will always be fixed during a calculation, this ambiguity in the notation causes no difficulties.

For any modulus m, $\operatorname{ord} 1 = 1$.

Example 3 Modulo 8 we have $\operatorname{ord} 3 = \operatorname{ord} 5 = \operatorname{ord} 7 = 2$. □

Example 4 Let us compute the order of the numbers 2, 3 and 6 modulo 7. Our calculations before Example 1 show that $\operatorname{ord} 3 = 6$. Since $2^2 \equiv 4 \pmod 7$ and $2^3 \equiv 1 \pmod 7$, $\operatorname{ord} 2 = 3$, and since $6^2 \equiv 1 \pmod 7$, $\operatorname{ord} 6 = 2$. □

The following theorem is an immediate consequence of the fact that the ideal A is generated by $h = \operatorname{ord} a$.

Theorem 15.3 *Assume $(a, m) = 1$ and write $h = \operatorname{ord} a$ modulo m. Then*

(i) $a^n \equiv 1 \pmod m$ *if and only if* $h \mid n$;

(ii) $h \mid \phi(m)$;

(iii) $a^j \equiv a^k \pmod{m}$ *if and only if* $j \equiv k \pmod{h}$*;*

(iv) the numbers 1, a*,* a^2*, ...,* a^{h-1} *are incongruent modulo* m*, and each power* a^n *is congruent to one of these modulo* m*;*

(v) $\operatorname{ord} a^k = h/(h,k)$*.*

Proof. (i) follows from the definition of a generator of an ideal.

(ii) follows from (i) and Euler's theorem.

(iii) Assume $k \geq j \geq 0$; then $a^k \equiv a^j \pmod{m}$ holds if and only if $a^{k-j} \equiv 1 \pmod{m}$, because we may divide the former congruence by a^j since $(a, m) = 1$. The conclusion now follows from (i).

(iv) is of course a consequence of (iii).

(v) By (i), $(a^k)^n \equiv 1 \pmod{m} \Leftrightarrow kn \equiv 0 \pmod{h}$. We can divide the right hand congruence by k provided we change the modulus to $h/(h,k)$. Thus,

$$(a^k)^n \equiv 1 \pmod{m} \Leftrightarrow n \equiv 0 \pmod{h/(h,k)}.$$

The smallest positive number n satisfying the last congruence is $n = h/(h,k)$; by definition, this is the order of a^k modulo m. □

Theorem 15.3 (ii) implies that $\operatorname{ord} a \leq \phi(m)$ for every number a which is relatively prime to m. An obvious question now arises: For which m does there exist an integer whose order is as large as possible, namely $\phi(m)$? This question motivates the following definition.

Definition 15.4 Assume that $(g, m) = 1$. If the order of g modulo m equals $\phi(m)$, then g is called a *primitive root modulo* m*,* or a *primitive root of* m*.*

Example 5 In Example 4 we calculated the order of 3 modulo 7 and found that $\operatorname{ord} 3 = 6 = \phi(7)$. Consequently, 3 is a primitive root modulo 7. □

Example 6 Not every integer has a primitive root. If $m = 8$, then $a^2 \equiv 1$ for every odd integer and hence $\operatorname{ord} a \leq 2 < 4 = \phi(8)$ for every a relatively prime to 8, that is 8 has no primitive roots. □

Theorem 15.5 *Suppose* g *is a primitive root modulo* m*. Then*

(i) $\{1, g, g^2, \dots, g^{\phi(m)-1}\}$ *is a reduced residue system modulo* m*;*

(ii) $g^j \equiv g^k \pmod{m}$ *if and only if* $j \equiv k \pmod{\phi(m)}$*;*

(iii) g^k *is a primitive root modulo* m *if and only if* $(k, \phi(m)) = 1$*.*
In particular, if there exists a primitive root modulo m*, then there are precisely* $\phi(\phi(m))$ *primitive roots.*

Proof. Theorem 15.5 is just a special case of Theorem 15.3. □

Example 7 We have found that 3 is a primitive root modulo 7. Since $\phi(\phi(7)) = \phi(6) = 2$, there are 2 primitive roots. The other primitive root is 3^5, i.e. 5 (mod 7). □

We will show that the only positive integers having primitive roots are 1, 2, 4, p^k and $2p^k$, where p is an odd prime and k an arbitrary positive integer. We start by proving that each prime has primitive roots; for this we will need the following two lemmas.

Lemma 15.6 *If a has order h and b has order k modulo m, and if $(h, k) = 1$, then ab has order hk modulo m.*

Proof. Let r be the order of ab. Since $(ab)^{hk} = (a^h)^k (b^k)^h \equiv 1^k \cdot 1^h = 1 \pmod{m}$, we conclude that $r \mid hk$. To complete the proof, we have to show that $hk \mid r$. Note that $b^{rh} \equiv (a^h)^r b^{rh} = (ab)^{rh} \equiv 1 \pmod{m}$, and hence $k \mid rh$. Since $(h, k) = 1$ it follows that $k \mid r$. In a similar way, we show that $h \mid r$. Since $(h, k) = 1$ it now follows that $hk \mid r$. □

EXAMPLE 8 Working modulo 7 we have $\operatorname{ord} 2 = 3$ and $\operatorname{ord} 6 = 2$. Consequently, since $2 \cdot 6 \equiv 5 \pmod 7$, $\operatorname{ord} 5 = \operatorname{ord}(2 \cdot 6) = 2 \cdot 3 = 6$. □

Lemma 15.7 *Let p and q be primes, and suppose that $q^k \mid (p - 1)$. Then there exists a number a of order q^k modulo p.*

Proof. By Corollary 9.9, the congruence $x^{q^k} \equiv 1 \pmod{p}$ has exactly q^k roots. By Theorem 15.3 (i), the order of such a root is a divisor of q^k. If a is a root of order less than q^k, then a is the root of the congruence $x^{q^{k-1}} \equiv 1 \pmod{p}$, but this congruence has exactly q^{k-1} roots. Hence, there are exactly $q^k - q^{k-1}$ incongruent numbers of order precisely q^k.

□

Theorem 15.8 *If p is a prime, then there exist exactly $\phi(p - 1)$ primitive roots modulo p.*

Proof. By the last statement of Theorem 15.5, it is enough to show that there exists at least one primitive root modulo p. Let $p - 1 = q_1^{k_1} q_2^{k_2} \cdots q_r^{k_r}$ be the factorization of $p - 1$ into distinct primes. By Lemma 15.7 there are integers a_i of order $q_i^{k_i}$ for $i = 1, 2, \ldots, r$. The numbers $q_i^{k_i}$ are pairwise relatively prime, so by repeated use of Lemma 15.6 we see that $g = a_1 a_2 \cdots a_r$ has order $p - 1$, that is g is a primitive root modulo p. □

Suppose that g is a primitive root modulo m. If $(a, m) = 1$, then Theorem 15.5 implies that there is a unique integer i, with $0 \le i \le \phi(m) - 1$ such that $g^i \equiv a \pmod{m}$. This fact allows us to make the following definition.

Definition 15.9 Let g be a primitive root of m, and suppose $(a, m) = 1$. The smallest nonnegative integer i such that $g^i \equiv a \pmod{m}$ is called the *index of a (to the base g)* and is denoted by $\operatorname{ind} a$.

The index depends on both the modulus m and the root g, but since m and g are usually fixed, the notation should cause no confusion.

There is a strong similarity between logarithms and indices, and the following theorem states the most important properties. The proof is simple and is left to the reader.

Theorem 15.10 *Suppose g is a primitive root modulo m, and let* $\operatorname{ind} a$ *denote the index of a to the base g.*

(i) $\operatorname{ind} 1 = 0$ *and* $\operatorname{ind} g = 1$.

(ii) $a \equiv b \pmod{m}$ *if and only if* $\operatorname{ind} a = \operatorname{ind} b$.

(iii) $\operatorname{ind} ab \equiv \operatorname{ind} a + \operatorname{ind} b \pmod{\phi(m)}$.

(iv) $\operatorname{ind} a^k \equiv k \operatorname{ind} a \pmod{\phi(m)}$, *for all nonnegative integers k.*

Theorem 15.11 *Let m be a positive integer having a primitive root, and suppose $(a, m) = 1$. Then the congruence $x^n \equiv a \pmod m$ has a solution if and only if*

$$a^{\phi(m)/(n,\phi(m))} \equiv 1 \pmod m. \tag{1}$$

If the congruence $x^n \equiv a \pmod m$ is solvable, then it has exactly $(n, \phi(m))$ incongruent solutions.

Proof. Let g be a primitive root modulo m, and let $d = (n, \phi(m))$. Taking indices, we see that the congruence $x^n \equiv a \pmod m$ holds if and only if $n \operatorname{ind} x \equiv \operatorname{ind} a \pmod{\phi(m)}$. By Theorem 5.1, this congruence is solvable if and only if $d \mid \operatorname{ind} a$, and if solutions exist, then there are exactly d incongruent solutions.

To complete the proof, we show that (1) holds if and only if $d \mid \operatorname{ind} a$. Taking indices, we see that (1) is equivalent to $(\phi(m)/d) \operatorname{ind} a \equiv 0 \pmod{\phi(m)}$, which holds if and only if $d \mid \operatorname{ind} a$. □

If m has a primitive root, then the solutions of a solvable congruence $x^n \equiv a \pmod m$ can be found using indices, provided we compute (or have available) a table of indices for the given modulus m. See Example 1.

Since every prime modulus has a primitive root, we have the following corollary of Theorem 15.11.

Corollary 15.12 *Suppose p is prime and $(a, p) = 1$. Then the congruence $x^n \equiv a \pmod p$ is solvable if and only if*

$$a^{(p-1)/(n,p-1)} \equiv 1 \pmod p.$$

Remark. The corollary gives an efficient procedure for determining whether the congruence $x^n \equiv a \pmod p$ is solvable, but to actually find a solution is more difficult. However, if $(n, p - 1) = 1$, this is relatively easy. Use the Euclidean Algorithm to find positive integers s and t such that $sn = t(p - 1) + 1$; then $a^{sn} = a^{t(p-1)}a \equiv a \pmod p$, that is a^s is a solution of the congruence $x^n \equiv a \pmod p$.

The following corollary is a generalization of Corollary 9.9.

Corollary 15.13 *Suppose that m has a primitive root and that $n \mid \phi(m)$. Then the congruence $x^n - 1 \equiv 0 \pmod m$ has exactly n roots.*

Proof. The congruence $x^n \equiv 1 \pmod m$ is obviously solvable. Hence, by Theorem 15.11 it has $(n, \phi(m)) = n$ incongruent solutions. □

We next show that if p is an odd prime, then p^k has primitive roots for each k.

Theorem 15.14 *Suppose that p is an odd prime.*

(i) If g is a primitive root modulo p, then $g + np$ is a primitive root modulo p^2 for exactly $p - 1$ values of n modulo p.

(ii) If g is a primitive root modulo p^2, then g is a primitive root modulo p^k for all $k \geq 2$.

Proof. Let h denote the order of $g + np$ modulo p^2. (h may depend on n.) Then $h \mid \phi(p^2)$, that is $h \mid p(p - 1)$.

But $(g+np)^h \equiv 1 \pmod{p^2}$ implies $(g+np)^h \equiv 1 \pmod{p}$, and by the binomial theorem, $(g+np)^h = g^h + \sum_{j=1}^{h}\binom{h}{j}(np)^j g^{h-j} \equiv g^h \pmod{p}$, and hence $g^h \equiv 1 \pmod{p}$. Since g has order $p-1$, it follows that $(p-1) \mid h$.

Thus $h = p-1$ or $h = p(p-1)$. In the latter case, $g+np$ is a primitive root of p^2, and in the former case it is not. We will prove that the former case arises only for one of the p possible values of n.

Let $f(x) = x^{p-1}-1$; then g is a root of the congruence $f(x) \equiv 0 \pmod{p}$ and $f'(g) = (p-1)g^{p-2} \not\equiv 0 \pmod{p}$, since $(g^{p-2}, p) = 1$. Hence, by Theorem 10.1 there is a unique root of the form $g+np$ of the congruence $f(x) \equiv 0 \pmod{p^2}$. This proves our claim.

(ii) It suffices to prove that if g is a primitive root modulo p^k, $k \geq 2$, then g is also a primitive root modulo p^{k+1}. Let h be the order of g modulo p^{k+1}; then $h \mid \phi(p^{k+1})$, that is $h \mid p^k(p-1)$. Because $g^h \equiv 1 \pmod{p^{k+1}}$ implies $g^h \equiv 1 \pmod{p^k}$ and g is a primitive root modulo p^k, $\phi(p^k)$ must divide h, that is $p^{k-1}(p-1) \mid h$.

Thus either $h = p^{k-1}(p-1)$ or $h = p^k(p-1) = \phi(p^{k+1})$. In the latter case, g is a primitive root modulo p^{k+1} as claimed. We must show that the former case is excluded.

Let $t = \phi(p^{k-1})$; then $g^t \equiv 1 \pmod{p^{k-1}}$ by Euler's theorem, and therefore $g^t = 1 + np^{k-1}$ for some integer n. If $p \mid n$ then we would have $g^t \equiv 1 \pmod{p^k}$, which contradicts the fact that g is primitive root modulo p^k. Thus, $p \nmid n$.

By the binomial theorem

$$\begin{aligned} g^{pt} &= (g^t)^p = (1+np^{k-1})^p = 1 + np^k + \frac{p(p-1)}{2}n^2p^{2k-2} + \dots \\ &\equiv 1 + np^k \pmod{p^{k+1}}. \end{aligned}$$

Here, we have used that fact that the integer $\frac{p(p-1)}{2}n^2p^{2k-2} = \frac{p-1}{2}n^2p^{2k-1}$ is divisible by p^{k+1}, because $2k-1 \geq k+1$ when $k \geq 2$, and the remaining omitted terms in the expansion contain even higher powers of p.

Since $p \nmid n$, we now conclude that

$$g^{pt} \not\equiv 1 \pmod{p^{k+1}}.$$

Therefore, $h \neq pt = p\phi(p^{k-1}) = p^{k-1}(p-1)$, and the proof is complete. □

Example 9 Since $2^2 \equiv -1 \not\equiv 1 \pmod{5}$, we conclude that the order of 2 modulo 5 must be 4, that is 2 is a primitive root of 5. By Theorem 15.14, $2+5n$ is a primitive root of 25 for exactly four values of n, $0 \leq n \leq 4$. Since $\phi(25) = 20$, the primitive roots of 25 have order 20. The order h modulo 25 of an arbitrary number a is a divisor of 20. If $h < 20$, then either $h \mid 4$ or $h \mid 10$, so it follows that $a^4 \equiv 1 \pmod{25}$ or $a^{10} \equiv 1 \pmod{25}$. Hence, to find whether a number a has order 20 it is enough to compute a^4 and a^{10} modulo 25; the order is 20 if and only if none of these two powers are congruent to 1. For $a = 2$ we obtain $2^2 \equiv 4$, $2^4 \equiv 16$, $2^8 \equiv 6$ and $2^{10} \equiv 24$. Hence, the order of 2 is 20, i.e. 2 is a primitive root of 25.

For $a = 7$ we obtain $7^2 \equiv -1$ and $7^4 \equiv 1 \pmod{25}$, that is the order of 7 is 4, and 7 is not a primitive root of 25. Of course, it now follows that 12, 17 and 22 are primitive roots of 25.

By Theorem 15.14 (ii), 2 is a primitive root of 5^k for all k. □

Theorem 15.15 *Suppose that p is an odd prime, and let g be a primitive root modulo p^k. If g is odd, then g is also a primitive root modulo $2p^k$, and if g is even, then $g + p^k$ is a primitive root modulo $2p^k$.*

Proof. If g is odd, then $g^j \equiv 1 \pmod 2$ for every $j \geq 1$. Thus $g^j \equiv 1 \pmod{2p^k}$ if and only if $g^j \equiv 1 \pmod{p^k}$, and hence the order of g modulo $2p^k$ is equal to the order of g modulo p^k, namely $\phi(p^k)$. Since $\phi(2p^k) = \phi(p^k)$, g is a primitive root of $2p^k$.

If g is even, then g cannot be a primitive root of $2p^k$, for a primitive root is always relatively prime to the modulus. But $g + p^k$ is odd and, since it is congruent to g modulo p^k, it is also a primitive root modulo p^k. Hence, $g + p^k$ is a primitive root of $2p^k$ by the preceding argument. □

Example 10 By Example 9, 2 is a primitive root of 5^k for each k. Hence, $2+5^k$ is a primitive root of $2 \cdot 5^k$ for each k. In particular 7 is a primitive root of 10, and 27 is a primitive root of 50. By the same example, 17 is also a primitive root of 5^k for each k. Since 17 is odd, it follows that 17 is a primitive root of $2 \cdot 5^k$ for each k. □

Theorem 15.16 *There exists a primitive root modulo m if and only if $m = 1$, 2, 4, p^k, or $2p^k$, where p is an odd prime and k is an arbitrary positive integer.*

Proof. First note that 1, 2, and 4 have primitive roots (1, 1, and 3, respectively). Theorems 15.8, 15.14, and 15.15 imply that p^k and $2p^k$ have primitive roots whenever p is an odd prime and k is an arbitrary positive integer.

Conversely, to prove that these are the only positive integers having primitive roots, assume $m > 2$ has a primitive root. By Corollary 15.13, the congruence $x^2 \equiv 1 \pmod m$ has exactly 2 incongruent solutions (because $2 \mid \phi(m)$ for all $m \geq 3$). Theorem 11.5 now implies that m must be either 4, p^k, or $2p^k$, where p is an odd prime. □

Concluding remarks. Readers with a basic knowledge of group theory may have noticed that most of the notions in this section are special cases of general group notions.

If G is a general finite group with identity element e, then the order $\operatorname{ord} a$ of an element a is defined to be the smallest positive integer n satisfying $a^n = e$, while the order of the group, $\operatorname{ord} G$, is defined to be the number of elements in G. If $h = \operatorname{ord} a$, then $h | \operatorname{ord} G$ and $\{e, a, a^2, \dots, a^{h-1}\}$ is a subgroup of G. This subgroup coincides with G if $\operatorname{ord} a = \operatorname{ord} G$, and G is then called a *cyclic* group with a as a *generator*.

Applying these general notions to the specific case when G is the group $\mathbf{Z}_m^*$ of all residue classes modulo m that are relatively prime to m, we see that the order h of an integer a modulo m coincides with the order of the residue class $\overline{a}$ in $\mathbf{Z}_m^*$, that $h|\phi(m)$, that a number g is a primitive root modulo m if and only if the residue class $\overline{g}$ generates $\mathbf{Z}_m^*$, and that there exists a primitive root modulo m if and only if the group $\mathbf{Z}_m^*$ is a cyclic group. Using the language of groups we can state Theorem 15.16 as follows: The group $\mathbf{Z}_m^*$ is cyclic if and only if $m = 1$, 2, 4, p^k or $2p^k$, where p is an odd prime and k is an arbitrary positive integer.

16 Arithmetic Functions

Functions that are defined for all positive integers and whose range is a subset of **R** (or more generally **C**) are called *arithmetic functions.*

We have already considered one very important arithmetic functions − the *Euler ϕ-function.* Other important arithmetic functions to be considered in this section are

- $\tau(n)$, the number of positive divisors of n;
- $\sigma(n)$, the sum of the positive divisors of n;
- $\sigma_k(n)$, the sum of the k^{th} powers of the positive divisors of n.

We will use the following sum and product conventions. $\sum_{d|n} f(d)$ and $\prod_{d|n} f(d)$ denote the sum and product, respectively, of $f(d)$ over all positive divisors d of n. For example, $\sum_{d|12} f(d) = f(1)+f(2)+f(3)+f(4)+f(6)+f(12)$.

Using this notation, we have

$$\tau(n) = \sum_{d|n} 1, \qquad \sigma(n) = \sum_{d|n} d, \qquad \sigma_k(n) = \sum_{d|n} d^k.$$

Note that the divisor functions $\tau(n)$ and $\sigma(n)$ are special cases of $\sigma_k(n)$, since $\tau(n) = \sigma_0(n)$ and $\sigma(n) = \sigma_1(n)$.

Definition 16.1 An arithmetic function $f(n)$ is called *multiplicative* if it is not identically zero and satisfies $f(mn) = f(m)f(n)$ for every pair of relatively prime positive integers m and n. If $f(mn) = f(m)f(n)$ for each pair m and n, relatively prime or not, then $f(n)$ is said to be *completely multiplicative.*

If f is a multiplicative function, then $f(n) = f(n)f(1)$ for every positive integer n, and since there is an n for which $f(n) \neq 0$, it follows that $f(1) = 1$. Using mathematical induction, it is easy to prove that if $m_1, m_2, \dots, m_r$ are pairwise relatively prime positive integers, then

$$f(m_1 m_2 \cdots m_r) = f(m_1)f(m_2)\cdots f(m_r).$$

In particular, this holds whenever the integers $m_1, m_2, \dots, m_r$ are powers of distinct primes. Thus, if $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ is the canonical factorization of the integer $n > 1$ as a product of powers of distinct primes, then $f(n) = f(p_1^{k_1}) f(p_2^{k_2}) \cdots f(p_r^{k_r})$. Therefore, the value of $f(n)$ for every n is completely determined by the values $f(p^k)$ for all prime powers.

We already know that $\phi(n)$ is multiplicative (Theorem 6.2), and we have used this fact to obtain a formula for $\phi(n)$.

Our next theorem yields a general method for constructing multiplicative functions.

Theorem 16.2 *Let $f(n)$ be a multiplicative function, and let $F(n) = \sum_{d|n} f(d)$. Then $F(n)$ is multiplicative.*

Proof. Let $(m, n) = 1$. If $d \mid mn$, then $d = d_1 d_2$, where $d_1 \mid m$ and $d_2 \mid n$. Moreover, $d_1 = (m, d)$, $d_2 = (n, d)$ and $(d_1, d_2) = 1$, and the factorization is unique. Consequently,

$$\begin{aligned}F(mn) &= \sum_{d|mn} f(d) = \sum_{d_1|m}\sum_{d_2|n} f(d_1d_2) = \sum_{d_1|m}\sum_{d_2|n} f(d_1)f(d_2)\\ &= \sum_{d_1|m} f(d_1)\sum_{d_2|n} f(d_2) = F(m)F(n).\end{aligned}$$ □

Corollary 16.3 *(i) The functions $\tau(n)$, $\sigma(n)$, and more generally, $\sigma_k(n)$ are multiplicative.*

(ii) If $n = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$, then

$$\tau(n) = \prod_{j=1}^{r}(k_j+1) \qquad \text{and} \qquad \sigma(n) = \prod_{j=1}^{r}\left(\frac{p_j^{k_j+1}-1}{p_j-1}\right).$$

Proof. (i) Since $\sigma_k(n) = \sum_{d|n} d^k$, and the function $f(n) = n^k$ is (completely) multiplicative, it follows from the previous theorem that $\sigma_k(n)$ is multiplicative. $\tau(n)$ and $\sigma(n)$ are special cases.

(ii) The positive divisors of p^k are 1, p, p^2, …, p^k, and hence $\tau(p^k) = k+1$ and $\sigma(p^k) = \sum_{j=0}^{k} p^j = (p^{k+1}-1)/(p-1)$. The formulas for $\tau(n)$ and $\sigma(n)$ follow from this. □

Theorem 16.4 *For every positive integer n, $\sum_{d|n}\phi(d) = n$.*

Proof. Write $F(n) = \sum_{d|n}\phi(d)$; then $F(n)$ is multiplicative by Theorem 16.2. Since the function $G(n) = n$ is also multiplicative, it suffices to verify that $F(p^k) = p^k$ for all prime powers p^k in order to prove that $F(n) = n$ for all n. But $\phi(p^j) = p^j - p^{j-1}$ for $j \geq 1$, and hence

$$F(p^k) = \sum_{d|p^k}\phi(d) = \sum_{j=0}^{k}\phi(p^j) = 1 + \sum_{j=1}^{k}(p^j - p^{j-1}) = p^k.$$ □

Let $f(n)$ be an arithmetic function, and define $F(n) = \sum_{d|n} f(n)$. Is the function f uniquely determined by the function F? We have

$$\begin{cases} F(1) = f(1) \\ F(2) = f(1) + f(2) \\ F(3) = f(1) \qquad\quad + f(3) \\ F(4) = f(1) + f(2) \qquad\quad + f(4) \\ F(5) = f(1) \qquad\qquad\qquad\qquad + f(5) \\ \quad\vdots \\ F(n) = f(1) + \qquad\qquad \dots \qquad\qquad + f(n) \end{cases}$$

This can be viewed as a triangular system of linear equations with $f(1)$, $f(2)$, …, $f(n)$ as unknowns. It is now obvious that $f(n)$ is a linear combination of $F(1)$, $F(2)$, …, $F(n)$ with integral coefficients. In particular, the function f is uniquely determined by the function F. Our next objective is to derive a formula for $f(n)$, and for this we will need the following function.

Definition 16.5 Define

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \text{ is divisible by } p^2 \text{ for some prime } p \\ (-1)^r & \text{if } n = p_1 p_2 \cdots p_r, \text{ where } p_1, p_2, \dots, p_r \text{ are distinct primes.} \end{cases}$$

The function μ is called *Möbius' μ-function.*

Theorem 16.6 *The function $\mu(n)$ is multiplicative and*

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \textit{if } n = 1 \\ 0 & \textit{if } n > 1. \end{cases}$$

Proof. Multiplicativity is obvious. Define $F(n) = \sum_{d|n} \mu(d)$; then $F(n)$ is multiplicative by Theorem 16.2. Since $\mu(p) = -1$ and $\mu(p^j) = 0$ for $j \ge 2$, we have $F(p^k) = \sum_{j=0}^{k} \mu(p^j) = \mu(1) + \mu(p) = 1 - 1 = 0$, for all primes p and all $k \ge 1$. Hence, $F(n) = 0$ for all $n > 1$, and $F(1) = \mu(1) = 1$. □

Theorem 16.7 (Möbius' inversion formula) *Let f be an arbitrary arithmetic function. If $F(n) = \sum_{d|n} f(d)$ for every positive integer n, then*

$$f(n) = \sum_{d|n} \mu(d) F(n/d).$$

Proof. Using the definition of F we obtain

$$\sum_{d|n} \mu(d) F(n/d) = \sum_{d|n} \mu(d) \sum_{k|(n/d)} f(k) = \sum_{\text{all } d,\, k \text{ with } dk|n} \mu(d) f(k).$$

Now we can reverse the order of summation and write the last sum in the form

$$\sum_{\text{all } d,\, k \text{ with } dk|n} \mu(d) f(k) = \sum_{k|n} f(k) \sum_{d|(n/k)} \mu(d).$$

By Theorem 16.6, $\sum_{d|(n/k)} \mu(d) = 0$ except for $k = n$, when the value is 1. Hence, $\sum_{d|n} \mu(d) F(n/d) = \sum_{k|n} f(k) \sum_{d|(n/k)} \mu(d) = f(n)$. □

The following converse is also true.

Theorem 16.8 *If $f(n) = \sum_{d|n} \mu(d) F(n/d)$ for every positive integer n, then $F(n) = \sum_{d|n} f(d)$.*

Proof. Define $G(n) = \sum_{d|n} f(d)$; then $f(n) = \sum_{d|n} \mu(d) G(n/d)$ by Theorem 16.7. Thus,

$$\sum_{d|n} \mu(d) F(n/d) = \sum_{d|n} \mu(d) G(n/d) \tag{1}$$

holds for all n. We will use induction to show that this implies that $F(n) = G(n)$ for all positive integers n.

First of all, taking $n = 1$ in (1) we get $\mu(1)F(1/1) = \mu(1)G(1/1)$, that is $F(1) = G(1)$. Suppose that we have $F(m) = G(m)$ for all $m < n$. Since $n/d < n$ for all positive divisors d of n except for $d = 1$, (1) now simplifies to $\mu(1)F(n/1) = \mu(1)G(n/1)$, and we conclude that $F(n) = G(n)$. This completes the induction. □

17 Sums of Squares

In this section we will treat the problem of representing a positive integer as a sum of squares. In particular, we will determine which numbers can be written as the sum of two squares, and we will prove that every positive number is a sum of four squares.

By definition, a positive integer n is the sum of two squares if the equation $x^2 + y^2 = n$ has an integral solution x, y. Since $x^2 \equiv 0$ or $1 \pmod 4$ for all integers x, it is clear that the sum $x^2 + y^2$ of two squares is never congruent to 3 modulo 4. Hence no integer of the form $4m+3$ is the sum of two squares. For primes we have the following necessary and sufficient condition

Theorem 17.1 *Let p be a prime. Then p is the sum of two squares if and only if $p = 2$ or $p \equiv 1 \pmod 4$.*

Proof. By the preceding comments, no prime $\equiv 3 \pmod 4$ is the sum of two squares, and clearly $2 = 1^2+1^2$. It remains to prove that every prime congruent to 1 modulo 4 is the sum of two squares.

Assume $p \equiv 1 \pmod 4$ and write $N = [\sqrt{p}]$; then of course $N < \sqrt{p} < N+1$. The number -1 is a quadratic residue of p by Theorem 11.9, and hence there is an integer i such that $i^2 \equiv -1 \pmod p$. Let A be the set of all pairs (j,k), where j and k are integers in the interval $[0,N]$, let B denote the set of all residue classes modulo p, i.e. $B = \{\overline{0}, \overline{1}, \dots, \overline{p-1}\}$, and define a function $f\colon A \to B$ by $f(x,y) = \overline{x+iy}$. Since B has p elements and A has $(N+1)^2 > p$ elements, it is impossible for the function f to be injective. Hence there must be two distinct pairs (x_1,y_1) and (x_2,y_2) in A that are mapped onto the same equivalence class, that is $x_1 + iy_1 \equiv x_2 + iy_2 \pmod p$. Let $a = x_1 - x_2$ and $b = y_1 - y_2$; then a and b are not both zero and $a \equiv -ib \pmod p$, and by squaring we obtain $a^2 \equiv i^2b^2 \equiv -b^2 \pmod p$, that is a^2+b^2 is a multiple of p. But $|a| \le N$ and $|b| \le N$, and therefore $0 < a^2+b^2 \le 2N^2 < 2p$. It follows that $a^2+b^2 = p$. □

Lemma 17.2 *Suppose that $n = a^2 + b^2$ and that the prime factorization of n contains the prime factor q, where $q \equiv 3 \pmod 4$. Then*

(i) $q|a$ and $q|b$;

(ii) q must appear to an even power in the prime factorization of n.

Proof. (i) Assume $q \nmid a$. Then there is a number s such that $sa \equiv 1 \pmod q$, and by multiplying the congruence $a^2+b^2 \equiv 0 \pmod q$ by s^2 we obtain $(sb)^2 = s^2b^2 \equiv -s^2a^2 \equiv -1 \pmod q$, that is -1 is a quadratic residue modulo q. By Theorem 11.9, this contradicts the assumption that $q \equiv 3 \pmod 4$. Thus q divides a and by symmetry, q also divides b.

(ii) Since $q \mid a$, $q \mid b$ and $n = a^2 + b^2$, it follows that $q^2 \mid n$. We can thus divide the equation $n = a^2 + b^2$ by q^2 to get $n/q^2 = (a/q)^2 + (b/q)^2$, that is the number $n_1 = n/q^2$ is a sum of squares. If $q \mid n_1$ then the preceding argument shows that $q^2 \mid n_1$. Proceeding in this way, we see that n must be divisible by an even number of factors of q. □

Lemma 17.3 *If m and n are each a sum of two squares, then their product mn is also a sum of two squares.*

Proof. Assume $m = a^2 + b^2$ and $n = c^2 + d^2$; then

$$mn = (a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2. \qquad \square$$

Theorem 17.4 *Write the canonical factorization of n as*

$$n = 2^{\alpha} \prod_{j=1}^{r} p_j^{\alpha_j} \prod_{j=1}^{s} q_j^{\beta_j},$$

where the p_js and q_js are distinct primes, $p_j \equiv 1 \pmod 4$ and $q_j \equiv 3 \pmod 4$. Then n can be expressed as a sum of two squares if and only if all the exponents β_j are even.

Proof. The sufficiency of the condition follows by repeated application of Lemma 17.3, because 2 and the primes p_j can be represented as a sum of two squares by Theorem 17.1, and the square q^2 (of any number q) is obviously the sum of two squares ($= q^2 + 0^2$). The necessity follows immediately from Lemma 17.2. □

Having decided which numbers are sums of two squares, the next natural task is to determine which numbers are representable as sums of three squares. Since the quadratic residues of 8 are 0, 1, and 4, the sum $a^2 + b^2 + c^2$ of three squares can never be congruent to 7 modulo 8. Hence, no integer of the form $8m + 7$ is representable as a sum of three squares. It is not difficult to extend the argument to show that no number of the form $4^k(8m + 7)$ is the sum of three squares. Conversely, any other number can be written as a sum of three squares. The proof, which is due to Gauss, is complicated and will not be given here. Thus the complete characterization is as follows.

Theorem 17.5 *A positive integer can be expressed as a sum of three squares if and only if it is not of the form $4^k(8m + 7)$.*

When it comes to the question of representing numbers as sums of four squares, we have the following simple answer.

Theorem 17.6 *Every positive integer is the sum of four squares.*

The first complete proof of this result was given by Lagrange in 1770. We will base our proof on the following two lemmas.

Lemma 17.7 *Suppose $3m$ is a sum of four squares. Then m is also a sum of four squares.*

Proof. Let $3m = a^2 + b^2 + c^2 + d^2$. Since every square is congruent to 0 or to 1 modulo 3, there are only two possibilities: either all four of the squares a^2, b^2, c^2, and d^2 are congruent to 0 modulo 3, or one of them, a^2 say, is congruent to 0 whereas the other three are congruent to 1. In the first case, $a \equiv b \equiv c \equiv d \equiv 0 \pmod 3$, and in the second case $a \equiv 0$, $b \equiv \pm 1$, $c \equiv \pm 1$, and $d \equiv \pm 1 \pmod 3$. In any case we have, by changing signs of b, c, and d if necessary, $b \equiv c \equiv d \pmod 3$. It follows that the four numbers $b + c + d$, $a + b - c$, $a + c - d$, and $a - b + d$ are all divisible by 3, and by expanding and simplifying, we find that

$$\left(\frac{b+c+d}{3}\right)^2+\left(\frac{a+b-c}{3}\right)^2+\left(\frac{a+c-d}{3}\right)^2+\left(\frac{a-b+d}{3}\right)^2$$
$$=\frac{3a^2+3b^2+3c^2+3d^2}{9}=\frac{9m}{9}=m,$$

that is m is the sum of four squares. □

Lemma 17.8 *Let n be a square-free integer, i.e. n is not divisible by p^2 for any prime p. Then there exist integers a and b such that $a^2+b^2 \equiv -1 \pmod{n}$.*

Proof. We will first prove that the result holds when n is a prime p. For $p=2$ and $p \equiv 1 \pmod 4$, it follows from Theorem 11.9 that there is a number a such that $a^2 \equiv -1 \pmod p$; thus we can take $b=0$. The case $p \equiv 3 \pmod 4$ remains, and we will give a proof for this case that works for all odd primes.

Let $m=(p-1)/2$, and define

$$A=\{0^2,1^2,2^2,\dots,m^2\} \quad\text{and}\quad B=\{-1-0^2,-1-1^2,-1-2^2,\dots,-1-m^2\}.$$

Any two elements of A are incongruent modulo p. For let $0 \le i \le m$. The congruence $x^2 \equiv i^2 \pmod p$ has exactly two roots $\pm i$ modulo p, and the only number x in the interval $[0,m]$ that is congruent to $\pm i$ is $x=i$, since $p-i>m$. Similarly, any two elements of B are incongruent modulo p. Thus, each of the sets A and B contains $m+1=(p+1)/2$ incongruent integers. Since their union contains $p+1$ integers, it follows that there is an element a^2 of A and an element $-1-b^2$ of B such that $a^2 \equiv -1-b^2 \pmod p$, that is $a^2+b^2 \equiv -1 \pmod p$.

Suppose now that $n=p_1p_2\cdots p_r$ is a product of distinct primes. For each prime p_j, choose a_j, b_j such that $a_j^2+b_j^2 \equiv -1 \pmod{p_j}$. By the Chinese Remainder Theorem, there are numbers a and b such that $a \equiv a_j \pmod{p_j}$ and $b \equiv b_j \pmod{p_j}$ for all j. It follows that $a^2+b^2 \equiv -1 \pmod{p_j}$ holds for all j, and this implies that $a^2+b^2 \equiv -1 \pmod n$. □

Proof of Theorem 17.6. Let n be a positive integer and write $n=k^2m$, where m is square-free. If m is the sum of four squares, say $m=a^2+b^2+c^2+d^2$, then $n=(ak)^2+(bk)^2+(ck)^2+(dk)^2$ is a sum of four squares, too.

Hence, we may as well assume that n is square-free. Then, by Lemma 17.8, there are two integers a and b such that $a^2+b^2 \equiv -1 \pmod n$. Consider all ordered pairs $(ax+by-z, bx-ay-w)$, where x, y, z, and w range over all integers from 0 to $[\sqrt{n}\,]$. There are $(1+[\sqrt{n}\,])^4>n^2$ choices for the quadruple (x,y,z,w) but only n^2 distinct ordered pairs modulo n. Hence there exist distinct ordered quadruples (x_1,y_1,z_1,w_1) and (x_2,y_2,z_2,w_2), with all entries lying in the interval from 0 to $[\sqrt{n}\,]$, such that $ax_1+by_1-z_1 \equiv ax_2+by_2-z_2 \pmod n$ and $bx_1-ay_1-w_1 \equiv bx_2-ay_2-w_2 \pmod n$.

Let $x=x_1-x_2$, $y=y_1-y_2$, $z=z_1-z_2$, and $w=w_1-w_2$. Then $ax+by \equiv z \pmod n$ and $bx-ay \equiv w \pmod n$. Therefore, $(ax+by)^2+(bx-ay)^2 \equiv z^2+w^2 \pmod n$. But $(ax+by)^2+(bx-ay)^2=(a^2+b^2)(x^2+y^2) \equiv -(x^2+y^2) \pmod n$. It follows that $x^2+y^2+z^2+w^2 \equiv 0 \pmod n$, that is $x^2+y^2+z^2+w^2=kn$ for some integer k. Clearly, $|x|$, $|y|$, $|z|$, and $|w|$ are all less than or equal to $[\sqrt{n}\,]$, and they are not all 0, since the ordered quadruples (x_1,y_1,z_1,w_1) and (x_2,y_2,z_2,w_2) are distinct. It follows that $0<x^2+y^2+z^2+w^2 \le 4[\sqrt{n}\,]^2<4n$, and thus $k=1$, 2, or 3.

If $k = 1$, we are done, while if $k = 3$, then $3n$ is a sum of four squares, and n itself is a sum of four squares by Lemma 17.7. Now suppose $k = 2$; since $2n$ is even, either zero, two, or four of the numbers x, y, z, and w are even. If exactly two are even, we may suppose that they are x and y. In each case, the numbers $x \pm y$ and $z \pm w$ are even, and hence $(x \pm y)/2$ and $(z \pm w)/2$ are integers. But

$$\begin{aligned}\left(\frac{x+y}{2}\right)^2 + \left(\frac{x-y}{2}\right)^2 + \left(\frac{z+w}{2}\right)^2 + \left(\frac{z-w}{2}\right)^2 &= \frac{2x^2 + 2y^2 + 2z^2 + 2w^2}{4} \\ &= \frac{4n}{4} = n,\end{aligned}$$

and hence n is a sum of four squares. This completes the proof. □

The same year as Lagrange proved the four squares theorem, Waring made the following broader conjecture: Every number is the sum of 4 squares, 9 cubes, 19 fourth powers, and in general, a sum of a fixed number of k^{th} powers. This conjecture was finally settled in 1909 when David Hilbert proved the following theorem.

Theorem 17.9 *For any $k \geq 2$, there is a smallest positive integer $s(k)$ such that every positive integer can be expressed as a sum of $s(k)$ nonnegative k^{th} powers.*

Hilbert's proof was a pure existence proof and gave no method of determining $s(k)$. A natural problem is to determine $s(k)$ explicitely. Lagrange's theorem and the fact that 7 is not the sum of three squares shows that $s(2) = 4$. To determine $s(k)$ for $k \geq 3$ turned out to be very difficult. It is now known that $s(3) = 9$, $s(4) = 19$, and $s(5) = 37$.

It is rather easy to show that the number $n = 2^k[(3/2)^k] - 1$ can not be expressed as a sum with fewer than $2^k + [(3/2)^k] - 2$ kth powers. Hence, $s(k) \geq 2^k + [(3/2)^k] - 2$ for every $k \geq 2$, and it has been conjectured that in fact $s(k) = 2^k + [(3/2)^k] - 2$ for every value of k. This equality is now known to hold for every $k \leq 471\,600\,000$ and also for all sufficiently large k, so the conjecture is very likely true.

18 Pythagorean Triples

In this section we will consider the problem of finding right triangles having sides of integral length, that is of finding integer solutions of the equation

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

This problem was considered in Egypt long before Pythagoras, but he is credited for having found a formula generating infinitely many solutions.

Definition 18.1 If x, y and z are positive integers satisfying $x^2 + y^2 = z^2$, then (x, y, z) is called a *Pythagorean triple*. If, in addition, the three numbers are pairwise relatively prime, then (x, y, z) is called a *primitive triple.*

Proposition 18.2 *If (x, y, z) is a Pythagorean triple, then $(x, y) = (x, z) = (y, z)$.*

Proof. Suppose $d \mid x$. If $d \mid y$, then $d^2 \mid (x^2+y^2)$, and hence $d \mid z$. Similarly, if $d \mid z$, then $d^2 \mid (z^2-x^2)$, so $d \mid y$. It follows that x and y have the same common divisors as x and z, and in particular the two pairs have the same greatest common divisor, i.e. $(x,y)=(x,z)$. Similarly, $(x,y)=(y,z)$. □

In particular, if (x,y,z) is a Pythagorean triple and two of the three numbers are relatively prime, then the three numbers are pairwise relatively prime, that is (x,y,z) is a primitive triple.

Theorem 18.3 *Every Pythagorean triple is a multiple of a primitive Pythagorean triple. Conversely, every multiple of a Pythagorean triple is a Pythagorean triple.*

Proof. If (x,y,z) is a Pythagorean triple and $d=(x,y)$ is the greatest common divisor of x and y, then clearly $(x/d,y/d,z/d)$ is a primitive Pythagorean triple by Proposition 18.2, and hence (x,y,z) is a multiple of a primitive triple. The converse is obvious. □

Proposition 18.4 *Suppose (x,y,z) is a primitive Pythagorean triple. Then x and y are of opposite parity, i.e. one of the numbers is odd and the other is even.*

Proof. Since $(x,y)=1$, both numbers cannot be even. Suppose x and y are both odd. Then $x^2 \equiv y^2 \equiv 1 \pmod 4$ and hence $z^2 \equiv 2 \pmod 4$, which is impossible. Therefore, x and y are of opposite parity. □

In order to find all Pythagorean triples it suffices, by Theorem 18.3, to find all primitive triples (x,y,z), and when we are looking for primitive triples, it is no restriction to assume that x is odd and y is even, because (x,y,z) is a Pythagorean triple if and only if (y,x,z) is so.

All primitive Pythagorean triples (x,y,z) with even y are generated as follows.

Theorem 18.5 *A triple (x,y,z), with y even, is a primitive Pythagorean triple if and only if it is of the form*

$$x=a^2-b^2, \quad y=2ab, \quad z=a^2+b^2,$$

where a and b are relatively prime positive integers of opposite parity with $a>b$.

Proof. If x, y, and z are defined as above, then $x^2+y^2=(a^2-b^2)^2+4(ab)^2=(a^2+b^2)^2=z^2$, that is (x,y,z) is a Pythagorean triple. To see that it is primitive, assume that $(x,z)>1$. Then, x and z have a common prime divisor p which must be odd, since x and z are both odd. Note that $z+x=2a^2$ and $z-x=2b^2$, and hence $p \mid 2a^2$ and $p \mid 2b^2$. Since p is odd, it follows that $p \mid a$ and $p \mid b$, which contradicts the assumption $(a,b)=1$. Therefore, $(x,z)=1$ and it follows from Proposition 18.2 that (x,y,z) is a primitive triple.

Conversely, assume (x,y,z) is a primitive Pythagorean triple with y even. Then x and z must be odd. Thus $z+x$ and $z-x$ are even. Let $r=(z+x)/2$ and $s=(z-x)/2$. Then $z=r+s$ and $x=r-s$, so any common divisor of r and s is a common divisor of x and z. Since $(x,z)=1$ it follows that $(r,s)=1$. Since $y^2=z^2-x^2=(z+x)(z-x)=4rs$, we have $(y/2)^2=rs$; but r and s

are relatively prime, so r and s must each be perfect squares. Write $r = a^2$ and $s = b^2$, where a and b are positive integers; then $a > b$ and $x = r - s = a^2 - b^2$, $z = r + s = a^2 + b^2$, and $y = 2\sqrt{rs} = 2ab$. It is clear that a and b are of opposite parity, because otherwise x would be even. Finally, $(a, b) = 1$, since any common divisor of a and b divides x and z, and $(x, z) = 1$. □

19 Fermat's Last Theorem

"I have discovered a truly remarkable proof, but the margin is too narrow to contain it." This famous note written by Fermat in 1637 in the margin of his copy of Diophantus's *Arithmetica* accompanied a statement by him which in modern terms reads as follows.

Theorem 19.1 (Fermat's Last Theorem) *The equation $x^n + y^n = z^n$ has no solution in nonzero integers if $n \geq 3$.*

It is very likely that Fermat had a proof for the case $n = 4$ and that he mistakenly thought that his argument could be generalized to cover the general case. For more than three and a half centuries a great many mathematicians tried, unsuccessfully to prove Fermat's conjecture, and during this search for a proof many new useful mathematical concepts and theories were invented. In the beginning of the nineteen nineties, the conjecture was known to be true for all n containing an odd prime factor less than 10^6. In June 1993, Andrew Wiles announced that he had a proof of Fermat's theorem, but his original proof turned out to contain some gaps; these were corrected one year later by Wiles and Richard Taylor. Fermat's *conjecture* was finally promoted to a *theorem.* The proof is very long and uses many deep results of algebraic geometry.

A popular account of the fascinating hunt for the solution of Fermat's conjecture is given in the book *Fermats gåta* by Simon Singh, MånPocket, 1999.

We will give a proof for the case $n = 4$ of Fermat's Last Theorem. In fact, we will prove the following slightly stronger result.

Theorem 19.2 *The equation $x^4 + y^4 = z^2$ has no solution in nonzero integers.*

Proof. Assume the contrary; then there is a solution with positive integers x, y and z, since any change of sign obviously still yields a solution. Let x, y, and z be a positive solution, where z is as small as possible. We will derive a contradiction by proving that there is another positive solution (x_1, y_1, z_1) with $z_1 < z$.

Suppose $(x, y) > 1$; then there is a prime p dividing both of x and y. It follows that $p^4 \mid (x^4 + y^4)$, that is $p^4 \mid z^2$, and hence $p^2 \mid z$. Thus $(x/p)^4 + (y/p)^4 = (z/p^2)^2$, and we have found a positive solution with a smaller value of z. This would contradict our original choice of (x, y, z), and we conclude that $(x, y) = 1$. It follows that $(x^2, y^2) = 1$ and hence (x^2, y^2, z) is a primitive Pythagorean triple. We may of course assume that x^2 is odd and y^2 is even, and by Theorem 18.5 there exist relatively prime numbers u and v such that

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

In particular, (x, v, u) is a primitive Pythagorean triple with x odd. Therefore, there exist relatively prime integers s and t such that

$$x = s^2 - t^2, \quad v = 2st, \quad u = s^2 + t^2.$$

Since $(s,t) = 1$ it follows from the last equality that u, s, and t are pairwise relatively prime. But $(y/2)^2 = uv/2 = ust$, so the product ust is a perfect square, and this implies that u, s, and t are all perfect squares. Hence, there exist positive integers a, b, and c such that $s = a^2$, $t = b^2$, and $u = c^2$. Since $u = s^2 + t^2$, it follows that $a^4 + b^4 = c^2$, i.e. (a,b,c) is a positive solution of our original equation. But this contradics our minimality assumption on z, because $c = \sqrt{u} \le u^2 < u^2 + v^2 = z$. This completes the proof. □

Corollary 19.3 *The equation $x^4 + y^4 = z^4$ has no solution in nonzero integers.*

Proof. If (x,y,z) is such a solution, then (x,y,z^2) is a solution of the equation in Theorem 19.2. This is a contradiction. □

20 Continued Fractions

In this and the following section, we will describe a technique for writing any real number as an iterated sequence of quotients. For example, the rational number 157/30 can be expanded as follows

$$\frac{157}{30} = 5 + \frac{7}{30} = 5 + \cfrac{1}{\cfrac{30}{7}} = 5 + \cfrac{1}{4 + \cfrac{2}{7}} = 5 + \cfrac{1}{4 + \cfrac{1}{\cfrac{7}{2}}} = 5 + \cfrac{1}{4 + \cfrac{1}{3 + \cfrac{1}{2}}}$$

and the last expression is called a finite continued fraction. To expand an irrational number, we need infinite continued fractions; for example

$$\begin{aligned}\sqrt{2} + 1 &= 2 + (\sqrt{2} - 1) = 2 + \frac{1}{\sqrt{2} + 1} = 2 + \cfrac{1}{2 + \cfrac{1}{\sqrt{2} + 1}} \\ &= 2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\sqrt{2} + 1}}} = 2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}}\end{aligned}$$

We start by giving a formal definition of finite continued fractions. Though we are mainly interested in continued fractions whose terms are integers, it is convenient with a more general definition.

Definition 20.1 Let $a_0, a_1, \ldots, a_n$ be real numbers, all positive except possibly

a_0. The expression

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_{n-1} + \cfrac{1}{a_n}}}}}$$

is called a *finite continued fraction* and is denoted by $\langle a_0, a_1, \dots, a_n\rangle$. The numbers a_k are called the *terms* or the *partial quotients* of the continued fraction.

If the reader does not like the dots in the above definition, the following recursive definition should satisfy her completely:

$$\begin{aligned}\langle a_0\rangle &= a_0\\ \langle a_0, a_1\rangle &= \langle a_0 + 1/a_1\rangle\\ \langle a_0, a_1, \dots, a_n\rangle &= \langle a_0, a_1, \dots, a_{n-2}, a_{n-1} + \frac{1}{a_n}\rangle, \quad \text{if } n \geq 2.\end{aligned}$$

The reason for assuming $a_k > 0$ for $k \geq 1$ in the above definition is that this guarantees that no division by zero will occur.

A continued fraction with $n+1$ terms can be compressed by viewing it as composed of two shorter continued fractions as follows, which is very useful in induction proofs.

Proposition 20.2 *Let $1 \leq k \leq n$. Then*

(i) $\langle a_0, a_1, \dots, a_n\rangle = \langle a_0, a_1, \dots, a_{k-1}, \langle a_k, a_{k+1}, \dots, a_n\rangle\rangle$, *and*

(ii) $\langle a_0, a_1, \dots, a_n\rangle = a_0 + \dfrac{1}{\langle a_1, \dots, a_n\rangle}$.

Proof. The formulas should be obvious from the very definition of continued fractions. For a formal proof of (i), use induction on the number m of terms in the innermost continued fraction $\langle a_k, a_{k+1}, \dots, a_n\rangle$. If $m = 1$, that is $k = n$, then $\langle a_n\rangle = a_n$, and there is nothing to prove. If $m = 2$, then $\langle a_{n-1}, a_n\rangle = a_{n-1} + 1/a_n$, and the identity (i) coincides with the recursive definition of $\langle a_0, a_1, \dots, a_n\rangle$.

Now suppose inductively that the identity (i) holds whenever the innermost continued fraction has m terms, and consider the case when $\langle a_k, a_{k+1}, \dots, a_n\rangle$ has $m+1$ terms. By the induction hypothesis applied twice and the case $m = 2$ applied once, we obtain

$$\begin{aligned}\langle a_0, a_1, \dots, a_n\rangle &= \langle a_0, a_1, \dots, a_{k-1}, a_k, \langle a_{k+1}, \dots, a_n\rangle\rangle\\ &= \langle a_0, a_1, \dots, a_{k-1}, \langle a_k, \langle a_{k+1}, \dots, a_n\rangle\rangle\rangle\\ &= \langle a_0, a_1, \dots, a_{k-1}, \langle a_k, a_{k+1}, \dots, a_n\rangle\rangle.\end{aligned}$$

This completes the induction argument.

(ii) is a special case of (i), obtained by taking $k = 1$. □

Infinite continued fractions are defined as limits of finite continued fractions in a straightforward way.

Definition 20.3 Let $(a_n)_{n=0}^{\infty}$ be a sequence of real numbers, all positive exept possibly a_0. The sequence $(\langle a_0, a_1, \dots, a_n\rangle)_{n=0}^{\infty}$ is called an *infinite continued fraction* and is denoted by $\langle a_0, a_1, a_2, \dots\rangle$. The infinite continued fraction is said to *converge* if the limit

$$\lim_{n\to\infty} \langle a_0, a_1, \dots, a_n\rangle$$

exits, and in that case the limit is also denoted by $\langle a_0, a_1, a_2, \dots\rangle$.

In order to determine the convergence of a given infinite continued fraction we need to consider the finite continued fractions $\langle a_0, a_1, \dots, a_n\rangle$ for increasing values of n. Suppose now that we have computed the value of $\langle a_0, a_1, \dots, a_n\rangle$ and want to compute the value of $\langle a_0, a_1, \dots, a_n, a_{n+1}\rangle$ without having to repeat the whole computation from scratch. The recursion formula (ii) in Proposition 20.2 will then be of no use, since it defines $\langle a_0, a_1, \dots, a_n, a_{n+1}\rangle$ in terms of a_0 and $\langle a_1, \dots, a_n, a_{n+1}\rangle$ and not in terms of $\langle a_0, a_1, \dots, a_n\rangle$ and a_{n+1}. Fortunately, there is an easy way to compute the continued fractions $\langle a_0, a_1, \dots, a_n\rangle$ in succession, and we will now describe this method.

Definition 20.4 Let $(a_n)_{n=0}^{N}$ be a finite ($N \in \mathbf{N}$) or infinite ($N = \infty$) sequence of real numbers, all positive except possibly a_0, and define two sequences $(p_n)_{n=-2}^{N}$ and $(q_n)_{n=-2}^{N}$ recursively as follows:

$$\begin{aligned} &p_{-2} = 0, \;\; p_{-1} = 1, \;\; p_n = a_n p_{n-1} + p_{n-2} \quad \text{if } n \geq 0,\\ &q_{-2} = 1, \;\; q_{-1} = 0, \;\; q_n = a_n q_{n-1} + q_{n-2} \quad \text{if } n \geq 0. \end{aligned}$$

The pair (p_n, q_n), as well as the quotient p_n/q_n (where $n \geq 0$), is called the *nth convergent* of the given sequence $(a_n)_{n=0}^{N}$ or, equivalently, of the corresponding continued fraction.

Obviously, $q_0 = 1$, and $q_n > 0$ for all $n \geq 0$. Thus, $(q_n)_{n=0}^{N}$ is a positive sequence.

The connection between continued fractions and convergents is given by the next theorem, which also contains some crucial identities.

Theorem 20.5 *Let $(a_n)_{n=0}^{N}$ be a sequence of real numbers, all positive except possibly a_0, with corresponding convergents (p_n, q_n), and write $c_n = p_n/q_n$. Then*

(i) $\langle a_0, a_1, \dots, a_n\rangle = c_n$, *for all $n \geq 0$;*

(ii) $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$, *if $n \geq -1$;*

(iii) $c_n - c_{n-1} = (-1)^{n-1}/q_{n-1}q_n$, *if $n \geq 1$;*

(iv) $p_n q_{n-2} - p_{n-2} q_n = (-1)^n a_n$, *if $n \geq 0$;*

(v) $c_n - c_{n-2} = (-1)^n a_n/q_{n-2}q_n$, *if $n \geq 2$.*

Proof. (i): The case $n = 0$ is trivial, because $c_0 = p_0/q_0 = a_0/1 = a_0$.

Suppose inductively that (i) holds for all continued fractions with n terms, and let $\langle a_0, a_1, \dots, a_n\rangle$ be a continued fraction with $n+1$ terms. Since

$$\langle a_0, a_1, \dots, a_n\rangle = \langle a_0, a_1, \dots, a_{n-2}, a_{n-1} + 1/a_n\rangle,$$

and since the latter continued fraction has n terms and its $(n-1)$st convergent equals $((a_{n-1}+1/a_n)p_{n-2}+p_{n-3}), (a_{n-1}+1/a_n)q_{n-2}+q_{n-3})$, we conclude that

$$\begin{aligned}\langle a_0, a_1, \dots, a_n\rangle &= \frac{(a_{n-1}+1/a_n)p_{n-2}+p_{n-3}}{(a_{n-1}+1/a_n)q_{n-2}+q_{n-3}} = \frac{a_n(a_{n-1}p_{n-2}+p_{n-3})+p_{n-2}}{a_n(a_{n-1}q_{n-2}+q_{n-3})+q_{n-2}}\\ &= \frac{a_np_{n-1}+p_{n-2}}{a_nq_{n-1}+q_{n-2}} = \frac{p_n}{q_n}.\end{aligned}$$

This completes the induction step.

(ii) Write $z_n = p_nq_{n-1} - p_{n-1}q_n$. Using the recursive definitions, we obtain

$$\begin{aligned}z_n &= p_nq_{n-1} - p_{n-1}q_n = (a_np_{n-1}+p_{n-2})q_{n-1} - p_{n-1}(a_nq_{n-1}+q_{n-2})\\ &= p_{n-2}q_{n-1} - p_{n-1}q_{n-2} = -z_{n-1},\end{aligned}$$

for $n \geq 0$, and it follows at once that $z_n = (-1)^{n-1}z_{-1}$. But $z_{-1} = 1$, since $p_{-1} = q_{-2} = 1$ and $p_{-2} = q_{-1} = 0$. Hence, $z_n = (-1)^{n-1}$, as required.

(iii) follows from (ii) upon division by $q_{n-1}q_n$, which is nonzero for $n \geq 1$.

(iv) Using the recursive definition of p_n and q_n and equality (ii), we obtain

$$\begin{aligned}p_nq_{n-2} - p_{n-2}q_n &= (a_np_{n-1}+p_{n-2})q_{n-2} - p_{n-2}(a_nq_{n-1}+q_{n-2})\\ &= a_n(p_{n-1}q_{n-2} - p_{n-2}q_{n-1}) = a_n(-1)^{n-2} = (-1)^na_n.\end{aligned}$$

(v) follows from (iv) upon division by $q_{n-2}q_n$. □

Example 1 We use Theorem 20.5 to evaluate the continued fraction

$$\langle -2, 5, 4, 3, 2, 1\rangle.$$

The computations are easily caried out by using the following table:

n	-2	-1	0	1	2	3	4	5
a_n			-2	5	4	3	2	1
p_n	0	1	-2	-9	-38	-123	-284	-407
q_n	1	0	1	5	21	68	157	225

The entries are computed according to the recursive formulas given in Definition 20.4. For example, to find $p_4 = a_4p_3 + p_2$, multiply $a_4 = 2$ by the last computed p-value p_3 $(= -123)$ and add the preceding term p_2 $(= -38)$ to obtain $p_4 = 2(-123) + (-38) = -284$. Finally, note that $\langle -2, 5, 4, 3, 2, 1\rangle = p_5/q_5 = -407/225$. The successive convergents are -2, $-9/5$, $-38/21$, $-123/68$, $-284/157$, and $-407/225$. □

Corollary 20.6 *Let $(a_n)_{n=0}^N$ be a finite or infinite sequence of real numbers, all positive except possibly a_0, with convergents $c_n = p_n/q_n$. The convergents c_{2i} with even indices form a strictly increasing sequence and the convergents c_{2j+1} with odd indices form a strictly decreasing sequence, and $c_{2i} < c_{2j+1}$, that is*

$$c_0 < c_2 < \cdots < c_{2i} < \cdots < c_{2j+1} < \cdots < c_3 < c_1.$$

Proof. We have $c_n - c_{n-2} = (-1)^na_n/q_nq_{n-2}$, by Theorem 20.5 (v). Hence, if $n \geq 2$ is even, then $c_n - c_{n-2} > 0$ and if $n \geq 3$ is odd, then $c_n - c_{n-2} < 0$. Finally, by Theorem 20.5 (iii), $c_{2k+1} - c_{2k} = 1/q_{2k}q_{2k+1} > 0$. Thus, if $i \geq j$, then $c_{2j} < c_{2i} < c_{2i+1}$ and $c_{2i} < c_{2i+1} < c_{2j+1}$. □

Example 2 In Example 1 we computed the continued fraction $\langle -2, 5, 4, 3, 2, 1\rangle$ and its successive convergents. It is easily verified that

$$-2 < -\frac{38}{21} < -\frac{284}{157} < -\frac{407}{225} < -\frac{123}{68} < -\frac{9}{5}$$

in accordance with Corollary 20.6. □

Let $(a_n)_{n=0}^{\infty}$ be a sequence of real numbers, all positive except possibly a_0, with convergents $c_n = p_n/q_n$. By Theorem 20.5, $c_n = \langle a_0, a_1, \dots, a_n\rangle$. Corollary 20.6 implies that the sequence $(c_{2k})_{k=0}^{\infty}$ of convergents with even indices is strictly increasing and bounded above by c_1. Hence, the limit $c' = \lim_{k\to\infty} c_{2k}$ exists. Similarly, the sequence $(c_{2k+1})_{k=0}^{\infty}$ is strictly decreasing and bounded below by c_0. Therefore, the limit $c'' = \lim_{k\to\infty} c_{2k+1}$ exists, too, and obviously $c_{2k} < c' \le c'' < c_{2k+1}$ for all k.

The limit

$$c = \lim_{n\to\infty} c_n = \lim_{n\to\infty} \langle a_0, a_1, \dots, a_n\rangle$$

exists if and only if $c' = c''$, that is if and only if $c_{2k+1} - c_{2k} \to 0$ as $k \to \infty$. By Theorem 20.5, $0 < c_{2k+1} - c_{2k} < 1/q_{2k}q_{2k+1}$. Therefore, $\lim_{n\to\infty} q_n = \infty$ is a sufficient condition for the existence of the limit c, i.e. for the convergence of the infinite continued fraction $\langle a_0, a_1, a_2, \dots\rangle$. Our next proposition gives a condition on the sequence $(a_n)_{n=0}^{\infty}$ which will guarantee that $q_n \to \infty$.

Proposition 20.7 *Let $(a_n)_{n=0}^{\infty}$ be a sequence with convergents (p_n, q_n) and assume that there is a constant $\alpha > 0$ such that $a_n \ge \alpha$ for all $n \ge 1$. Then $q_n \to \infty$ as $n \to \infty$. More precisely, there is a constant $r > 1$ and a positive constant C such that $q_n \ge Cr^n$ for all $n \ge 0$. The sequence $(q_n)_{n=1}^{\infty}$ is strictly increasing if $a_n \ge 1$ for all $n \ge 1$.*

Proof. By assumption, $q_n = a_nq_{n-1} + q_{n-2} \ge \alpha q_{n-1} + q_{n-2}$ for all $n \ge 1$. Let r denote the positive root of the quadratic equation $x^2 = \alpha x + 1$, that is $r = \alpha/2 + \sqrt{1 + \alpha^2/4}$, and let C denote the smallest of the two numbers 1 and a_1/r. Then $q_0 = 1 \ge Cr^0$ and $q_1 = a_1 \ge Cr^1$. We claim that $q_n \ge Cr^n$ for all $n \ge 0$. This follows by induction, because if $q_k \ge Cr^k$ for $0 \le k \le n-1$, then $q_n \ge \alpha Cr^{n-1} + Cr^{n-2} = Cr^{n-2}(\alpha r + 1) = Cr^{n-2} \cdot r^2 = Cr^n$. Obviously, $r > 1$, so it follows that $q_n \to \infty$ as $n \to \infty$.

If $a_n \ge 1$ for all $n \ge 1$, then $q_n = a_nq_{n-1} + q_{n-2} \ge q_{n-1} + q_{n-2} > q_{n-1}$ for all $n \ge 2$, which means that the sequence $(q_n)_{n=1}^{\infty}$ is strictly increasing. □

Definition 20.8 A sequence $(a_n)_{n=0}^{\infty}$ of real numbers will be called *admissible* if there is a positive constant α such that $a_n \ge \alpha$ for all $n \ge 1$.

A sequence $(a_n)_{n=0}^{\infty}$ consisting of integers, all positive except possibly a_0, is obviously admissible with $\alpha = 1$. In particular, for such sequences the corresponding sequence $(q_n)_{n=1}^{\infty}$ is strictly increasing and unbounded.

The discussion preceeding Proposition 20.7 may now be summarized as follows:

Theorem 20.9 *Let $(a_n)_{n=0}^{\infty}$ be an admissible sequence with convergents $c_n = p_n/q_n$. The infinite continued fraction $\xi = \langle a_0, a_1, a_2, \dots\rangle$ is then convergent,* *and it satisfies*

(1) $$c_{2n} < \xi < c_{2n+1} \qquad \text{and}$$

(2) $$\frac{a_{n+2}}{q_n q_{n+2}} < |\xi - c_n| < \frac{1}{q_n q_{n+1}}$$

for all $n \geq 0$.

Proof. It only remains to prove (2). By (1), for each $n \geq 0$, the number ξ belongs to the interval with endpoints c_n and c_{n+1}, and hence

$$|\xi - c_n| < |c_{n+1} - c_n| = \frac{1}{q_n q_{n+1}},$$

where the last equality follows from Theorem 20.5 (iii).

Moreover, the number c_{n+2} lies strictly between the numbers c_n and ξ. Consequently,

$$|\xi - c_n| > |c_{n+2} - c_n| = \frac{a_{n+2}}{q_n q_{n+2}},$$

where the last equality is a consequence of Theorem 20.5 (v). This completes the proof of the theorem. □

Is is often useful to regard an infinite continued fractions as a finite continued fraction with an infinite continued fraction as its last term (cf. Proposition 20.2).

Theorem 20.10 *Let $(a_n)_{n=0}^{\infty}$ be an admissible sequence of real numbers, let k be a positive integer, and write $\xi_k = \langle a_k, a_{k+1}, a_{k+2}, \dots \rangle$. Then*

$$\langle a_0, a_1, a_2, \dots \rangle = \langle a_0, a_1, \dots, a_{k-1}, \xi_k \rangle.$$

Proof. Write $\xi = \langle a_0, a_1, a_2, \dots \rangle = \lim_{n\to\infty} \langle a_0, a_1, \dots, a_n \rangle$. By letting $n \to \infty$ in the relation

$$\langle a_0, a_1, \dots, a_n \rangle = a_0 + \frac{1}{\langle a_1, a_2, \dots, a_n \rangle},$$

we obtain

$$\xi = a_0 + 1/\xi_1 = \langle a_0, \xi_1 \rangle.$$

(Note that $\xi_1 > a_1 > 0$.) This proves the case $k = 1$. In particular, we have $\xi_k = \langle a_k, \xi_{k+1} \rangle$ for each k.

The general case now follows by induction. Assume that the theorem holds for a certain $k \geq 1$; then

$$\begin{aligned}\xi &= \langle a_0, a_1, \dots, a_{k-1}, \xi_k \rangle = \langle a_0, a_1, \dots, a_{k-1}, \langle a_k, \xi_{k+1} \rangle \rangle \\ &= \langle a_0, a_1, \dots, a_{k-1}, a_k, \xi_{k+1} \rangle,\end{aligned}$$

where the last equality follows from Proposition 20.2. This completes the induction step.

□

Example 3 Let us use Theorem 20.10 to compute the periodic infinite continued fraction $\xi = \langle 1, 2, 3, 1, 2, 3, \dots \rangle = \langle \overline{1, 2, 3} \rangle$, where the bar over 1, 2, 3 indicates that this block of integers is repeated indefinitely. By periodicity, $\xi = \langle 1, 2, 3, \xi_3 \rangle$ with $\xi_3 = \xi$, that is $\xi = \langle 1, 2, 3, \xi \rangle$. To compute the value of this finite continued fraction we use convergents, which are computed in the following table:

n	-2	-1	0	1	2	3
a_n			1	2	3	ξ
p_n	0	1	1	3	10	$10\xi+3$
q_n	1	0	1	2	7	$7\xi+2$

It follows that

$$\xi = \langle 1, 2, 3, \xi \rangle = \frac{p_3}{q_3} = \frac{10\xi + 3}{7\xi + 2}.$$

Solving for ξ we obtain the quadratic equation $7\xi^2 - 8\xi - 3 = 0$ with the roots $(4 \pm \sqrt{37})/7$. Since $\xi > 0$, we conclude that $\xi = (4 + \sqrt{37})/7$. □

Example 4 To compute the infinite periodic continued fraction

$$\eta = \langle 0, 1, \overline{1, 2, 3}\,\rangle,$$

we start by writing $\eta = \langle 0, 1, \xi \rangle$, where $\xi = \langle \overline{1, 2, 3}\,\rangle$, and $\eta = 0 + 1/(1 + 1/\xi) = \xi/(\xi + 1)$. The value of ξ was computed in the previous example. Inserting $\xi = (4 + \sqrt{37})/7$ into the expression for η, we obtain $\eta = (1 + \sqrt{37})/12$. □

Example 5 $\xi = \langle 1, 1, 1, \dots \rangle = \langle \overline{1}\,\rangle$ is the simplest possible infinite continued fraction. We will see later that this number plays a special role when it comes to approximation of irrational numbers by rational numbers. Since $\xi = \langle 1, \xi \rangle$, ξ satisfies the equation $\xi = 1 + 1/\xi$, that is $\xi^2 = \xi + 1$. This quadratic equation has the roots $(1 \pm \sqrt{5})/2$, and since ξ is positive we conclude that $\langle 1, 1, 1, \dots \rangle = (1 + \sqrt{5})/2$. □

21 Simple Continued Fractions

Definition 21.1 A finite or infinite continued fraction is called *simple*, if all its terms are integers.

We recall that all terms of a continued fraction, except possibly the first term a_0, are by default supposed to be positive. In particular, all terms of a simple continued fraction, except possibly the first one, are positive integers. This means that the terms of an infinite simple continued fraction form an admissible sequence (with $\alpha = 1$), so there are no convergence problems: The infinite simple continued fractions are automatically convergent.

The value of a finite simple continued fraction is a rational number. Of course, this follows easily from the recursive definition of finite continued fractions, but we can also deduce it from the following theorem.

Theorem 21.2 *Let (p_n, q_n) be the nth convergent of a finite or infinite simple continued fraction. The numbers p_n and q_n are then relatively prime integers for each n. Thus, the fractions $c_n = p_n/q_n$, $n \geq 0$, are rational numbers in reduced form.*

Proof. It follows at once from their defining recursive relations that p_n and q_n are integers when the terms a_n of the continued fraction are integers. Relative primeness is a consequence of the identity $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$. □

Corollary 21.3 *Every finite simple continued fraction $\langle a_0, a_1, \dots, a_n\rangle$ is a rational number.*

Proof. Because $\langle a_0, a_1, \dots, a_n\rangle = p_n/q_n$. □

Theorem 21.4 *The value of an infinite simple continued fraction is irrational.*

Proof. Assume $\xi = \langle a_0, a_1, a_2, \dots\rangle$ is rational and write $\xi = a/b$ with integers a and b. By Theorem 20.9, $0 < |a/b - p_n/q_n| < 1/q_nq_{n+1}$. Multiplying by bq_n we obtain

$$0 < |aq_n - bp_n| < \frac{b}{q_{n+1}}.$$

By choosing n so large that $b/q_{n+1} < 1$, which is possible since $q_{n+1} \to \infty$, we obtain the inequality $0 < |aq_n - bp_n| < 1$. Since $aq_n - bp_n$ is an integer, this is a contradiction. □

Theorem 21.5 *Every real number can be expressed as a simple continued fraction. The fraction is finite if and only if the real number is rational.*

Proof. Let ξ be a real number, and define $a_0 = [\xi]$. We use the following recursive algorithm to define a (possibly empty) finite or infinite sequence a_1, a_2, ... of positive integers.

Step 0: If $\xi = a_0$, then $\xi = \langle a_0\rangle$, and the algorithm stops. Otherwise, $0 < \xi - a_0 < 1$, and we define $\xi_1 = 1/(\xi - a_0)$, noting that $\xi_1 > 1$ and that $\xi = \langle a_0, \xi_1\rangle$. We then proceed to step 1.

Step k for $k = 1, 2, \dots$: Suppose the positive integers $a_1, a_2, \dots, a_{k-1}$ and the real number $\xi_k > 1$ have been defined and that $\xi = \langle a_0, a_1, \dots, a_{k-1}, \xi_k\rangle$. Define $a_k = [\xi_k]$.

If $\xi_k = a_k$, then $\xi = \langle a_0, a_1, \dots, a_k\rangle$ and the algorithm stops. Otherwise, define $\xi_{k+1} = 1/(\xi_k - a_k)$, which is then a real number > 1, note that $\xi_k = \langle a_k, \xi_{k+1}\rangle$, and $\xi = \langle a_0, a_1, \dots, a_k, \xi_{k+1}\rangle$, and proceed to step $k + 1$.

If the algorithm stops, then ξ is a finite simple continued fraction. Otherwise it defines an infinite sequence $(a_n)_{n=0}^{\infty}$. Define $\eta = \langle a_0, a_1, a_2, \dots\rangle$, and let $c_n = p_n/q_n$ denote the nth convergent of the infinite continued fraction η. Since $\xi = \langle a_0, a_1, \dots, a_n, \xi_{n+1}\rangle$, the numbers c_{n-1} and c_n are also convergents of ξ. It therefore follows from Theorem 20.9 and Corollary 20.6 that ξ and η both lie between the numbers c_{n-1} and c_n. Hence,

$$|\xi - \eta| < |c_n - c_{n-1}| = \frac{1}{q_{n-1}q_n}.$$

Since $q_n \to \infty$ as $n \to \infty$, we conclude that $\xi = \eta = \langle a_0, a_1, a_2, \dots\rangle$. □

Example 1 Using the algorithm of Theorem 21.5 we compute the continued fraction expansion of $\sqrt{2}$ as follows:

$$\begin{aligned} a_0 &= [\sqrt{2}] = 1, & \xi_1 &= 1/(\xi - a_0) \ = 1/(\sqrt{2} - 1) = \sqrt{2} + 1; \\ a_1 &= [\xi_1] = 2, & \xi_2 &= 1/(\xi_1 - a_1) = 1/(\sqrt{2} - 1) = \sqrt{2} + 1 = \xi_1. \end{aligned}$$

Since $\xi_2 = \xi_1$, we conclude that $a_2 = a_1$ and $\xi_3 = \xi_2$, etc. Hence, $a_n = a_1 = 2$ for all $n \geq 1$. Therefore,

$$\sqrt{2} = \langle 1, 2, 2, 2, \dots\rangle = \langle 1, \overline{2}\,\rangle.$$

□

Since $k = k - 1 + 1/1$, any integer k can be written in two ways as a simple continued fraction: $k = \langle k\rangle = \langle k-1, 1\rangle$. It follows that every rational number has at least two different representations as finite simple continued fractions, because if $\langle a_0, a_1, \dots, a_n\rangle$ is a representation with $a_n > 1$, then

$$\langle a_0, a_1, \dots, a_n - 1, 1\rangle$$

is a different representation ending in 1. Conversely, if $\langle a_0, a_1, \dots, a_n, 1\rangle$ is a continued fraction ending in 1, then $\langle a_0, a_1, \dots, a_n, 1\rangle = \langle a_0, a_1, \dots, a_n + 1\rangle$. However, these are the only different ways to represent a rational number as a simple continued fraction. For the proof of this fact we shall need the following lemma.

Lemma 21.6 *Let a_0, b_0 be integers, let $a_1, a_2, \dots, a_n$ be positive integers, and let x, y be two real numbers ≥ 1. Then*

$$(1) \qquad b_0 = \langle a_0, x\rangle \Rightarrow x = 1 \text{ and } a_0 = b_0 - 1$$

$$(2) \qquad a_0 \neq b_0 \Rightarrow \langle a_0, x\rangle \neq \langle b_0, y\rangle$$

$$(3) \qquad \langle a_0, a_1, \dots, a_n, x\rangle = \langle a_0, a_1, \dots, a_n, y\rangle \Rightarrow x = y$$

Proof. (1): Suppose $b_0 = \langle a_0, x\rangle$ and $x > 1$. Then

$$a_0 < \langle a_0, x\rangle = b_0 = a_0 + 1/x < a_0 + 1,$$

which is a contradiction, since b_0 is an integer. Hence, $x = 1$, and $b_0 = a_0 + 1$.

(2): Suppose $a_0 < b_0$; then $\langle a_0, x\rangle = a_0 + 1/x \leq a_0 + 1 \leq b_0 < \langle b_0, y\rangle$.

(3): If $\langle a_0, x\rangle = \langle a_0, y\rangle$, then obviously $x = y$, so the assertion holds when $n = 0$. Now suppose that the implication is true with n replaced by $n - 1$, and assume that $\langle a_0, a_1, \dots, a_n, x\rangle = \langle a_0, a_1, \dots, a_n, y\rangle$. Since

$$\langle a_0, a_1, \dots, a_n, x\rangle = \langle a_0, a_1, \dots, a_{n-1}, \langle a_n, x\rangle\rangle,$$

and the other continued fraction may be shortened analogously, it follows from our induction hypothesis that first $\langle a_n, x\rangle = \langle a_n, y\rangle$, and then $x = y$. □

Theorem 21.7 *Each integer k has exactly two representations as simple continued fractions, viz. $\langle k\rangle$ and $\langle k-1, 1\rangle$. Each nonintegral rational number has exactly two representations as simple continued fractions, and they are of the form $\langle a_0, a_1, \dots, a_n\rangle$ and $\langle a_0, a_1, \dots, a_n - 1, 1\rangle$, where $n \geq 1$ and $a_n > 1$. Each irrational number has a unique representation as an infinite simple continued fraction.*

Proof. We have already noted that each rational number has two different representations as finite simple continued fractions, and that each irrational has one representation as infinite simple continued fraction, so it suffices to prove that these representations are the only one.

First assume that k is an integer and $k = \langle a_0, a_1, \dots, a_n\rangle = \langle a_0, \langle a_1, \dots, a_n\rangle\rangle$, with $n \geq 1$. It then follows from Lemma 21.6 (1) that $a_0 = k - 1$ and $x = \langle a_1, \dots, a_n\rangle = 1$. If $n \geq 2$, then $x > a_1 \geq 1$, which is impossible. Hence $n = 1$ and $a_1 = 1$, that is $k = \langle k\rangle$ and $k = \langle k-1, 1\rangle$ are the only representations of k as a simple continued fraction.

Let now $\langle a_0, a_1, \dots, a_n\rangle = \langle b_0, b_1, \dots, b_m\rangle$ be two representations of a non-integral rational number, and assume that $m \geq n$. Suppose there is an index $k < n$ such that $a_k \neq b_k$, and let k denote the least such index. Writing the continued fraction $\langle a_0, a_1, \dots, a_n\rangle$ as $\langle a_0, \dots, a_{k-1}, \langle a_k, \dots, a_n\rangle\rangle$ and similarly for $\langle b_0, b_1, \dots, b_m\rangle$, we then conclude, using Lemma 21.6 (3), that $\langle a_k, \dots, a_n\rangle = \langle b_k, \dots, b_m\rangle$, or equivalently that

$$\langle a_k, \langle a_{k+1}, \dots, a_n\rangle\rangle = \langle b_k, \langle b_{k+1}, \dots, b_m\rangle\rangle.$$

However, this is impossible because of (2). Thus, $a_k = b_k$ for all $k < n$ and we conclude using (3) that $a_n = \langle b_n, \dots, b_m\rangle$. But a_n is an integer, and we already know that there are only two possible representations of integers as simple continued fractions; either $m = n$ and $a_n = b_n$, or $m = n+1$, $b_n = a_n - 1$ and $b_{n+1} = 1$.

Let finally ξ be an irrational number, and suppose

$$\xi = \langle a_0, a_1, a_2, \dots\rangle = \langle b_0, b_1, b_2, \dots\rangle$$

are two different representations of ξ. Then there is a first index k such that $a_k \neq b_k$, and we conclude from (3) that $\langle a_k, a_{k+1}, a_{k+2}, \dots\rangle = \langle b_k, b_{k+1}, b_{k+2}, \dots\rangle$. However, this contradicts (2). □

22 Rational Approximations to Irrational Numbers

Let ξ be an irrational number. Given a positive integer b, we let a denote the integer that is nearest to $b\xi$, that is a is either equal to $[b\xi]$ or $[b\xi] + 1$. Then $|b\xi - a| < 1/2$, and dividing by b we obtain

$$|\xi - \frac{a}{b}| < \frac{1}{2b}\,.$$

Since b can be taken arbitrarily large, it follows that ξ can be approximated arbitrarily well by rational numbers a/b. This is sometimes expressed by saying that the rational numbers are dense in the set of real numbers.

The inequality above gives a bound on $|\xi - a/b|$ in terms of the denominator b. A natural question now arises: How well can we approximate ξ with rational numbers a/b given that there is a prescribed upper bound on the size of b?

It follows from Theorem 20.9, that $c_n = p_n/q_n$, the nth convergent of the expansion of ξ as an infinite simple continued fraction, satisfies

$$|\xi - \frac{p_n}{q_n}| < \frac{1}{q_n^2}\,.$$

Thus, the approximation error for convergents p_n/q_n is considerably smaller than what can be expected for general rational numbers a/b.

We will prove that we can do a bit better; there are infinitely many rational numbers a/b such that $|\xi - a/b| < 1/\sqrt{5}\,b^2$ (Theorem 22.7). This result is sharp in the sense that the constant $\sqrt{5}$ can not be replaced by any bigger constant (Theorem 22.8). The rational numbers a/b satisfying this inequality have to be convergents, because we will also prove that if $|\xi - a/b| < 1/2b^2$, then necessarily a/b is a convergent (Theorem 22.4).

Thus, continued fractions and convergents play a very important role in the theory of rational approximation.

In the sequel, we will use both $|\xi - a/b|$ and $|b\xi - a|$ as measures of how well a/b approximates ξ. For convenience, we first reformulate inequality (2) of Theorem 20.9 slightly:

Let p_n/q_n be the nth convergent of the irrational number $\xi = \langle a_0, a_1, a_2, \dots \rangle$. Then

$$\frac{a_{n+2}}{q_{n+2}} < |q_n\xi - p_n| < \frac{1}{q_{n+1}}. \tag{1}$$

Inequality (1) is of course obtained from Theorem 20.9 (2) by multiplying through by q_n.

Theorem 22.1 *Let (p_n, q_n) denote the nth convergent of the simple continued fraction expansion of the irrational number ξ. Then*

$$\left|\xi - \frac{p_n}{q_n}\right| < \left|\xi - \frac{p_{n-1}}{q_{n-1}}\right| \qquad \textit{and} \tag{2}$$

$$|q_n\xi - p_n| < |q_{n-1}\xi - p_{n-1}| \tag{3}$$

for every $n \geq 1$.

Proof. We start by proving the second stronger inequality. Suppose

$$\xi = \langle a_0, a_1, a_2, \dots \rangle.$$

Using inequality (1) twice, we obtain

$$|q_{n-1}\xi - p_{n-1}| > a_{n+1}/q_{n+1} \geq 1/q_{n+1} > |q_n\xi - p_n|,$$

which proves (3).

Inequality (2) is an easy consequence of (3) and the fact that $q_n \geq q_{n-1}$ for $n \geq 1$:

$$\left|\xi - \frac{p_n}{q_n}\right| = \frac{1}{q_n}\left|q_n\xi - p_n\right| < \frac{1}{q_n}\left|q_{n-1}\xi - p_{n-1}\right| \leq \frac{1}{q_{n-1}}\left|q_{n-1}\xi - p_{n-1}\right| = \left|\xi - \frac{p_{n-1}}{q_{n-1}}\right|.$$

□

Convergents have some interesting extremal properties which we are now going to study. The proofs of these will rely on the following simple observation:

If $r_1 = a_1/b_1$ and $r_2 = a_2/b_2$ are two rational numbers with positive denominators and $r_1 \neq r_2$, then

$$|r_1 - r_2| \geq \frac{1}{b_1 b_2}.$$

This is obvious because $r_1 - r_2 = (a_1b_2 - a_2b_1)/b_1b_2$ and the numerator $a_1b_2 - a_2b_1$ is a nonzero integer and its absolute value is thus at least one.

On the other hand, if $c_n = p_n/q_n$ and $c_{n+1} = p_{n+1}/q_{n+1}$ are two consecutive convergents of a number ξ, then $|c_{n+1} - c_n| = 1/q_{n+1}q_n$, according to Theorem 20.5.

In the formulation of the following theorems we will use the notation

$$\min_{1 \leq t \leq B} f(s, t)$$

to denote the minimum of $f(s, t)$ when s ranges over all integers and t ranges over all integers in the interval $[1, B]$.

Theorem 22.2 *Let ξ be an irrational number, let B be a positive integer, and consider the value of $|t\xi - s|$ for all integers t in the interval $[1, B]$ and all integers s. Suppose that the minimum is achieved for $s = a$ and $t = b$, that is*

$$|b\xi - a| = \min_{1\le t\le B} |t\xi - s|.$$

Then, a and b are relatively prime, and (a, b) is a convergent of the simple continued fraction expansion of ξ.

Proof. Let d be a common divisor of a and b, and suppose $d > 1$. Write $a' = a/d$ and $b' = b/d$; then $1 \le b' \le B$ and $|b'\xi - a'| = |b\xi - a|/d < |b\xi - a|$, which contradicts the choice of a and b. Thus $d = 1$, and a and b are relatively prime.

Let now $c_n = p_n/q_n$ denote the nth convergent of ξ and write $r = a/b$. We shall prove that $r = c_n$ for some n; since both fractions a/b and p_n/q_n are in reduced form, it will then follow that $a = p_n$ and $b = q_n$.

First suppose $r < c_0$. Since $c_0 < \xi$, $|\xi - r| > |c_0 - r| \ge 1/bq_0$. Multiplying through by b, we obtain $|b\xi - a| = b|\xi - r| > 1/q_0 \ge 1/q_1 > |q_0\xi - p_0|$. Since $q_0 = 1 \le b$, this contradict the minimality assumption on a and b. Thus, $r \ge c_0$.

Next suppose that $r > c_1$. Since $c_1 > \xi$, $|\xi - r| > |c_1 - r| \ge 1/bq_1$. Multiplying through by b, we again obtain $|b\xi - a| > 1/q_1 > |q_0\xi - p_0|$, which is impossible.

Hence, $c_0 \le r \le c_1$. Since (c_{2k}) is an increasing sequence and (c_{2k+1}) is a decreasing sequence, both with limit ξ, the rational number r lies between c_{n-1} and c_{n+1} for some integer n. If r is either c_{n-1} or c_{n+1}, we are finished. Otherwise, note that these two convergents are on the same side of ξ, and that c_n lies on the other side. It follows that $|r - c_{n-1}| < |c_n - c_{n-1}|$. Since the left hand side of this inequality is $\ge 1/bq_{n-1}$ and the right hand side equals $1/q_nq_{n-1}$, we conclude that $1/bq_{n-1} < 1/q_nq_{n-1}$, that is $q_n < b$.

We also have $|\xi - r| > |c_{n+1} - r| \ge 1/bq_{n+1}$, and multiplying both sides by b we obtain $|b\xi - a| > 1/q_{n+1} > |q_n\xi - p_n|$. Since $q_n < b$, this contradicts the assumption that $|t\xi - s|$ is minimized when $t = b$ and $s = a$. The proof is now complete. □

Theorem 22.2 tells us that if a/b is the “best” approximation to ξ in the sense that $|b\xi - a|$ cannot be made smaller by replacing a/b with any other rational number s/t with $1 \le t \le b$, then a/b is necessarily a convergent of ξ. By combining this result with Theorem 22.1 we obtain the following more precise information:

Theorem 22.3 *Let ξ be irrational with convergents (p_n, q_n). Then*

$$|q_n\xi - p_n| = \min_{1\le t<q_{n+1}} |t\xi - s| \tag{4}$$

$$\left|\xi - \frac{p_n}{q_n}\right| = \min_{1\le t\le q_n} \left|\xi - \frac{s}{t}\right|. \tag{5}$$

Proof. By Theorem 22.2, there is a convergent (p_m, q_m) of ξ such that

$$|q_m\xi - p_m| = \min_{1\le t<q_{n+1}} |t\xi - s|.$$

Since $q_k \ge q_{n+1}$ for all $k \ge n+1$ and since $|q_k\xi - p_k|$ decreases when k increases, it follows that $m = n$. This proves (4).

To prove (5), assume that $1 \le t \le q_n$ and let s be arbitrary. Using (4), we obtain

$$\big|\xi - \frac{p_n}{q_n}\big| = \frac{1}{q_n}\big|q_n\xi - p_n\big| \le \frac{1}{q_n}\big|t\xi - s\big| = \frac{t}{q_n}\big|\xi - \frac{s}{t}\big| \le \frac{q_n}{q_n}\big|\xi - \frac{s}{t}\big| = \big|\xi - \frac{s}{t}\big|.$$

Hence,

$$\big|\xi - \frac{p_n}{q_n}\big| = \min_{1\le t\le q_n}\big|\xi - \frac{s}{t}\big|.$$

□

Remark. If $q_{n+1} > 2q_n$, the following stronger result holds:

$$\big|\xi - \frac{p_n}{q_n}\big| = \min_{1\le t\le q_{n+1}/2}\big|\xi - \frac{s}{t}\big|.$$

Proof. Assume $|\xi - s/t| < |\xi - p_n/q_n|$. Using the triangle inequality and Theorem 20.9, we then obtain

$$1/tq_n \le |s/t - p_n/q_n| \le |s/t - \xi| + |\xi - p_n/q_n| < 2|\xi - p_n/q_n| < 2/q_nq_{n+1}.$$

It follows that $t > q_{n+1}/2$. □

Example 1 Using the algorithm in Theorem 21.5 and the decimal expansion of π, one easily finds that $\pi = \langle 3, 7, 15, 1, 292, 1, 1, 1, 2, \dots\rangle$. To compute the first five convergents we use the following table:

n	-2	-1	0	1	2	3	4
a_n		3	7	15	1	292	1
p_n	0	1	3	22	333	355	103 993
q_n	1	0	1	7	106	113	33 102

The convergent p_1/q_1 is the familiar approximation 22/7, first given by Archimedes; it is the best approximation among all rationals with denominator not exceeding 7. The approximation 355/113 is remarkably accurate; by Theorem 20.9,

$$\left|\pi - \frac{355}{113}\right| < \frac{1}{113 \cdot 33102} < 3 \cdot 10^{-7}.$$

Using the remark following Theorem 22.3 we see that in order to obtain a better approximation we need a rational a/b with $b > 33102/2 = 16\,551$. In fact, 355/113 is the best approximation to π among all rationals with denominators not exceeding 16 603. □

In Theorems 22.2 and 22.3, the convergents appear as solutions to certain minimum problems. Therefore, it should not come as a surprise that "best" rational approximations to irrationals have to be convergents. The following theorem gives a precise meaning to this statement.

Theorem 22.4 *Let ξ be irrational, and let a and b be integers with b positive. If*

$$\left|\xi - \frac{a}{b}\right| < \frac{1}{2b^2},$$

then a/b equals one of the convergents of the simple continued fraction expansion of ξ.

Remark. The fraction a/b is not necessarily reduced.

Proof. If the fraction a/b is not reduced, then the reduced fraction a'/b' obviously satisfies the same inequality. Therefore, we may as well assume from the beginning that a/b is reduced, i.e. that a and b are relatively prime. Under this assumption we will prove that the inequality

$$|b\xi - a| \leq |t\xi - s| \tag{6}$$

holds for all integers s and t with $1 \leq t \leq b$. It then follows from Theorem 22.2 that a/b is a convergent.

Assume (6) is false for some integers s and t with $1 \leq t \leq b$. Then

$$|t\xi - s| < |b\xi - a|, \tag{7}$$

and it follows that

$$\left|\xi - \frac{s}{t}\right| < \frac{1}{t}\,|b\xi - a| = \frac{b}{t}\left|\xi - \frac{a}{b}\right| < \frac{b}{t}\cdot\frac{1}{2b^2} = \frac{1}{2bt}.$$

Using the triangle inequality, we thus obtain

$$\left|\frac{a}{b} - \frac{s}{t}\right| \leq \left|\frac{a}{b} - \xi\right| + \left|\xi - \frac{s}{t}\right| < \frac{1}{2b^2} + \frac{1}{2bt} = \frac{1}{2bt}\Bigl(1 + \frac{t}{b}\Bigr) \leq \frac{1}{bt}.$$

Multiply through by bt; this results in $|at - bs| < 1$, and since $at - bs$ is an integer we conclude that $at - bs = 0$, that is $a/b = s/t$. Since the fraction a/b is reduced, $t \geq b$. But $t \leq b$, and hence $t = b$ and $s = a$. This is a contradiction because of (7). □

It remains to prove that there are fractions a/b that satisfy the inequality in Theorem 22.4. By Theorem 20.9, the convergents p/q of an irrational number ξ satisfy the inequality $|\xi - p/q| < 1/q^2$. The following theorem shows that of any two successive convergents at least one will satisfy the stronger inequality in Theorem 22.4.

Theorem 22.5 *Of any two successive convergents of the simple continued fraction expansion of the irrational number ξ, at least one convergent p/q will satisfy the inequality*

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{2q^2}\,.$$

Proof. Assume the theorem is false. Then there are two successive convergents $c_n = p_n/q_n$ and $c_{n+1} = p_{n+1}/q_{n+1}$ such that $|\xi - c_n| > 1/2q_n^2$ and $|\xi - c_{n+1}| > 1/2q_{n+1}^2$. (The inequalities are strict since ξ is irrational.) Since the two convergents are on opposite sides of ξ, it follows that

$$\frac{1}{q_nq_{n+1}} = |c_{n+1} - c_n| = |c_{n+1} - \xi| + |\xi - c_n| > \frac{1}{2q_{n+1}^2} + \frac{1}{2q_n^2}.$$

Multiplying through by $q_{n+1}q_n$, we obtain

$$1 > \frac{1}{2}\Bigl(\frac{q_n}{q_{n+1}} + \frac{q_{n+1}}{q_n}\Bigr). \tag{8}$$

To conclude the proof we note that $x + 1/x = 2 + (\sqrt{x} - 1/\sqrt{x})^2 \geq 2$ for all positive numbers x. Therefore, the right hand side of inequality (8) is certainly ≥ 1, and this is a contradiction. This proves the theorem. □

The result of Theorem 22.5 can be improved, as Theorem 22.7 will show. We shall need the following simple lemma.

Lemma 22.6 *Let x be a positive real number, and suppose $x + \dfrac{1}{x} < \sqrt{5}$. Then*

$$x < \frac{\sqrt{5}+1}{2} \qquad \textit{and} \qquad \frac{1}{x} > \frac{\sqrt{5}-1}{2}.$$

Proof.

$$\begin{aligned} x + 1/x < \sqrt{5} &\Leftrightarrow x^2 - \sqrt{5}x + 1 < 0 \Leftrightarrow \big(x - (\sqrt{5}+1)/2\big)\big(x - (\sqrt{5}-1)/2\big) < 0 \\ &\Leftrightarrow (\sqrt{5}-1)/2 < x < (\sqrt{5}+1)/2, \end{aligned}$$

and if $x < (\sqrt{5}+1)/2$, then $1/x > (\sqrt{5}-1)/2$. □

Theorem 22.7 *If ξ is irrational, then there exist infinitely many rational numbers a/b such that*

$$\left|\xi - \frac{a}{b}\right| < \frac{1}{\sqrt{5}\, b^2}.$$

Indeed, out of three successive convergents of ξ, at least one will satisfy the inequality.

Proof. Suppose on the contrary that none of the convergents $c_k = p_k/q_k$, $k = n-1$, n, $n+1$, satisfies the inequality. Then $|\xi - c_k| \ge 1/\sqrt{5}\, q_k^2$ for $k = n-1$, n, and $n+1$. The successive convergents c_{n-1} and c_n lie on opposite sides of ξ; hence

$$\frac{1}{q_n q_{n-1}} = |c_n - c_{n-1}| = |c_n - \xi| + |\xi - c_{n-1}| \ge \frac{1}{\sqrt{5}}\Big(\frac{1}{q_n^2} + \frac{1}{q_{n-1}^2}\Big).$$

Multiplying through by $q_n q_{n-1}$, we obtain $q_n/q_{n-1} + q_{n-1}/q_n < \sqrt{5}$ (the inequality is strict since the number on the left hand side is rational), and using Lemma 22.6, we conclude that $q_n/q_{n-1} < (\sqrt{5}+1)/2$ and $q_{n-1}/q_n > (\sqrt{5}-1)/2$.

Analogously, we must have $q_{n+1}/q_n < (\sqrt{5}+1)/2$, and hence

$$\begin{aligned} \frac{\sqrt{5}+1}{2} > \frac{q_{n+1}}{q_n} &= \frac{a_n q_n + q_{n-1}}{q_n} \ge \frac{q_n + q_{n-1}}{q_n} = 1 + \frac{q_{n-1}}{q_n} \\ &> 1 + \frac{\sqrt{5}-1}{2} = \frac{\sqrt{5}+1}{2}. \end{aligned}$$

This contradiction proves the theorem. □

Theorem 22.8 *The constant $\sqrt{5}$ in the preceding theorem is best possible, because if $C > \sqrt{5}$ and $\xi = \langle 1, 1, 1, \dots \rangle = (\sqrt{5}+1)/2$, then the inequality*

$$\left|\xi - \frac{a}{b}\right| < \frac{1}{Cb^2}$$

holds for only finitely many integers a and b.

Proof. By Theorem 22.4, any fraction a/b satisfying the inequality must be a convergent, so it suffices to prove that only finitely many convergents p_n/q_n of ξ satisfy the inequality.

First note that $\xi^{-1} = (\sqrt{5}-1)/2$, $\xi + \xi^{-1} = \sqrt{5}$, and $\xi - \xi^{-1} = 1$.

Since $p_n = p_{n-1} + p_{n-2}$ and $q_n = q_{n-1} + q_{n-2}$, where $p_{-2} = 0$ and $p_{-1} = 1$, whereas $q_{-1} = 0$ and $q_0 = 1$, we conclude that $q_n = p_{n-1}$ for all $n \geq -1$, and that $(q_n)_0^\infty$ is the ordinary Fibonacci sequence. Moreover, using induction it is easy to show that

$$p_n = A\xi^n + B(-\xi)^{-n},$$

where the constants A and B are determined by the condition

$$\begin{cases} p_{-1} = A\xi^{-1} - B\xi = 1 \\ \quad p_0 = A + B = 1 \end{cases}$$

Solving for A and B, we find

$$A = \frac{1+\xi}{\xi+\xi^{-1}}, \quad B = \frac{\xi^{-1}-1}{\xi+\xi^{-1}}, \quad \text{and} \quad AB(\xi+\xi^{-1}) = \frac{\xi^{-1}-\xi}{\xi+\xi^{-1}} = -\frac{1}{\sqrt{5}}.$$

Hence,

$$\begin{aligned} |q_n\xi - p_n| &= |p_{n-1}\xi - p_n| = |A\xi^n - B(-\xi)^{2-n} - A\xi^n - B(-\xi)^{-n}| \\ &= |B(\xi^2+1)\xi^{-n}| = -B(\xi^2+1)\xi^{-n}. \end{aligned}$$

It follows that

$$\begin{aligned} q_n^2\,|\xi - \frac{p_n}{q_n}| &= |q_n\xi - p_n|\,q_n = -B(\xi^2+1)\xi^{-n}(A\xi^{n-1} + B(-\xi)^{1-n}) \\ &= -AB(\xi+\xi^{-1}) + (-1)^n B^2(\xi^2+1)\xi^{1-2n} \\ &= \frac{1}{\sqrt{5}} + B^2(-1)^n(\xi^2+1)\xi^{1-2n}. \end{aligned}$$

Let n tend to ∞; then $\xi^{1-2n} \to 0$ since $\xi > 1$, and we conclude that

$$\lim_{n\to\infty} q_n^2\,|\xi - \frac{p_n}{q_n}| = \frac{1}{\sqrt{5}}.$$

Since $1/\sqrt{5} > 1/C$, it follows that

$$q_n^2|\xi - \frac{p_n}{q_n}| > \frac{1}{C},$$

for all but finitely many n. Thus, the inequality in Theorem 22.8 holds for only finitely many convergents p_n/q_n. □

23 Periodic Continued Fractions

In section 20 we computed some periodic simple continued fractions and found that they were roots of quadratic equations with integer coefficients. The goal of this section is to prove that this property characterizes the periodic simple continued fractions, that is an irrational number has a periodic continued fraction expansion if and only if it satifies a quadratic equation with integer coefficients.

Definition 23.1 An infinite sequence $(a_n)_{n=0}^{\infty}$ is called *periodic* if there is a nonzero integer p and an integer m such that

$$a_n = a_{n+p} \quad \text{for all } n \geq m.$$

The integer p is called a *period* of the sequence.

If p and q are two different periods for the sequence, then $p - q$ is a period, too, because $a_{n+p-q} = a_{n+p-q+q} = a_{n+p} = a_n$ for all sufficiently large integers n. Thus, the set of all periods together with the number 0 is an ideal in $\mathbf{Z}$. It follows that there exists a *smallest* positive integer r such that all periods of the sequence are multiples of r. This uniquely determined number is called *the period* and *the period length* of the sequence.

A periodic sequence with period $p > 0$ can be written in the form

$$\begin{aligned} &b_0, b_1, \dots, b_{m-1}, c_0, c_1, \dots, c_{p-1}, c_0, c_1, \dots, c_{p-1}, \dots \\ &\quad = b_0, b_1, \dots, b_{m-1}, \overline{c_0, c_1, \dots, c_{p-1}} \end{aligned}$$

where the bar over the c_0, c_1, …, c_{p-1} indicates that this block of numbers is repeated indefinitely.

A periodic sequence $(a_n)_{n=0}^{\infty}$ with period $p > 0$ is called *purely periodic* if $a_n = a_{n+p}$ holds for all $n \geq 0$. Purely periodic sequences are of the form $\overline{a_0, a_1, \dots, a_{p-1}}$.

Definition 23.2 An infinite continued fraction $\langle a_0, a_1, a_2, \dots \rangle$ is called *(purely) periodic* if the corresponding sequence $(a_n)_{n=0}^{\infty}$ of terms is (purely) periodic. Of course, the period of a periodic continued fraction is by definition the period of the sequence of terms.

Let $\xi = \langle a_0, a_1, a_2, \dots \rangle$ be a continued fraction and write

$$\xi_k = \langle a_k, a_{k+1}, a_{k+2}, \dots \rangle.$$

If ξ is a periodic continued fraction with period p, then obviously there is an integer m such that $\xi_n = \xi_{n+p}$ holds for all $n \geq m$. Conversely, if $\xi_{n+p} = \xi_n$ holds for some number n, then ξ is a periodic continued fraction with p as a period (and *the* period r is some divisor of p).

Definition 23.3 An irrational number ξ is called a *quadratic irrational* (or *algebraic of degree two*) if it is the root of a quadratic polynomial with integer coefficients, that is if $a\xi^2 + b\xi + c = 0$ for suitable integer coefficients a, b, and c with $a \neq 0$.

Proposition 23.4 *A real number ξ is a quadratic irrational if and only if it has the form $\xi = r + s\sqrt{d}$, where d is a positive integer that is not a perfect square, r and s are rational numbers and $s \neq 0$.*

Proof. Any real irrational solution of a quadratic equation $ax^2 + bx + c = 0$ obviously has this form. Conversely, a real number of this form is irrational and satisfies the quadratic equation $(x - r)^2 = s^2 d$, which can be turned into a quadratic equation with integer coefficients upon multiplication by the squares of the denominators of r and s. □

Definition 23.5 Let d be a positive integer that is not a perfect square. We define $\mathbf{Q}[\sqrt{d}\,]$ to be the set of all real numbers ξ of the form $\xi = r + s\sqrt{d}$, with r and s rational. The number $\xi' = r - s\sqrt{d}$ is called the *conjugate* of ξ.

The simple proofs of the following two propositions are left to the reader.

Proposition 23.6 *$\mathbf{Q}[\sqrt{d}\,]$ is a number field, that is if ξ and η are numbers in $\mathbf{Q}[\sqrt{d}\,]$, then their sum $\xi+\eta$, difference $\xi-\eta$, product $\xi\eta$, and quotient ξ/η also belong to $\mathbf{Q}[\sqrt{d}\,]$, the quotient of course provided $\eta \neq 0$.*

Proposition 23.7 *Suppose ξ, $\eta \in \mathbf{Q}[\sqrt{d}\,]$. Then $(\xi+\eta)' = \xi' + \eta'$, $(\xi-\eta)' = \xi' - \eta'$, $(\xi\eta)' = \xi'\eta'$, and $(\xi/\eta)' = \xi'/\eta'$.*

Proposition 23.8 *If the number ξ has a periodic simple continued fraction expansion, then ξ is a quadratic irrational.*

Proof. Being an infinite continued fraction, ξ is irrational. We will prove that $\xi \in \mathbf{Q}[\sqrt{d}\,]$ for a suitable positive integer d that is not a perfect square.

Assume

$$\xi = \langle b_0, b_1, \dots, b_{m-1}, \overline{c_0, c_1, \dots, c_{r-1}}\,\rangle,$$

and let $\eta = \langle\, \overline{c_0, c_1, \dots, c_{r-1}}\,\rangle$. Then $\eta = \langle c_0, c_1, \dots, c_{r-1}, \eta\rangle$.

Let (p_k, q_k) be the convergents of the continued fraction $\langle c_0, c_1, \dots, c_{r-1}\rangle$. Then

$$\eta = \langle c_0, c_1, \dots, c_{r-1}, \eta\rangle = \frac{\eta p_{r-1} + p_{r-2}}{\eta q_{r-1} + q_{r-2}},$$

and solving for η we see that η satisfies a quadratic equation with integer coefficients. Hence, η is a quadratic irrational, that is $\eta \in \mathbf{Q}[\sqrt{d}\,]$ for a suitable positive integer d that is not a perfect square.

Similarly, in terms of the convergents (P_k, Q_k) of $\langle b_0, b_1, \dots, b_{m-1}\rangle$, we have

$$\xi = \langle b_0, b_1, \dots, b_{m-1}, \eta\rangle = \frac{\eta P_{m-1} + P_{m-2}}{\eta Q_{m-1} + Q_{m-2}},$$

so by Proposition 23.6, ξ belongs to $\mathbf{Q}[\sqrt{d}\,]$. □

The converse of Proposition 23.8 is true, that is every quadratic irrational has a periodic simple continued fraction expansion. The proof of this needs some preparatory work.

Lemma 23.9 *If ξ is a quadratic irrational, then ξ can be written in the form*

$$\xi = \frac{u + \sqrt{d}}{v},$$

where d is an integer that is not a perfect square, u and v are integers, and $v \mid (d - u^2)$.

Proof. By Proposition 23.4, $\xi = r + s\sqrt{D}$, where D is an integer that is not a perfect square, r and s are rational numbers and $s \neq 0$. We can obviously write $r = a/c$ and $s = b/c$, where a, b, and c are integers and $b > 0$. Then

$$\xi = \frac{a + b\sqrt{D}}{c} = \frac{a|c| + \sqrt{b^2c^2D}}{c|c|} = \frac{u + \sqrt{d}}{v},$$

and the integers $u = a|c|$, $v = c|c|$ and $d = b^2c^2D$ satisfy the requirement $v \mid (d - u^2)$. □

Suppose ξ_0 is a quadratic irrational. Using Lemma 23.9, we first write

$$\xi_0 = (u_0 + \sqrt{d})/v_0,$$

where d is an integer that is not a perfect square, and u_0 and v_0 are integers, and $v_0 \mid (d - u_0^2)$.

We then recall the recursive algorithm in Theorem 21.5 for obtaining the continued fraction expansion of $\langle a_0, a_1, a_2, \dots \rangle$ of ξ_0. The terms a_n are given by

$$a_0 = [\xi_0], \quad \xi_{n+1} = \frac{1}{\xi_n - a_n}, \quad \text{and} \quad a_{n+1} = [\xi_{n+1}] \qquad \text{for } n = 0,\, 1,\, 2,\, \dots,$$

and we have $\xi_0 = \langle a_0, a_1, \dots, a_n, \xi_{n+1} \rangle$ for all n.

Now suppose inductively that $\xi_n = (u_n + \sqrt{d})/v_n$, with integers u_n and v_n that satisfy $v_n \mid (d - u_n^2)$. Then

$$\xi_{n+1} = \frac{1}{\xi_n - a_n} = \frac{1}{\dfrac{\sqrt{d} - (a_nv_n - u_n)}{v_n}} = \frac{\sqrt{d} + (a_nv_n - u_n)}{\dfrac{d - (a_nv_n - u_n)^2}{v_n}} = \frac{u_{n+1} + \sqrt{d}}{v_{n+1}},$$

where $u_{n+1} = a_nv_n - u_n$ and $v_{n+1} = (d - u_{n+1}^2)/v_n$.

Clearly, u_{n+1} is an integer and $u_{n+1} \equiv -u_n \pmod{v_n}$. Hence by the induction assumption, $d - u_{n+1}^2 \equiv d - u_n^2 \equiv 0 \pmod{v_n}$, that is v_n divides $d - u_{n+1}^2$. Therefore, v_{n+1} is also an integer, and $v_{n+1} \mid (d - u_{n+1}^2)$, because $v_nv_{n+1} = d - u_{n+1}^2$.

By induction, we have thus proved the validity of the following algorithm:

Theorem 23.10 *Suppose $\xi_0 = (u_0 + \sqrt{d})/v_0$, where d is a positive integer that is not a perfect square, u_0 and v_0 are integers and $v_0 \mid (d - u_0^2)$. Define recursively the sequences $(u_n)_0^\infty$, $(v_n)_0^\infty$, $(a_n)_0^\infty$ and $(\xi_n)_0^\infty$ as follows:*

$$\begin{aligned} \xi_n &= \frac{u_n + \sqrt{d}}{v_n}, & a_n &= [\xi_n] \\ u_{n+1} &= a_nv_n - u_n, & v_{n+1} &= \frac{d - u_{n+1}^2}{v_n}, \qquad \textit{for } n \ge 0. \end{aligned}$$

Then u_n and v_n are integers, $v_n \mid (d - u_n^2)$, and $\xi_0 = \langle a_0, a_1, \dots, a_n, \xi_{n+1} \rangle$ for all n, and

$$\xi_0 = \langle a_0, a_1, a_2, \dots \rangle.$$

Example 1 Let us compute the continued fraction expansion of the number $(1 - \sqrt{5})/3$ using the algorithm of Theorem 23.10. Since $3 \nmid (5 - 1^2)$, we first have to put the number in the form of Lemma 23.9. Multiplying numerator and denominator by -3, we obtain

$$\xi_0 = \frac{-3 + \sqrt{45}}{-9}, \quad \text{that is} \quad u_0 = -3, \quad v_0 = -9, \quad \text{and} \quad d = 45.$$

Now $v_0 \mid (d - u_0^2)$, so we can start the algorithm. The result of the computations is shown in the following table:

n	0	1	2	3	4	5	6	7	8	9
u_n	-3	12	-1	5	5	3	6	6	3	5
v_n	-9	11	4	5	4	9	1	9	4	5
a_n	-1	1	1	2	2	1	12	1	2	2

Since $(u_9, v_9) = (u_3, v_3)$, we conclude that $\xi_9 = \xi_3$. Thus,

$$\frac{1-\sqrt{5}}{3} = \langle -1, 1, 1, \overline{2, 2, 1, 12, 1, 2}\,\rangle.$$

□

Lemma 23.11 *Let ξ be a quadratic irrational and define ξ_n as in Theorem 23.10. If the conjugate $\xi'_k < 0$ for some index k, then $-1 < \xi'_n < 0$ for all $n > k$.*

Proof. By induction, it suffices to prove that $\xi'_n < 0$ implies $-1 < \xi'_{n+1} < 0$. So assume $\xi'_n < 0$. Using the relation $\xi_{n+1} = 1/(\xi_n - a_n)$ and taking conjugates, we obtain $\xi'_{n+1} = 1/(\xi'_n - a_n)$. Since $a_n \geq 1$, the denominator $\xi'_n - a_n$ is strictly less than -1, so it follows that $-1 < \xi'_{n+1} < 0$. □

Lemma 23.12 *Let ξ be a quadratic irrational, and define ξ_n and $a_n = [\xi_n]$ as above. If $-1 < \xi'_n < 0$, then $a_n = [-1/\xi'_{n+1}]$.*

Proof. We have $\xi'_{n+1} = 1/(\xi'_n - a_n)$, whence $-1/\xi'_{n+1} = a_n - \xi'_n$. Since $0 < -\xi'_n < 1$, it follows that $[-1/\xi'_{n+1}] = [a_n - \xi'_n] = a_n$. □

Lemma 23.13 *If ξ is a quadratic irrational, then there exists an index k such that $\xi'_k < 0$.*

Proof. Let (p_k, q_k) denote the kth convergent of ξ. Since

$$\xi = \langle a_0, a_1, \dots, a_{n-1}, \xi_n \rangle,$$

we have

$$\xi = \frac{p_{n-1}\xi_n + p_{n-2}}{q_{n-1}\xi_n + q_{n-2}},$$

and solving for ξ_n we obtain

$$\xi_n = \frac{q_{n-2}\xi - p_{n-2}}{p_{n-1} - q_{n-1}\xi} = -\frac{q_{n-2}}{q_{n-1}}\Big(\frac{\xi - p_{n-2}/q_{n-2}}{\xi - p_{n-1}/q_{n-1}}\Big).$$

By taking conjugates, we get

$$\xi'_n = -\frac{q_{n-2}}{q_{n-1}}\Big(\frac{\xi' - p_{n-2}/q_{n-2}}{\xi' - p_{n-1}/q_{n-1}}\Big).$$

We now use the fact that the convergents p_n/q_n converge to ξ as n tends to infinity and that $\xi' \neq \xi$. It follows that the expression within parenthesis converges to $(\xi' - \xi)/(\xi' - \xi)$, that is to 1, as n tends to infinity. Consequently, the expression within parenthesis is certainly > 0 when n is big enough, that is ξ'_n has the same sign as $-q_{n-2}/q_{n-1}$, which is negative since q_n is positive for all $n \geq 0$. □

Theorem 23.14 *A real number ξ has a periodic simple continued fraction expansion if and only if it is a quadratic irrational.*

Proof. We have already proved that a periodic continued fraction is a quadratic irrational (Proposition 23.8). To prove the converse, let $\xi = \xi_0$ be a quadratic irrational and write

$$\xi_n = \frac{u_n + \sqrt{d}}{v_n}$$

as in Theorem 23.10. By Lemma 23.13, there is an index k such that $\xi'_k < 0$, and by Lemma 23.11, $-1 < \xi'_n < 0$ for all $n > k$. Since $\xi_n > 1$ for all $n \geq 1$, we conclude that

$$1 < \xi_n - \xi'_n = \frac{2\sqrt{d}}{v_n} \quad \text{and} \quad 0 < \xi_n + \xi'_n = \frac{2u_n}{v_n}$$

for all $n > k$. Hence $0 < v_n < 2\sqrt{d}$ and $u_n > 0$ if $n > k$. Moreover, using the relation $d - u_{n+1}^2 = v_n v_{n+1} > 0$, we obtain $u_{n+1}^2 < d$, that is $u_{n+1} < \sqrt{d}$ for $n > k$. Thus, if $n > k+1$, then $0 < u_n < \sqrt{d}$ and $0 < v_n < 2\sqrt{d}$. Hence, the ordered pairs (u_n, v_n) can assume only a fixed number of possible pair values, and so there are distinct integers i and j with $j > i$ such that $u_j = u_i$ and $v_j = v_i$. This implies that $\xi_i = \xi_j = \xi_{i+(j-i)}$, and hence ξ has a periodic continued fraction expansion. □

We will next characterize the purely periodic continued fractions.

Definition 23.15 A quadratic irrational $\xi = r + s\sqrt{d}$ is called *reduced* it $\xi > 1$ and its conjugate $\xi' = r - s\sqrt{d}$ satisfies $-1 < \xi' < 0$.

Theorem 23.16 *The simple continued fraction expansion of the real quadratic irrational number ξ is purely periodic if and only if ξ is reduced. Also, if $\xi = \langle \overline{a_0, a_1, \dots, a_{r-1}} \rangle$, then $-1/\xi' = \langle \overline{a_{r-1}, a_{r-2}, \dots, a_1, a_0} \rangle$.*

Proof. Suppose $\xi = \xi_0$ is a reduced quadratic irrational, and use Theorem 23.10 to write $\xi_n = (u_n + \sqrt{d})/v_n$. Since $-1 < \xi'_0 < 0$ by assumption, we have $-1 < \xi'_n < 0$ and $a_n = [-1/\xi'_{n+1}]$ for all $n \geq 0$ by Lemma 23.11 and Lemma 23.12.

We know from Theorem 23.14 that ξ has a simple periodic continued fraction expansion. Let r be the period length; then there is a smallest number $m \geq 0$ such that

$$\xi_{n+r} = \xi_n \quad \text{for all } n \geq m.$$

We must prove that $m = 0$.

Assume $m \geq 1$. Starting from $\xi_m = \xi_{m+r}$ we first obtain $\xi'_m = \xi'_{m+r}$ by taking conjugates, and hence $a_{m-1} = [-1/\xi'_m] = [-1/\xi'_{m+r}] = a_{m+r-1}$. Since

$$\frac{1}{\xi_{m-1} - a_{m-1}} = \xi_m = \xi_{m+r} = \frac{1}{\xi_{m+r-1} - a_{m+r-1}},$$

we then conclude that $\xi_{m-1+r} = \xi_{m-1}$, which violates the definition of m. Thus $m = 0$, and ξ is purely periodic.

Conversely, suppose that ξ is purely periodic, say $\xi = \langle\overline{a_0, a_1, \dots, a_{r-1}}\,\rangle$, where a_0, a_1, ..., a_{r-1} are positive integers. Then $\xi > a_0 \geq 1$. Let (p_n, q_n) denote the nth convergent of ξ; then

$$\xi = \langle a_0, a_1, \dots, a_{r-1}, \xi\rangle = \frac{p_{r-1}\xi + p_{r-2}}{q_{r-1}\xi + q_{r-2}}.$$

Thus ξ satisfies the quadratic equation

$$f(x) = q_{r-1}x^2 + (q_{r-2} - p_{r-1})x - p_{r-2} = 0.$$

This equation has two roots, ξ and its conjugate ξ'. Since $\xi > 1$, we need only prove that $f(x)$ has a root between -1 and 0 to establish that $-1 < \xi' < 0$. We will do this by showing that $f(0) < 0$ and $f(-1) > 0$.

Note that p_n is positive for all $n \geq -1$ (since $a_0 > 0$). Hence, $f(0) = -p_{r-2} < 0$. Next we see that

$$\begin{aligned} f(-1) &= q_{r-1} - q_{r-2} + p_{r-1} - p_{r-2} = (a_{r-1} - 1)(q_{r-2} + p_{r-2}) + q_{r-3} + p_{r-3} \\ &\geq q_{r-3} + p_{r-3} > 0. \end{aligned}$$

Thus, ξ is reduced.

Finally, to prove that $-1/\xi'$ has the stated continued fraction expansion, we suppose that $\xi = \langle\overline{a_0, a_1, \dots, a_{r-1}}\,\rangle$. Taking conjugates in the relation $\xi_n = 1/(\xi_{n-1} - a_{n-1})$ we obtain $\xi'_n = 1/(\xi'_{n-1} - a_{n-1})$, which can be rewritten as

$$-1/\xi'_n = a_{n-1} + \frac{1}{-1/\xi'_{n-1}} \quad \text{for all } n \geq 1.$$

Since $-1/\xi'_n > 1$ for all n, the above equation can be expressed as a continued fraction expansion

$$-1/\xi'_n = \langle a_{n-1}, -1/\xi'_{n-1}\rangle.$$

Starting with $-1/\xi'_r$, iterating and using the fact that $\xi = \xi_0 = \xi_r$, we thus obtain

$$\begin{aligned} -1/\xi' = -1/\xi'_0 = -1/\xi'_r &= \langle a_{r-1}, -1/\xi'_{r-1}\rangle = \langle a_{r-1}, a_{r-2}, -1/\xi'_{r-2}\rangle = \dots \\ &= \langle a_{r-1}, a_{r-2}, \dots, a_1, a_0, -1/\xi'_0\rangle. \end{aligned}$$

Hence, $-1/\xi' = \langle\overline{a_{r-1}, a_{r-2}, \dots, a_1, a_0}\,\rangle$. □

Example 2 The quadratic irrational $(2 + \sqrt{10})/3$ is reduced. Its continued fraction expansion is easily computed with the aid of Theorem 23.10. Since $3 \mid (10 - 2^2)$, we can start with $u_0 = 2$, $v_0 = 3$ and $d = 10$. The computations are summarized in the following table:

n	0	1	2	3
u_n	2	1	2	2
v_n	3	3	2	3
a_n	1	1	2	1

Since $(u_3, v_3) = (u_0, v_0)$, the period is 3 and $(2 + \sqrt{10})/3 = \langle\overline{1, 1, 2}\,\rangle$. □

24 Continued Fraction Expansion of $\sqrt{\mathrm{d}}$

Theorem 24.1 *Let d be a positive integer that is not a perfect square. The simple continued fraction expansion of $\sqrt{d}$ is of the form*

$$\langle a_0, \overline{a_1, a_2, \ldots, a_{r-1}, 2a_0}\,\rangle,$$

where $a_0 = [\sqrt{d}\,]$ and $a_j = a_{r-j}$ for $j = 1, 2, \ldots, r-1$.

Proof. Let $a_0 = [\sqrt{d}\,]$ and $\xi = a_0 + \sqrt{d}$. Then ξ is reduced, because $\xi > 1$ and $\xi' = a_0 - \sqrt{d}$ satisfies $-1 < \xi' < 0$. By Theorem 23.16, ξ has a purely periodic continued fraction expansion starting with $[\xi] = 2a_0$, say

$$(1) \qquad \xi = a_0 + \sqrt{d} = \langle \overline{2a_0, a_1, a_2, \ldots, a_{r-1}}\,\rangle = \langle 2a_0, \overline{a_1, a_2, \ldots, a_{r-1}, 2a_0}\,\rangle.$$

If we subtract a_0 from each side, we get

$$\sqrt{d} = \langle a_0, \overline{a_1, a_2, \ldots, a_{r-1}, 2a_0}\,\rangle.$$

To prove that the sequence $a_1, a_2, \ldots, a_{r-1}$ is "symmetric", we note that

$$\xi = a_0 + \sqrt{d} = 2a_0 + \sqrt{d} - a_0 = 2a_0 - \xi' = 2a_0 + \frac{1}{-1/\xi'} = \langle 2a_0, -1/\xi' \rangle.$$

By Theorem 23.16,

$$-1/\xi' = \langle \overline{a_{r-1}, a_{r-2}, \ldots, a_1, 2a_0}\,\rangle,$$

and hence

$$\xi = \langle 2a_0, \overline{a_{r-1}, a_{r-2}, \ldots, a_1, 2a_0}\,\rangle.$$

A comparison with (1) gives $a_j = a_{r-j}$ for $1 \le j \le r-1$. □

Example 1 To compute the continued fraction expansion of $\sqrt{19}$ we use Theorem 23.10 with $u_0 = 0$, $v_0 = 1$ and $d = 19$. We get the following table:

n	0	1	2	3	4	5	6	7
u_n	0	4	2	3	3	2	4	4
v_n	1	3	5	2	5	3	1	3
a_n	4	2	1	3	1	2	8	2

It follows that the expansion has period length 6, and that

$$\sqrt{19} = \langle 4, \overline{2, 1, 3, 1, 2, 8}\,\rangle.$$

□

Theorem 24.2 *Let (p_n, q_n) denote the nth convergent of $\sqrt{d}$, let the integers u_n and v_n be defined for the number $\xi = \sqrt{d}$ as in Theorem 23.10, that is $\xi_n = (u_n + \sqrt{d})/v_n$ with $v_n \mid (d - u_n^2)$, and let r be the period length of the continued fraction expansion of $\sqrt{d}$. Then*

(i) $p_n^2 - dq_n^2 = (-1)^{n-1} v_{n+1}$ *for every* $n \ge -1$*;*

(ii) $v_n > 0$ *for every* $n \ge 0$*;*

(iii) $v_n = 1$ *if and only if* $r \mid n$*.*

Proof. Write $\sqrt{d} = \langle a_0, a_1, a_2, \dots \rangle = \langle a_0, a_1, \dots, a_n, \xi_{n+1} \rangle$.

(i) We have

$$\sqrt{d} = \frac{\xi_{n+1}p_n + p_{n-1}}{\xi_{n+1}q_n + q_{n-1}} = \frac{(u_{n+1} + \sqrt{d})p_n + v_{n+1}p_{n-1}}{(u_{n+1} + \sqrt{d})q_n + v_{n+1}q_{n-1}},$$

which can also be written as

$$u_{n+1}p_n + v_{n+1}p_{n-1} - dq_n - (u_{n+1}q_n + v_{n+1}q_{n-1} - p_n)\sqrt{d} = 0.$$

Since $\sqrt{d}$ is irrational, it follows that

$$\begin{cases} u_{n+1}p_n + v_{n+1}p_{n-1} - dq_n = 0 \\ u_{n+1}q_n + v_{n+1}q_{n-1} - p_n = 0 \end{cases}$$

Eliminating u_{n+1} from this system, we obtain

$$p_n^2 - dq_n^2 = v_{n+1}(p_nq_{n-1} - q_np_{n-1}) = (-1)^{n-1}v_{n+1},$$

where we used Theorem 20.5 to get the last equality.

(ii) The convergents p_n/q_n are $> \sqrt{d}$ if n is odd and $< \sqrt{d}$ if n is even. Therefore, $p_n^2 - dq_n^2$ has the same sign as $(-1)^{n-1}$, so it follows from (i) that v_{n+1} is positive for every $n \geq -1$.

(iii) Since $\xi = \sqrt{d}$ has period length r, $\xi_{kr+1} = \xi_1$ for all positive integers k. It follows that

$$\xi_{kr} - a_{kr} = \frac{1}{\xi_{kr+1}} = \frac{1}{\xi_1} = \xi_0 - a_0 = -a_0 + \sqrt{d},$$

that is $\xi_{kr} = a_{kr} - a_0 + \sqrt{d}$. Hence, $v_{kr} = 1$ (and $u_{kr} = a_{kr} - a_0$).

Conversely, assume $v_n = 1$; then $\xi_n = u_n + \sqrt{d}$, so $a_n = [\xi_n] = u_n + [\sqrt{d}\,] = u_n + a_0$ and $\xi_n - a_n = \sqrt{d} - a_0 = \xi_0 - a_0$, that is $\xi_{n+1} = 1/(\xi_n - a_n) = 1/(\xi_0 - a_0) = \xi_1$. It follows from this that n is a multiple of the period length r. □

The reader may have noted in the few examples of continued fraction expansion of $\sqrt{d}$ that we have given, that the numbers appearing in the period of $\sqrt{d}$ were all less than or equal to a_0 except for the last one, which equals $2a_0$. This holds in general.

Proposition 24.3 *Let $\sqrt{d} = \langle a_0, \overline{a_1, \dots, a_{r-1}, 2a_0}\,\rangle$. Then $a_n \leq a_0$ for $1 \leq n \leq r - 1$.*

Proof. With $\xi = \xi_0 = \sqrt{d}$, let $\xi_n = (u_n + \sqrt{d})/v_n$ be as in Theorem 23.10 and suppose $1 \leq n \leq r - 1$. Then $v_n \geq 2$ by the previous theorem, and using Lemma 23.11 we conclude that $\xi_n' = (u_n - \sqrt{d})/v_n < 0$, because $\xi_0' = -\sqrt{d} < 0$. It follows that $u_n - \sqrt{d} < 0$, that is $u_n < \sqrt{d}$ and hence $\xi_n < 2\sqrt{d}/v_n \leq \sqrt{d}$. Finally, $a_n = [\xi_n] \leq [\sqrt{d}\,] = a_0$. □

25 Pell's Equation

The equation $x^2 - dy^2 = N$, with given nonzero integers d and N, is called *Pell's equation.* If d is negative, Pell's equation can have only a finite number of solutions in integers, since $x^2 \le N$ and $y^2 \le -N/d$.

If $d = a^2$ is a perfect square, then we have $(x + ay)(x - ay) = N$, and again there is only a finite number of integral solutions to Pell's equation, since there is only a finite number of ways to factor N.

We will therefore suppose that *d is a positive integer that is not a perfect square.* We will show that in that case there is either no solution at all or infinitely many solutions in integers. When $N = \pm 1$, we will give a complete description of the set of solutions.

If (u, v) is an integral solution of Pell's equation $x^2 - dy^2 = N$, then $(\pm u, \pm v)$ is also a solution for every combination of the signs. Thus, in order to find all integral solutions it suffices to find all *positive* solutions, that is all solutions (u, v) with $u > 0$ and $v > 0$. If N is a perfect square, there will of course be two additional trivial solutions $(\pm\sqrt{N}, 0)$, and if $-N/d$ happens to be an integer that is a perfect square, $(0, \pm\sqrt{-N/d})$ are two trivial solutions of Pell's equation.

If (x_1, y_1) and (x_2, y_2) are two positive solutions of $x^2 - dy^2 = N$, then $x_1^2 - x_2^2 = d(y_1^2 - y_2^2)$, and hence $x_1 < x_2$ if and only if $y_1 < y_2$. Thus, if we order the positive solutions according to increasing x-value or according to increasing y-value we will get the same result.

If there is a positive solution in integers of Pell's equation, then there is obviously a positive solution (x_1, y_1) with a least positive x-value. This solution has also the least y-value among all positive solutions. Since it plays a special role we introduce the following definition.

Definition 25.1 Suppose Pell's equation $x^2 - dy^2 = N$ has positive integral solutions. The *fundamental solution*, or *least positive solution*, is the positive solution (x_1, y_1) such that $x_1 < u$ and $y_1 < v$ for every other positive solution (u, v).

The following theorem gives a connection between Pell's equation and continued fractions.

Theorem 25.2 *Let d be a positive integer that is not a perfect square, and suppose $|N| < \sqrt{d}$. If (u, v) is a positive solution in integers of $x^2 - dy^2 = N$, then there is a convergent (p_n, q_n) of the simple continued fraction expansion of $\sqrt{d}$ such that $u/v = p_n/q_n$.*

Remark. The numbers u and v need not be relatively prime, but if c is their greatest common divisor, then obviously $c^2 \mid N$. Hence, if N is square-free, and in particular if $N = \pm 1$, then u and v are necessarily relatively prime. That means that there is an index n such that $u = p_n$ and $v = q_n$.

Proof. We will consider a more general situation. Let d and N be any positive real numbers, not necessarily integers, such that $\sqrt{d}$ is irrational and $N < \sqrt{d}$, and assume that u and v are positive integers satisfying $u^2 - dv^2 = N$.

Since

$$\left(\frac{u}{v} - \sqrt{d}\right)\left(\frac{u}{v} + \sqrt{d}\right) = \frac{u^2 - dv^2}{v^2} = \frac{N}{v^2}$$

and the second factor of the left hand side is positive, we first conclude that $u/v - \sqrt{d} > 0$, and consequently $u/v + \sqrt{d} > 2\sqrt{d}$. Hence

$$0 < \frac{u}{v} - \sqrt{d} = \frac{N}{v^2(u/v + \sqrt{d})} < \frac{\sqrt{d}}{2v^2\sqrt{d}} = \frac{1}{2v^2}.$$

By Theorem 22.4, u/v is a convergent of $\sqrt{d}$.

Let now d and N be as in the statement of the theorem. The case $N > 0$ is a special case of what we have just proved.

If $N < 0$, we rewrite the equation as $y^2 - (1/d)x^2 = -N/d$. Since $0 < -N/d < \sqrt{d}/d = \sqrt{1/d}$, we can apply the general case above, and we conclude that v/u is a convergent of $1/\sqrt{d}$. Suppose $\sqrt{d}$ has the continued fraction expansion $\langle a_0, a_1, a_2, \dots \rangle$. Then $1/\sqrt{d} = \langle 0, \sqrt{d} \rangle = \langle 0, a_0, a_1, a_2, \dots \rangle$. Hence, there is an n such that

$$\frac{v}{u} = \langle 0, a_0, a_1, \dots, a_n \rangle = \frac{1}{\langle a_0, a_1, \dots, a_n \rangle},$$

that is $u/v = \langle a_0, a_1, \dots, a_n \rangle$ is a convergent of $\sqrt{d}$. □

By combining the theorem above with Theorem 24.2, we get a complete description of the solution set of Pell's equation in the case $N = \pm 1$.

Theorem 25.3 *Suppose d is a positive integer that is not a perfect square, let r be the period length of the simple continued fraction expansion of $\sqrt{d}$, and let $(p_n, q_n)_{n=0}^{\infty}$ be the corresponding sequence of convergents.*

(i) Suppose r is even. Then

(a) $x^2 - dy^2 = -1$ has no solutions in integers;

(b) all positive integral solutions of $x^2 - dy^2 = 1$ are given by $x = p_{kr-1}$, $y = q_{kr-1}$ for $k = 1, 2, 3, \dots$, with $x = p_{r-1}$ and $y = q_{r-1}$ as the fundamental solution.

(ii) Suppose r is odd Then

(a) all positive integral solutions of $x^2 - dy^2 = -1$ are given by $x = p_{kr-1}$, $y = q_{kr-1}$ for $k = 1, 3, 5, \dots$, with $x = p_{r-1}$ and $y = q_{r-1}$ as the fundamental solution;

(b) all positive integral solutions of $x^2 - dy^2 = 1$ are given by $x = p_{kr-1}$, $y = q_{kr-1}$ for $k = 2, 4, 6, \dots$, with $x = p_{2r-1}$ and $y = q_{2r-1}$ as the fundamental solution.

Proof. By the previous theorem, the positive integral solutions of $x^2 - dy^2 = \pm 1$ are to be found among the convergents (p_n, q_n). Furthermore, $a_0 = [\sqrt{d}] \geq 1$, so the sequence $(p_n)_{n=0}^{\infty}$ is strictly increasing. Therefore, the first solution that appears in the sequence (p_n, q_n) will be the fundamental solution.

According to Theorem 24.2, $p_n^2 - dq_n^2 = (-1)^{n-1}v_{n+1}$, where $v_n \geq 1$ for all n and $v_n = 1$ if and only if $r \mid n$. Thus, $|p_n^2 - dq_n^2| \geq 2$ except when $n = kr - 1$ for some nonnegative integer k, in which case

$$p_n^2 - dq_n^2 = (-1)^{kr}.$$

If r is even, then $(-1)^{kr} = 1$ for all k, and hence (p_{kr-1}, q_{kr-1}) is a solution of $x^2 - dy^2 = 1$ for all k, whereas the equation $x^2 - dy^2 = -1$ has no positive solution, and of course no solution at all in integers. This proves part (i). If the period length r is odd, then $(-1)^{kr} = 1$ for k even, and $= -1$ for k odd, and this proves part (ii). □

EXAMPLE 1 We shall use Theorem 25.3 to find the fundamental solution of the equation $x^2 - 19y^2 = 1$.

The continued fraction expansion $\sqrt{19} = \langle 4, \overline{2, 1, 3, 1, 2, 8}\,\rangle$ was found in the previous section. Since the period length is 6, the fundamental solution is $(x, y) = (p_5, q_5)$. The convergents are computed in the following table:

n	-2	-1	0	1	2	3	4	5
a_n			4	2	1	3	1	2
p_n	0	1	4	9	13	48	61	170
q_n	1	0	1	2	3	11	14	39

Thus, the fundamental solution is $(x, y) = (170, 39)$. □

Theorem 25.3 gives a method for computing the successive solutions of Pell's equation but it is tedious to compute convergents (p_n, q_n). Having found the fundamental solution, we can find the remaining positive solutions by a simpler method, which will be described in Theorem 25.6 below.

Lemma 25.4 *Let (x_1, y_1) be an arbitrary integral solution of $x^2 - dy^2 = M$ and (x_2, y_2) an arbitrary integral solution of $x^2 - dy^2 = N$, and define the integers u and v by*

$$(x_1 + y_1\sqrt{d})(x_2 + y_2\sqrt{d}) = u + v\sqrt{d},$$

that is $u = x_1x_2 + y_1y_2d$, $v = x_1y_2 + x_2y_1$. Then (u, v) is a solution of $x^2 - dy^2 = MN$. If (x_1, y_1) and (x_2, y_2) are positive solutions, then (u, v) is also positive.

Proof. By taking conjugates we have $(x_1 - y_1\sqrt{d})(x_2 - y_2\sqrt{d}) = u - v\sqrt{d}$, and hence

$$\begin{aligned} u^2 - dv^2 &= (u + v\sqrt{d})(u - v\sqrt{d}) \\ &= (x_1 + y_1\sqrt{d})(x_2 + y_2\sqrt{d})(x_1 - y_1\sqrt{d})(x_2 - y_2\sqrt{d}) \\ &= (x_1^2 - dy_1^2)(x_2^2 - dy_2^2) = MN. \end{aligned}$$

The solution (u, v) will obviously be positive if the original ones are positive. □

Corollary 25.5 *If the equation $x^2 - dy^2 = N$ has an integral solution, then it has infinitely many integral solutions.*

Proof. Suppose the equation $x^2 - dy^2 = N$ has at least one integral solution. This solution multiplied by any solution of $x^2 - dy^2 = 1$ yields another solution of $x^2 - dy^2 = N$. Since the equation $x^2 - dy^2 = 1$ has infinitely many integral solutions, there will also be infinitely many integral solutions of $x^2 - dy^2 = N$. □

Theorem 25.6 *Let (x_1, y_1) be the fundamental solution of $x^2 - dy^2 = 1$. Then all positive integral solutions are given by (x_n, y_n), $n \geq 1$, where the integers x_n*

and y_n are recursively defined by

$$x_{n+1} = x_1x_n + y_1y_nd, \qquad y_{n+1} = x_1y_n + y_1x_n.$$

Proof. Note that $x_{n+1} + y_{n+1}\sqrt{d} = (x_1 + y_1\sqrt{d})(x_n + y_n\sqrt{d}) = (x_1 + y_1\sqrt{d})^{n+1}$. Hence by Lemma 25.4 with $M = N = 1$, if (x_n, y_n) is a positive solution of Pell's equation $x^2 - dy^2 = 1$, then (x_{n+1}, y_{n+1}) will also be a positive solution. It therefore follows by induction, that (x_n, y_n) is a solution for all n.

It remains to show that every positive integral solution is obtained in this way. Suppose there is a positive solution (u, v) that is not of the form (x_n, y_n). Since x_n forms an increasing sequence, there must be some integer m such that $x_m \le u < x_{m+1}$. It follows that $y_m \le v < y_{m+1}$, because we get the same result if positive solutions are ordered according to their x-value or y-value. We cannot have equality, because $u = x_m$ would imply $v = y_m$. Now $(x_m, -y_m)$ is of course also a (non-positive) solution of $x^2 - dy^2 = 1$, so by Lemma 25.4 we will obtain another solution (s, t) by defining

$$s + t\sqrt{d} = (u + v\sqrt{d})(x_m - y_m\sqrt{d}) = \frac{u + v\sqrt{d}}{x_m + y_m\sqrt{d}}.$$

Since $x_m + y_m\sqrt{d} < u + v\sqrt{d} < x_{m+1} + y_{m+1}\sqrt{d}$, we have

$$1 < s + t\sqrt{d} < \frac{x_{m+1} + y_{m+1}\sqrt{d}}{x_m + y_m\sqrt{d}} = x_1 + y_1\sqrt{d}.$$

But $s - t\sqrt{d} = 1/(s + t\sqrt{d})$ and hence $0 < s - t\sqrt{d} < 1$. It now follows that

$$\begin{aligned} s &= \tfrac{1}{2}(s + t\sqrt{d}) + \tfrac{1}{2}(s - t\sqrt{d}) > \tfrac{1}{2} + 0 > 0 \\ t\sqrt{d} &= \tfrac{1}{2}(s + t\sqrt{d}) - \tfrac{1}{2}(s - t\sqrt{d}) > \tfrac{1}{2} - \tfrac{1}{2} = 0, \end{aligned}$$

so (s, t) is a positive solution. Therefore, $s > x_1$ and $t > y_1$, but this contradicts $s + t\sqrt{d} < x_1 + y_1\sqrt{d}$. This contradiction shows that every integral solution (u, v) must be of the form (x_n, y_n). □

EXAMPLE 2 In Example 1, we showed that the fundamental solution of

$$x^2 - 19y^2 = 1$$

is $(x_1, y_1) = (170, 39)$. Using the recursion formulas

$$x_n = x_1x_n + 19y_1y_n, \;\; y_n = x_1y_n + y_1x_n,$$

we can compute the next positive solutions. They are

$$\begin{aligned} (x_2, y_2) &= (57\,799, 13\,260) \\ (x_3, y_3) &= (19\,651\,490, 4\,508\,361) \\ (x_4, y_4) &= (6\,681\,448\,801, 1\,532\,829\,480) \end{aligned}$$ □

Just as in the case of $x^2 - dy^2 = 1$, further solutions of the equation $x^2 - dy^2 = -1$ can be found from its fundamental solution. We leave the proof of the following result to the reader.

Theorem 25.7 *Suppose that $x^2 - dy^2 = -1$ has an integral solution, and let (x_1, y_1) denote the fundamental solution. For $n \geq 1$, define positive integers x_n and y_n recursively as in Theorem 25.6, i.e. $(x_n + y_n\sqrt{d}) = (x_1 + y_1\sqrt{d})^n$. Then all positive integral solutions of $x^2 - dy^2 = -1$ are given by (x_n, y_n) with n odd, and all positive integral solutions of $x^2 - dy^2 = 1$ are given by (x_n, y_n) with n even. In particular, (x_2, y_2) is the fundamental solution of $x^2 - dy^2 = 1$.*