

II

Divisibility

1 Greatest Common Divisors

In the set \mathbb{N} of all positive integers we can perform two basic operations: addition and multiplication. In this chapter we will be primarily concerned with the second operation.

Multiplication has the following properties:

- (M1) if $ab = ac$, then $b = c$; (cancellation law)
- (M2) $ab = ba$ for all a, b ; (commutative law)
- (M3) $(ab)c = a(bc)$ for all a, b, c ; (associative law)
- (M4) $1a = a$ for all a . (identity element)

For any $a, b \in \mathbb{N}$ we say that b divides a , or that b is a factor of a , or that a is a multiple of b if $a = ba'$ for some $a' \in \mathbb{N}$. We write $b|a$ if b divides a and $b \nmid a$ if b does not divide a . For example, $2|6$, since $6 = 2 \times 3$, but $4 \nmid 6$. (We sometimes use \times instead of \cdot for the product of positive integers.) The following properties of divisibility follow at once from the definition:

- (i) $a|a$ and $1|a$ for every a ;
- (ii) if $b|a$ and $c|b$, then $c|a$;
- (iii) if $b|a$, then $b|ac$ for every c ;
- (iv) $bc|ac$ if and only if $b|a$;
- (v) if $b|a$ and $a|b$, then $b = a$.

For any $a, b \in \mathbb{N}$ we say that d is a common divisor of a and b if $d|a$ and $d|b$. We say that a common divisor d of a and b is a greatest common divisor if every common divisor of a and b divides d . The greatest common divisor of a and b is uniquely determined, if it exists, and will be denoted by (a, b) .

The greatest common divisor of a and b is indeed the numerically greatest common divisor. However, it is preferable not to define greatest common divisors in this way, since the concept is then available for algebraic structures in which there is no relation of magnitude and only the operation of multiplication is defined.

Proposition 1 Any $a, b \in \mathbb{N}$ have a greatest common divisor (a, b) .

Proof Without loss of generality we may suppose $a \geq b$. If b divides a , then $(a, b) = b$. Assume that there exists a pair a, b without greatest common divisor and choose one for which a is a minimum. Then $1 < b < a$, since b does not divide a . Since also $1 \leq a - b < a$, the pair $a - b, b$ has a greatest common divisor d . Since any common divisor of a and b divides $a - b$, and since d divides $(a - b) + b = a$, it follows that d is a greatest common divisor of a and b . But this is a contradiction. \square

The proof of Proposition 1 uses not only the multiplicative structure of the set \mathbb{N} , but also its ordering and additive structure. To see that there is a reason for this, consider the set S of all positive integers of the form $4k + 1$. The set S is closed under multiplication, since

$$(4j + 1)(4k + 1) = 4(4jk + j + k) + 1,$$

and we can define divisibility and greatest common divisors in S by simply replacing \mathbb{N} by S in our previous definitions. However, although the elements 693 and 189 of S have the common divisors 9 and 21, they have no greatest common divisor according to this definition.

In the following discussion we use the result of Proposition 1, but make no further appeal to either addition or order.

For any $a, b \in \mathbb{N}$ we say that h is a *common multiple* of a and b if $a|h$ and $b|h$. We say that a common multiple h of a and b is a *least common multiple* if h divides every common multiple of a and b . The least common multiple of a and b is uniquely determined, if it exists, and will be denoted by $[a, b]$.

It is evident that, for every a ,

$$\begin{aligned} (a, 1) &= 1, & [a, 1] &= a, \\ (a, a) &= a = [a, a]. \end{aligned}$$

Proposition 2 Any $a, b \in \mathbb{N}$ have a least common multiple $[a, b]$. Moreover,

$$(a, b)[a, b] = ab.$$

Furthermore, for all $a, b, c \in \mathbb{N}$,

$$\begin{aligned} (ac, bc) &= (a, b)c, & [ac, bc] &= [a, b]c, \\ ([a, b], [a, c]) &= [a, (b, c)], & [(a, b), (a, c)] &= (a, [b, c]). \end{aligned}$$

Proof We show first that $(ac, bc) = (a, b)c$. Put $d = (a, b)$. Clearly cd is a common divisor of ac and bc , and so $(ac, bc) = qcd$ for some $q \in \mathbb{N}$. Thus $ac = qcda'$, $bc = qcdb'$ for some $a', b' \in \mathbb{N}$. It follows that $a = qda'$, $b = qdb'$. Thus qd is a common divisor of a and b . Hence qd divides d , which implies $q = 1$.

If g is any common multiple of a and b , then ab divides ga and gb , and hence ab also divides (ga, gb) . But, by what we have just proved,

$$(ga, gb) = (a, b)g = dg.$$

Hence $h := ab/d$ divides g . Since h is clearly a common multiple of a and b , it follows that $h = [a, b]$. Replacing a, b by ac, bc , we now obtain

$$[ac, bc] = acbc/(ac, bc) = abc/(a, b) = hc.$$

If we put

$$A = ([a, b], [a, c]), \quad B = [a, (b, c)],$$

then by what we have already proved,

$$\begin{aligned} A &= (ab/(a, b), ac/(a, c)), \\ B &= a(b, c)/(a, (b, c)) = (ab/(a, (b, c)), ac/(a, (b, c))). \end{aligned}$$

Since any common divisor of $ab/(a, b)$ and $ac/(a, c)$ is also a common divisor of $ab/(a, (b, c))$ and $ac/(a, (b, c))$, it follows that A divides B . On the other hand, a divides A , since a divides $[a, b]$ and $[a, c]$, and similarly (b, c) divides A . Hence B divides A . Thus $B = A$.

The remaining statement of the proposition is proved in the same way, with greatest common divisors and least common multiples interchanged. \square

The last two statements of Proposition 2 are referred to as the distributive laws, since if the greatest common divisor and least common multiple of a and b are denoted by $a \wedge b$ and $a \vee b$ respectively, they take the form

$$(a \vee b) \wedge (a \vee c) = a \vee (b \wedge c), \quad (a \wedge b) \vee (a \wedge c) = a \wedge (b \vee c).$$

Properties (i), (ii) and (v) at the beginning of the section say that divisibility is a *partial ordering* of the set \mathbb{N} with 1 as least element. The existence of greatest common divisors and least common multiples says that \mathbb{N} is a *lattice* with respect to this partial ordering. The distributive laws say that \mathbb{N} is actually a *distributive* lattice.

We say that $a, b \in \mathbb{N}$ are *relatively prime*, or *coprime*, if $(a, b) = 1$. Divisibility properties in this case are much simpler:

Proposition 3 For any $a, b, c \in \mathbb{N}$ with $(a, b) = 1$,

- (i) if $a|c$ and $b|c$, then $ab|c$;
- (ii) if $a|bc$, then $a|c$;
- (iii) $(a, bc) = (a, c)$;
- (iv) if also $(a, c) = 1$, then $(a, bc) = 1$;
- (v) $(a^m, b^n) = 1$ for all $m, n \geq 1$.

Proof To prove (i), note that $[a, b]$ divides c and $[a, b] = ab$. To prove (ii), note that a divides $(ac, bc) = (a, b)c = c$. To prove (iii), note that any common divisor of a and bc divides c , by (ii). Obviously (iii) implies (iv), and (v) follows by induction. \square

Proposition 4 If $a, b \in \mathbb{N}$ and $(a, b) = 1$, then any divisor of ab can be uniquely expressed in the form de , where $d|a$ and $e|b$. Conversely, any product of this form is a divisor of ab .

Proof The proof is based on Proposition 3. Suppose c divides ab and put $d = (a, c)$, $e = (b, c)$. Then $(d, e) = 1$ and hence de divides c . If $a = da'$ and $c = dc'$, then $(a', c') = 1$ and $e|c'$. On the other hand, $c'|a'b$ and hence $c'|b$. Since $e = (b, c)$, it follows that $c' = e$ and $c = de$.

Suppose $de = d'e'$, where d, d' divide a and e, e' divide b . Then $d|d'$, since $(d, e') = 1$, and similarly $d'|d$, since $(d', e) = 1$. Hence $d' = d$ and $e' = e$.

The final statement of the proposition is obvious. \square

It follows from Proposition 4 that if $c^n = ab$, where $(a, b) = 1$, then $a = d^n$ and $b = e^n$ for some $d, e \in \mathbb{N}$.

The greatest common divisor and least common multiple of any finite set of elements of \mathbb{N} may be defined in the same way as for sets of two elements. By induction we easily obtain:

Proposition 5 *Any $a_1, \dots, a_n \in \mathbb{N}$ have a greatest common divisor (a_1, \dots, a_n) and a least common multiple $[a_1, \dots, a_n]$. Moreover,*

- (i) $(a_1, a_2, \dots, a_n) = (a_1, (a_2, \dots, a_n))$, $[a_1, a_2, \dots, a_n] = [a_1, [a_2, \dots, a_n]]$;
- (ii) $(a_1c, \dots, a_nc) = (a_1, \dots, a_n)c$, $[a_1c, \dots, a_nc] = [a_1, \dots, a_n]c$;
- (iii) $(a_1, \dots, a_n) = a/[a/a_1, \dots, a/a_n]$, $[a_1, \dots, a_n] = a/(a/a_1, \dots, a/a_n)$, where $a = a_1 \cdots a_n$.

We can use the distributive laws to show that

$$([a, b], [a, c], [b, c]) = [(a, b), (a, c), (b, c)].$$

In fact the left side is equal to $\{a \vee (b \wedge c)\} \wedge (b \vee c)$, whereas the right side is equal to

$$\begin{aligned} (b \wedge c) \vee \{a \wedge (b \vee c)\} &= \{(b \wedge c) \vee a\} \wedge \{(b \wedge c) \vee (b \vee c)\} \\ &= \{a \vee (b \wedge c)\} \wedge (b \vee c). \end{aligned}$$

If

$$a = (a_1, \dots, a_m), \quad b = (b_1, \dots, b_n),$$

then ab is the greatest common divisor of all products $a_j b_k$, since $(a_j b_1, \dots, a_j b_n) = a_j b$ and $(a_1 b, \dots, a_m b) = ab$.

Similarly, if

$$a = [a_1, \dots, a_m], \quad b = [b_1, \dots, b_n],$$

then ab is the least common multiple of all products $a_j b_k$.

It is easily shown by induction that if $(a_i, a_j) = 1$ for $1 \leq i < j \leq m$, then

$$(a_1 \cdots a_m, c) = (a_1, c) \cdots (a_m, c), \quad [a_1 \cdots a_m, c] = [a_1, \dots, a_m, c].$$

Proposition 6 *If $a \in \mathbb{N}$ has two factorizations*

$$a = b_1 \cdots b_m = c_1 \cdots c_n,$$

then these factorizations have a common refinement, i.e. there exist $d_{jk} \in \mathbb{N}$ ($1 \leq j \leq m, 1 \leq k \leq n$) such that

$$b_j = \prod_{k=1}^n d_{jk}, \quad c_k = \prod_{j=1}^m d_{jk}.$$

Proof We show first that if $a = a_1 \cdots a_n$ and $d|a$, then $d = d_1 \cdots d_n$, where $d_i|a_i$ ($1 \leq i \leq n$). We may suppose that $n > 1$ and that the assertion holds for products of less than n elements of \mathbb{N} . Put $a' = a_1 \cdots a_{n-1}$ and $d' = (a', d)$. Then $d' = d_1 \cdots d_{n-1}$, where $d_i|a_i$ ($1 \leq i < n$). Moreover $a'' = a'/d'$ and $d'' = d/d'$ are coprime. Since $d'' = d/d'$ divides $a''a_n = a/d'$, the greatest common divisor $a_n = (a_na'', a_nd'')$ is divisible by d'' . Thus we can take $d_n = d''$.

We return now to the proposition. Since $c_1|\prod_j b_j$, we can write $c_1 = \prod_j d_{j1}$, where $d_{j1}|b_j$. Put $b'_j = b_j/d_{j1}$. Then

$$\prod_j b'_j = a/c_1 = c_2 \cdots c_n.$$

Hence we can write $c_2 = \prod_j d_{j2}$, where $d_{j2}|b'_j$. Proceeding in this way, we obtain factorizations $c_k = \prod_j d_{jk}$ such that $\prod_k d_{jk}$ divides b_j . In fact, since

$$\prod_{j,k} d_{jk} = a = \prod_j b_j,$$

we must have $b_j = \prod_k d_{jk}$. □

Instead of defining divisibility and greatest common divisors in the set \mathbb{N} of all positive integers, we can define them in the set \mathbb{Z} of all integers by simply replacing \mathbb{N} by \mathbb{Z} in the previous definitions. The properties (i)–(v) at the beginning of this section continue to hold, provided that in (iv) we require $c \neq 0$ and in (v) we alter the conclusion to $b = \pm a$. We now list some additional properties:

- (i)' $a|0$ for every a ;
- (ii)' if $0|a$, then $a = 0$;
- (iii)' if $c|a$ and $c|b$, then $c|ax + by$ for all x, y .

Greatest common divisors and least common multiples still exist, but uniqueness holds only up to sign. With this understanding, Propositions 2–4 continue to hold, and so also do Propositions 5 and 6 if we require $a \neq 0$. It is evident that, for every a ,

$$(a, 0) = a, \quad [a, 0] = 0.$$

More generally, we can define divisibility in any *integral domain*, i.e. a commutative ring in which $a \neq 0$ and $b \neq 0$ together imply $ab \neq 0$. The properties (i)–(v) at the beginning of the section continue to hold, provided that in (iv) we require $c \neq 0$ and in (v) we alter the conclusion to $b = ua$, where u is a *unit*, i.e. $u|1$. The properties (i)'–(iii)' above also remain valid.

We define a *GCD domain* to be an integral domain in which any pair of elements has a greatest common divisor. This implies that any pair of elements also has a least common multiple. Uniqueness now holds only up to unit multiples. With this understanding Propositions 2–6 continue to hold in any GCD domain in the same way as for \mathbb{Z} .

An important example, which we will consider in Section 3, of a GCD domain other than \mathbb{Z} is the *polynomial ring* $K[t]$, consisting of all polynomials in t with coefficients from an arbitrary field K . The units in this case are the nonzero elements of K .

Another example, which we will meet in §4 of Chapter VI, is the valuation ring R of a non-archimedean valued field. In this case, for any $a, b \in R$, either $a|b$ or $b|a$ and so (a, b) is either a or b .

In the same way that the ring \mathbb{Z} of integers may be embedded in the field \mathbb{Q} of rational numbers, any integral domain R may be embedded in a field K , its *field of fractions*, so that any nonzero $c \in K$ has the form $c = ab^{-1}$, where $a, b \in R$ and $b \neq 0$. If R is a GCD domain we can further require $(a, b) = 1$, and a, b are then uniquely determined apart from a common unit multiple. The field of fractions of the polynomial ring $K[t]$ is the field $K(t)$ of *rational functions*.

In our discussion of divisibility so far we have avoided all mention of prime numbers. A positive integer $a \neq 1$ is said to be *prime* if 1 and a are its only positive divisors, and otherwise is said to be *composite*.

For example, 2, 3 and 5 are primes, but $4 = 2 \times 2$ and $6 = 2 \times 3$ are composite. The significance of the primes is that, as far as multiplication is concerned, they are the ‘atoms’ and the composite integers are the ‘molecules’. This is made precise in the following so-called *fundamental theorem of arithmetic*:

Proposition 7 *If $a \in \mathbb{N}$ and $a \neq 1$, then a can be represented as a product of finitely many primes. Moreover, the representation is unique, except for the order of the factors.*

Proof Assume, on the contrary, that some composite $a_1 \in \mathbb{N}$ is not a product of finitely many primes. Since a_1 is composite, it has a factorization $a_1 = a_2 b_2$, where $a_2, b_2 \in \mathbb{N}$ and $a_2, b_2 \neq 1$. At least one of a_2, b_2 must be composite and not a product of finitely many primes, and we may choose the notation so that a_2 has these properties. The preceding argument can now be repeated with a_2 in place of a_1 . Proceeding in this way, we obtain an infinite sequence (a_k) of positive integers such that a_{k+1} divides a_k and $a_{k+1} \neq a_k$ for each $k \geq 1$. But then the sequence (a_k) has no least element, which contradicts Proposition I.3.

Suppose now that

$$a = p_1 \cdots p_m = q_1 \cdots q_n$$

are two representations of a as a product of primes. Then, by Proposition 6, there exist $d_{jk} \in \mathbb{N}$ ($1 \leq j \leq m, 1 \leq k \leq n$) such that

$$p_j = \prod_{k=1}^n d_{jk}, \quad q_k = \prod_{j=1}^m d_{jk}.$$

Since p_1 is a prime, we must have $d_{1k_1} = p_1$ for some $k_1 \in \{1, \dots, n\}$, and since q_{k_1} is a prime, we must have $q_{k_1} = d_{1k_1} = p_1$. The same argument can now be applied to

$$a' = \prod_{j \neq 1} p_j = \prod_{k \neq k_1} q_k.$$

It follows that $m = n$ and q_1, \dots, q_n is a permutation of p_1, \dots, p_m . □

It should be noted that factorization into primes would not be unique if we admitted 1 as a prime. The fundamental theorem of arithmetic may be reformulated in the following way: any $a \in \mathbb{N}$ can be uniquely represented in the form

$$a = \prod_p p^{\alpha_p},$$

where p runs through the primes and the α_p are non-negative integers, only finitely many of which are nonzero. It is easily seen that if $b \in \mathbb{N}$ has the analogous representation

$$b = \prod_p p^{\beta_p},$$

then $b|a$ if and only if $\beta_p \leq \alpha_p$ for all p . It follows that the greatest common divisor and least common multiple of a and b have the representations

$$(a, b) = \prod_p p^{\gamma_p}, \quad [a, b] = \prod_p p^{\delta_p},$$

where

$$\gamma_p = \min\{\alpha_p, \beta_p\}, \quad \delta_p = \max\{\alpha_p, \beta_p\}.$$

The fundamental theorem of arithmetic extends at once from \mathbb{N} to \mathbb{Q} : any nonzero rational number a can be uniquely represented in the form

$$a = u \prod_p p^{\alpha_p},$$

where $u = \pm 1$ is a unit, p runs through the primes and the α_p are integers (not necessarily non-negative), only finitely many of which are nonzero.

The following property of primes was already established in Euclid's *Elements* (Book VII, Proposition 30):

Proposition 8 *If p is a prime and $p|bc$, then $p|b$ or $p|c$.*

Proof If p does not divide b , we must have $(p, b) = 1$. But then p divides c , by Proposition 3(ii). \square

The property in Proposition 8 actually characterizes primes. For if a is composite, then $a = bc$, where $b, c \neq 1$. Thus $a|bc$, but $a \nmid b$ and $a \nmid c$.

We consider finally the extension of these notions to an arbitrary integral domain R . For any nonzero $a, b \in R$, we say that a divisor b of a is a *proper divisor* if a does not divide b (i.e., if a and b do not differ only by a unit factor). We say that $p \in R$ is *irreducible* if p is neither zero nor a unit and if every proper divisor of p is a unit. We say that $p \in R$ is *prime* if p is neither zero nor a unit and if $p|bc$ implies $p|b$ or $p|c$.

By what we have just said, the notions of 'prime' and 'irreducible' coincide if $R = \mathbb{Z}$, and the same argument applies if R is any GCD domain. However, in an arbitrary integral domain R , although any prime element is irreducible, an irreducible element need not be prime. (For example, in the integral domain R consisting of all complex numbers of the form $a + b\sqrt{-5}$, where $a, b \in \mathbb{Z}$, it may be seen that

$6 = 2 \times 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$ has two essentially distinct factorizations into irreducibles, and thus none of these irreducibles is prime.)

The proof of Proposition 7 shows that, in an arbitrary integral domain R , every element which is neither zero nor a unit can be represented as a product of finitely many irreducible elements if and only if the following *chain condition* is satisfied:

(#) *there exists no infinite sequence (a_n) of elements of R such that a_{n+1} is a proper divisor of a_n for every n .*

Furthermore, the representation is *essentially unique* (i.e. unique except for the order of the factors and for multiplying them by units) if and only if R is also a GCD domain.

An integral domain R is said to be *factorial* (or a ‘unique factorization domain’) if the ‘fundamental theorem of arithmetic’ holds in R , i.e. if every element which is neither zero nor a unit has such an essentially unique representation as a product of finitely many irreducibles. By the above remarks, an integral domain R is factorial if and only if it is a GCD domain satisfying the chain condition (#).

For future use, we define an element of a factorial domain to be *square-free* if it is neither zero nor a unit and if, in its representation as a product of irreducibles, no factor is repeated. In particular, a positive integer is square-free if and only if it is a nonempty product of distinct primes.

2 The Bézout Identity

If a, b are arbitrary integers with $a \neq 0$, then there exist unique integers q, r such that

$$b = qa + r, \quad 0 \leq r < |a|.$$

In fact qa is the greatest multiple of a which does not exceed b . The integers q and r are called the *quotient* and *remainder* in the ‘division’ of b by a .

(For $a > 0$ this was proved in Proposition I.14. It follows that if a and n are positive integers, any positive integer b less than a^n has a unique representation ‘to the base a ’:

$$b = b_0 + b_1a + \cdots + b_{n-1}a^{n-1},$$

where $0 \leq b_j < a$ for all j . In fact b_{n-1} is the quotient in the division of b by a^{n-1} , b_{n-2} is the quotient in the division of the remainder by a^{n-2} , and so on.)

If a, b are arbitrary integers with $a \neq 0$, then there exist also integers q, r such that

$$b = qa + r, \quad |r| \leq |a|/2.$$

In fact qa is the nearest multiple of a to b . Thus q and r are not uniquely determined if b is midway between two consecutive multiples of a .

Both these *division algorithms* have their uses. We will be impartial and merely use the fact that

$$b = qa + r, \quad |r| < |a|.$$

An *ideal* in the commutative ring \mathbb{Z} of all integers is defined to be a nonempty subset J such that if $a, b \in J$ and $x, y \in \mathbb{Z}$, then also $ax + by \in J$.

For example, if a_1, \dots, a_n are given elements of \mathbb{Z} , then the set of all linear combinations $a_1x_1 + \dots + a_nx_n$ with $x_1, \dots, x_n \in \mathbb{Z}$ is an ideal, the ideal *generated* by a_1, \dots, a_n . An ideal generated by a single element, i.e. the set of all multiples of that element, is said to be a *principal ideal*.

Lemma 9 *Any ideal J in the ring \mathbb{Z} is a principal ideal.*

Proof If 0 is the only element of J , then 0 generates J . Otherwise there is a nonzero $a \in J$ with minimum absolute value. For any $b \in J$, we can write $b = qa + r$, for some $q, r \in \mathbb{Z}$ with $|r| < |a|$. By the definition of an ideal, $r \in J$ and so, by the definition of a , $r = 0$. Thus a generates J . \square

Proposition 10 *Any $a, b \in \mathbb{Z}$ have a greatest common divisor $d = (a, b)$. Moreover, for any $c \in \mathbb{Z}$, there exist $x, y \in \mathbb{Z}$ such that*

$$ax + by = c$$

if and only if d divides c .

Proof Let J be the ideal generated by a and b . By Lemma 9, J is generated by a single element d . Since $a, b \in J$, d is a common divisor of a and b . On the other hand, since $d \in J$, there exist $u, v \in \mathbb{Z}$ such that $d = au + bv$. Hence any common divisor of a and b also divides d . Thus $d = (a, b)$. The final statement of the proposition follows immediately since, by definition, $c \in J$ if and only if there exist $x, y \in \mathbb{Z}$ such that $ax + by = c$. \square

It is readily shown that if the ‘linear Diophantine’ equation $ax + by = c$ has a solution $x_0, y_0 \in \mathbb{Z}$, then all solutions $x, y \in \mathbb{Z}$ are given by the formula

$$x = x_0 + kb/d, \quad y = y_0 - ka/d,$$

where $d = (a, b)$ and k is an arbitrary integer.

Proposition 10 provides a new proof for the existence of greatest common divisors and, in addition, it shows that the greatest common divisor of two integers can be represented as a linear combination of them. This representation is usually referred to as the *Bézout identity*, although it was already known to Bachet (1624) and even earlier to the Hindu mathematicians Aryabhata (499) and Brahmagupta (628).

In exactly the same way that we proved Proposition 10 – or, alternatively, by induction from Proposition 10 – we can prove

Proposition 11 *Any finite set a_1, \dots, a_n of elements of \mathbb{Z} has a greatest common divisor $d = (a_1, \dots, a_n)$. Moreover, for any $c \in \mathbb{Z}$, there exist $x_1, \dots, x_n \in \mathbb{Z}$ such that*

$$a_1x_1 + \dots + a_nx_n = c$$

if and only if d divides c .

The proof which we gave for Proposition 10 is a pure existence proof – it does not help us to find the greatest common divisor. The following constructive proof was already given in Euclid’s *Elements* (Book VII, Proposition 2). Let a, b be arbitrary

integers. Since $(0, b) = b$, we may assume $a \neq 0$. Then there exist integers q, r such that

$$b = qa + r, \quad |r| < |a|.$$

Put $a_0 = b, a_1 = a$ and repeatedly apply this procedure:

$$a_0 = q_1 a_1 + a_2, \quad |a_2| < |a_1|,$$

$$a_1 = q_2 a_2 + a_3, \quad |a_3| < |a_2|,$$

...

$$a_{N-2} = q_{N-1} a_{N-1} + a_N, \quad |a_N| < |a_{N-1}|,$$

$$a_{N-1} = q_N a_N.$$

The process must eventually terminate as shown, because otherwise we would obtain an infinite sequence of positive integers with no least element. We claim that a_N is a greatest common divisor of a and b . In fact, working forwards from the first equation we see that any common divisor c of a and b divides each a_k and so, in particular, a_N . On the other hand, working backwards from the last equation we see that a_N divides each a_k and so, in particular, a and b .

The Bézout identity can also be obtained in this way, although Euclid himself lacked the necessary algebraic notation. Define sequences $(x_k), (y_k)$ by the recurrence relations

$$x_{k+1} = x_{k-1} - q_k x_k, \quad y_{k+1} = y_{k-1} - q_k y_k \quad (1 \leq k < N),$$

with the starting values

$$x_0 = 0, \quad x_1 = 1, \quad \text{resp. } y_0 = 1, \quad y_1 = 0.$$

It is easily shown by induction that $a_k = ax_k + by_k$ and so, in particular, $a_N = ax_N + by_N$.

The Euclidean algorithm is quite practical. For example, the reader may use it to verify that 13 is the greatest common divisor of 2171 and 5317, and that

$$49 \times 5317 - 120 \times 2171 = 13.$$

However, the first proof given for Proposition 10 also has its uses: there is some advantage in separating the conceptual from the computational and the proof actually rests on more general principles, since there are quadratic number fields whose ring of integers is a 'principal ideal domain' that does not possess any Euclidean algorithm.

It is not visibly obvious that the binomial coefficients

$${}^{m+n}C_n = (m+1) \cdots (m+n)/1 \cdot 2 \cdots n$$

are integers for all positive integers m, n , although it is apparent from their combinatorial interpretation. However, the property is readily proved by induction, using the relation

$${}^{m+n}C_n = {}^{m+n-1}C_n + {}^{m+n-1}C_{n-1}.$$

Binomial coefficients have other arithmetic properties. Hermite observed that $^{m+n}C_n$ is divisible by the integers $(m+n)/(m, n)$ and $(m+1)/(m+1, n)$. In particular, the Catalan numbers $(n+1)^{-1} {}^{2n}C_n$ are integers. The following proposition is a substantial generalization of these results and illustrates the application of Proposition 10.

Proposition 12 *Let (a_n) be a sequence of nonzero integers such that, for all $m, n \geq 1$, every common divisor of a_m and a_n divides a_{m+n} , and every common divisor of a_m and a_{m+n} divides a_n . Then, for all $m, n \geq 1$,*

- (i) $(a_m, a_n) = a_{(m,n)}$;
- (ii) $A_{m,n} := a_{m+1} \cdots a_{m+n} / a_1 \cdots a_n \in \mathbb{Z}$;
- (iii) $A_{m,n}$ is divisible by $a_{m+n}/(a_m, a_n)$, by $a_{m+1}/(a_{m+1}, a_n)$ and by $a_{n+1}/(a_m, a_{n+1})$;
- (iv) $(A_{m,n-1}, A_{m+1,n}, A_{m-1,n+1}) = (A_{m-1,n}, A_{m+1,n-1}, A_{m,n+1})$.

Proof The hypotheses imply that

$$(a_m, a_n) = (a_m, a_{m+n}) \quad \text{for all } m, n \geq 1.$$

Since $a_m = (a_m, a_m)$, it follows by induction that $a_m | a_{km}$ for all $k \geq 1$. Moreover,

$$(a_{km}, a_{(k+1)m}) = a_m,$$

since every common divisor of a_{km} and $a_{(k+1)m}$ divides a_m .

Put $d = (m, n)$. Then $m = dm'$, $n = dn'$, where $(m', n') = 1$. Thus there exist integers u, v such that $m'u - n'v = 1$. By replacing u, v by $u + tn'$, $v + tm'$ with any $t > \max\{|u|, |v|\}$, we may assume that u and v are both positive. Then

$$(a_{mu}, a_{nv}) = (a_{(n'v+1)d}, a_{n'vd}) = a_d.$$

Since a_d divides (a_m, a_n) and (a_m, a_n) divides (a_{mu}, a_{nv}) , this implies $(a_m, a_n) = a_d$. This proves (i).

Since $a_1 | a_{m+1}$, it is evident that $A_{m,1} \in \mathbb{Z}$ for all $m \geq 1$. We assume that $n > 1$ and $A_{m,n} \in \mathbb{Z}$ for all smaller values of n and all $m \geq 1$. Since it is trivial that $A_{0,n} \in \mathbb{Z}$, we assume also that $m \geq 1$ and $A_{m,n} \in \mathbb{Z}$ for all smaller values of m . By Proposition 10, there exist $x, y \in \mathbb{Z}$ such that

$$a_m x + a_n y = a_{m+n},$$

since (a_m, a_n) divides a_{m+n} . Since

$$A_{m,n} = \frac{a_{m+1} \cdots a_{m+n}}{a_1 \cdots a_n} = \frac{a_m a_{m+1} \cdots a_{m+n-1}}{a_1 \cdots a_n} x + \frac{a_{m+1} \cdots a_{m+n-1}}{a_1 \cdots a_{n-1}} y,$$

our induction hypotheses imply that $A_{m,n} \in \mathbb{Z}$. This proves (ii).

Since

$$a_{m+n} A_{m,n-1} = a_n A_{m,n},$$

a_{m+n} divides $(a_n, a_{m+n}) A_{m,n}$ and, since $(a_n, a_{m+n}) = (a_m, a_n)$, this in turn implies that $a_{m+n}/(a_m, a_n)$ divides $A_{m,n}$.

Similarly, since

$$a_{m+1}A_{m+1,n} = a_{m+n+1}A_{m,n}, \quad a_{m+1}A_{m+1,n-1} = a_nA_{m,n},$$

a_{m+1} divides $(a_n, a_{m+n+1})A_{m,n}$ and, since $(a_n, a_{m+n+1}) = (a_{m+1}, a_n)$, it follows that $a_{m+1}/(a_{m+1}, a_n)$ divides $A_{m,n}$. In the same way, since

$$a_{n+1}A_{m,n+1} = a_{m+n+1}A_{m,n}, \quad a_{n+1}A_{m-1,n+1} = a_mA_{m,n},$$

a_{n+1} divides $(a_m, a_{m+n+1})A_{m,n}$ and hence $a_{n+1}/(a_m, a_{n+1})$ divides $A_{m,n}$. This proves (iii).

By multiplying by $a_1 \cdots a_{n+1}/a_{m+2} \cdots a_{m+n-1}$, we see that (iv) is equivalent to

$$\begin{aligned} & (a_n a_{n+1} a_{m+1}, a_{n+1} a_{m+n} a_{m+n+1}, a_m a_{m+1} a_{m+n}) \\ &= (a_{n+1} a_m a_{m+1}, a_n a_{n+1} a_{m+n}, a_{m+1} a_{m+n} a_{m+n+1}). \end{aligned}$$

Since here the two sides are interchanged when m and n are interchanged, it is sufficient to show that any common divisor e of the three terms on the right is also a common divisor of the three terms on the left. We have

$$\begin{aligned} (a_{n+1} a_m a_{m+1}, a_n a_{n+1} a_{m+n}) &= a_{n+1} a_{m+1} (a_m, a_n) = a_{n+1} a_{m+1} (a_m, a_{m+n}) \\ &= (a_{n+1} a_m a_{m+1}, a_{m+1} a_{n+1} a_{m+n}), \end{aligned}$$

and similarly

$$\begin{aligned} (a_n a_{n+1} a_{m+n}, a_{n+1} a_{m+n} a_{m+n+1}) &= (a_n a_{n+1} a_{m+n}, a_{m+1} a_{n+1} a_{m+n}), \\ (a_{m+1} a_{m+n} a_{m+n+1}, a_m a_{m+1} a_{m+n}) &= (a_{m+1} a_{m+n} a_{m+n+1}, a_{m+1} a_{n+1} a_{m+n}). \end{aligned}$$

Hence if we put $g = a_{m+1} a_{n+1} a_{m+n}$, then

$$(e, g) = (e, a_n a_{n+1} a_{m+1}) = (e, a_{n+1} a_{m+n} a_{m+n+1}) = (e, a_m a_{m+1} a_{m+n})$$

and if we put $f = (e, g)$, then

$$1 = (e/f, a_n a_{n+1} a_{m+1}/f) = (e/f, a_{n+1} a_{m+n} a_{m+n+1}/f) = (e/f, a_m a_{m+1} a_{m+n}/f).$$

Hence $(e/f, P/f^3) = 1$, where

$$P = a_n a_{n+1} a_{m+1} \cdot a_{n+1} a_{m+n} a_{m+n+1} \cdot a_m a_{m+1} a_{m+n}.$$

But P is divisible by e^3 , since we can also write

$$P = a_{n+1} a_m a_{m+1} \cdot a_n a_{n+1} a_{m+n} \cdot a_{m+1} a_{m+n} a_{m+n+1}.$$

Hence the previous relation implies $e/f = 1$. Thus $e = f$ is a common divisor of $a_n a_{n+1} a_{m+1}$, $a_{n+1} a_{m+n} a_{m+n+1}$ and $a_m a_{m+1} a_{m+n}$, as we wished to show. \square

For the binomial coefficient case, i.e. $a_n = n$, the property (iv) of Proposition 12 was discovered empirically by Gould (1972) and then proved by Hillman and Hoggatt (1972). It states that if in the *Pascal triangle* one picks out the hexagon surrounding a particular element, then the greatest common divisor of three alternately

chosen vertices is equal to the greatest common divisor of the remaining three vertices. Hillman and Hoggatt also gave generalizations along the lines of Proposition 12.

The hypotheses of Proposition 12 are also satisfied if $a_n = q^n - 1$, for some integer $q > 1$, since in this case $a_{m+n} = a_m a_n + a_m + a_n$. The corresponding q -binomial coefficients were studied by Gauss and, as mentioned in Chapter XIII, they play a role in the theory of partitions.

We may also take (a_n) to be the sequence defined recurrently by

$$a_1 = 1, \quad a_2 = c, \quad a_{n+2} = ca_{n+1} + ba_n \quad (n \geq 1),$$

where b and c are coprime positive integers. Indeed it is easily shown by induction that

$$(a_n, a_{n+1}) = (b, a_{n+1}) = 1 \quad \text{for all } n \geq 1.$$

By induction on m one may also show that

$$a_{m+n} = a_{m+1}a_n + ba_m a_{n-1} \quad \text{for all } m \geq 1, n > 1.$$

It follows that the hypotheses of Proposition 12 are satisfied. In particular, for $b = c = 1$, they are satisfied by the sequence of *Fibonacci numbers*.

We consider finally extensions of our results to more general algebraic structures. An integral domain R is said to be a *Bézout domain* if any $a, b \in R$ have a common divisor of the form $au + bv$ for some $u, v \in R$. Since such a common divisor is necessarily a greatest common divisor, any Bézout domain is a GCD domain. It is easily seen, by induction on the number of generators, that an integral domain is a Bézout domain if and only if every finitely generated ideal is a principal ideal. Thus Propositions 10 and 11 continue to hold if \mathbb{Z} is replaced by any Bézout domain.

An integral domain R is said to be a *principal ideal domain* if every ideal is a principal ideal.

Lemma 13 *An integral domain R is a principal ideal domain if and only if it is a Bézout domain satisfying the chain condition*

(#) *there exists no infinite sequence (a_n) of elements of R such that a_{n+1} is a proper divisor of a_n for every n .*

Proof It is obvious that any principal ideal domain is a Bézout domain. Suppose R is a Bézout domain, but not a principal ideal domain. Then R contains an ideal J which is not finitely generated. Hence there exists a sequence (b_n) of elements of J such that b_{n+1} is not in the ideal J_n generated by b_1, \dots, b_n . But J_n is a principal ideal. If a_n generates J_n , then a_{n+1} is a proper divisor of a_n for every n . Thus the chain condition is violated.

Suppose now that R is a Bézout domain containing a sequence (a_n) such that a_{n+1} is a proper divisor of a_n for every n . Let J denote the set of all elements of R which are divisible by at least one term of this sequence. Then J is an ideal. For if $a_j | b$ and $a_k | c$, where $j \leq k$, then also $a_k | b$ and hence $a_k | bx + cy$ for all $x, y \in R$. If J were generated by a single element a , we would have $a | a_n$ for every n . On the other hand, since $a \in J$, $a_N | a$ for some N . Hence $a_N | a_{N+1}$. Since a_{N+1} is a proper divisor of a_N , this is a contradiction. Thus R is not a principal ideal domain. \square

It follows from the remarks at the end of Section 1 that a principal ideal domain is factorial, i.e. any element which is neither zero nor a unit can be represented as a product of finitely many irreducibles and the representation is essentially unique.

In the next section we will show that the ring $K[t]$ of all polynomials in one indeterminate t with coefficients from an arbitrary field K is a principal ideal domain.

It may be shown that the ring of all algebraic integers is a Bézout domain, and likewise the ring of all functions which are holomorphic in a nonempty connected open subset G of the complex plane \mathbb{C} . However, neither is a principal ideal domain. In the former case there are no irreducibles, since any algebraic integer a has the factorization $a = \sqrt{a} \cdot \sqrt{a}$. In the latter case $z - \zeta$ is an irreducible for any $\zeta \in G$, but the chain condition is violated. For example, take

$$a_n(z) = f(z)/(z - \zeta_1) \cdots (z - \zeta_n),$$

where $f(z)$ is a non-identically vanishing function which is holomorphic in G and has infinitely many zeros ζ_1, ζ_2, \dots in G .

3 Polynomials

In this section we study the most important example of a principal ideal domain other than \mathbb{Z} , namely the ring $K[t]$ of all polynomials in t with coefficients from an arbitrary field K (e.g., $K = \mathbb{Q}$ or \mathbb{C}).

The attitude adopted towards polynomials in algebra is different from that adopted in analysis. In analysis we regard ' t ' as a variable which can take different values; in algebra we regard ' t ' simply as a symbol, an 'indeterminate', on which we can perform various algebraic operations. Since the concept of function is so pervasive, the algebraic approach often seems mysterious at first sight and it seems worthwhile taking the time to give a precise meaning to an 'indeterminate'.

Let R be an integral domain (e.g., $R = \mathbb{Z}$ or \mathbb{Q}). A polynomial with coefficients from R is defined to be a sequence $f = (a_0, a_1, a_2, \dots)$ of elements of R in which at most finitely many terms are nonzero. The sum and product of two polynomials

$$f = (a_0, a_1, a_2, \dots), \quad g = (b_0, b_1, b_2, \dots)$$

are defined by

$$\begin{aligned} f + g &= (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots), \\ fg &= (a_0b_0, a_0b_1 + a_1b_0, a_0b_2 + a_1b_1 + a_2b_0, \dots). \end{aligned}$$

It is easily verified that these are again polynomials and that the set $R[t]$ of all polynomials with coefficients from R is a commutative ring with $O = (0, 0, 0, \dots)$ as zero element. (By dropping the requirement that at most finitely many terms are nonzero, we obtain the ring $R[[t]]$ of all formal power series with coefficients from R .)

We define the degree $\partial(f)$ of a polynomial $f = (a_0, a_1, a_2, \dots) \neq O$ to be the greatest integer n for which $a_n \neq 0$ and we put

$$|f| = 2^{\partial(f)}, \quad |O| = 0.$$

It is easily verified that, for all polynomials f, g ,

$$|f + g| \leq \max\{|f|, |g|\}, \quad |fg| = |f||g|.$$

Since $|f| \geq 0$, with equality if and only if $f = O$, the last property implies that $R[t]$ is an integral domain. Thus we can define divisibility in $R[t]$, as explained in Section 1.

The set of all polynomials of the form $(a_0, 0, 0, \dots)$ is a subdomain isomorphic to R . By identifying this set with R , we may regard R as embedded in $R[t]$. The only units in $R[t]$ are the units in R , since $1 = ef$ implies $1 = |e||f|$ and hence $|e| = 1$.

If we put $t = (0, 1, 0, 0, \dots)$, then

$$t^2 = tt = (0, 0, 1, 0, \dots), \quad t^3 = tt^2 = (0, 0, 0, 1, \dots), \dots$$

Hence if the polynomial $f = (a_0, a_1, a_2, \dots)$ has degree n , then it can be uniquely expressed in the form

$$f = a_0 + a_1t + \dots + a_nt^n \quad (a_n \neq 0).$$

We refer to the elements a_0, a_1, \dots, a_n of R as the *coefficients* of f . In particular, a_0 is the *constant* coefficient and a_n the *highest* coefficient. We say that f is *monic* if its highest coefficient $a_n = 1$.

If also

$$g = b_0 + b_1t + \dots + b_mt^m \quad (b_m \neq 0),$$

then the sum and product assume their familiar forms:

$$\begin{aligned} f + g &= (a_0 + b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2 + \dots, \\ fg &= a_0b_0 + (a_0b_1 + a_1b_0)t + (a_0b_2 + a_1b_1 + a_2b_0)t^2 + \dots. \end{aligned}$$

Suppose now that $R = K$ is a field, and let

$$\begin{aligned} f &= a_0 + a_1t + \dots + a_nt^n \quad (a_n \neq 0), \\ g &= b_0 + b_1t + \dots + b_mt^m \quad (b_m \neq 0) \end{aligned}$$

be any two nonzero elements of $K[t]$. If $|g| < |f|$, i.e. if $m < n$, then $g = qf + r$, with $q = O$ and $r = g$. Suppose on the other hand that $|f| \leq |g|$. Then

$$g = a_n^{-1}b_mt^{m-n}f + g^\dagger,$$

where $g^\dagger \in K[t]$ and $|g^\dagger| < |g|$. If $|f| \leq |g^\dagger|$, the process can be repeated with g^\dagger in place of g . Continuing in this way, we obtain $q, r \in K[t]$ such that

$$g = qf + r, \quad |r| < |f|.$$

Moreover, q and r are uniquely determined, since if also

$$g = q_1f + r_1, \quad |r_1| < |f|,$$

then

$$(q - q_1)f = r_1 - r, \quad |r_1 - r| < |f|,$$

which is only possible if $q = q_1$.

Ideals in $K[t]$ can be defined in the same way as for \mathbb{Z} and the proof of Lemma 9 remains valid. Thus $K[t]$ is a principal ideal domain and, *a fortiori*, a GCD domain.

The Euclidean algorithm can also be applied in $K[t]$ in the same way as for \mathbb{Z} and again, from the sequence of polynomials f_0, f_1, \dots, f_N which it provides to determine the greatest common divisor f_N of f_0 and f_1 we can obtain polynomials u_k, v_k such that

$$f_k = f_1 u_k + f_0 v_k \quad (0 \leq k \leq N).$$

We can actually say more for polynomials than for integers, since if

$$f_{k-1} = q_k f_k + f_{k+1}, \quad |f_{k+1}| < |f_k|,$$

then $|f_{k-1}| = |q_k| |f_k|$ and hence, by induction,

$$|f_{k-1}| |u_k| = |f_0|, \quad |f_{k-1}| |v_k| = |f_1| \quad (1 \leq k \leq N).$$

It may be noted in passing that the Euclidean algorithm can also be applied in the ring $K[t, t^{-1}]$ of *Laurent polynomials*. A Laurent polynomial $f \neq O$, with coefficients from the field K , has the form

$$f = a_m t^m + a_{m+1} t^{m+1} + \dots + a_n t^n,$$

where $m, n \in \mathbb{Z}$ with $m \leq n$ and $a_j \in K$ with $a_m a_n \neq 0$. Thus we can write $f = t^m f_0$, where $f_0 \in K[t]$. Put

$$|f| = 2^{n-m}, \quad |O| = 0;$$

then the division algorithm for ordinary polynomials implies one for Laurent polynomials: for any $f, g \in K[t, t^{-1}]$ with $f \neq O$, there exist $q, r \in K[t, t^{-1}]$ such that $g = qf + r$, $|r| < |f|$.

We return now to ordinary polynomials. The general definition for integral domains in Section 1 means, in the present case, that a polynomial $p \in K[t]$ is *irreducible* if it has positive degree and if every proper divisor has degree zero.

It follows that any polynomial of degree 1 is irreducible. However, there may exist also irreducible polynomials of higher degree. For example, we will show shortly that the polynomial $t^2 - 2$ is irreducible in $\mathbb{Q}[t]$. For $K = \mathbb{C}$, however, every irreducible polynomial has degree 1, by the fundamental theorem of algebra (Theorem I.30) and Proposition 14 below. It follows that, for $K = \mathbb{R}$, every irreducible polynomial has degree 1 or 2. (For if a real polynomial $f(t)$ has a root $\alpha \in \mathbb{C} \setminus \mathbb{R}$, its conjugate $\bar{\alpha}$ is also a root and $f(t)$ has the real irreducible factor $(t - \alpha)(t - \bar{\alpha})$.)

It is obvious that the chain condition (#) of Section 1 holds in the integral domain $K[t]$, since if g is a proper divisor of f , then $|g| < |f|$. It follows that any polynomial of positive degree can be represented as a product of finitely many irreducible polynomials and that the representation is essentially unique.

We now consider the connection between polynomials in the sense of algebra (polynomial forms) and polynomials in the sense of analysis (polynomial functions). Let K be a field and $f \in K[t]$:

$$f = a_0 + a_1t + \cdots + a_nt^n.$$

If we replace ' t ' by $c \in K$ we obtain an element of K , which we denote by $f(c)$:

$$f(c) = a_0 + a_1c + \cdots + a_nc^n.$$

A rapid procedure ('Horner's rule') for calculating $f(c)$ is to use the recurrence relations

$$f_0 = a_n, \quad f_j = f_{j-1}c + a_{n-j} \quad (j = 1, \dots, n).$$

It is readily shown by induction that

$$f_j = a_nc^j + a_{n-1}c^{j-1} + \cdots + a_{n-j},$$

and hence $f(c) = f_n$ is obtained with just n multiplications and n additions.

It is easily seen that $f = g + h$ implies $f(c) = g(c) + h(c)$, and $f = gh$ implies $f(c) = g(c)h(c)$. Thus the mapping $f \rightarrow f(c)$ is a 'homomorphism' of $K[t]$ into K . A simple consequence is the so-called *remainder theorem*:

Proposition 14 *Let K be a field and $c \in K$. If $f \in K[t]$, then*

$$f = (t - c)g + f(c),$$

for some $g \in K[t]$.

In particular, f is divisible by $t - c$ if and only if $f(c) = 0$.

Proof We already know that there exist $q, r \in K[t]$ such that

$$f = (t - c)q + r, \quad |r| \leq 1.$$

Thus $r \in K$ and the homomorphism properties imply that $f(c) = r$. □

We say that $c \in K$ is a *root* of the polynomial $f \in K[t]$ if $f(c) = 0$.

Proposition 15 *Let K be a field. If $f \in K[t]$ is a polynomial of degree $n \geq 0$, then f has at most n distinct roots in K .*

Proof If f is of degree 0, then $f = c$ is a nonzero element of K and f has no roots. Suppose now that $n \geq 1$ and the result holds for polynomials of degree less than n . If c is a root of f then, by Proposition 14, $f = (t - c)g$ for some $g \in K[t]$. Since g has degree $n - 1$, it has at most $n - 1$ roots. But every root of f distinct from c is a root of g . Hence f has at most n roots. □

We consider next properties of the integral domain $R[t]$, when R is an integral domain rather than a field (e.g., $R = \mathbb{Z}$). The famous Pythagorean proof that $\sqrt{2}$ is irrational is considerably generalized by the following result:

Proposition 16 *Let R be a GCD domain and K its field of fractions. Let*

$$f = a_0 + a_1t + \cdots + a_nt^n$$

be a polynomial of degree $n > 0$ with coefficients $a_j \in R$ ($0 \leq j \leq n$). If $c \in K$ is a root of f and $c = ab^{-1}$, where $a, b \in R$ and $(a, b) = 1$, then $b|a_n$ and $a|a_0$.

In particular, if f is monic, then $c \in R$.

Proof We have

$$a_0b^n + a_1ab^{n-1} + \cdots + a_{n-1}a^{n-1}b + a_na^n = 0.$$

Hence $b|a_na^n$ and $a|a_0b^n$. Since $(a^n, b) = (a, b^n) = 1$, by Proposition 3(v), the result follows from Proposition 3(ii). \square

The polynomial $t^2 - 2$ has no integer roots, since $0, 1, -1$ are not roots and if $c \in \mathbb{Z}$ and $c \neq 0, 1, -1$, then $c^2 \geq 4$. Consequently, by Proposition 16, the polynomial $t^2 - 2$ also has no rational roots. It now follows from Proposition 14 that $t^2 - 2$ is irreducible in $\mathbb{Q}[t]$, since it has no divisors of degree 1.

Proposition 16 was known to Euler (1774) for the case $R = \mathbb{Z}$. In this case it shows that to obtain all rational roots of a polynomial with rational coefficients we need test only a finite number of possibilities, which can be explicitly enumerated. For example, if $z \in \mathbb{Z}$, the cubic polynomial $t^3 + zt + 1$ has no rational roots unless $z = 0$ or $z = -2$.

It was shown by Gauss (1801), again for the case $R = \mathbb{Z}$, that Proposition 16 may itself be considerably generalized. His result may be formulated in the following way:

Proposition 17 *Let $f, g \in R[t]$, where R is a GCD domain with field of fractions K . Then g divides f in $R[t]$ if and only if g divides f in $K[t]$ and the greatest common divisor of the coefficients of g divides the greatest common divisor of the coefficients of f .*

Proof For any polynomial $f \in R[t]$, let $c(f)$ denote the greatest common divisor of its coefficients. We say that f is *primitive* if $c(f) = 1$. We show first that the product $f = gh$ of two primitive polynomials g, h is again primitive.

Let

$$g = b_0 + b_1t + \cdots, \quad h = c_0 + c_1t + \cdots, \quad f = a_0 + a_1t + \cdots,$$

and assume on the contrary that the coefficients a_i have a common divisor d which is not a unit. Then d does not divide all the coefficients b_j , nor all the coefficients c_k . Let b_m, c_n be the first coefficients of g, h which are not divisible by d . Then

$$a_{m+n} = \sum_{j+k=m+n} b_j c_k$$

and d divides every term on the right, except possibly $b_m c_n$. In fact, since $d|a_{m+n}$, d must also divide $b_m c_n$. Hence we cannot have both $(d, b_m) = 1$ and $(d, c_n) = 1$.

Consequently we can replace d by a proper divisor d' , again not a unit, for which $m' + n' > m + n$. Since there exists a divisor d for which $m + n$ is a maximum, this yields a contradiction.

Now let f, g be polynomials in $R[t]$ such that g divides f in $K[t]$. Thus $f = gH$, where $H \in K[t]$. We can write $H = ab^{-1}h_0$, where a, b are coprime elements of R and h_0 is a primitive polynomial in $R[t]$. Also

$$f = c(f)f_0, \quad g = c(g)g_0,$$

where f_0, g_0 are primitive polynomials in $R[t]$. Hence

$$bc(f)f_0 = ac(g)g_0h_0.$$

Since g_0h_0 is primitive, it follows that

$$bc(f) = ac(g).$$

If $H \in R[t]$, then $b = 1$ and so $c(g)|c(f)$. On the other hand, if $c(g)|c(f)$, then $bc(f)/c(g) = a$. Since $(a, b) = 1$, this implies that $b = 1$ and $H \in R[t]$. \square

Corollary 18 *If R is a GCD domain, then $R[t]$ is also a GCD domain. If, moreover, R is a factorial domain, then $R[t]$ is also a factorial domain.*

proof Let K denote the field of fractions of R . Since $K[t]$ is a GCD domain and $R[t] \subseteq K[t]$, $R[t]$ is certainly an integral domain. If $f, g \in R[t]$, then there exists a primitive polynomial $h_0 \in R[t]$ which is a greatest common divisor of f and g in $K[t]$. It follows from Proposition 17 that

$$h = (c(f), c(g))h_0$$

is a greatest common divisor of f and g in $R[t]$.

This proves the first statement of the corollary. It remains to show that if R also satisfies the chain condition (#), then $R[t]$ does likewise. But if $f_n \in R[t]$ and $f_{n+1}|f_n$ for every n , then f_n must be of constant degree for all large n . The second statement of the corollary now also follows from Proposition 17 and the chain condition in R . \square

It follows by induction that in the statement of Corollary 18 we may replace $R[t]$ by the ring $R[t_1, \dots, t_m]$ of all polynomials in finitely many indeterminates t_1, \dots, t_m with coefficients from R . In particular, if K is a field, then any polynomial $f \in K[t_1, \dots, t_m]$ such that $f \notin K$ can be represented as a product of finitely many irreducible polynomials and the representation is essentially unique.

It is now easy to give examples of GCD domains which are not Bézout domains. Let R be a GCD domain which is not a field (e.g., $R = \mathbb{Z}$). Then some $a_0 \in R$ is neither zero nor a unit. By Corollary 18, $R[t]$ is a GCD domain and, by Proposition 17, the greatest common divisor in $R[t]$ of the polynomials a_0 and t is 1. If there existed $g, h \in R[t]$ such that

$$a_0g + th = 1,$$

where $g = b_0 + b_1t + \cdots$, then by equating constant coefficients we would obtain $a_0b_0 = 1$, which is a contradiction. Thus $R[t]$ is not a Bézout domain.

As an application of the preceding results we show that if a_1, \dots, a_n are distinct integers, then the polynomial

$$f = \prod_{j=1}^n (t - a_j) - 1$$

is irreducible in $\mathbb{Q}[t]$. Assume, on the contrary, that $f = gh$, where $g, h \in \mathbb{Q}[t]$ and have positive degree. We may suppose without loss of generality that $g \in \mathbb{Z}[t]$ and that the greatest common divisor of the coefficients of g is 1. Since $f \in \mathbb{Z}[t]$, it then follows from Proposition 17 that also $h \in \mathbb{Z}[t]$. Thus $g(a_j)$ and $h(a_j)$ are integers for every j . Since $g(a_j)h(a_j) = -1$, it follows that $g(a_j) = -h(a_j)$. Thus the polynomial $g + h$ has the distinct roots a_1, \dots, a_n . Since $g + h$ has degree less than n , this is possible only if $g + h = 0$. Hence $f = -g^2$. But, since the highest coefficient of f is 1, this is a contradiction.

In general, it is not an easy matter to determine if a polynomial with rational coefficients is irreducible in $\mathbb{Q}[t]$. However, the following *irreducibility criterion*, due to Eisenstein (1850), is sometimes useful:

Proposition 19 *If*

$$f(t) = a_0 + a_1t + \cdots + a_{n-1}t^{n-1} + t^n$$

is a monic polynomial of degree n with integer coefficients such that a_0, a_1, \dots, a_{n-1} are all divisible by some prime p , but a_0 is not divisible by p^2 , then f is irreducible in $\mathbb{Q}[t]$.

Proof Assume on the contrary that f is reducible. Then there exist polynomials $g(t), h(t)$ of positive degrees l, m with integer coefficients such that $f = gh$. If

$$\begin{aligned} g(t) &= b_0 + b_1t + \cdots + b_lt^l, \\ h(t) &= c_0 + c_1t + \cdots + c_mt^m, \end{aligned}$$

then $a_0 = b_0c_0$. The hypotheses imply that exactly one of b_0, c_0 is divisible by p . Without loss of generality, assume it to be b_0 . Since p divides $a_1 = b_0c_1 + b_1c_0$, it follows that $p|b_1$. Since p divides $a_2 = b_0c_2 + b_1c_1 + b_2c_0$, it now follows that $p|b_2$. Proceeding in this way, we see that p divides b_j for every $j \leq l$. But, since $b_lc_m = 1$, this yields a contradiction. \square

It follows from Proposition 19 that, for any prime p , the p -th *cyclotomic polynomial*

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

is irreducible in $\mathbb{Q}[x]$. For $\Phi_p(x) = (x^p - 1)/(x - 1)$ and, if we put $x = 1 + t$, the transformed polynomial

$$\{(1+t)^p - 1\}/t = t^{p-1} + {}^pC_{p-1}t^{p-2} + \cdots + {}^pC_2t + p$$

satisfies the hypotheses of Proposition 19.

For any field K , we define the *formal derivative* of a polynomial $f \in K[t]$,

$$f = a_0 + a_1t + \cdots + a_nt^n,$$

to be the polynomial

$$f' = a_1 + 2a_2t + \cdots + na_nt^{n-1}.$$

If the field K is of *characteristic* 0 (see Chapter I, §8), then $\partial(f') = \partial(f) - 1$.

Formal derivatives share the following properties with the derivatives of real analysis:

- (i) $(f + g)' = f' + g'$;
- (ii) $(cf)' = cf'$ for any $c \in K$;
- (iii) $(fg)' = f'g + fg'$;
- (iv) $(f^k)' = kf^{k-1}f'$ for any $k \in \mathbb{N}$.

The first two properties are easily established and the last two properties then need only be verified for monomials $f = t^m$, $g = t^n$.

We can use formal derivatives to determine when a polynomial is *square-free*:

Proposition 20 *Let f be a polynomial of positive degree with coefficients from a field K . If f is relatively prime to its formal derivative f' , then f is a product of irreducible polynomials, no two of which differ by a constant factor. Conversely, if f is such a product and if K has characteristic 0, then f is relatively prime to f' .*

Proof If $f = g^2h$ for some polynomials $g, h \in K[t]$ with $\partial(g) > 0$ then, by the rules above,

$$f' = 2gg'h + g^2h'.$$

Hence $g|f'$ and f, f' are not relatively prime.

On the other hand, if $f = p_1 \cdots p_m$ is a product of essentially distinct irreducible polynomials p_j , then

$$f' = p'_1p_2 \cdots p_m + p_1p'_2p_3 \cdots p_m + \cdots + p_1 \cdots p_{m-1}p'_m.$$

If the field K has characteristic 0, then p'_1 is of lower degree than p_1 and is not the zero polynomial. Thus the first term on the right is not divisible by p_1 , but all the other terms are. Therefore $p_1 \nmid f'$, and hence $(f', p_1) = 1$. Similarly, $(f', p_j) = 1$ for $1 < j \leq m$. Since essentially distinct irreducible polynomials are relatively prime, it follows that $(f', f) = 1$. \square

For example, it follows from Proposition 20 that the polynomial $t^n - 1 \in K[t]$ is square-free if the characteristic of the field K does not divide the positive integer n .

4 Euclidean Domains

An integral domain R is said to be *Euclidean* if it possesses a Euclidean algorithm, i.e. if there exists a map $\delta: R \rightarrow \mathbb{N} \cup \{0\}$ such that, for any $a, b \in R$ with $a \neq 0$, there exist $q, r \in R$ with the properties

$$b = qa + r, \quad \delta(r) < \delta(a).$$

It follows that $\delta(a) > \delta(0)$ for any $a \neq 0$. For there exist $q_1, a_1 \in R$ such that

$$0 = q_1a + a_1, \quad \delta(a_1) < \delta(a),$$

and if $a_n \neq 0$ there exist $q_{n+1}, a_{n+1} \in R$ such that

$$0 = q_{n+1}a_n + a_{n+1}, \quad \delta(a_{n+1}) < \delta(a_n).$$

Repeatedly applying this process, we must arrive at $a_N = 0$ for some N , since the sequence $\{\delta(a_n)\}$ cannot decrease forever, and we then have $\delta(0) = \delta(a_N) < \dots < \delta(a_1) < \delta(a)$.

By replacing δ by $\delta - \delta(0)$ we may, and will, assume that $\delta(0) = 0$, $\delta(a) > 0$ if $a \neq 0$.

Since the proof of Lemma 9 remains valid if \mathbb{Z} is replaced by R and $|a|$ by $\delta(a)$, any Euclidean domain is a principal ideal domain.

The polynomial ring $K[t]$ is a Euclidean domain with $\delta(a) = |a| = 2^{\delta(a)}$. Polynomial rings are characterized among all Euclidean domains by the following result:

Proposition 21 *For a Euclidean domain R , the following conditions are equivalent:*

- (i) *for any $a, b \in R$ with $a \neq 0$, there exist unique $q, r \in R$ such that $b = qa + r$, $\delta(r) < \delta(a)$;*
- (ii) *for any $a, b, c \in R$ with $c \neq 0$,*

$$\delta(a + b) \leq \max\{\delta(a), \delta(b)\}, \quad \delta(a) \leq \delta(ac).$$

Moreover, if one or other of these two conditions holds, then either R is a field and $\delta(a) = \delta(1)$ for every $a \neq 0$, or $R = K[t]$ for some field K and δ is an increasing function of $||$.

Proof Suppose first that (i) holds. If $a \neq 0$, $c \neq 0$, then from $0 = 0a - 0 = ca - ac$, we obtain $\delta(ac) \geq \delta(a)$, and this holds also if $a = 0$. If we take $c = -1$ and replace a by $-a$, we get $\delta(-a) = \delta(a)$. Since $b = 0(a + b) + b = 1(a + b) + (-a)$, it follows that either $\delta(b) \geq \delta(a + b)$ or $\delta(a) \geq \delta(a + b)$. Thus (i) \Rightarrow (ii).

Suppose next that (ii) holds. Assume that, for some $a, b \in R$ with $a \neq 0$, there exist pairs q, r and q', r' such that

$$b = qa + r = q'a + r', \quad \max\{\delta(r), \delta(r')\} < \delta(a).$$

From (ii) we obtain first $\delta(-r) = \delta(r)$ and then $\delta(r' - r) \leq \max\{\delta(r), \delta(r')\} < \delta(a)$. Since $r' - r = a(q - q')$, this implies $q - q' = 0$ and hence $r' - r = 0$. Thus (ii) \Rightarrow (i).

Suppose now that (i) and (ii) both hold. Then $\delta(1) \leq \delta(a)$ for any $a \neq 0$, since $a = 1a$. Furthermore, $\delta(a) = \delta(ae)$ for any unit e , since

$$\delta(a) \leq \delta(ae) \leq \delta(aee^{-1}) = \delta(a).$$

On the other hand, $\delta(a) = \delta(ae)$ for some $a \neq 0$ implies that e is a unit. For from

$$a = qae + r, \quad \delta(r) < \delta(ae),$$

we obtain $r = (1 - qe)a$, $\delta(r) < \delta(a)$, and hence $1 - qe = 0$. In particular, $\delta(e) = \delta(1)$ if and only if e is a unit.

The set K of all $a \in R$ such that $\delta(a) \leq \delta(1)$ thus consists of 0 and all units of R . Since $a, b \in K$ implies $a - b \in K$, it follows that K is a field. We assume that $K \neq R$, since otherwise we have the first alternative of the proposition.

Choose $x \in R \setminus K$ so that

$$\delta(x) = \min_{a \in R \setminus K} \delta(a).$$

For any $a \in R \setminus K$, there exist $q_0, r_0 \in R$ such that

$$a = q_0x + r_0, \quad \delta(r_0) < \delta(x),$$

i.e. $r_0 \in K$. Then $\delta(q_0) < \delta(q_0x) = \delta(a - r_0) \leq \delta(a)$. If $\delta(q_0) \geq \delta(x)$, i.e. if $q_0 \in R \setminus K$, then in the same way there exist $q_1, r_1 \in R$ such that

$$q_0 = q_1x + r_1, \quad r_1 \in K, \quad \delta(q_1) < \delta(q_0).$$

After finitely many repetitions of this process we must arrive at some $q_{n-1} \in K$. Putting $r_n = q_{n-1}$, we obtain

$$a = r_nx^n + r_{n-1}x^{n-1} + \cdots + r_0,$$

where $r_0, \dots, r_n \in K$ and $r_n \neq 0$. Since $\delta(r_jx^j) = \delta(x^j)$ if $r_j \neq 0$ and $\delta(x^j) < \delta(x^{j+1})$ for every j , it follows that $\delta(a) = \delta(x^n)$. Since the representation $a = qx^n + r$ with $\delta(r) < \delta(x^n)$ is unique, it follows that r_0, \dots, r_n are uniquely determined by a . Define a map $\psi: R \rightarrow K[t]$ by

$$\psi(r_nx^n + r_{n-1}x^{n-1} + \cdots + r_0) = r_nt^n + r_{n-1}t^{n-1} + \cdots + r_0.$$

Then ψ is a bijection and actually an isomorphism, since it preserves sums and products. Furthermore $\delta(a) >, =, \text{ or } < \delta(b)$ according as $|\psi(a)| >, =, \text{ or } < |\psi(b)|$. \square

Some significant examples of principal ideal domains are provided by quadratic fields, which will be studied in Chapter III. Any quadratic number field has the form $\mathbb{Q}(\sqrt{d})$, where $d \in \mathbb{Z}$ is square-free and $d \neq 1$. The set \mathcal{O}_d of all algebraic integers in $\mathbb{Q}(\sqrt{d})$ is an integral domain. In the equivalent language of binary quadratic forms, it was known to Gauss that \mathcal{O}_d is a principal ideal domain for nine negative values of d , namely

$$d = -1, -2, -3, -7, -11, -19, -43, -67, -163.$$

Heilbronn and Linfoot (1934) showed that there was at most one additional negative value of d for which \mathcal{O}_d is a principal ideal domain. Stark (1967) proved that this additional value does not in fact exist, and soon afterwards it was observed that a gap in a previous proof by Heegner (1952) could be filled without difficulty. It is conjectured that \mathcal{O}_d is a principal ideal domain for infinitely many positive values of d , but this remains unproved.

Much work has been done on determining for which quadratic number fields $\mathbb{Q}(\sqrt{d})$ the ring of integers \mathcal{O}_d is a Euclidean domain. Although we regard being Euclidean more as a useful property than as an important concept, we report here the results which have been obtained for their intrinsic interest.

The ring \mathcal{O}_d is said to be *norm-Euclidean* if it is Euclidean when one takes $\delta(a)$ to be the absolute value of the *norm* of a . It has been shown that \mathcal{O}_d is norm-Euclidean for precisely the following values of d :

$$d = -11, -7, -3, -2, -1, 2, 3, 5, 6, 7, 11, 13, 17, 19, 21, 29, 33, 37, 41, 57, 73.$$

It is known that, for $d < 0$, \mathcal{O}_d is Euclidean only if it is norm-Euclidean. Comparing the two lists, we see that for $d = -19, -43, -67, -163$, \mathcal{O}_d is a principal ideal domain, but not a Euclidean domain. On the other hand it is also known that, for $d = 69$, \mathcal{O}_d is Euclidean but not norm-Euclidean.

5 Congruences

The invention of a new notation often enables one to replace a long, involved argument by simple and mechanical algebraic operations. This is well illustrated by the congruence notation.

Two integers a and b are said to be *congruent modulo* a third integer m if m divides $a - b$, and this is denoted by $a \equiv b \pmod{m}$. For example,

$$13 \equiv 4 \pmod{3}, \quad 13 \equiv -7 \pmod{5}, \quad 19 \equiv 7 \pmod{4}.$$

The notation is a modification by Gauss of the notation $a = b \pmod{m}$ used by Legendre, as Gauss explicitly acknowledged (*D.A.*, §2). (If a and b are not congruent modulo m , we write $a \not\equiv b \pmod{m}$.) Congruence has, in fact, many properties in common with equality:

- | | |
|--|--------------------|
| (C1) $a \equiv a \pmod{m}$ for all a, m ; | (reflexive law) |
| (C2) if $a \equiv b \pmod{m}$, then $b \equiv a \pmod{m}$; | (symmetric law) |
| (C3) if $a \equiv b$ and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$; | (transitive law) |
| (C4) if $a \equiv a'$ and $b \equiv b' \pmod{m}$, then $a + b \equiv a' + b'$ and $ab \equiv a'b' \pmod{m}$. | (replacement laws) |

The proofs of these properties are very simple. For any a, m we have $a - a = 0 = m \cdot 0$. If m divides $a - b$, then it also divides $b - a = -(a - b)$. If m divides both $a - b$ and $b - c$, then it also divides $(a - b) + (b - c) = a - c$. Finally, if m divides both $a - a'$ and $b - b'$, then it also divides $(a - a') + (b - b') = (a + b) - (a' + b')$ and $(a - a')b + a'(b - b') = ab - a'b'$.

The properties (C1)–(C3) state that congruence mod m is an *equivalence relation*. Since $a = b$ implies $a \equiv b \pmod{m}$, it is a coarsening of the equivalence relation of

equality (but coincides with it if $m = 0$). The corresponding equivalence classes are called *residue classes*. The set \mathbb{Z} with equality replaced by congruence mod m will be denoted by $\mathbb{Z}_{(m)}$. If $m > 0$, $\mathbb{Z}_{(m)}$ has cardinality m , since an arbitrary integer a can be uniquely represented in the form $a = qm + r$, where $r \in \{0, 1, \dots, m-1\}$ and $q \in \mathbb{Z}$. The particular r which represents a given $a \in \mathbb{Z}$ is referred to as the *least non-negative residue* of a mod m .

The replacement laws imply that the associative, commutative and distributive laws for addition and multiplication are inherited from \mathbb{Z} by $\mathbb{Z}_{(m)}$. Hence $\mathbb{Z}_{(m)}$ is a commutative ring, with 0 as an identity element for addition and 1 as an identity element for multiplication. However, $\mathbb{Z}_{(m)}$ is not an integral domain if m is composite, since if $m = m'm''$ with $1 < m' < m$, then

$$m'm'' \equiv 0, \text{ but } m' \not\equiv 0, m'' \not\equiv 0 \pmod{m}.$$

On the other hand, if $ab \equiv ac \pmod{m}$ and $(a, m) = 1$, then $b \equiv c \pmod{m}$, by Proposition 3(ii). Thus factors which are relatively prime to the modulus can be cancelled.

In algebraic terms, $\mathbb{Z}_{(m)}$ is the *quotient ring* $\mathbb{Z}/m\mathbb{Z}$ of \mathbb{Z} with respect to the ideal $m\mathbb{Z}$ generated by m , and the elements of $\mathbb{Z}_{(m)}$ are the *cosets* of this ideal. For convenience, rather than necessity, we suppose from now on that $m > 1$.

Congruences enter implicitly into many everyday problems. For example, the ring $\mathbb{Z}_{(2)}$ contains two distinct elements, 0 and 1, with the addition and multiplication tables

$$\begin{aligned} 0 + 0 &= 1 + 1 = 0, & 0 + 1 &= 1 + 0 = 1, \\ 0 \cdot 0 &= 0 \cdot 1 = 1 \cdot 0 = 0, & 1 \cdot 1 &= 1. \end{aligned}$$

This is the arithmetic of *odds* (1) and *evens* (0), which is used by electronic computers.

Again, to determine the day of the week on which one was born, from the date and day of the week today, is an easy calculation in the arithmetic of $\mathbb{Z}_{(7)}$ (remembering that $366 \equiv 2 \pmod{7}$).

The well-known tests for divisibility of an integer by 3 or 9 are easily derived by means of congruences. Let the positive integer a have the decimal representation

$$a = a_0 + a_1 10 + \dots + a_n 10^n,$$

where $a_0, a_1, \dots, a_n \in \{0, 1, \dots, 9\}$. Since $10 \equiv 1 \pmod{m}$, where $m = 3$ or 9 , the replacement laws imply that $10^k \equiv 1 \pmod{m}$ for any positive integer k and hence

$$a \equiv a_0 + a_1 + \dots + a_n \pmod{m}.$$

Thus a is divisible by 3 or 9 if and only if the sum of its digits is so divisible.

This can be used to check the accuracy of arithmetical calculations. Any equation involving only additions and multiplications must remain valid when equality is replaced by congruence mod m . For example, suppose we wish to check if

$$7714 \times 3036 = 23,419,804.$$

Taking congruences mod 9, we have on the left side $19 \times 12 \equiv 1 \times 3 \equiv 3$ and on the right side $5 + 14 + 12 \equiv 5 + 5 + 3 \equiv 4$. Since $4 \not\equiv 3 \pmod{9}$, the original equation is incorrect (the 8 should be a 7).

Since the distinct squares in $\mathbb{Z}_{(4)}$ are 0 and 1, it follows that an integer $a \equiv 3 \pmod{4}$ cannot be represented as the sum of two squares of integers. Similarly, since the distinct squares in $\mathbb{Z}_{(8)}$ are 0, 1, 4, an integer $a \equiv 7 \pmod{8}$ cannot be represented as the sum of three squares of integers.

The oldest known work on number theory is a Babylonian cuneiform text, from at least as early as 1600 B.C., which contains a list of right-angled triangles whose side lengths are all exact multiples of the unit length. By Pythagoras' theorem, the problem is to find positive integers x, y, z such that

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

For example, 3, 4, 5 and 5, 12, 13 are solutions. The number of solutions listed suggests that the Babylonians not only knew the theorem of Pythagoras, but also had some rule for finding such *Pythagorean triples*. There are in fact infinitely many, and a rule for finding them all is given by Euclid in his *Elements* (Book X, Lemma 1 following Proposition 28). This rule will now be derived.

We may assume that x and y are relatively prime since, if x, y, z is a Pythagorean triple for which x and y have greatest common divisor d , then $d^2 | z^2$ and hence $d | z$, so that $x/d, y/d, z/d$ is also a Pythagorean triple. If x and y are relatively prime, then they are not both even and without loss of generality we may assume that x is odd. If y were also odd, we would have

$$z^2 = x^2 + y^2 \equiv 1 + 1 \equiv 2 \pmod{4},$$

which is impossible. Hence y is even and z is odd. Then 2 is a common divisor of $z + x$ and $z - x$, and is actually their greatest common divisor, since $(x, y) = 1$ implies $(x, z) = 1$. Since

$$(y/2)^2 = (z + x)/2 \cdot (z - x)/2$$

and the two factors on the right are relatively prime, they are also squares:

$$(z + x)/2 = a^2, \quad (z - x)/2 = b^2,$$

where $a > b > 0$ and $(a, b) = 1$. Then

$$x = a^2 - b^2, \quad y = 2ab, \quad z = a^2 + b^2.$$

Moreover a and b cannot both be odd, since z is odd.

Conversely, if x, y, z are defined by these formulas, where a and b are relatively prime positive integers with $a > b$ and either a or b even, then x, y, z is a Pythagorean triple. Moreover x is odd, since z is odd and y even, and it is easily verified that $(x, y) = 1$. For given x and z , a^2 and b^2 are uniquely determined, and hence a and b are also. Thus different couples a, b give different solutions x, y, z .

To return to congruences, we now consider the structure of the ring $\mathbb{Z}_{(m)}$. If $a \equiv a' \pmod{m}$ and $(a, m) = 1$, then also $(a', m) = 1$. Hence we may speak of an element of $\mathbb{Z}_{(m)}$ as being relatively prime to m . The set of all elements of $\mathbb{Z}_{(m)}$ which are relatively prime to m will be denoted by $\mathbb{Z}_{(m)}^\times$. If a is a unit of the ring $\mathbb{Z}_{(m)}$, then clearly $a \in \mathbb{Z}_{(m)}^\times$. The following proposition shows that, conversely, if $a \in \mathbb{Z}_{(m)}^\times$, then a is a unit of the ring $\mathbb{Z}_{(m)}$.

Proposition 22 *The set $\mathbb{Z}_{(m)}^\times$ is a commutative group under multiplication.*

Proof By Proposition 3(iv), $\mathbb{Z}_{(m)}^\times$ is closed under multiplication. Since multiplication is associative and commutative, it only remains to show that any $a \in \mathbb{Z}_{(m)}^\times$ has an inverse $a^{-1} \in \mathbb{Z}_{(m)}^\times$.

The elements of $\mathbb{Z}_{(m)}^\times$ may be taken to be the positive integers c_1, \dots, c_h which are less than m and relatively prime to m , and we may choose the notation so that $c_1 = 1$. Since $ac_j \equiv ac_k \pmod{m}$ implies $c_j \equiv c_k \pmod{m}$, the elements ac_1, \dots, ac_h are distinct elements of $\mathbb{Z}_{(m)}^\times$ and hence are a permutation of c_1, \dots, c_h . In particular, $ac_i \equiv c_1 \pmod{m}$ for one and only one value of i . (The existence of inverses also follows from the Bézout identity $au + mv = 1$, since this implies $au \equiv 1 \pmod{m}$. Hence the Euclidean algorithm provides a way of calculating a^{-1} .) \square

Corollary 23 *If p is a prime, then $\mathbb{Z}_{(p)}$ is a finite field with p elements.*

Proof We already know that $\mathbb{Z}_{(p)}$ is a commutative ring, whose distinct elements are represented by the integers $0, 1, \dots, p-1$. Since p is a prime, $\mathbb{Z}_{(p)}^\times$ consists of all nonzero elements of $\mathbb{Z}_{(p)}$. Since $\mathbb{Z}_{(p)}^\times$ is a multiplicative group, by Proposition 22, it follows that $\mathbb{Z}_{(p)}$ is a field. \square

The finite field $\mathbb{Z}_{(p)}$ will be denoted from now on by the more usual notation \mathbb{F}_p . Corollary 23, in conjunction with Proposition 15, implies that if p is a prime and f a polynomial of degree $n \geq 1$, then the congruence

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

has at most n mutually incongruent solutions mod p . This is no longer true if the modulus is not a prime. For example, the congruence $x^2 - 1 \equiv 0 \pmod{8}$ has the distinct solutions $x \equiv 1, 3, 5, 7 \pmod{8}$.

The *order* of the group $\mathbb{Z}_{(m)}^\times$, i.e. the number of positive integers less than m and relatively prime to m , is traditionally denoted by $\varphi(m)$, with the convention that $\varphi(1) = 1$. For example, if p is a prime, then $\varphi(p) = p - 1$. More generally, for any positive integer k ,

$$\varphi(p^k) = p^k - p^{k-1},$$

since the elements of $\mathbb{Z}_{(p^k)}$ which are not in $\mathbb{Z}_{(p^{k-1})}^\times$ are the multiples jp with $0 \leq j < p^{k-1}$. By Proposition 4, if $m = m'm''$, where $(m', m'') = 1$, then $\varphi(m) = \varphi(m')\varphi(m'')$. Together with what we have just proved, this implies that if an arbitrary positive integer m has the factorization

$$m = p_1^{k_1} \cdots p_s^{k_s}$$

as a product of positive powers of distinct primes, then

$$\varphi(m) = p_1^{k_1-1}(p_1 - 1) \cdots p_s^{k_s-1}(p_s - 1).$$

In other words,

$$\varphi(m) = m \prod_{p|m} (1 - 1/p).$$

The function $\varphi(m)$ was first studied by Euler and is known as Euler's *phi*-function (or 'totient' function), although it was Gauss who decided on the letter φ . Gauss (*D.A.*, §39) also established the following property:

Proposition 24 *For any positive integer n ,*

$$\sum_{d|n} \varphi(d) = n,$$

where the summation is over all positive divisors d of n .

Proof Let d be a positive divisor of n and let S_d denote the set of all positive integers $m \leq n$ such that $(m, n) = d$. Since $(m, n) = d$ if and only if $(m/d, n/d) = 1$, the cardinality of S_d is $\varphi(n/d)$. Moreover every positive integer $m \leq n$ belongs to exactly one such set S_d . Hence

$$n = \sum_{d|n} \varphi(n/d) = \sum_{d|n} \varphi(d),$$

since n/d runs through the positive divisors of n at the same time as d . □

Much of the significance of Euler's function stems from the following property:

Proposition 25 *If m is a positive integer and a an integer relatively prime to m , then*

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

Proof Let c_1, \dots, c_h , where $h = \varphi(m)$, be the distinct elements of $\mathbb{Z}_{(m)}^\times$. As we saw in the proof of Proposition 22, the elements ac_1, \dots, ac_h of $\mathbb{Z}_{(m)}^\times$ are just a permutation of c_1, \dots, c_h . Forming their product, we obtain $a^h c_1 \cdots c_h \equiv c_1 \cdots c_h \pmod{m}$. Since the c 's are relatively prime to m , they can be cancelled and we are left with $a^h \equiv 1 \pmod{m}$. □

Corollary 26 *If p is a prime and a an integer not divisible by p , then $a^{p-1} \equiv 1 \pmod{p}$.*

Corollary 26 was stated without proof by Fermat (1640) and is commonly known as 'Fermat's little theorem'. The first published proof was given by Euler (1736), who later (1760) proved the general Proposition 25.

Proposition 25 is actually a very special case of Lagrange's theorem that the order of a subgroup of a finite group divides the order of the whole group. In the present case the whole group is $\mathbb{Z}_{(m)}^\times$ and the subgroup is the cyclic group generated by a .

Euler gave also another proof of Corollary 26, which has its own interest. For any two integers a, b and any prime p we have, by the binomial theorem,

$$(a + b)^p = \sum_{k=0}^p {}^pC_k a^k b^{p-k},$$

where the binomial coefficients

$${}^pC_k = (p - k + 1) \cdots p / 1 \cdot 2 \cdots k$$

are integers. Moreover p divides pC_k for $0 < k < p$, since p divides ${}^pC_k \cdot k!$ and is relatively prime to $k!$ It follows that

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

In particular, $(a + 1)^p \equiv a^p + 1 \pmod{p}$, from which we obtain by induction $a^p \equiv a \pmod{p}$ for every integer a . If p does not divide a , the factor a can be cancelled to give $a^{p-1} \equiv 1 \pmod{p}$.

The first part of the second proof actually shows that *in any commutative ring R , of prime characteristic p , the map $a \rightarrow a^p$ is a homomorphism:*

$$(a + b)^p = a^p + b^p, \quad (ab)^p = a^p b^p.$$

(As defined in §8 of Chapter I, R has *characteristic k* if k is the least positive integer such that the sum of k 1's is 0, and has *characteristic zero* if there is no such positive integer.) By way of illustration, we give one important application of this result.

We showed in §3 that, for any prime p , the polynomial

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

is irreducible in $\mathbb{Q}[x]$. The roots in \mathbb{C} of $\Phi_p(x)$ are the p -th roots of unity, other than 1. By a quite different argument we now show that, for any positive integer n , the ‘primitive’ n -th roots of unity are the roots of a monic polynomial $\Phi_n(x)$ with integer coefficients which is irreducible in $\mathbb{Q}[x]$. The uniquely determined polynomial $\Phi_n(x)$ is called the *n -th cyclotomic polynomial*.

Let ζ be a *primitive n -th root of unity*, i.e. $\zeta^n = 1$ but $\zeta^k \neq 1$ for $0 < k < n$. It follows from Corollary 18 that ζ is a root of some monic irreducible polynomial $f(x) \in \mathbb{Z}[x]$ which divides $x^n - 1$. If p is a prime which does not divide n , then ζ^p is also a primitive n -th root of unity and, for the same reason, ζ^p is a root of some monic irreducible polynomial $g(x) \in \mathbb{Z}[x]$ which divides $x^n - 1$.

We show first that $g(x) = f(x)$. Assume on the contrary that $g(x) \neq f(x)$. Then

$$x^n - 1 = f(x)g(x)h(x)$$

for some $h(x) \in \mathbb{Z}[x]$. Since ζ is a root of $g(x^p)$, we also have

$$g(x^p) = f(x)k(x)$$

for some $k(x) \in \mathbb{Z}[x]$. If $\bar{f}(x), \dots$ denotes the polynomial in $\mathbb{F}_p[x]$ obtained from $f(x), \dots$ by reducing the coefficients mod p ,

then

$$x^n - 1 = \bar{f}(x)\bar{g}(x)\bar{h}(x), \quad \bar{g}(x^p) = \bar{f}(x)\bar{k}(x).$$

But $\bar{g}(x^p) = \bar{g}(x)^p$, since $\mathbb{F}_p[x]$ is a ring of characteristic p and $a^p = a$ for every $a \in \mathbb{F}_p$. Hence any irreducible factor $\bar{e}(x)$ of $\bar{f}(x)$ in $\mathbb{F}_p[x]$ also divides $\bar{g}(x)$. Consequently $\bar{e}(x)^2$ divides $x^n - 1$ in $\mathbb{F}_p[x]$. But $x^n - 1$ is relatively prime to its formal derivative nx^{n-1} , since $p \nmid n$, and so is square-free. This is the desired contradiction.

By applying this repeatedly for the same or different primes p , we see that ζ^m is a root of $f(x)$ for any positive integer m less than n and relatively prime to n . If ω is any n -th root of unity, then $\omega = \zeta^k$ for a unique k such that $0 \leq k < n$. If $(k, n) \neq 1$, then $\omega^d = 1$ for some proper divisor d of n (cf. Lemma 31 below). If such an ω were a root of $f(x)$, then $f(x)$ would divide $x^d - 1$, which is impossible since ζ is not a root of $x^d - 1$. Hence $f(x)$ does not depend on the original choice of primitive n -th root of unity, its roots being all the primitive n -th roots of unity. The polynomial $f(x)$ will now be denoted by $\Phi_n(x)$. Since $x^n - 1$ is square-free, we have

$$x^n - 1 = \prod_{d|n} \Phi_d(x).$$

This yields a new proof of Proposition 24, since $\Phi_d(x)$ has degree $\phi(d)$.

As an application of Fermat's little theorem (Corollary 26) we now prove

Proposition 27 *If p is a prime, then $(p-1)! + 1$ is divisible by p .*

Proof Since $1! + 1 = 2$, we may suppose that the prime p is odd. By Corollary 26, the polynomial $f(t) = t^{p-1} - 1$ has the distinct roots $1, 2, \dots, p-1$ in the field \mathbb{F}_p . But the polynomial $g(t) = (t-1)(t-2)\cdots(t-p+1)$ has the same roots. Since $f(t) - g(t)$ is a polynomial of degree less than $p-1$, it follows from Proposition 15 that $f(t) - g(t)$ is the zero polynomial. In particular, $f(t)$ and $g(t)$ have the same constant coefficient. Since $(-1)^{p-1} = 1$, this yields the result. \square

Proposition 27 is known as *Wilson's theorem*, although the first published proof was given by Lagrange (1773). Lagrange observed also that $(n-1)! + 1$ is divisible by n only if n is prime. For suppose $n = n'n''$, where $1 < n', n'' < n$. If $n' \neq n''$, then both n' and n'' occur as factors in $(n-1)!$ and hence n divides $(n-1)!$. If $n' = n'' > 2$ then, since $n > 2n'$, both n' and $2n'$ occur as factors in $(n-1)!$ and again n divides $(n-1)!$. Finally, if $n = 4$, then n divides $(n-1)! + 2$.

As another application of Fermat's little theorem, we prove *Euler's criterion for quadratic residues*. If p is a prime and a an integer not divisible by p , we say that a is a *quadratic residue*, or *quadratic nonresidue*, of p according as there exists, or does not exist, an integer c such that $c^2 \equiv a \pmod{p}$. Thus a is a quadratic residue of p if and only if it is a square in \mathbb{F}_p^\times . Euler's criterion is the first statement of the following proposition:

Proposition 28 *If p is an odd prime and a an integer not divisible by p , then*

$$a^{(p-1)/2} \equiv 1 \text{ or } -1 \pmod{p},$$

according as a is a quadratic residue or nonresidue of p .

Moreover, exactly half of the integers $1, 2, \dots, p-1$ are quadratic residues of p .

Proof If a is a quadratic residue of p , then $a \equiv c^2 \pmod{p}$ for some integer c and hence, by Fermat's little theorem,

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

Since the polynomial $t^{(p-1)/2} - 1$ has at most $(p-1)/2$ roots in the field \mathbb{F}_p , it follows that there are at most $r := (p-1)/2$ distinct quadratic residues of p . On the other hand, no two of the integers $1^2, 2^2, \dots, r^2$ are congruent mod p , since $u^2 \equiv v^2 \pmod{p}$ implies $u \equiv v$ or $u \equiv -v \pmod{p}$. Hence there are exactly $(p-1)/2$ distinct quadratic residues of p and, if b is a quadratic nonresidue of p , then $b^{(p-1)/2} \not\equiv 1 \pmod{p}$. Since $b^{p-1} \equiv 1 \pmod{p}$, and

$$b^{p-1} - 1 = (b^{(p-1)/2} - 1)(b^{(p-1)/2} + 1),$$

we must have $b^{(p-1)/2} \equiv -1 \pmod{p}$. □

Corollary 29 *If p is an odd prime, then -1 is a quadratic residue of p if $p \equiv 1 \pmod{4}$ and a quadratic nonresidue of p if $p \equiv 3 \pmod{4}$.*

Euler's criterion may also be used to determine for what primes 2 is a quadratic residue:

Proposition 30 *For any odd prime p , 2 is a quadratic residue of p if $p \equiv \pm 1 \pmod{8}$ and a quadratic nonresidue if $p \equiv \pm 3 \pmod{8}$.*

Proof Let A denote the set of all even integers a such that $p/2 < a < p$, and let B denote the set of all even integers b such that $0 < b < p/2$. Since $A \cup B$ is the set of all positive even integers less than p , it has cardinality $r := (p-1)/2$. Evidently $a \in A$ if and only if $p-a$ is odd and $0 < p-a < p/2$. Hence the integers $1, 2, \dots, r$ are just the elements of B , together with the integers $p-a$ ($a \in A$). If we denote the cardinality of A by $\#A$, it follows that

$$\begin{aligned} r! &= \prod_{a \in A} (p-a) \prod_{b \in B} b \\ &\equiv (-1)^{\#A} \prod_{a \in A} a \prod_{b \in B} b \pmod{p} \\ &= (-1)^{\#A} 2^r r! \end{aligned}$$

Thus $2^r \equiv (-1)^{\#A} \pmod{p}$ and hence, by Proposition 28, 2 is a quadratic residue or nonresidue of p according as $\#A$ is even or odd. But $\#A = k$ if $p = 4k + 1$ and $\#A = k + 1$ if $p = 4k + 3$. The result follows. □

We now introduce some simple group-theoretical concepts. Let G be a finite group and $a \in G$. Then there exist $j, k \in \mathbb{N}$ with $j < k$ such that $a^j = a^k$. Thus $a^{k-j} = 1$, where 1 is the identity element of G . The *order* of a is the least positive integer d such that $a^d = 1$.

Lemma 31 *Let G be a finite group of order n and a an element of G of order d . Then*

- (i) *for any $k \in \mathbb{N}$, $a^k = 1$ if and only if d divides k ;*

- (ii) for any $k \in \mathbb{N}$, a^k has order $d/(k, d)$;
 (iii) $H = \{1, a, \dots, a^{d-1}\}$ is a subgroup of G and d divides n .

Proof Any $k \in \mathbb{N}$ can be written in the form $k = qd + r$, where $q \geq 0$ and $0 \leq r < d$. Since $a^{qd} = (a^d)^q = 1$, we have $a^k = 1$ if and only if $a^r = 1$, i.e. if and only if $r = 0$, by the definition of d .

It follows that if a^k has order e , then $ke = [k, d]$. Since $[k, d] = kd/(k, d)$, this implies $e = d/(k, d)$. In particular, a^k again has order d if and only if $(k, d) = 1$.

If $0 \leq j, k < d$, put $i = j + k$ if $j + k < d$ and $i = j + k - d$ if $j + k \geq d$. Then $a^j a^k = a^i$, and so H contains the product of any two of its elements. If $0 < k < d$, then $a^k a^{d-k} = 1$, and so H contains also the inverse of any one of its elements. Finally d divides n , by Lagrange's theorem that the order of a subgroup divides the order of the whole group. \square

The subgroup H in Lemma 31 is the *cyclic subgroup generated by a* . For $G = \mathbb{Z}_{(m)}^\times$, the case which we will be interested in, there is no need to appeal to Lagrange's theorem, since $\mathbb{Z}_{(m)}^\times$ has order $\varphi(m)$ and d divides $\varphi(m)$, by Proposition 25 and Lemma 31(i).

A group G is *cyclic* if it coincides with the cyclic subgroup generated by one of its elements. For example, the n -th roots of unity in \mathbb{C} form a cyclic group generated by $e^{2\pi i/n}$. In fact the generators of this group are just the primitive n -th roots of unity.

Our next result provides a sufficient condition for a finite group to be cyclic.

Lemma 32 *A finite group G of order n is cyclic if, for each positive divisor d of n , there are at most d elements of G whose order divides d .*

Proof If H is a cyclic subgroup of G , then its order d divides n . Since all its elements are of order dividing d , the hypothesis of the lemma implies that any element of G whose order divides d must be in H . Furthermore, H contains exactly $\varphi(d)$ elements of order d since, if a generates H , a^k has order d if and only if $(k, d) = 1$.

For each divisor d of n , let $\psi(d)$ denote the number of elements of G of order d . Then, by what we have just proved, either $\psi(d) = 0$ or $\psi(d) = \varphi(d)$. But $\sum_{d|n} \psi(d) = n$, since the order of each element is a divisor of n , and $\sum_{d|n} \varphi(d) = n$, by Proposition 24. Hence we must have $\psi(d) = \varphi(d)$ for every $d|n$. In particular, the group G has $\psi(n) = \varphi(n)$ elements of order n . \square

The condition of Lemma 32 is also necessary. For let G be a finite cyclic group of order n , generated by the element a , and let d be a divisor of n . An element $x \in G$ has order dividing d if and only if $x^d = 1$. Thus the elements a^k of G of order dividing d are given by $k = jn/d$, with $j = 0, 1, \dots, d-1$.

We now return from group theory to number theory.

Proposition 33 *For any prime p , the multiplicative group \mathbb{F}_p^\times of the field \mathbb{F}_p is cyclic.*

Proof Put $G = \mathbb{F}_p^\times$ and denote the order of G by n . For any divisor d of n , the polynomial $t^d - 1$ has at most d roots in \mathbb{F}_p . Hence there are at most d elements of G whose order divides d . The result now follows from Lemma 32. \square

The same argument shows that, for an arbitrary field K , any finite subgroup of the multiplicative group of K is cyclic.

In the terminology of number theory, an integer which generates $\mathbb{Z}_{(m)}^\times$ is said to be a *primitive root* of m . Primitive roots may be used to replace multiplications mod m by additions mod $\varphi(m)$ in the same way that logarithms were once used in analysis. If g is a primitive root of m , then the elements of $\mathbb{Z}_{(m)}^\times$ are precisely $1, g, g^2, \dots, g^{n-1}$, where $n = \varphi(m)$. Thus for each $a \in \mathbb{Z}_{(m)}^\times$ we have $a \equiv g^\alpha \pmod{m}$ for a unique index α ($0 \leq \alpha < n$). We can construct a table of these indices once and for all. If $a \equiv g^\alpha$ and $b \equiv g^\beta$, then $ab \equiv g^{\alpha+\beta}$. By replacing $\alpha + \beta$ by its least non-negative residue γ mod n and going backwards in our table we can determine c such that $ab \equiv c \pmod{m}$.

For any prime p , an essentially complete proof for the existence of primitive roots of p was given by Euler (1774). Jacobi (1839) constructed tables of indices for all primes less than 1000.

We now use primitive roots to prove a general property of polynomials with coefficients from a finite field:

Proposition 34 *If $f(x_1, \dots, x_n)$ is a polynomial of degree less than n in n variables with coefficients from the finite field \mathbb{F}_p , then the number of zeros of f in \mathbb{F}_p^n is divisible by the characteristic p . In particular, $(0, \dots, 0)$ is not the only zero of f if f has no constant term.*

Proof Put $K = \mathbb{F}_p$ and $g = 1 - f^{p-1}$. If $\alpha = (a_1, \dots, a_n)$ is a zero of f , then $g(\alpha) = 1$. If α is not a zero of f , then $f(\alpha)^{p-1} = 1$ and $g(\alpha) = 0$. Hence the number N of zeros of f satisfies

$$N \equiv \sum_{\alpha \in K^n} g(\alpha) \pmod{p}.$$

We will complete the proof by showing that

$$\sum_{\alpha \in K^n} g(\alpha) = 0.$$

Since g has degree less than $n(p-1)$, it is a constant linear combination of polynomials of the form $x_1^{k_1} \cdots x_n^{k_n}$, where $k_1 + \cdots + k_n < n(p-1)$. Thus $k_j < p-1$ for at least one j . Since

$$\sum_{\alpha \in K^n} a_1^{k_1} \cdots a_n^{k_n} = \left(\sum_{a_1 \in K} a_1^{k_1} \right) \cdots \left(\sum_{a_n \in K} a_n^{k_n} \right),$$

it is enough to show that $S_k := \sum_{a \in K} a^k$ is zero for $0 \leq k < p-1$. If $k = 0$, then $a^k = 1$ and $S_0 = p \cdot 1 = 0$. Suppose $1 \leq k < p-1$ and let b be a generator for the multiplicative group K^\times of K . Then $c := b^k \neq 1$ and

$$S_k = \sum_{j=1}^{p-1} c^j = c(c^{p-1} - 1)/(c - 1) = 0. \quad \square$$

The general case of Proposition 34 was first proved by Warning (1936), after the particular case had been proved by Chevalley (1936). As an illustration, the particular case implies that, for any integers a, b, c and any prime p , the congruence $ax^2 + by^2 + cz^2 \equiv 0 \pmod{p}$ has a solution in integers x, y, z not all divisible by p .

If m is not a prime, then $\mathbb{Z}_{(m)}$ is not a field. However, we now show that the group $\mathbb{Z}_{(m)}^\times$ is cyclic also if $m = p^2$ is the square of a prime.

Let g be a primitive root of p . It follows from the binomial theorem that

$$(g + p)^p \equiv g^p \pmod{p^2}.$$

Hence, if $g^p \equiv g \pmod{p^2}$, then $(g + p)^p \not\equiv g + p \pmod{p^2}$. Thus, by replacing g by $g + p$ if necessary, we may assume that $g^{p-1} \not\equiv 1 \pmod{p^2}$. If the order of g in $\mathbb{Z}_{(p^2)}^\times$ is d , then d divides $\phi(p^2) = p(p-1)$. But $\phi(p) = p-1$ divides d , since $g^d \equiv 1 \pmod{p^2}$ implies $g^d \equiv 1 \pmod{p}$ and g is a primitive root of p . Since p is prime and $d \neq p-1$, it follows that $d = p(p-1)$, i.e. $\mathbb{Z}_{(p^2)}^\times$ is cyclic with g as generator.

We briefly state some further results about primitive roots, although we will not use them. Gauss (*D.A.*, §§89–92) showed that *the group $\mathbb{Z}_{(m)}^\times$ is cyclic if and only if $m \in \{2, 4, p^k, 2p^k\}$, where p is an odd prime and $k \in \mathbb{N}$. Evidently 1 is a primitive root of 2 and 3 is a primitive root of 4. If g is a primitive root of p^2 , where p is an odd prime, then g is a primitive root of p^k for every $k \in \mathbb{N}$; and if $g' = g$ or $g + p^k$, according as g is odd or even, then g' is a primitive root of $2p^k$.*

By Fermat's little theorem, if p is prime, then $a^{p-1} \equiv 1 \pmod{p}$ for every $a \in \mathbb{Z}$ such that $(a, p) = 1$. With the aid of primitive roots we will now show that there exist also composite integers n such that $a^{n-1} \equiv 1 \pmod{n}$ for every $a \in \mathbb{Z}$ such that $(a, n) = 1$.

Proposition 35 *For any integer $n > 1$, the following two statements are equivalent:*

- (i) $a^{n-1} \equiv 1 \pmod{n}$ for every integer a such that $(a, n) = 1$;
- (ii) n is a product of distinct primes and, for each prime $p|n$, $p-1$ divides $n-1$.

Proof Suppose first that (i) holds and assume that, for some prime p , $p^2|n$. As we have just proved, there exists a primitive root g of p^2 . Evidently $p \nmid g$. It is easily seen that there exists $c \in \mathbb{N}$ such that $a = g + cp^2$ is relatively prime to n ; in fact we can take c to be the product of the distinct prime factors of n , other than p , which do not divide g . Since n divides $a^{n-1} - 1$, also p^2 divides $a^{n-1} - 1$. But a , like g , is a primitive root of p^2 , and so its order in $\mathbb{Z}_{(p^2)}^\times$ is $\phi(p^2) = p(p-1)$. Hence $p(p-1)$ divides $n-1$. But this contradicts $p|n$.

Now let p be any prime divisor of n and let g be a primitive root of p . In the same way as before, there exists $c \in \mathbb{N}$ such that $a = g + cp$ is relatively prime to n . Arguing as before, we see that $\phi(p) = p-1$ divides $n-1$. This proves that (i) implies (ii).

Suppose next that (ii) holds and let a be any integer relatively prime to n . If p is a prime factor of n , then $p \nmid a$ and hence $a^{p-1} \equiv 1 \pmod{p}$. Since $p-1$ divides $n-1$, it follows that $a^{n-1} \equiv 1 \pmod{p}$. Thus $a^{n-1} - 1$ is divisible by each prime factor of n and hence, since n is squarefree, also by n itself. \square

Proposition 35 was proved by Carmichael (1910), and a composite integer n with the equivalent properties stated in the proposition is said to be a *Carmichael number*.

Any Carmichael number n must be odd, since it has an odd prime factor p such that $p - 1$ divides $n - 1$. Furthermore a Carmichael number must have more than two prime factors. For assume $n = pq$, where $1 < p < q < n$ and $q - 1$ divides $n - 1$. Since $q \equiv 1 \pmod{q - 1}$, it follows that

$$0 \equiv pq - 1 \equiv p - 1 \pmod{q - 1},$$

which contradicts $p < q$.

The composite integer $561 = 3 \times 11 \times 17$ is a Carmichael number, since 560 is divisible by 2, 10 and 16, and it is in fact the smallest Carmichael number. The taxicab number 1729, which Hardy reckoned to Ramanujan was uninteresting, is also a Carmichael number, since $1729 = 7 \times 13 \times 19$. Indeed it is not difficult to show that if p , $2p - 1$ and $3p - 2$ are all primes, with $p > 3$, then their product is a Carmichael number. Recently Alford, Granville and Pomeroy (1994) confirmed a long-standing conjecture by proving that there are infinitely many Carmichael numbers.

Our next topic is of greater importance. Many arithmetical problems require for their solution the determination of an integer which is congruent to several given integers according to various given moduli. We consider first a simple, but important, special case.

Proposition 36 *Let $m = m'm''$, where m' and m'' are relatively prime integers. Then, for any integers a' , a'' , there exists an integer a , which is uniquely determined mod m , such that*

$$a \equiv a' \pmod{m'}, \quad a \equiv a'' \pmod{m''}.$$

Moreover, a is relatively prime to m if and only if a' is relatively prime to m' and a'' is relatively prime to m'' .

Proof By Proposition 22, there exist integers c' , c'' such that

$$c'm'' \equiv 1 \pmod{m'}, \quad c''m' \equiv 1 \pmod{m''}.$$

Thus $e' := c'm''$ is congruent to 1 mod m' and congruent to 0 mod m'' . Similarly $e'' := c''m'$ is congruent to 0 mod m' and congruent to 1 mod m'' . It follows that $a = a'e' + a''e''$ is congruent to $a' \pmod{m'}$ and congruent to $a'' \pmod{m''}$.

It is evident that if $b \equiv a \pmod{m}$, then also $b \equiv a' \pmod{m'}$ and $b \equiv a'' \pmod{m''}$. Conversely, if b satisfies these two congruences, then $b - a \equiv 0 \pmod{m'}$ and $b - a \equiv 0 \pmod{m''}$. Hence $b - a \equiv 0 \pmod{m}$, by Proposition 3(i).

Since m' and m'' are relatively prime, it follows from Proposition 3(iv) that $(a, m) = 1$ if and only if $(a, m') = (a, m'') = 1$. Since $a \equiv a' \pmod{m'}$ implies $(a, m') = (a', m')$, and $a \equiv a'' \pmod{m''}$ implies $(a, m'') = (a'', m'')$, this proves the last statement of the proposition. \square

In algebraic terms, Proposition 36 says that if $m = m'm''$, where m' and m'' are relatively prime integers, then the ring $\mathbb{Z}_{(m)}$ is (isomorphic to) the direct sum of the rings $\mathbb{Z}_{(m')}$ and $\mathbb{Z}_{(m'')}$. Furthermore, the group $\mathbb{Z}_{(m)}^\times$ is (isomorphic to) the direct product of the groups $\mathbb{Z}_{(m')}^\times$ and $\mathbb{Z}_{(m'')}^\times$.

Proposition 36 can be considerably generalized:

Proposition 37 *For any integers m_1, \dots, m_n and a_1, \dots, a_n , the simultaneous congruences*

$$x \equiv a_1 \pmod{m_1}, \dots, x \equiv a_n \pmod{m_n}$$

have a solution x if and only if

$$a_j \equiv a_k \pmod{(m_j, m_k)} \quad \text{for } 1 \leq j < k \leq n.$$

Moreover, y is also a solution if and only if

$$y \equiv x \pmod{[m_1, \dots, m_n]}.$$

proof The necessity of the conditions is trivial. For if x is a solution and if $d_{jk} = (m_j, m_k)$ is the greatest common divisor of m_j and m_k , then $a_j \equiv x \equiv a_k \pmod{d_{jk}}$. Also, if y is another solution, then $y - x$ is divisible by m_1, \dots, m_n and hence also by their least common multiple $[m_1, \dots, m_n]$.

We prove the sufficiency of the conditions by induction on n . Suppose first that $n = 2$ and $a_1 \equiv a_2 \pmod{d}$, where $d = (m_1, m_2)$. By the Bézout identity,

$$d = x_1 m_1 - x_2 m_2$$

for some $x_1, x_2 \in \mathbb{Z}$. Since $a_1 - a_2 = kd$ for some $k \in \mathbb{Z}$, it follows that

$$x := a_1 - kx_1 m_1 = a_2 - kx_2 m_2$$

is a solution.

Suppose next that $n > 2$ and the result holds for all smaller values of n . Then there exists $x' \in \mathbb{Z}$ such that

$$x' \equiv a_i \pmod{m_i} \quad \text{for } 1 \leq i < n,$$

and x' is uniquely determined mod m' , where $m' = [m_1, \dots, m_{n-1}]$. Since any solution of the two congruences

$$x \equiv x' \pmod{m'}, x \equiv a_n \pmod{m_n}$$

is a solution of the given congruences, we need only show that $x' \equiv a_n \pmod{(m', m_n)}$. But, by the distributive law connecting greatest common divisors and least common multiples,

$$(m', m_n) = [(m_1, m_n), \dots, (m_{n-1}, m_n)].$$

Since $x' \equiv a_i \equiv a_n \pmod{(m_i, m_n)}$ for $1 \leq i < n$, it follows that $x' \equiv a_n \pmod{(m', m_n)}$. \square

Corollary 38 *Let m_1, \dots, m_n be integers, any two of which are relatively prime, and let $m = m_1 \cdots m_n$ be their product. Then, for any given integers a_1, \dots, a_n , there is a unique integer $x \pmod{m}$ such that*

$$x \equiv a_1 \pmod{m_1}, \dots, x \equiv a_n \pmod{m_n}.$$

Moreover, x is relatively prime to m if and only if a_i is relatively prime to m_i for $1 \leq i \leq n$.

Corollary 38 can also be proved by an extension of the argument used to prove Proposition 36. Both Proposition 37 and Corollary 38 are referred to as the *Chinese remainder theorem*. Sunzi (4th century A.D.) gave a procedure for obtaining the solution $x = 23$ of the simultaneous congruences

$$x \equiv 2 \pmod{3}, \quad x \equiv 3 \pmod{5}, \quad x \equiv 2 \pmod{7}.$$

Qin Jiushao (1247) gave a general procedure for solving simultaneous congruences, the moduli of which need not be pairwise relatively prime, although he did not state the necessary condition for the existence of a solution. The problem appears to have its origin in the construction of calendars.

6 Sums of Squares

Which positive integers n can be represented as a sum of two squares of integers? The question is answered completely by the following proposition, which was stated by Girard (1625). Fermat (1645) claimed to have a proof, but the first published proof was given by Euler (1754).

Proposition 39 *A positive integer n can be represented as a sum of two squares if and only if for each prime $p \equiv 3 \pmod{4}$ that divides n , the highest power of p dividing n is even.*

Proof We observe first that, since

$$(x^2 + y^2)(u^2 + v^2) = (xu + yv)^2 + (xv - yu)^2,$$

any product of sums of two squares is again a sum of two squares.

Suppose $n = x^2 + y^2$ for some integers x, y and that n is divisible by a prime $p \equiv 3 \pmod{4}$. Then $x^2 \equiv -y^2 \pmod{p}$. But -1 is not a square in the field \mathbb{F}_p , by Corollary 29. Consequently we must have $y^2 \equiv x^2 \equiv 0 \pmod{p}$. Thus p divides both x and y . Hence p^2 divides n and $(n/p)^2 = (x/p)^2 + (y/p)^2$. It follows by induction that the highest power of p which divides n is even.

Thus the condition in the statement of the proposition is necessary. Suppose now that this condition is satisfied. Then $n = qm^2$, where q is square-free and the only possible prime divisors of q are 2 and primes $p \equiv 1 \pmod{4}$. Since $m^2 = m^2 + 0^2$ and $2 = 1^2 + 1^2$, it follows from our initial observation that n is a sum of two squares if every prime $p \equiv 1 \pmod{4}$ is a sum of two squares. Following Gauss (1832), we will prove this with the aid of complex numbers.

A complex number $\gamma = a + bi$ is said to be a *Gaussian integer* if $a, b \in \mathbb{Z}$. The set of all Gaussian integers will be denoted by \mathcal{G} . Evidently $\gamma \in \mathcal{G}$ implies $\bar{\gamma} \in \mathcal{G}$, where $\bar{\gamma} = a - bi$ is the complex conjugate of γ . Moreover $\alpha, \beta \in \mathcal{G}$ implies $\alpha \pm \beta \in \mathcal{G}$ and $\alpha\beta \in \mathcal{G}$. Thus \mathcal{G} is a commutative ring. In fact \mathcal{G} is an integral domain, since it is a subset of the field \mathbb{C} . We are going to show that \mathcal{G} can be given the structure of a Euclidean domain.

Define the *norm* of a complex number $\gamma = a + bi$ to be

$$N(\gamma) = \gamma \bar{\gamma} = a^2 + b^2.$$

Then $N(\gamma) \geq 0$, with equality if and only if $\gamma = 0$, and $N(\gamma_1\gamma_2) = N(\gamma_1)N(\gamma_2)$. If $\gamma \in \mathcal{G}$, then $N(\gamma)$ is an ordinary integer. Furthermore, γ is a unit in \mathcal{G} , i.e. γ divides 1 in \mathcal{G} , if and only if $N(\gamma) = 1$.

We wish to show that if $\alpha, \beta \in \mathcal{G}$ and $\alpha \neq 0$, then there exist $\kappa, \rho \in \mathcal{G}$ such that

$$\beta = \kappa\alpha + \rho, \quad N(\rho) < N(\alpha).$$

We have $\beta\alpha^{-1} = r + si$, where $r, s \in \mathbb{Q}$. Choose $a, b \in \mathbb{Z}$ so that

$$|r - a| \leq 1/2, \quad |s - b| \leq 1/2.$$

If $\kappa = a + bi$, then $\kappa \in \mathcal{G}$ and

$$N(\beta\alpha^{-1} - \kappa) \leq 1/4 + 1/4 = 1/2 < 1.$$

Hence if $\rho = \beta - \kappa\alpha$, then $\rho \in \mathcal{G}$ and $N(\rho) < N(\alpha)$.

It follows that we can apply to \mathcal{G} the whole theory of divisibility in a Euclidean domain. Now let p be a prime such that $p \equiv 1 \pmod{4}$. We will show that p is a sum of two squares by constructing $\beta \in \mathcal{G}$ for which $N(\beta) = p$.

By Corollary 29, there exists an integer a such that $a^2 \equiv -1 \pmod{p}$. Put $\alpha = a + i$. Then $N(\alpha) = a\bar{a} = a^2 + 1$ is divisible by p in \mathbb{Z} and hence also in \mathcal{G} . However, neither α nor \bar{a} is divisible by p in \mathcal{G} , since αp^{-1} and $\bar{a} p^{-1}$ are not in \mathcal{G} . Thus p is not a prime in \mathcal{G} and consequently, since \mathcal{G} is a Euclidean domain, it has a factorization $p = \beta\gamma$, where neither β nor γ is a unit. Hence $N(\beta) > 1$, $N(\gamma) > 1$. Since

$$N(\beta)N(\gamma) = N(p) = p^2,$$

it follows that $N(\beta) = N(\gamma) = p$. □

Proposition 39 solves the problem of representing a positive integer as a sum of two squares. What if we allow more than two squares? When congruences were first introduced in §5, it was observed that a positive integer $a \equiv 7 \pmod{8}$ could not be represented as a sum of three squares. It was first completely proved by Gauss (1801) that a positive integer can be represented as a sum of three squares if and only if it is not of the form $4^n a$, where $n \geq 0$ and $a \equiv 7 \pmod{8}$. The proof of this result is more difficult, and will be given in Chapter VII.

It was conjectured by Bachet (1621) that *every* positive integer can be represented as a sum of four squares. Fermat claimed to have a proof, but the first published proof was given by Lagrange (1770), using earlier ideas of Euler (1751). The proof of the four-squares theorem we will give is similar to that just given for the two-squares theorem, with complex numbers replaced by quaternions.

Proposition 40 *Every positive integer n can be represented as a sum of four squares.*

Proof A quaternion $\gamma = a + bi + cj + dk$ will be said to be a *Hurwitz integer* if a, b, c, d are either all integers or all halves of odd integers. The set of all Hurwitz integers will be denoted by \mathcal{H} . Evidently $\gamma \in \mathcal{H}$ implies $\bar{\gamma} \in \mathcal{H}$, where $\bar{\gamma} = a - bi - cj - dk$. Moreover $\alpha, \beta \in \mathcal{H}$ implies $\alpha \pm \beta \in \mathcal{H}$. We will show that $\alpha, \beta \in \mathcal{H}$ also implies $\alpha\beta \in \mathcal{H}$.

Evidently $\gamma \in \mathcal{H}$ if and only if it can be written in the form $\gamma = a_0h + a_1i + a_2j + a_3k$, where $a_0, a_1, a_2, a_3 \in \mathbb{Z}$ and $h = (1 + i + j + k)/2$. It is obvious that the product of h with i, j or k is again in \mathcal{H} and it is easily verified that $h^2 = h - 1$. It follows that \mathcal{H} is closed under multiplication and hence is a ring.

Define the *norm* of a quaternion $\gamma = a + bi + cj + dk$ to be

$$N(\gamma) = \gamma \bar{\gamma} = a^2 + b^2 + c^2 + d^2.$$

Then $N(\gamma) \geq 0$, with equality if and only if $\gamma = 0$. Moreover, since $\overline{\gamma_1\gamma_2} = \bar{\gamma}_2\bar{\gamma}_1$,

$$N(\gamma_1\gamma_2) = \gamma_1\gamma_2\bar{\gamma}_2\bar{\gamma}_1 = \gamma_1\bar{\gamma}_1\gamma_2\bar{\gamma}_2 = N(\gamma_1)N(\gamma_2).$$

If $\gamma \in \mathcal{H}$, then $N(\gamma) = \gamma \bar{\gamma} \in \mathcal{H}$ and hence $N(\gamma)$ is an ordinary integer. Furthermore, γ is a unit in \mathcal{H} , i.e. γ divides 1 in \mathcal{H} , if and only if $N(\gamma) = 1$.

We now show that a Euclidean algorithm may be defined on \mathcal{H} . Suppose $\alpha, \beta \in \mathcal{H}$ and $\alpha \neq 0$. Then

$$\beta\alpha^{-1} = r_0 + r_1i + r_2j + r_3k,$$

where $r_0, r_1, r_2, r_3 \in \mathbb{Q}$. If $\kappa = a_0h + a_1i + a_2j + a_3k$, then

$$\begin{aligned} \beta\alpha^{-1} - \kappa &= (r_0 - a_0/2) + (r_1 - a_0/2 - a_1)i + (r_2 - a_0/2 - a_2)j \\ &\quad + (r_3 - a_0/2 - a_3)k. \end{aligned}$$

We can choose $a_0 \in \mathbb{Z}$ so that $|2r_0 - a_0| \leq 1/2$ and then choose $a_v \in \mathbb{Z}$ so that $|r_v - a_0/2 - a_v| \leq 1/2$ ($v = 1, 2, 3$). Then $\kappa \in \mathcal{H}$ and

$$N(\beta\alpha^{-1} - \kappa) \leq 1/16 + 3/4 = 13/16 < 1.$$

Thus if we set $\rho = \beta - \kappa\alpha$, then $\rho \in \mathcal{H}$ and

$$N(\rho) = N(\beta\alpha^{-1} - \kappa)N(\alpha) < N(\alpha).$$

By repeating this division process finitely many times we see that any $\alpha, \beta \in \mathcal{H}$ have a *greatest common right divisor* $\delta = (\alpha, \beta)_r$. Furthermore, there is a *left Bézout identity*: $\delta = \xi\alpha + \eta\beta$ for some $\xi, \eta \in \mathcal{H}$.

If a positive integer n is a sum of four squares, say $n = a^2 + b^2 + c^2 + d^2$, then $n = \gamma \bar{\gamma}$, where $\gamma = a + bi + cj + dk \in \mathcal{H}$. Since the norm of a product is the product of the norms, it follows that any product of sums of four squares is again a sum of four squares. Hence to prove the proposition we need only show that any prime p is a sum of four squares.

We show first that there exist integers a, b such that $a^2 + b^2 \equiv -1 \pmod{p}$. This follows from the illustration given for Proposition 34, but we will give a direct proof.

If $p = 2$, we can take $a = 1, b = 0$. If $p \equiv 1 \pmod{4}$ then, by Corollary 29, there exists an integer a such that $a^2 \equiv -1 \pmod{p}$ and we can take $b = 0$. Suppose now that $p \equiv 3 \pmod{4}$. Let c be the least positive quadratic non-residue of p . Then $c \geq 2$ and $c - 1$ is a quadratic residue of p . On the other hand, -1 is a quadratic non-residue of p , by Corollary 29. Hence, by Proposition 28, $-c$ is a quadratic residue. Thus there exist integers a, b such that

$$a^2 \equiv -c, b^2 \equiv c - 1 \pmod{p},$$

and then $a^2 + b^2 \equiv -1 \pmod{p}$.

Put $\alpha = 1 + ai + bj$. Then p divides $N(\alpha) = \alpha\bar{\alpha} = 1 + a^2 + b^2$ in \mathbb{Z} and hence also in \mathcal{H} . However, p does not divide either α or $\bar{\alpha}$ in \mathcal{H} , since αp^{-1} and $\bar{\alpha} p^{-1}$ are not in \mathcal{H} .

Let $\gamma = (p, \alpha)_r$. Then $p = \beta\gamma$ for some $\beta \in \mathcal{H}$. If β were a unit, p would be a right divisor of γ and hence also of α , which is a contradiction. Therefore $N(\beta) > 1$. Evidently $\gamma\bar{\alpha}$ is a common right divisor of $p\bar{\alpha}$ and $\alpha\bar{\alpha}$, and the Bézout representation for γ implies that $\gamma\bar{\alpha} = (p\bar{\alpha}, \alpha\bar{\alpha})_r$. Since $p\bar{\alpha} = \bar{\alpha}p$ and p divides $\alpha\bar{\alpha}$, it follows that p is a right divisor of $\gamma\bar{\alpha}$. Since p does not divide $\bar{\alpha}$, γ is not a unit and hence $N(\gamma) > 1$. Since

$$N(\beta)N(\gamma) = N(p) = p^2,$$

we must have $N(\beta) = N(\gamma) = p$.

Thus if $\gamma = c_0 + c_1i + c_2j + c_3k$, then $c_0^2 + c_1^2 + c_2^2 + c_3^2 = p$. If c_0, \dots, c_3 are all integers, we are finished. Otherwise c_0, \dots, c_3 are all halves of odd integers. Hence we can write $c_v = 2d_v + e_v$, where $d_v \in \mathbb{Z}$ and $e_v = \pm 1/2$. If we put

$$\delta = d_0 + d_1i + d_2j + d_3k, \quad \varepsilon = e_0 + e_1i + e_2j + e_3k,$$

then $\gamma = 2\delta + \varepsilon$ and $N(\varepsilon) = 1$. Hence $\theta := \gamma\bar{\varepsilon} = 2\delta\bar{\varepsilon} + 1$ has all its coordinates integers and $N(\theta) = N(\gamma) = p$. \square

In his *Meditationes Algebraicae*, which also contains the first statement in print of Wilson's theorem, Waring (1770) stated that every positive integer is a sum of at most 4 positive integral squares, of at most 9 positive integral cubes and of at most 19 positive integral fourth powers. The statement concerning squares was proved by Lagrange in the same year, as we have seen. The statement concerning cubes was first proved by Wieferich (1909), with a gap filled by Kempner (1912), and the statement concerning fourth powers was first proved by Balasubramanian, Deshouillers and Dress (1986).

In a later edition of his book, Waring (1782) raised the same question for higher powers. *Waring's problem* was first solved by Hilbert (1909), who showed that, for each $k \in \mathbb{N}$, there exists $\gamma_k \in \mathbb{N}$ such that every positive integer is a sum of at most γ_k k -th powers. The least possible value of γ_k is traditionally denoted by $g(k)$. For example, $g(2) = 4$, since $7 = 2^2 + 3 \cdot 1^2$ is not a sum of less than 4 squares.

A lower bound for $g(k)$ was already derived by Euler (c. 1772). Let $m = \lfloor (3/2)^k \rfloor$ denote the greatest integer $\leq (3/2)^k$ and take

$$n = 2^k m - 1.$$

Since $1 \leq n < 3^k$, the only k -th powers of which n can be the sum are 0^k , 1^k and 2^k . Since the number of powers 2^k must be less than m , and since $n = (m - 1)2^k + (2^k - 1)1^k$, the least number of k -th powers with sum n is $m + 2^k - 2$. Hence $g(k) \geq w(k)$, where

$$w(k) = \lfloor (3/2)^k \rfloor + 2^k - 2.$$

In particular,

$$w(2) = 4, \quad w(3) = 9, \quad w(4) = 19, \quad w(5) = 37, \quad w(6) = 73.$$

By the results stated above, $g(k) = w(k)$ for $k = 2, 3, 4$ and this has been shown to hold also for $k = 5$ by Chen (1964) and for $k = 6$ by Pillai (1940).

Hilbert's method of proof yielded rather large upper bounds for $g(k)$. A completely new approach was developed in the 1920's by Hardy and Littlewood, using their analytic 'circle' method. They showed that, for each $k \in \mathbb{N}$, there exists $\Gamma_k \in \mathbb{N}$ such that every sufficiently large positive integer is a sum of at most Γ_k k -th powers. The least possible value of Γ_k is traditionally denoted by $G(k)$. For example, $G(2) = 4$, since no positive integer $n \equiv 7 \pmod{8}$ is a sum of less than four squares. Davenport (1939) showed that $G(4) = 16$, but these are the only two values of k for which today $G(k)$ is known exactly.

It is obvious that $G(k) \leq g(k)$, and in fact $G(k) < g(k)$ for all $k > 2$. In particular, Dickson (1939) showed that 23 and 239 are the only positive integers which require the maximum 9 cubes. Hardy and Littlewood obtained the upper bound $G(k) \leq (k-2)2^{k-1} + 5$, but this has been repeatedly improved by Hardy and Littlewood themselves, Vinogradov and others. For example, Wooley (1992) has shown that $G(k) \leq k(\log k + \log \log k + O(1))$.

By using the upper bound for $G(k)$ of Vinogradov (1935), it was shown by Dickson, Pillai and Niven (1936–1944) that $g(k) = w(k)$ for any given $k > 6$, *provided that*

$$(3/2)^k - \lfloor (3/2)^k \rfloor \leq 1 - \lfloor (3/2)^k \rfloor / 2^k.$$

It is possible that this inequality holds for every $k \in \mathbb{N}$. For a given k , it may be checked by direct calculation, and Kubina and Wunderlich (1990) have verified in this way that the inequality holds if $k \leq 471600000$. Furthermore, using a p -adic extension by Ridout (1957) of the theorem of Roth (1955) on the approximation of algebraic numbers by rationals, Mahler (1957) proved that there exists $k_0 \in \mathbb{N}$ such that the inequality holds for all $k \geq k_0$. However, the proof does not provide a means of estimating k_0 .

Thus we have the bizarre situation that $G(k)$ is known for only two values of k , that $g(k)$ is known for a vast number of values of k and is given by a simple formula, probably for all k , but the information about $g(k)$ is at present derived from information about $G(k)$. Is it too much to hope that an examination of the numerical data will reveal some pattern in the fractional parts of $(3/2)^k$?

7 Further Remarks

There are many good introductory books on the theory of numbers, e.g. Davenport [4], LeVeque [28] and Scholz [41]. More extensive accounts are given in Hardy and Wright [15], Hua [18], Narkiewicz [33] and Niven *et al.* [34].

Historical information is provided by Dickson [5], Smith [42] and Weil [46], as well as the classics Euclid [11], Gauss [13] and Dirichlet [6]. Gauss's masterpiece is quoted here and in the text as '*D.A.*'

The reader is warned that, besides its use in §1, the word 'lattice' also has quite a different mathematical meaning, which will be encountered in Chapter VIII.

The basic theory of divisibility is discussed more thoroughly than in the usual texts by Stieltjes [43]. For Proposition 6, see Prüfer [35]. In the theory of groups, Schreier's

refinement theorem and the Jordan–Hölder theorem may be viewed as generalizations of Propositions 6 and 7. These theorems are stated and proved in Chapter I, §3 of Lang [23]. The fundamental theorem of arithmetic (Proposition 7) is usually attributed to Gauss (*D.A.*, §16). However, it is really contained in Euclid’s *Elements* (Book VII, Proposition 31 and Book IX, Proposition 14), except for the appropriate terminology. Perhaps this is why Euler and his contemporaries simply assumed it without proof.

Generalizations of the fundamental theorem of arithmetic to other algebraic structures are discussed in Chap. 2 of Jacobson [21]. For factorial domains, see Samuel [39].

Our discussion of the fundamental theorem did not deal with the practical problems of deciding if a given integer is prime or composite and, in the latter case, of obtaining its factorization into primes. Evidently if the integer a is composite, its least prime factor p satisfies $p^2 \leq a$. In former days one used this observation in conjunction with tables, such as [24], [25], [26]. With new methods and supercomputers, the primality of integers with hundreds of digits can now be determined without difficulty. The progress in this area may be traced through the survey articles [48], [7] and [27]. Factorization remains a more difficult problem, and this difficulty has found an important application in *public-key cryptography*; see Rivest *et al.* [37].

For Proposition 12, cf. Hillman and Hoggatt [17]. A proof that the ring of all algebraic integers is a Bézout domain is given on p. 86 of Mann [31]. The ring of all functions which are holomorphic in a given region was shown to be a Bézout domain by Wedderburn (1915); see Narasimhan [32].

For Gauss’s version of Proposition 17, see *D.A.*, §42. It is natural to ask if Corollary 18 remains valid if the polynomial ring $R[t]$ is replaced by the ring $R[[t]]$ of formal power series. The ring $K[[t_1, \dots, t_m]]$ of all formal power series in finitely many indeterminates with coefficients from an arbitrary field K is indeed a factorial domain. However, if R is a factorial domain, the integral domain $R[[t]]$ of all formal power series in t with coefficients from R need not be factorial. For an example in which R is actually a complete local ring, see Salmon [38].

For generalizations of Eisenstein’s irreducibility criterion (Proposition 19), see Gao [12]. Proposition 21 is proved in Rhai [36]. Euclidean domains are studied further in Samuel [40]. Quadratic fields $\mathbb{Q}(\sqrt{d})$ whose ring of integers \mathcal{O}_d is Euclidean are discussed in Clark [3], Dubois and Steger [8] and Eggleton *et al.* [9].

Congruences are discussed in all the books on number theory cited above. In connection with Lemma 32 we mention a result of Frobenius (1895). Frobenius proved that if G is a finite group of order n and if d is a positive divisor of n , then the number of elements of G whose order divides d is a multiple of d . He conjectured that if the number is exactly d , then these elements form a (normal) subgroup of G . The conjecture can be reduced to the case where G is simple, since a counterexample of minimal order must be a noncyclic simple group. By appealing to the recent classification of all finite simple groups (see Chapter V, §7), the proof of the conjecture was completed by Iiyori and Yamaki [20].

There is a table of primitive roots on pp. 52–56 of Hua [18]. For more extensive tables, see Western and Miller [47].

It is easily seen that an even square is never a primitive root, that an odd square (including 1) is a primitive root only for the prime $p = 2$, and that -1 is a primitive root only for the primes $p = 2, 3$. Artin (1927) conjectured that if the integer a is not

a square or -1 , then it is a primitive root for infinitely many primes p . (A quantitative form of the conjecture is considered in Chapter IX.) If the conjecture is not true, then it is almost true, since it has been shown by Heath-Brown [16] that there are at most 3 square-free positive integers a for which it fails.

A finite subgroup of the multiplicative group of a division ring need not be cyclic. For example, if \mathbb{H} is the division ring of Hamilton's quaternions, \mathbb{H}^\times contains the non-cyclic subgroup $\{\pm 1, \pm i, \pm j, \pm k\}$ of order 8. All possible finite subgroups of the multiplicative group of a division ring have been determined (with the aid of *class field theory*) by Amitsur [2].

For Carmichael numbers, see Alford *et al.* [1].

Galois (1830) showed that there were other finite fields besides \mathbb{F}_p and indeed, as Moore (1893) later proved, he found them all. Finite fields have the following basic properties:

- (i) The number of elements in a finite field is a prime power p^n , where $n \in \mathbb{N}$ and the prime p is the characteristic of the field.
- (ii) For any prime power $q = p^n$, there is a finite field \mathbb{F}_q containing exactly q elements. Moreover the field \mathbb{F}_q is unique, up to isomorphism, and is the splitting field of the polynomial $t^q - t$ over \mathbb{F}_p .
- (iii) For any finite field \mathbb{F}_q , the multiplicative group \mathbb{F}_q^\times of nonzero elements is cyclic.
- (iv) If $q = p^n$, the map $\sigma : a \rightarrow a^p$ is an automorphism of \mathbb{F}_q and the distinct automorphisms of \mathbb{F}_q are the powers σ^k ($k = 0, 1, \dots, n-1$).

The theorem of Chevalley and Warning (Proposition 34) extends immediately to arbitrary finite fields. Proofs and more detailed information on finite fields may be found in Lidl and Niederreiter [30] and in Joly [22].

A celebrated theorem of Wedderburn (1905) states that any finite division ring is a field, i.e. the commutative law of multiplication is a consequence of the other field axioms if the number of elements is finite. Here is a purely algebraic proof.

Assume there exists a finite division ring which is not a field and let D be one of minimum cardinality. Let C be the centre of D and $a \in D \setminus C$. The set M of all elements of D which commute with a is a field, since it is a division ring but not the whole of D . Evidently M is a maximal subfield of D which contains a . If $[D : C] = n$ and $[M : C] = m$ then, by Proposition I.32, $[D : M] = m$ and $n = m^2$. Thus m is independent of a .

If C has cardinality q , then D has cardinality q^n , M has cardinality q^m and the number of conjugates of a in D is $(q^n - 1)/(q^m - 1)$. Since this holds for every $a \in D \setminus C$, the partition of the multiplicative group of D into conjugacy classes shows that

$$q^n - 1 = q - 1 + r(q^n - 1)/(q^m - 1)$$

for some positive integer r . Hence $q - 1$ is divisible by

$$(q^n - 1)/(q^m - 1) = 1 + q^m + \dots + q^{m(m-1)}.$$

Since $n > m > 1$, this is a contradiction.

For the history of the Chinese remainder theorem (not only in China), see Libbrecht [29].

We have developed the arithmetic of quaternions only as far as is needed to prove the four-squares theorem. A fuller account was given in the original (1896) paper of Hurwitz [19]. For more information about sums of squares, see Grosswald [14] and also Chapter XIII. For Waring's problem, see Waring [45], Ellison [10] and Vaughan [44].

8 Selected References

- [1] W.R. Alford, A. Granville and C. Pomerance, There are infinitely many Carmichael numbers, *Ann. Math.* **139** (1994), 703–722.
- [2] S.A. Amitsur, Finite subgroups of division rings, *Trans. Amer. Math. Soc.* **80** (1955), 361–386.
- [3] D.A. Clark, A quadratic field which is Euclidean but not norm-Euclidean, *Manuscripta Math.* **83** (1994), 327–330.
- [4] H. Davenport, *The higher arithmetic*, 7th ed., Cambridge University Press, 1999.
- [5] L.E. Dickson, *History of the theory of numbers*, 3 vols., Carnegie Institute, Washington, D.C., 1919–1923. [Reprinted, Chelsea, New York, 1992.]
- [6] P.G.L. Dirichlet, *Lectures on number theory*, with supplements by R. Dedekind, English transl. by J. Stillwell, American Mathematical Society, Providence, R.I., 1999. [German original, 1894.]
- [7] J.D. Dixon, Factorization and primality tests, *Amer. Math. Monthly* **91** (1984), 333–352.
- [8] D.W. Dubois and A. Steger, A note on division algorithms in imaginary quadratic fields, *Canad. J. Math.* **10** (1958), 285–286.
- [9] R.B. Eggleton, C.B. Lacampagne and J.L. Selfridge, Euclidean quadratic fields, *Amer. Math. Monthly* **99** (1992), 829–837.
- [10] W.J. Ellison, Waring's problem, *Amer. Math. Monthly* **78** (1971), 10–36.
- [11] Euclid, *The thirteen books of Euclid's elements*, English translation by T.L. Heath, 2nd ed., reprinted in 3 vols., Dover, New York, 1956.
- [12] S. Gao, Absolute irreducibility of polynomials via Newton polytopes, *J. Algebra* **237** (2001), 501–520.
- [13] C.F. Gauss, *Disquisitiones arithmeticae*, English translation by A.A. Clarke, revised by W.C. Waterhouse, Springer, New York, 1986. [Latin original, 1801.]
- [14] E. Grosswald, *Representations of integers as sums of squares*, Springer-Verlag, New York, 1985.
- [15] G.H. Hardy and E.M. Wright, *An introduction to the theory of numbers*, 6th ed., Oxford University Press, 2008.
- [16] D.R. Heath-Brown, Artin's conjecture for primitive roots, *Quart. J. Math. Oxford Ser. (2)* **37** (1986), 27–38.
- [17] A.P. Hillman and V.E. Hoggatt, Exponents of primes in generalized binomial coefficients, *J. Reine Angew. Math.* **262/3** (1973), 375–380.
- [18] L.K. Hua, *Introduction to number theory*, English translation by P. Shiu, Springer-Verlag, Berlin, 1982.
- [19] A. Hurwitz, Über die Zahlentheorie der Quaternionen, *Mathematische Werke, Band II*, pp. 303–330, Birkhäuser, Basel, 1933.
- [20] N. Iiyori and H. Yamaki, On a conjecture of Frobenius, *Bull. Amer. Math. Soc. (N.S)* **25** (1991), 413–416.
- [21] N. Jacobson, *Basic Algebra I*, 2nd ed., W.H. Freeman, New York, 1985.
- [22] J.-R. Joly, Équations et variétés algébriques sur un corps fini, *Enseign. Math. (2)* **19** (1973), 1–117.

- [23] S. Lang, *Algebra*, corrected reprint of 3rd ed., Addison-Wesley, Reading, Mass., 1994.
- [24] D.H. Lehmer, *Guide to tables in the theory of numbers*, National Academy of Sciences, Washington, D.C., reprinted 1961.
- [25] D.N. Lehmer, *List of prime numbers from 1 to 10,006,721*, reprinted, Hafner, New York, 1956.
- [26] D.N. Lehmer, *Factor table for the first ten millions*, reprinted, Hafner, New York, 1956.
- [27] A.K. Lenstra, Primality testing, *Proc. Symp. Appl. Math.* **42** (1990), 13–25.
- [28] W.J. LeVeque, *Fundamentals of number theory*, reprinted Dover, Mineola, N.Y., 1996.
- [29] U. Libbrecht, *Chinese mathematics in the thirteenth century*, MIT Press, Cambridge, Mass., 1973.
- [30] R. Lidl and H. Niederreiter, *Finite fields*, 2nd ed., Cambridge University Press, 1997.
- [31] H.B. Mann, *Introduction to algebraic number theory*, Ohio State University, Columbus, Ohio, 1955.
- [32] R. Narasimhan, *Complex analysis in one variable*, Birkhäuser, Boston, Mass., 1985.
- [33] W. Narkiewicz, *Number theory*, English translation by S. Kanemitsu, World Scientific, Singapore, 1983.
- [34] I. Niven, H.S. Zuckerman and H.L. Montgomery, *An introduction to the theory of numbers*, 5th ed., Wiley, New York, 1991.
- [35] H. Prüfer, Untersuchungen über Teilbarkeitseigenschaften, *J. Reine Angew. Math.* **168** (1932), 1–36.
- [36] T.-S. Rhai, A characterization of polynomial domains over a field, *Amer. Math. Monthly* **69** (1962), 984–986.
- [37] R.L. Rivest, A. Shamir and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, *Comm. ACM* **21** (1978), 120–126.
- [38] P. Salmon, Sulla fattorialità delle algebre graduate e degli anelli locali, *Rend. Sem. Mat. Univ. Padova* **41** (1968), 119–138.
- [39] P. Samuel, Unique factorization, *Amer. Math. Monthly* **75** (1968), 945–952.
- [40] P. Samuel, About Euclidean rings, *J. Algebra* **19** (1971), 282–301.
- [41] A. Scholz, *Einführung in die Zahlentheorie*, revised and edited by B. Schoeneberg, 5th ed., de Gruyter, Berlin, 1973.
- [42] H.J.S. Smith, Report on the theory of numbers, *Collected mathematical papers, Vol. 1*, pp. 38–364, reprinted, Chelsea, New York, 1965. [Original, 1859–1865.]
- [43] T.J. Stieltjes, Sur la théorie des nombres, *Ann. Fac. Sci. Toulouse* **4** (1890), 1–103. [Reprinted in Tome 2, pp. 265–377 of T.J. Stieltjes, *Oeuvres complètes*, 2 vols., Noordhoff, Groningen, 1914–1918.]
- [44] R.C. Vaughan, *The Hardy–Littlewood method*, 2nd ed., Cambridge Tracts in Mathematics **125**, Cambridge University Press, 1997.
- [45] E. Waring, *Meditationes algebraicae*, English transl. of 1782 edition by D. Weeks, Amer. Math. Soc., Providence, R.I., 1991.
- [46] A. Weil, *Number theory: an approach through history*, Birkhäuser, Boston, Mass., 1984.
- [47] A.E. Western and J.C.P. Miller, *Tables of indices and primitive roots*, Royal Soc. Math. Tables, Vol. 9, Cambridge University Press, London, 1968.
- [48] H.C. Williams, Primality testing on a computer, *Ars Combin.* **5** (1978), 127–185.

Additional References

- M. Agarwal, N. Kayal and N. Saxena, PRIMES is in P, *Ann. of Math.* **160** (2004), 781–793. [An unconditional deterministic polynomial-time algorithm for determining if an integer > 1 is prime or composite.]
- A. Granville, It is easy to determine whether a given integer is prime, *Bull. Amer. Math. Soc. (N.S.)* **42** (2005), 3–38.



<http://www.springer.com/978-0-387-89485-0>

Number Theory

An Introduction to Mathematics

Coppel, W.A.

2009, XIV, 610 p. 17 illus., Softcover

ISBN: 978-0-387-89485-0