1
Euclidean Arithmetic

Preview

Euclid's *Elements*, from around 300 BCE, is the source of many basic parts of modern mathematics, such as geometry, the axiomatic method, and the theory of real numbers. It is also the source of *arithmetic* as mathematicians know it: the theory of addition and multiplication of natural numbers, with emphasis on the concepts of divisibility and primes.

For Euclid, a natural number b is a **divisor** of a natural number a if

$$a = bc \quad \text{for some natural number } c.$$

Then a natural number $p > 1$ is *prime* if its only divisors are itself and 1. These concepts lead, as Euclid showed by a short but ingenious proof, to the discovery that there are infinitely many primes.

Even more ingeniously, Euclid proved the **prime divisor property**: If a prime p divides a product ab, then p divides a or p divides b. His proof is based on the famous **Euclidean algorithm** for finding the greatest common divisor of two natural numbers. The prime divisor property easily implies what we now call the **fundamental theorem of arithmetic**, or **unique prime factorization**: Every natural number greater than 1 may be expressed uniquely (up to the order of factors) as a product of primes.

Unique prime factorization is so useful that mathematicians would like it to hold wherever the concept of "factorization" makes sense. In fact, as we will see in later chapters, even when it is lost they will try to recover it. In this chapter we prepare to explore more general domains for factorization by introducing the concepts (and some examples) of **ring** and **field**.

1.1 Divisors and Primes

In this chapter we will be working mainly with the set $\mathbb{N} = \{0, 1, 2, 3, 4, 5, \ldots\}$ of **natural numbers**. These are the numbers obtained from 0 by "counting": that is, by repeatedly adding 1. It follows (informally) that from any natural number n we can reach 0 in a finite number of steps by "counting backwards," and hence that *any set of natural numbers has a least member*. Since Euclid, this so-called **well-ordering** property of $\mathbb{N}$ has been the basis of virtually all reasoning about the natural numbers, so it is usually taken as an axiom. In this section we will use it, as Euclid did, to prove results about divisibility and primes.

We have already said what it means for a natural number b to divide a natural number a; namely, $a = bc$ for some natural number c. So if b does *not* divide a, we necessarily have, for any natural number q,

$$a = bq + r, \quad \text{with } r > 0.$$

When r is least possible, we call q the **quotient** (of a by b) and r the **remainder**. It then follows that $0 < r < b$, because if $r = b + r'$, we would have

$$a = b(q + 1) + r', \quad \text{contrary to the assumption that } r \text{ is the least remainder.}$$

The two cases, where b does and does not divide a, can be combined in the following **division property**: For any natural numbers, there are natural numbers q and r such that

$$a = bq + r, \quad \text{where} \quad 0 \leq r < b. \tag{*}$$

This property is often misleadingly called the "division algorithm." (It is not an algorithm, but it paves the way for the very important Euclidean algorithm, as we will see in the next section.) Finding the quotient and remainder for a given pair a, b is called **division with remainder**.

Another easy application of well-ordering of $\mathbb{N}$ tells us that *every natural number greater than 1 is divisible by a prime*. Start with any natural number $a > 1$. If a is not prime, then $a = bc$ for some smaller numbers b and c. Then if b is not prime, we have $b = de$ for some smaller natural numbers d and e, and so on. Since natural numbers cannot decrease forever, this process must halt – necessarily with a prime p that divides a. It follows, by repeatedly finding prime divisors, that *every natural number has a prime factorization*.

With these easy properties of divisors and primes, we are now ready for something ingenious: Euclid's proof that there are infinitely many primes.

Infinitude of primes. *For any prime numbers* $p_1, p_2, \ldots, p_k$*, there is a prime number* $p_{k+1} \neq p_1, p_2, \ldots, p_k$.

Proof. Consider the number $N = (p_1 \cdot p_2 \cdots p_k) + 1$. None of $p_1, p_2, \ldots, p_k$ divide N because they each leave remainder 1. But *some* prime divides N because $N > 1$. This prime is the p_{k+1} we seek. □

The beauty of this proof is that it avoids having to find any pattern in the sequence of primes, or finding divisors of a number, both of which are hard problems.

1.1.1 The Euclidean Algorithm

Although it is hard to find the divisors of a given (large) natural number, it is surprisingly quick and easy to find *common* divisors of two natural numbers. This can be done by the **Euclidean algorithm** for finding the **greatest common divisor** $\gcd(a,b)$ of two natural numbers a and b. As Euclid described it, (*Elements*, Book VII, Proposition 1) the algorithm "repeatedly subtracts the lesser number from the greater." More formally, it repeatedly replaces the pair $\{a,b\}$, where $a > b$, by the pair $\{b, a-b\}$ until the members of the pair become equal – at which stage each member is $\gcd(a,b)$.

For example, if we begin with the pair $\{34,21\}$, the pairs produced by the algorithm are the following

$$\{34,21\} \to \{21,13\} \to \{13,8\} \to \{8,5\} \to \{5,3\} \to \{3,2\} \to \{2,1\} \to \{1,1\}.$$

And we conclude that $\gcd(34,21) = 1$.

In general, the correctness of the Euclidean algorithm is guaranteed by the following theorem.

Euclidean algorithm produces the gcd. *If the Euclidean algorithm is applied to two natural numbers* $a, b > 0$*, then it terminates in a finite number of steps with the pair whose members are both* $\gcd(a,b)$.

Proof. Suppose that d is any common divisor of a and b, where $a > b$. This means that $a = a'd$ and $b = b'd$ for some $a', b' > 0$, and hence that

$$a - b = (a' - b')d.$$

Thus, d is also a divisor of $a - b$. There is a similar proof that any common divisor of two numbers is also a divisor of their sum, so a divisor of b and $a-b$ is also a divisor of $b + (a-b) = a$. It follows that *each* pair produced by the Euclidean algorithm has the same common divisors, and hence the same gcd.

Now, as long as the pairs produced by the algorithm are unequal, subtraction occurs, and it will decrease the sum of the two members of the pair. By the well-ordering of $\mathbb{N}$, the sum cannot decrease forever, so the algorithm necessarily halts with a pair of equal numbers. Being equal, they equal their own gcd; hence they each equal $\gcd(a,b)$. □

In practice it is usual to speed up the Euclidean algorithm by doing **division with remainder** instead of subtraction. That is, we replace the pair $\{a,b\}$, where $a > b$, with the pair $\{b,r\}$, where r is the remainder when a is divided by b. This process is simply a shortening of repeated subtraction, because r can be found by subtracting b repeatedly from a. However, the usual "long division" process generally finds r more quickly than repeated subtraction.

In fact, by using division with remainder, we can be sure that the number of steps required for the Euclidean algorithm to halt is roughly proportional to the number of decimal digits in a. The example above, incidentally, is one where each division with remainder is actually the same as a single subtraction. This happens whenever a and b are a pair of consecutive **Fibonacci numbers**: the numbers $0,1,1,2,3,5,8,13,21,34,55,89,144,233,\ldots$ defined by

$$F_0 = 0, \quad F_{n+2} = F_{n+1} + F_n.$$

This is the case where the Euclidean algorithm runs most slowly. But even here, the number of steps is roughly proportional to the number of decimal digits.

Exercises

1. Explain why the Euclidean algorithm, applied to the pair $\{F_{n+2}, F_{n+1}\}$, yields all preceding pairs of consecutive Fibonacci numbers.
2. Deduce that $\gcd(F_{n+2}, F_{n+1}) = 1$.

Division with remainder is the preferred way to run the Euclidean algorithm in practice, because it is generally faster. But it also has advantages in theory, since it applies in situations (such as division of polynomials) where division with remainder is *not* achievable by repeated subtraction. In the case of ordinary positive integers a,b, the process of repeated division with remainder can be elegantly "frozen in time" by the so-called **continued fraction** for a/b.

Given positive integers $a > b$, the continued fraction process finds $q_1 > 0$ and $r_1 \geq 0$ ("quotient" and "remainder") such that $a = bq_1 + r_1$ with $r_1 < b_1$, and we write down the equivalent equation

$$\frac{a}{b} = q_1 + \frac{r_1}{b}.$$

If $r_1 = 0$, then the process ends there, because we have found that b divides a and hence that $\gcd(a,b) = b$.

If $r_1 > 0$, then we rewrite the above equation as

$$\frac{a}{b} = q_1 + \frac{1}{b/r_1}$$

and repeat the process on the fraction b/r_1 (which we can do since $b > r_1 > 0$). In this way we can simulate the action of the Euclidean algorithm on a pair (a,b) by the process of "continuing" a fraction a/b.

3. Explain why the continued fraction process terminates for any positive integers a, b.
4. Applying the continued fraction process to 23 and 5, show that

$$\frac{23}{5} = 4 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2}}}$$

Division with remainder also has a neat representation by 2 × 2 matrices, in which division with remainder corresponds to *extracting a matrix factor* from a column vector. In this setup, the pair $\{a,b\}$ is represented by the column vector

$$\begin{pmatrix} a \\ b \end{pmatrix}, \quad \text{where } a > b.$$

5. If $a = q_1 b + r_1$, show that $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b \\ r_1 \end{pmatrix}$.

Then, if $b > r_1 \neq 0$, one can repeat the process on the column vector $\begin{pmatrix} b \\ r_1 \end{pmatrix}$.

6. Show in particular that $\begin{pmatrix} 23 \\ 5 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

1.2 The Form of the gcd

The correctness of the Euclidean algorithm says that $\gcd(a,b)$ results from the pair $\{a,b\}$ by repeated subtraction. This implies that $\gcd(a,b)$ has a very simple symbolic form. Because subtraction is involved, the form involves **integers**; that is, natural numbers and their negatives. The system of integers is denoted by $\mathbb{Z}$, from the German word "Zahlen" for numbers.

Form of the gcd. *For any natural numbers $a, b > 0$, there are $m, n \in \mathbb{Z}$ such that*

$$\gcd(a,b) = ma + nb.$$

Proof. We show in fact that the numbers produced from a, b at *each step* of the Euclidean algorithm are of the form $ma + nb$. This is certainly true at the beginning, where $a = 1 \cdot a + 0 \cdot b$ and $b = 0 \cdot a + 1 \cdot b$.

And if the pair at some stage is $\{m_1a + n_1b, m_2a + n_2b\}$, then the pair at the next stage is $\{m_2a + n_2b, (m_1 - m_2)a + (n_1 - n_2)b\}$, which again consists of numbers of the required form.

Thus, the numbers at all stages are of the form $ma + nb$. In particular, this is true at the last stage, when each number is $\gcd(a,b)$. □

Given a pair of moderately sized numbers a, b (say, two-digit numbers), it may be hard to spot m and n such that $\gcd(a,b) = ma + nb$. However, m and n are easily computed by running the Euclidean algorithm on the letters a and b, doing exactly the same subtractions on the symbolic forms that we originally did on numbers. For example, here is what happens when we run the numerical and symbolic computations side by side in the case where $a = 34$ and $b = 21$.

$$\begin{array}{llll}
 & \{34,21\} & & \{a,b\} \\
\to & \{21, 34-21\} = \{21,13\} & \to & \{b, a-b\} \\
\to & \{13, 21-13\} = \{13,8\} & \to & \{a-b, b-(a-b)\} = \{a-b, -a+2b\} \\
\to & \{8, 13-8\} = \{8,5\} & \to & \{-a+2b, a-b-(-a+2b)\} = \{-a+2b, 2a-3b\} \\
\to & \{5, 8-5\} = \{5,3\} & \to & \{2a-3b, -a+2b-(2a-3b)\} = \{2a-3b, -3a+5b\} \\
\to & \{3, 5-3\} = \{3,2\} & \to & \{-3a+5b, 2a-3b-(-3a+5b)\} = \{-3a+5b, 5a-8b\} \\
\to & \{2, 3-2\} = \{2,1\} & \to & \{5a-8b, -3a+5b-(5a-8b)\} = \{5a-8b, -8a+13b\}.
\end{array}$$

From the last line we read off $1 = \gcd(a,b) = -8a + 13b$, and it can be checked that indeed $1 = -8 \cdot 34 + 13 \cdot 21$.

The symbolic form of the Euclidean algorithm, and hence of the gcd, was not known to Euclid. Indeed, written calculation with numbers did not develop until centuries after him, because numerical calculation could be done perfectly well with the abacus. And it was not until the sixteenth century that mathematicians realized that written calculation with symbols ("algebra") was a powerful idea – in fact more powerful than written calculation with numbers. Still, even with the primitive notation at his disposal, Euclid was able to prove the **prime divisor property**, the main result of the next section.

1.2.1 Linear Diophantine Equations

The equation $ax + by = c$, where a, b, c are integers, becomes interesting when integer solutions for x and y are sought. The equation obviously has no such solution when $\gcd(a,b)$ does not divide c, because in that case $\gcd(a,b)$ divides $ax+by$ but not c. However, this is the only obstruction.

Criterion for solvability. *If* $\gcd(a,b)$ *divides* c, *then* $ax+by=c$ *has an integer solution.*

Proof. It follows from the above that $\gcd(a,b)=ma+nb$ for some integers m and n. Then, if $c=d\cdot\gcd(a,b)$, it follows that $ax+by=c$ for $x=dm$ and $y=dn$. □

This criterion for solvability generalizes to linear equations in more than two variables. For example, $ax+by+cz=d$ has an integer solution ⇔ $\gcd(a,b,c)$ divides d. The (⇒) direction is clear, for the same reason as above. The (⇐) direction holds because

$$\gcd(a,b,c)=la+mb+nc \quad \text{for some integers } l,m,n,$$

which follows from the above because $\gcd(a,b,c)=\gcd(\gcd(a,b),c)$.

We also know that we can find the required m,n for $\gcd(a,b)$ by the extended Euclidean algorithm described above. Finally, we can find *all* solutions of $ax+by=c$ by adding to any single solution the solutions of $ax+by=0$, which are $x=kb/\gcd(a,b)$, $y=-ka/\gcd(a,b)$ for all integers k.

With these observations we can move on to Diophantine equations of higher degree. We begin in Section 1.5 with a quadratic equation in two variables. Other examples, of degree 2 and 3, are discussed in the next chapter. But first, let us see what the gcd can tell us about prime numbers.

Exercises

1. Using the symbolic Euclidean algorithm above, find integers m,n such that $13m+17n=1$.

The matrix version of division with remainder, explored in the previous set of exercises, can be very elegantly "inverted" to give the integers m and n such that $\gcd(a,b)=ma+nb$. Recall that $a=q_1b+r_1$ is represented by the matrix equation

$$\begin{pmatrix}a\\b\end{pmatrix}=\begin{pmatrix}q_1&1\\1&0\end{pmatrix}\begin{pmatrix}b\\r_1\end{pmatrix}.$$

2. Show that if repeated division with remainder on the pair a,b produces successive quotients $q_1,q_2,\ldots,q_n$ and $\gcd(a,b)=d$, then

$$\begin{pmatrix}a\\b\end{pmatrix}=\begin{pmatrix}q_1&1\\1&0\end{pmatrix}\begin{pmatrix}q_2&1\\1&0\end{pmatrix}\cdots\begin{pmatrix}q_n&1\\1&0\end{pmatrix}\begin{pmatrix}d\\0\end{pmatrix}.$$

3. Deduce that

$$\begin{pmatrix} d \\ 0 \end{pmatrix} = \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix}^{-1} \cdots \begin{pmatrix} q_2 & 1 \\ 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \end{pmatrix},$$

and show that

$$\begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & -q \end{pmatrix}.$$

4. Deduce from exercise 6 of Section 1.1 that

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -4 \end{pmatrix} \begin{pmatrix} 23 \\ 5 \end{pmatrix},$$

and hence express the gcd of 23 and 5 in the form $23m + 5n$.

Another way to prove $\gcd(a,b) = ma + nb$ is by considering the smallest positive value c of $ma + nb$ for $m, n \in \mathbb{Z}$. This idea will be used in Section 5.2 to prove that $\mathbb{Z}$ is a **principal ideal domain**.

5. Show that all values of $ma + nb$ are multiples of c (this part uses the division property of $\mathbb{Z}$).
6. Deduce that c divides a and b, and that any divisor of a and b divides c.
7. Conclude that $c = \gcd(a,b)$.

1.3 The Prime Divisor Property

The relevance of the Euclidean algorithm to the theory of primes becomes clear when we consider $\gcd(a, p)$, where p is prime. If p does not divide a, then we must have $\gcd(a, p) = 1$, because the only divisors of p are 1 and p itself. This leads to a crucial result.

Prime divisor property. *If a and b are natural numbers and p is a prime that divides ab, then p divides a or p divides b.*

Proof. Suppose that p does not divide a, so we must prove that p divides b. First, as we have just remarked, $\gcd(a, p) = 1$. Also, as we saw in the previous section, $\gcd(a, p) = ma + np$ for some integers m and n, so

$$1 = ma + np \quad \text{for some integers } m \text{ and } n.$$

Multiplying both sides of this equation by b, we get

$$b = mab + npb \quad \text{for some integers } m \text{ and } n.$$

Since p divides ab by hypothesis and p divides pb, obviously, b is a sum of terms divisible by p. Hence, b itself is divisible by p. □

In proving this prime divisor property, Euclid came as close as he probably could (given his poor notational resources) to proving what we now call the **fundamental theorem of arithmetic**, or **unique prime factorization**. Unique prime factorization easily follows from the prime divisor property if one has notation for arbitrary products of primes.

Unique prime factorization. *If $p_1, p_2, \ldots, p_k$ and $q_1, q_2, \ldots, q_l$ are prime numbers such that*

$$p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l,$$

then the same factors occur on each side, perhaps in a different order.

Proof. Since p_1 divides the left side of the equation, it also divides the right side, hence, it divides one of the factors q_i by the prime divisor property. It follows that $p_1 = q_i$, and we may cancel p_1 and q_i from the equation. Repeating the argument with the factors that remain, we eventually find that each p_j equals some q_k, and vice versa, so the factors on each side are exactly the same, though perhaps in a different order. □

We sometimes express this theorem by saying that factorization of a natural number greater than 1 into primes is unique "up to the order of factors." Later, we will see many other statements of unique prime factorization, and the "uniqueness" will be "up to order" and sometimes other trivial variations. For example, prime factorization of *integers* is unique not only "up to order" but also "up to sign" because, for example, $6 = 2 \cdot 3 = (-2) \cdot (-3)$.

The next section gives some applications of unique prime factorization. Due to its usefulness and simplicity, unique prime factorization has been sought in many other domains where "factorization" makes sense. In fact, a major theme of this book is the search for appropriate concepts of "prime" in domains where the obvious kind of factorization fails to be unique.

Exercises

In school you may have used prime factorization to find the gcd ("greatest common divisor") and the lcm ("least common multiple") of given positive integers. We can justify this idea with the help of unique prime factorization.

1. Find gcd of 60 and 84 by finding the common primes in their prime factorizations.

2. Also find $\operatorname{lcm}(60,84)$.
3. Given that $p_1,\ldots,p_k$ are the primes in the factorizations of a and b, so

$$\begin{aligned} a &= p_1^{m_1}\cdots p_k^{m_k}, \quad \text{and}\\ b &= p_1^{n_1}\cdots p_k^{n_k}, \quad \text{for some integers } m_1,n_1,\ldots,m_k,n_k \geq 0,\end{aligned}$$

explain why

$$\begin{aligned} \gcd(a,b) &= p_1^{\min(m_1,n_1)}\cdots p_k^{\min(m_k,n_k)}\\ \operatorname{lcm}(a,b) &= p_1^{\max(m_1,n_1)}\cdots p_k^{\max(m_k,n_k)}.\end{aligned}$$

4. Use these formulas for gcd and lcm to prove $\gcd(a,b)\operatorname{lcm}(a,b) = ab$.

Our proof of unique prime factorization in this section comes from the division property of $\mathbb{Z}$, via the prime divisor property. In the exercises to the last section we showed that the division property also implies the **principal ideal property** of $\mathbb{Z}$, according to which the numbers of the form $ma + nb$ are all multiples of a certain nonzero member c. We can also prove the prime divisor property from the principal ideal property, as the following exercises show.

5. Suppose that p divides ab, but p does not divide a. Given that the numbers of the form $mp + na$ are all multiples of some positive $c \neq 0$, show that $c = 1$.
6. Now deduce the prime divisor property.

1.4 Irrational Numbers

The numbers considered so far are the natural numbers and their close relatives the integers. A still larger class whose properties derive from those of the integers is the set $\mathbb{Q}$ of **rational** numbers: the ratios, or quotients, m/n of integers m,n with $n \neq 0$. It was once thought that *all* numbers are rational, but that hope was dashed (and serious mathematics began) when one of the followers of Pythagoras discovered that $\sqrt{2}$ is not. This discovery shocked the Pythagoreans, who sought a "rational" (number) explanation of everything, but who also knew that $\sqrt{2}$ was a fundamental quantity in geometry – the diagonal of the unit square. We give a proof of the irrationality of $\sqrt{2}$ by a method that extends to many other numbers.

Irrationality of $\sqrt{2}$. *For any natural numbers m and n, $m/n \neq \sqrt{2}$.*