1
Numbers

"$1 + 1 = 2$."
A. N. Whitehead and B. Russell,
Principia Mathematica, Vol. II, p. 83.

Historically, mathematics came into being to serve two purposes, counting and measuring. Both of these required the use of *numbers*, the positive integers $\mathbb{N}$ and the real numbers $\mathbb{R}$, respectively. The need to solve equations such as

$$2x = 3, \qquad x + 4 = 0, \qquad x^2 + 1 = 0$$

subsequently led to the appearance of more sophisticated number systems like the rational numbers $\mathbb{Q}$, the integers $\mathbb{Z}$, and the complex numbers $\mathbb{C}$. We are chiefly concerned here with properties of the positive integers and, at the same time, the means by which such properties are established. This revolves around the concept of a mathematical *proof*, of which we give examples of four kinds, finishing up with the most important for $\mathbb{N}$, proof by induction.

1.1 Arithmetic Progressions

Definition 1.1

An **arithmetic progression** is a sequence of numbers

$$a_0, a_1, a_2, \ldots, a_k, \ldots$$

whose consecutive terms differ by a constant called the **common difference**

$$d = a_1 - a_0 = a_2 - a_1 = \cdots = a_k - a_{k-1} = \cdots. \tag{1.1}$$

Thus, if the **first term** $a_0 = a$ say, then

$$a_1 = a + d, \qquad a_2 = a + 2d; \qquad a_k = a + kd \tag{1.2}$$

in general. We seek a formula for the sum of the first n terms of this sequence,

$$s_n = a_0 + a_1 + \cdots + a_{n-1}, \tag{1.3}$$

which is usually arrived at in the following way. Compare (1.3) with the same sum "written backwards",

$$s_n = a_{n-1} + a_{n-2} + \cdots + a_0 \tag{1.4}$$

and observe that the n pairs of terms aligned vertically all have the same sum,

$$a_{k-1} + a_{n-k} = a + (k-1)d + a + (n-k)d = 2a + (n-1)d, \tag{1.5}$$

$k = 1, \ldots, n$, so that n times this number is equal to twice the desired sum, that is,

$$2s_n = n(2a + (n-1)d). \tag{1.6}$$

Theorem 1.1

The sum s_n of the first n terms of the arithmetic progression with first term a and common difference d is given by the formula

$$s_n = na + \tfrac{1}{2}n(n-1)d. \tag{1.7}$$

□

While this process of derivation has a number of points in its favour, namely, it

(a) contains the key idea—reversing the sum,

(b) is fairly convincing—you believe the result,

(c) leads to the right answer—formula (1.7) is correct,

there are nevertheless some shortcomings, such as:

(i) the dots $\cdots$ in (1.1), (1.3), (1.4) are not defined precisely,

(ii) the formula $a_k = a + kd$ in (1.2) has not been proved, although it is almost "obvious",

(iii) tacit assumptions are made in the manipulations leading successively to (1.5), (1.6), (1.7).

We attempt to remedy each of these failings in turn.

First of all, the dots $\cdots$ in (1.1) will be eliminated if we replace all these equations by the single assertion

$$\text{for every positive integer } k,\ a_k - a_{k-1} = d.$$

Next, the dots $\cdots$ in (1.3) and (1.4) can be eliminated by passing to so-called **Σ-notation**, pronounced "sigma-notation", when they become

$$s_n = \sum_{k=0}^{n-1} a_k, \qquad s_n = \sum_{k=1}^{n} a_{n-k}. \tag{1.8}$$

In each of these equations, the right-hand side is evaluated by substituting the indicated values of k in the given expression and adding the resulting numbers. The fact that these two apparently different expressions for s_n are formally equal is a consequence of various rules of manipulation in Σ-notation like those in the exercises below. These rules are in turn the consequences of fundamental laws of arithmetic referred to after the next paragraph.

Turning to (ii), the general formula in (1.2) can be proved either by induction or by using the Σ-rules just mentioned. Since the latter are also proved by induction, the remedy for this failing must be deferred until Section 1.4.

Finally, the tacit assumptions referred to in (iii) are all of the type introduced as axioms (see below) in elementary algebra and more properly called **fundamental laws of arithmetic**. Perhaps the most prominent of these are the **commutative laws**:

$$a + b = b + a \quad \text{and} \quad ab = ba, \tag{1.9}$$

of addition and multiplication, respectively. The term "law" means that these equations hold *for all* a, b, c in $\mathbb{N}$. They generalize to sums and products of any number n of terms, and thus prove that the right-hand sides of (1.3) and (1.4) are equal.

Next, rather more subtly, notice that the right-hand side of (1.3) contains not only n terms but also $n-1$ *operations* (of addition), and, just as the former can be written in any order, the latter can be performed in any order. This is a consequence of the **associative laws**:

$$(a + b) + c = a + (b + c) \quad \text{and} \quad (ab)c = a(bc), \tag{1.10}$$

for all a, b, c in $\mathbb{N}$, of addition and multiplication, respectively.

Finally, in (1.5) and (1.7), we have made use of the **distributive laws**:

$$a(b + c) = ab + ac, \qquad (a + b)c = ac + bc, \tag{1.11}$$

valid for all a, b, c in $\mathbb{N}$. These two laws relate addition to multiplication and, like the commutative and associative laws, are commonly assumed and used without explicit reference. Note that when multiplication is commutative, each of the two distributive laws is a consequence of the other.

We have now assembled most of the axioms for a particularly important type of number system. To get the others, observe that the integers $\mathbb{Z}$ have an **additive identity** 0 and a **multiplicative identity** 1:

$$a + 0 = a, \qquad a1 = a, \tag{1.12}$$

for all integers a, and that every integer a has an **additive inverse** $-a$:

$$a + (-a) = 0. \tag{1.13}$$

A system closed under addition and multiplication for which (1.9)–(1.13) hold is called a **commutative ring-with-identity**. Removing the second equations in (1.9) and (1.12), we simply get a **ring**. Adjoining to (1.9)–(1.13) the extra axiom that there are no **divisors of zero**:

$$\text{if} \quad a \neq 0 \quad \text{and} \quad b \neq 0, \quad \text{then} \quad ab \neq 0, \tag{1.14}$$

gives an **integral domain**. The existence of **multiplicative inverses**:

$$\text{if} \quad a \neq 0, \quad \text{there is an } a^{-1} \text{ with } aa^{-1} = 1, \tag{1.15}$$

gives a **field**.

Thus, the integers $\mathbb{Z}$ form an integral domain but not a field. The rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ are all fields.

EXERCISES

Convince yourself that the following ten Σ-rules are all valid.

1.1 $\sum_{k=1}^{n} a = na.$

1.2 $a\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} aa_k.$

1.3 $\sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k = \sum_{k=1}^{n}(a_k + b_k).$

1.4 $\sum_{k=1}^{n} a_k = \sum_{i=1}^{n} a_i.$

1.5 $\sum_{k=1}^{n} a_k = \sum_{k=0}^{n-1} a_{k+1}.$

1.6 $\sum_{k=0}^{n-1} a_k = \sum_{k=1}^{n} a_{n-k}.$

1.7 $\sum_{k=1}^{n} a_k = \sum_{k=1}^{n-1} a_k + a_n.$

1.8 $\left(\sum_{i=1}^{m} a_i\right)\left(\sum_{j=1}^{n} b_j\right) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_i b_j\right).$

1.9 $\sum_{i=0}^{n-1}\left(\sum_{j=i+1}^{n} a_i b_j\right) = \sum_{j=1}^{n}\left(\sum_{i=0}^{j-1} a_i b_j\right)$.

1.10 $\left(\sum_{k=0}^{n} a_k\right)^2 = \sum_{k=0}^{n} a_k^2 + 2\sum_{i=0}^{n-1}\left(\sum_{j=i+1}^{n} a_i a_j\right)$.

1.11 Use these results to put together a more formal proof of formula (1.7).

1.12 The number of ways of ordering the n terms in the sum

$$s_n = \sum_{k=1}^{n} a_k$$

is, almost by definition, equal to $n!$, pronounced "n factorial". On the other hand, curiously, the number c_n of ways of ordering the $n-1$ operations in s_n, that is, the number of different bracketings, is not equal to $(n-1)!$ for $n \geq 4$. Calculate the value of c_n for some small values of n, and see what you can deduce about these numbers (called the **Catalan numbers**).

1.13 Prove that every field is an integral domain.

1.14 Prove that every finite integral domain is a field.

1.2 Proof by Contradiction

Suppose you wish to prove that a certain (mathematical) statement P is true. One way of doing this is to prove that P is not false, and this may be done as follows. Let $\sim P$ stand for the statement "P is false"; $\sim P$ is called the **negation** of P and pronounced "not P". Suppose you can deduce from $\sim P$ that a certain statement Q is both true and false; such a situation flies in the face of reason and a fundamental law of logic, and is called a **contradiction**. You can then conclude that your assumption $\sim P$ is wrong, that is, $\sim P$ is false, which is the same as saying that P is true. This method of proof is called "reductio an absurdum" in Euclidean geometry. It is more commonly known as **proof by contradiction**, and we shall give a more formal justification for its validity in Chapter 2. The example of this type of proof given below concerns numbers and requires a familiar definition.

Definition 1.2

A real number r is called **rational** if some non-zero integral multiple of r is an integer: $br = a$ for some integers a and b with $b \neq 0$, and then we write

$r = a/b$, pronounced "a over b". The expression a/b is called a **fraction** with **numerator** a and **denominator** b.

Examples of fractions are

$$1/2, \qquad 0/5, \qquad -2/3, \qquad 6/1, \qquad 3/6, \qquad 2/-3. \tag{1.16}$$

These fractions all represent rational numbers, and the main reason why the treatment of fractions in elementary school is so difficult is that this representation is *not unique*, that is, different fractions can represent the same rational number. This is inherent in the definition, for the equation $br = a$ can be multiplied by any non-zero integer m to give the equivalent equation $mbr = ma$, so that the fractions ma/mb all represent the same rational number r. The way round this is to choose the "simplest" one, namely the one where

$$\text{the denominator is the } \textit{smallest} \text{ positive integer multiple of } r \text{ to be an integer.} \tag{1.17}$$

In this case a and b can have no common factor (other than ± 1) and a/b is said to be **in lowest terms**.

The Ancient Greeks believed at first that all real numbers are rational. Imagine their horror when, as a consequence of Pythagoras' celebrated theorem, an irrational number appeared on the scene: the hypotenuse h of a right triangle with shorter sides both of length 1 satisfies the equation $h^2 = 2$; write $h = \sqrt{2}$, pronounced "root-two".

Theorem 1.2 (Euclid)

The number $h = \sqrt{2}$ is irrational.

Proof

Assume (for a contradiction) that the statement of Theorem 1.2 is false, that is, that h is rational. Then we can write $h = a/b$ in lowest terms, where a and b are integers. Then

$$h^2 = a^2/b^2 = 2,$$

whence $a^2 = 2b^2$. So a^2 is even and thus so is $a : a = 2c$, say. But then

$$4c^2 = a^2 = 2b^2,$$

whence $b^2 = 2c^2$. So b^2 is even and thus so is $b : b = 2d$, say. But then

$$h = a/b = 2c/2d = c/d,$$

which contradicts the fact that a/b is in lowest terms. We conclude that our original assumption, that h is rational, is false. Therefore, $h = \sqrt{2}$ is irrational, as required. □

In this proof, we used the statements

$$\begin{aligned} P &: \sqrt{2} \text{ is irrational,} \\ \sim P &: \sqrt{2} \text{ is rational,} \\ Q &: a/b \text{ is in lowest terms,} \\ \sim Q &: a/b \text{ is not in lowest terms.} \end{aligned}$$

We already know that Q is true, by (1.17). We then deduced $\sim Q$ from $\sim P$. The contradiction Q and $\sim Q$ enables us to conclude P.

EXERCISES

1.15 Write the fractions given in (1.16) in lowest terms.

1.16 Convince yourself that the representation of a rational number satisfying (1.17) is unique.

1.17 Let n be an integer. Prove that the remainder on division of n^2 by 4 is 0 or 1 according as n is even or odd. What remainders are possible when n^2 is divided by 8?

1.18 Let c be a real number satisfying $c^3 = 5$; write $c = \sqrt[3]{5}$, the "cube-root of five". Prove that c is irrational.

1.19 Convince yourself that statements P and $\sim\sim P$ are the same. Suppose that P is a statement from which you can deduce $\sim P$. Which, if any, of the following conclusions can you draw:

(a) P is true, (b) P is false, (c) $\sim P$ is true, (d) $\sim P$ is false.

1.20 Give a proof by contradiction of the statement:

P: the sum of the squares of three consecutive integers cannot leave remainder -1 on division by 12.

1.21 Prove by contradiction that the cube of the largest of three consecutive integers cannot be equal to the sum of the cubes of the other two.

1.22 Prove that the polynomial $f(x) = x^4 + 2x^2 + 2x + 1998$ cannot be written as the product of two quadratic polynomials with integer coefficients.

1.23 Prove that if m and n are odd integers, then the equation $x^2+2mx+2n=0$ has no rational root by deducing that

(a) such a root must be an integer, and

(b) that integer can be neither odd nor even.

1.24 Prove that at a party of at least two people, there are at least two who have the same number of friends at the party.

1.3 Proof by Contraposition

This method may be applied when you wish to prove a compound statement of the form "if P, then Q", where P and Q are themselves statements about the same thing or things. Such a statement is called an **implication**, with premise P and conclusion Q, and it can be put in many different ways:

- Q holds whenever P holds,
- given P, Q is true,
- Q is a consequence of P,
- P is a **sufficient condition** for Q,
- Q is a **necessary condition** for P,
- P **implies** Q, written $P \Rightarrow Q$.

We prefer the last of these and will study it in more detail in the next chapter.

For the moment, note first that the roles of P and Q in such a statement are completely different. To prove that $P \Rightarrow Q$, we would normally start off by assuming that P is true, go through a process of logical deduction, and finish up with the conclusion that Q is true. Now it may be that, having assumed P, the best way to establish Q is by getting a particular contradiction, namely, by assuming $\sim Q$ and deducing $\sim P$. The contradiction "P and $\sim P$" then allows us to conclude that our assumption $\sim Q$ was false, that is, Q is true.

So this proof boils down to showing that $\sim Q \Rightarrow \sim P$; this statement is called the **contrapositive** of $P \Rightarrow Q$. In the next chapter, we shall give a formal proof that the statements $P \Rightarrow Q$ and $\sim Q \Rightarrow \sim P$ are logically equivalent. On the other hand, the implication $Q \Rightarrow P$ is called the **converse** of $P \Rightarrow Q$ and is NOT logically equivalent to it. We sometimes write $P \Leftarrow Q$ instead of $Q \Rightarrow P$, and $P \Leftrightarrow Q$ pronounced "P if and only if Q" instead of "$P \Rightarrow Q$ and $P \Leftarrow Q$".

This prompts a word of warning. When proving an implication $P \Rightarrow Q$, you can NEVER assume that Q is true. It is, after all, what you are trying to

prove: if you want to climb a mountain, you don't start at the top. It would save time of course but wouldn't really achieve anything much.

To set up a good example of proof by contraposition, we shall need some useful notation and a familiar definition. Let a and b be integers and consider the condition "b is a multiple of a": $b = ma$ for some integer m. In elementary number theory, it is both convenient and customary to change the emphasis and state this in the form "a divides b", written $a \mid b$. Thus, 1 has only one positive divisor (itself), and every $n > 1$ has at least two (n and 1). Note that $a \mid b$ is *not* a number (like the fraction a/b), but a *statement* about a pair of numbers. Thus, $2 \mid 4$, $2 \mid 2$, $2 \mid 0$ are true, and $2 \mid 1$, $2 \mid 5$, $0 \mid 2$ are false.

Definition 1.3

A positive integer is called **prime** if it has exactly two divisors.

These are, of course, n and 1 (note that 1 is not prime). It is clear that if n is prime, then n is not divisible by any integer a in the range $2 \leq a \leq \sqrt{n}$. We shall prove by contraposition that this necessary condition for primeness is also sufficient.

Theorem 1.3

If $n \geq 2$ is a positive integer divisible by no integer d in the range $2 \leq d \leq \sqrt{n}$, then n is prime.

Proof

We have two statements about positive integers $n \geq 2$:

P n is divisible by no integer d in the range $2 \leq d \leq \sqrt{n}$,
Q n is prime.

Proceeding by contraposition, we assume $\sim Q$, that is, n has a divisor a with $1 \neq a \neq n$. Write $n = ab$ with $n \neq b \neq 1$ and consider two cases:

(i) $a \leq b$: then $a^2 \leq ab = n$ and $a \leq \sqrt{n}$,
(ii) $a > b$: then $b^2 < ab = n$ and $b \leq \sqrt{n}$.

Taking $d = a$ in case (i) and $d = b$ in case (ii), we have a divisor d of n in the range $2 \leq d \leq \sqrt{n}$. This is the negation of P, and we have proved $\sim Q \Rightarrow \sim P$. Hence, $P \Rightarrow Q$ by contraposition. □

EXERCISES

Consider the following statements about pairs of positive integers a, b:

$$\begin{aligned} A &: a = b, \\ B &: a < b, \\ C &: a \mid b, \\ D &: a \text{ and } b \text{ have no common prime divisor.} \end{aligned}$$

1.25 Write down four more statements A', B', C', D' by interchanging a and b.

1.26 Write down the negations of A, B, C, D.

1.27 Of the 16 implications $P' \Rightarrow \sim Q$, where P and Q range independently over A, B, C, D, exactly five are true. Which ones?

1.28 In each of the remaining 11 cases, write down a specific pair a, b for which P' is true but $\sim Q$ is false.

1.29 Write down the converse of Theorem 1.3. Is it true?

1.30 Jack said: "If there is a hole in this bucket, we won't get down unhurt." A few minutes later, Jill replied: "There must have been a hole in the bucket." Was this deduction correct or not? Explain.

Prove the following three statements by contraposition.

1.31 If n is an integer with n^2 even, then n is even.

1.32 If a and b are real numbers with $a \neq b$, then $\sqrt{ab} \neq (a+b)/2$.

1.33 For positive integers a and b, if $a \neq b$, then $ax^2 + bx + (b-a) = 0$ has no positive integer root.

1.4 Proof by Induction

The methods of proof described in the previous two sections are in common use throughout mathematics. In contrast, the method presented in this section is special: it applies only to statements about positive integers, more precisely, statements of the form "$P(n)$ is true for all positive integers n", where $P(n)$ is a statement involving n. The fact that the method "works" is a consequence of a fundamental property of $\mathbb{N}$ that is best treated as an axiom, although we shall give a "proof" later.

Principle of Mathematical Induction (PMI)

Let $P(n)$ be a statement about positive integers n. Suppose that

(a) $P(1)$ is true, and

(b) $P(n-1) \Rightarrow P(n)$ for all $n > 1$.

Then $P(n)$ is true for all n.

Here, n is called the **inductive variable**,

$P(1)$	the **inductive base**,
$P(n-1)$	the **inductive hypothesis** IH, and
$P(n-1) \Rightarrow P(n)$	the **inductive step**.

In many books, the inductive step (b) is written

$$P(n) \Rightarrow P(n+1) \quad \text{for all } n \geq 1, \tag{1.18}$$

which is the same as (b): merely replace n in (b) by $n+1$ throughout (cf. changing the variable of summation in Exercise 1.5). You are free to use either form, whichever best suits the context.

Our first example comprises something we already know, the important special case $a = d = 1$ of Theorem 1.1. It is interesting to compare the proofs in Section 1.1, Exercise 1.11, and the following.

Theorem 1.4

For every positive integer n,

$$P(n): \qquad \sum_{k=1}^{n} k = \frac{1}{2}n(n+1).$$

Proof

We need to (a) establish the base, (b) carry out the inductive step.

(a) $P(1)$ is true by inspection: both sides are equal to 1 when $n = 1$.

(b) Now suppose that $n > 1$ and assume $P(n-1)$:

$$\text{IH} \qquad \sum_{k=1}^{n-1} k = \frac{1}{2}(n-1)n.$$

Then

$$\begin{aligned}
\sum_{k=1}^{n} k &= \sum_{k=1}^{n-1} k + n, && \text{by Exercise 1.7,} \\
&= \frac{1}{2}(n-1)n + n, && \text{by IH,} \\
&= \frac{1}{2}(n-1+2)n, && \text{by the distributive laws,} \\
&= \frac{1}{2}n(n+1).
\end{aligned}$$

We have thus deduced $P(n)$ from $P(n-1)$. By the PMI, $P(n)$ is true for all positive integers n. □

In the next example, it is not easy to see how to proceed without the PMI.

Theorem 1.5

For every positive integer n,

$$P(n): \qquad \sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1).$$

Proof

(a) is true by inspection: both sides are equal to 1 when $n = 1$.
For (b), suppose that $n > 1$ and assume $P(n-1)$:

$$\text{IH} \qquad \sum_{k=1}^{n-1} k^2 = \frac{1}{6}(n-1)n(2n-1).$$

Then

$$\begin{aligned}
\sum_{k=1}^{n} k^2 &= \sum_{k=1}^{n-1} k^2 + n^2 \\
&= \frac{1}{6}(n-1)n(2n-1) + n^2 \\
&= \frac{1}{6}n(2n^2 - 3n + 1 + 6n) \\
&= \frac{1}{6}n(2n^2 + 3n + 1) \\
&= \frac{1}{6}n(n+1)(2n+1).
\end{aligned}$$

By the PMI, this completes the proof. □

The inductive method admits a number of useful variants, all based on the PMI. Here are four of them.

Variant 1: Changing the Base

It may be that we wish to prove a statement $P(n)$ for all values of $n \geq n_0$, where n_0 is an integer other than 1. In this case, it is sufficient to prove

$$\text{(a) } P(n_0), \qquad \text{and} \qquad \text{(b) } P(n-1) \Rightarrow P(n) \text{ for all } n > n_0.$$

This is no more than the PMI applied to the new statement $Q(n) = P(n + n_0 - 1)$, since

$$\text{(a) } Q(1) = P(n_0), \qquad \text{(b) } n > n_0 \Leftrightarrow n - n_0 + 1 > 1.$$

Theorem 1.6

For all $n \geq 5$, $2^n > n^2$.

Proof

Let $P(n)$ be the statement $2^n > n^2$. $P(5) : 32 > 25$.

Now let $n > 5$ and assume $P(n-1)$.

IH
$$2^{n-1} > (n-1)^2.$$

Then

$$\begin{aligned} 2^n &= 2 \,.\, 2^{n-1} \\ &> 2(n-1)^2 \\ &= 2n^2 - 4n + 2 \\ &= n^2 + n(n-4) + 2 \\ &> n^2 \text{ as } n > 5. \end{aligned}$$

□

Variant 2: Cumulative Induction

In carrying out the inductive step, it is sometimes convenient (even necessary) to assume *all* the statements $P(k)$ for $1 \leq k \leq n-1$ in order to conclude $P(n)$. This is allowed. Given that

(a) $P(1)$ is true, and

(b) $P(1)$ and $P(2)$ and ... and $P(n-1)$ together $\Rightarrow P(n)$ for all $n > 1$,

you can conclude that $P(n)$ is true for all $n \geq 1$. To see this, merely apply the PMI to the new statement $Q(n) = P(1)$ and $P(2)$ and ... and $P(n)$.

Theorem 1.7

Every integer $n \geq 2$ is a product of primes.

Proof

The inductive base $P(2)$ is clear: 2 is a prime. Now let $n > 2$ and assume $P(k)$ for all k in the range $2 \leq k \leq n-1$.

We need to distinguish two (mutually exclusive and exhaustive) cases: n is prime, n is not prime. In the first case there is nothing to prove, so assume that n is not prime. Then $n = lm$ with $1 < l, m < n$. By $P(l)$, l is a product of primes, and by $P(m)$, so is m. Thus, so also is their product $lm = n$. □

P.S. You may be worried about the use of the expression "product of primes" being allowed to include the case of a single prime. This is a convention: a product $t_1 t_2 \cdots t_n$ of n terms is just t_1 when $n = 1$. It is also convenient to attach a meaning when $n = 0$: the **empty product** is taken to be 1. Using the notation

$$t_1 t_2 \cdots t_n = \prod_{k=1}^{n} t_k = \pi_n,$$

we take $\pi_0 = 1$, just as in the case of sums,

$$t_1 + t_2 + \cdots + t_n = \sum_{k=1}^{n} t_k = \sigma_n,$$

we take $\sigma_0 = 0$.

Variant 3: Double Induction

Suppose you want to prove a statement of the form $P(m,n)$ for all positive integers m and n. Induction on n would involve:

(a) establishing the inductive base: $P(m,1)$ for all values of m, and

(b) carrying out the inductive step: for all $n > 1$,

$$P(m, n-1) \Longrightarrow P(m,n) \text{ for all values of } m.$$

Now either or both of these statements are candidates for proof by induction on m. Then again, the roles of m and n in the aforesaid can be reversed. There are thus a number of subvariants of proof by double induction, depending on the *order* in which the pairs (m, n) are taken (this idea will be developed further in Section 4.4). We consider three of these now, spelling out in each case the strategy for proving $P(m, n)$ for all positive integers m, n.

Strategy 1

(a) Prove $P(1, 1)$,

(b) assume $P(m, 1)$

(c) deduce $P(m + 1, 1)$,

(d) assume $P(m, n)$ for all values of m,

(e) deduce $P(m, n + 1)$ for all values of m.

A little thought shows that this is really two separate single inductions: (a) plus (b) and (c) establish (by induction on m) the base for an induction on n (with (b) and (c) forming the inductive step). For an example of this subvariant, see Exercise 1.43.

Strategy 2

(a) Prove $P(1, n)$, for all values of n,

(b) prove $P(m, 1)$ for all values of m,

(c) for $m, n > 1$, assume $P(m - 1, n)$ and $P(m, n - 1)$,

(d) deduce $P(m, n)$.

Here, (a) and/or (b) can be proved by (single) induction, and (c) and (d) may be thought of as a single induction on $m + n$. This method is illustrated in Exercises 1.47–1.49 below.

Strategy 3

(a) Prove $P(m, 1)$, for all values of m,

(b) for $n > 1$, assume $P(m, n - 1)$,

(c) prove $P(1,n)$,

(d) for $m > 1$, assume $P(m-1,n)$,

(e) prove $P(m,n)$.

This method differs only slightly from Strategy 2, but highlights the fact that the inductive step (in the induction on n) is being carried out by induction (on m). An example of this kind, but with m starting at 0, follows at once.

Theorem 1.8

For every positive integer n, the product of any n consecutive integers is divisible by $n!$.

Proof

Let us write

$$\pi(m,n) = \prod_{k=1}^{n}(m+k),\ n \geq 1,\ m \geq 0$$

for the product of the n consecutive integers starting from $m+1$.

Notice that if $n \geq 2$, the first $n-1$ terms comprise $\pi(m,n-1)$, so that we have

$$\pi(m,n) = (m+n)\,\pi(m,n-1). \tag{1.19}$$

If also $m \geq 1$, the last $n-1$ terms of $\pi(m-1,n)$, also comprise $\pi(m,n-1)$, so that

$$\pi(m-1,n) = m\pi(m,n-1). \tag{1.20}$$

Subtracting (1.20) from (1.19), we get the formula

$$\pi(m,n) - \pi(m-1,n) = n\,\pi(m,n-1), \quad m \geq 1,\ n \geq 2, \tag{1.21}$$

which will come in handy later. We are now ready to embark on the proof.

We want to prove the statement

$$P(m,n): \qquad n! \mid \pi(m,n) \quad \text{for all } n \geq 1,\ m \geq 0. \tag{1.22}$$

We induct first on $n \geq 1$. The inductive base,

$$P(m,1): 1! \mid \prod_{k=1}^{1}(m+k) \quad \text{for all } m \geq 0,$$

merely asserts that 1 is a divisor of $m+1$.

Now let $n > 1$ and assume the inductive hypothesis

$$(n-1)! \mid \pi(m, n-1) \quad \text{for all } m \geq 0. \tag{1.23}$$

We have to prove that

$$n! \mid \pi(m, n) \text{ for all } m \geq 0, \tag{1.24}$$

and to do this, we now induct on $m \geq 0$, keeping $n \geq 1$ fixed. The inductive base at $m = 0$ is again trivial: $n! \mid n!$, as $\pi(0, n) = n!$

Now let $m > 0$ and assume the (second) inductive hypothesis

$$n! \mid \pi(m-1, n). \tag{1.25}$$

Our goal is to prove

$$n! \mid \pi(m, n) \tag{1.26}$$

assuming two inductive hypotheses (1.23) and (1.25). The first of these implies that $n!$ divides the right-hand side of (1.21), and so also the left-hand side:

$$n! \mid (\pi(m, n) - \pi(m-1, n)).$$

Since $n!$ divides the second term by (1.25), it also divides the first, and we have proved (1.26). This completes the induction on m, and we have proved (1.24). This in turn completes the induction on n, as we have proved (1.22), as required. □

Variant 4: Simultaneous Induction

This final variant applies when $P(n)$ is a compound statement of the form $Q(n)$ and $R(n)$, that is, when we are trying to prove two statements simultaneously. The inductive base simply requires two proofs, of $P(1)$ and of $Q(1)$. For the inductive step,

$$P(n-1) \text{ and } Q(n-1) \Rightarrow P(n) \text{ and } Q(n) \quad \text{for all } n > 1,$$

a number of approaches are possible, depending on the relation between $P(n)$ and $Q(n)$. For example, it is enough to prove

$$P(n-1) \Rightarrow Q(n) \text{ and } Q(n-1) \Rightarrow P(n), \quad \text{both for all } n > 1,$$

or, as in our star example in the next section,

$$\begin{aligned} P(n-1) \text{ and } Q(n-1) &\Rightarrow P(n), \quad \text{and} \\ Q(n-1) \text{ and } P(n) &\Rightarrow Q(n), \quad \text{both for all } n > 1. \end{aligned}$$

Extreme examples of the use of induction are provided by recent important research by the Russian group theorists S. I. Adian and A. Yu. Ol'shanskii and their colleagues on the long-standing Burnside problem. In one case, the inductive base is 667 and there are as many as 92 inductive variables. This is "improved" in the other to around eight inductive variables, but at the expense of up to six simultaneous inductions in many places and an inductive base of around 10^{10}.

EXERCISES

1.34 Prove that for every positive integer n, $\sum_{k=1}^{n} k^3 = \frac{1}{4}n^2(n+1)^2$.

Can you think of any natural reason for the fact that this is equal to $\left(\sum_{k=1}^{n} k\right)^2$?

1.35 Consider the polynomials

$$s_0(x) = x,\ s_1(x) = \tfrac{1}{2}x(x+1),$$
$$s_2(x) = \tfrac{1}{6}x(x+1)(2x+1), s_3(x) = \tfrac{1}{4}x^2(x+1)^2.$$

Observe that for each $m = 1, 2, 3$,

$$s_m(x) = m \int s_{m-1}(x)\,dx + cx,$$

where c is chosen to make $s_m(1) = 1$. Apply this formula with $m = 4$ to get $s_4(x)$, and prove by induction on n that $\sum_{k=1}^{n} k^4 = s_4(n)$.

1.36 What is wrong with the following "proof" that all horses are the same colour? Let $P(n)$ be the statement: in any group of n horses, all are the same colour. This is clearly true when $n = 1$ as any horse is the same colour as itself. Next, take any group of n horses and exclude one. The remaining $n-1$ are the same colour by the IH. Now exclude a different horse, so that the remaining $n-1$ (including the one originally excluded) are the same colour, by the IH again. So all n are the same colour.

1.37 Find a formula for the sum of the first n terms of the geometric progression with first term a and common ratio $r \neq 1$, and prove it by induction using the form (1.18) of the inductive step.

1.38 Find the smallest value n_0 of n for which the statement $2^n > n^3$ is true, then prove it for all $n \geq n_0$ by induction.

1.39 Give a proof by contradiction that among the positive integers there are infinitely many primes.

1.40 Prove that $\sum_{k=1}^{n} k \,.\, k! = (n+1)! - 1$ for all $n \geq 1$.

1.41 Prove that for every integer $n \geq 2$

$$\prod_{k=2}^{n}(1 - 1/k^2) = (n+1)/2n.$$

1.42 Put together an overall proof strategy for a double induction in which the inductive base and the inductive step of the first induction are *both* proved by induction.

1.43 Taking the associative law for granted and using the form (1.18) of the inductive step, prove the commutative law

$$m + n = n + m \text{ for all positive integers } m \text{ and } n.$$

1.44 Can you think of a shorter proof of Theorem 1.8?

1.5 Inductive Definition

Suppose you wish to define a quantity $q(n)$ that depends on the positive integer n: such an object is called, in the terms of Chapter 5, a **map with domain** $\mathbb{N}$, or a **sequence**. Suppose further that

(a) $q(1)$ is known, and

(b) $q(n)$ can be expressed in terms of $q(n-1)$ for all $n > 1$.

Then a simple induction on n proves the statement $P(n)$: $q(n)$ is defined for every positive integer n.

For example, the rules,

(a) $q(1) = 1$,

(b) $q(n) = n \,.\, q(n-1)$ for $n \geq 2$

provide the definition of $n!$.

As with inductive proofs, the base need not always be 1. For example, the formulae for $s_0(x)$ and $s_m(x)$ in Exercise 1.35 above comprise an inductive definition of the polynomial $s_m(x)$ for all integers $m \geq 0$. A simpler and more basic example of this kind is as follows.

Definition 1.4

The **powers** of a variable x are defined by

$$\text{(a)} \quad x^0 = 1, \qquad \text{(b)} \quad x^n = x^{n-1}x \quad \text{for all } n \geq 1.$$

It is an inevitable fact that statements about inductively-defined quantities are proved by induction. The so-called *rules of indices* furnish a simple illustration.

Theorem 1.9

For all integers $m, n \geq 0$,

$$\text{(i)} \quad x^m x^n = x^{m+n}, \qquad \text{(ii)} \quad (x^m)^n = x^{mn}.$$

Proof

(i) Induct on n: both sides are equal to x^m when $n = 0$. Let $n \geq 1$ and assume the IH:

$$x^m \,.\, x^{n-1} = x^{m+n-1} \text{ for all } m \geq 0.$$

Then

$$\begin{aligned} x^m \,.\, x^n &= x^m \,.\, x^{n-1} \,.\, x && \text{by definition} \\ &= x^{n+n-1} \,.\, x && \text{by the IH} \\ &= x^{m+n-1+1} && \text{by definition} \\ &= x^{m+n} && \text{for all } m \geq 0. \end{aligned}$$

(ii) Induct on n: both sides are equal to 1 when $n = 0$. Let $n \geq 1$ and assume the IH:

$$(x^m)^{n-1} = x^{m(n-1)} \text{ for all } m \geq 0.$$

Then

$$\begin{aligned} (x^m)^n &= (x^m)^{n-1} x^m && \text{by definition} \\ &= x^{m(n-1)} x^m && \text{by the IH} \\ &= x^{m(n-1)+m} && \text{by part (i)} \\ &= x^{mn} && \text{for all } m. \end{aligned}$$

□

As with inductive proofs, more than one inductive variable may be involved in an inductive definition. In this case, a number of strategies are available, perhaps the cleanest being Strategy 2 on page 15. Thus, to define a quantity $q(m, n)$ for all $m, n \geq 0$, it is sufficient to

(a) specify $q(m,0)$ for all $m \geq 0$ and $q(0,n)$ for all $n \geq 0$, and

(b) express $q(m,n)$ in terms of $q(m-1,n)$ and $q(m,n-1)$ for all $m,n \geq 1$.

The following example is very natural, important and (I hope) familiar.

Definition 1.5

Let us define

(a) $b(m,0) = 1$ for all $m \geq 0$ and $b(0,n) = 1$ for all $n \geq 0$, and

(b) $b(m,n) = b(m-1,n) + b(m,n-1)$ for all $m,n \geq 1$.

Then the $b(m,n)$ are called **binomial coefficients**; it is customary to write $b(m,n) = \binom{m+n}{m}$, pronounced "$m+n$ choose m".

A little thought will convince you that this is nothing but a formal definition of Pascal's triangle.

The following properties of the binomial coefficients are all easy exercises. For all $m,n \geq 0$,

(i) $\binom{m+n}{m} = \binom{m+n}{n}$,

(ii) $\binom{m+n}{m} = \frac{(m+n)!}{m!\,n!}$,

(iii) $\binom{m+n}{m}$ is the number of ways of choosing m things from $m+n$.

Property (i) expresses the *symmetry* of $b(m,n)$, (ii) provides a *closed formula*, and (iii) indicates an alternative approach in terms of *combinations*. Arguably the most important manifestation of these numbers is in the famous theorem from which they get their name.

Theorem 1.10 (The Binomial Theorem)

For all $n \geq 0$,

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k. \tag{1.27}$$

Proof

The proof is (necessarily) by induction on n. To get the base, let $n = 0$. Then the left-hand side is equal to 1 (Definition 1.4(a)), and so is the right-hand side:

$$\sum_{k=0}^{0} \binom{0}{k} x^k = \binom{0}{0} x^0 = 1,$$

by Definitions 1.5(a) and 1.4(a).

For the inductive step, let $n \geq 1$ and assume the IH:

$$(1+x)^{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k.$$

Then

$$\begin{aligned}
(1+x)^n &= (1+x)^{n-1}(1+x) \text{ by Definition 1.4(b)} \\
&= \left(\sum_{k=0}^{n-1} \binom{n-1}{k} x^k\right)(1+x) \text{ by the IH} \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} x^k + \sum_{k=0}^{n-1} \binom{n-1}{k} x^{k+1} \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} x^k + \sum_{k=1}^{n} \binom{n-1}{k-1} x^k \\
&= \binom{n-1}{0} x^0 + \sum_{k=1}^{n-1} \left(\binom{n-1}{k} + \binom{n-1}{k-1}\right) x^k + \binom{n-1}{n-1} x^n.
\end{aligned}$$

In the second-last step, we replaced k by $k-1$ throughout the second sum, and in the last step we isolated the first term of the first sum and the last term of the second sum, then combined what was left into a single sum.

Comparing this expression with the right-hand side of (1.27) we have to prove that

(i) $\binom{n-1}{0} x^0 = \binom{n}{0} x^0$,

(ii) $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$ for $1 \leq k \leq n-1$, and

(iii) $\binom{n-1}{n-1} x^n = \binom{n}{n} x^n$.

Well, both sides of (i) are equal to 1 by Definitions 1.5(a) and 1.4(a), and (iii) also follows from Definition 1.5(a): $\binom{m}{m} = 1$ for all $m \geq 0$. To see (ii), write the formula in Definition 1.5(b) in the form

$$\binom{m+n}{m} = \binom{m+n-1}{m-1} + \binom{m+n-1}{m}.$$

Replacing m by k and n by $n-k$ does the trick. □

Finally, and again in analogy with inductive proofs, inductive definitions can be cumulative, that is to say, having got the base, $q(1)$ say, $q(n)$ for $n > 1$ can be defined in terms of any or all of the previous values $q(k)$, $1 \leq k \leq n-1$. We round off this section by giving two classical examples.

Definition 1.6

Define a sequence of numbers by setting

$$u_0 = 0, \quad u_1 = 1, \quad u_n = u_{n-2} + u_{n-1} \text{ for } n \geq 2.$$

The u_n, $n \geq 0$, are called the **Fibonacci numbers**. Note that two equations are required for the base because of the form of the inductive part of this definition.

As their name suggests, the Fibonacci numbers were first studied by Leonardo of Pisa at the time of the Renaissance. Since that time, these numbers have been a constant source of recreational mathematics, as well as providing insight into natural phenomena such as phyllotaxis. At the present time, the Fibonacci Society produces its journal, the Fibonacci Quarterly, once every three months. The u_n have an almost limitless number of interesting properties, of which we shall prove just one.

Theorem 1.11

The Fibonacci numbers u_n have the property

$$u_{n-1}\, u_{n+1} = u_n^2 + (-1)^n \text{ for all } n \geq 1.$$

Proof

When $n = 1$,

$$\text{lhs} = u_0\, u_2 = 0(0+1) = 0 = 1^2 + (-1)^1 = \text{rhs}$$

and we have the inductive base. Let $n \geq 2$ and assume the IH

$$u_{n-2}\, u_n = u_{n-1}^2 + (-1)^{n-1}.$$

Then, making free use of the definition,

$$\begin{aligned} u_{n-1}\, u_{n+1} - u_n^2 &= u_{n-1}(u_{n-1} + u_n) - (u_{n-2} + u_{n-1})u_n \\ &= u_{n-1}^2 - u_{n-2}\, u_n \\ &= (-1)^n, \text{ by the IH.} \end{aligned}$$

□

As a second example of definition by cumulative induction, consider the following.

Definition 1.7

Define a sequence of numbers by setting

$$c_1 = 1, \quad c_n = \sum_{k=1}^{n-1} c_k \, c_{n-k} \text{ for } n \geq 2.$$

The c_n, $n \geq 1$, are called the **Catalan numbers**. As mentioned in Exercise 1.12, c_n is just the number of ways of bracketing a sum or product of n terms in a given order. The closed formula for the c_n which follows provides a proof of the otherwise non-obvious fact that $(n+1) \mid \binom{2n}{n}$ for all $n \geq 0$ (cf. Theorem 1.8 above).

Theorem 1.12

The Catalan numbers c_n are given by

$$c_{n+1} = \frac{1}{n+1}\binom{2n}{n} \text{ for all } n \geq 0. \tag{1.28}$$

Proof

The first step in the proof is to solve the following apparently harder problem: find a formula for the number d_n of different bracketings of a product of n terms, say $x_1, x_2, \ldots, x_n$, *in any order*. It turns out that

$$d_1 = 1, \quad d_{n+1} = (4n-2)d_n, \; n \geq 1. \tag{1.29}$$

The first equation is obvious, and provides the base for an induction on $n \geq 1$. The case $n = 1$ is also obvious: $d_2 = 2$. The next case is more typical, and we spell it out now.

The new term x_3 can be introduced *inside* each of the products $x_1\,x_2$, $x_2\,x_1$ in four different ways. In $x_1\,x_2$ for example, it can precede or follow x_1, or precede or follow x_2, resulting in $(x_3\,x_1)x_2$, $(x_1\,x_3)x_2$, $x_1(x_3\,x_2)$, $x_1(x_2\,x_3)$ respectively. Further, it can occur *outside* $x_1\,x_2$ in two ways: $x_3(x_1\,x_2)$, $(x_1\,x_2)x_3$, giving six new products altogether from $x_1\,x_2$. From $x_2\,x_1$ we get another 6, so that $d_3 = 12$, as required.

The general case is similar. In any particular bracketed product of $x_1, x_2, \ldots, x_n$ there are $n-1$ multiplications. Inside each of these we can introduce x_{n+1} in four ways as above to give $4(n-1)$ new products, and there

are two ways of putting it outside. Thus, for each of the d_n products of n terms, we get $4n-2$ with $n+1$ terms. This establishes (1.29).

Next, to forge a link with the c_n, observe that for each bracketing, the terms can be written in any of $n!$ orders: $d_n = n!\,c_n$, $n \geq 1$. Hence, by (1.29),

$$c_{n+1} = \frac{d_{n+1}}{(n+1)!} = \frac{(4n-2)d_n}{(n+1)!} = \frac{(4n-2)\,n!\,c_n}{(n+1)!} = \frac{(4n-2)}{n+1}\,c_n. \tag{1.30}$$

This provides the key to the inductive step in proving (1.28), to which we now finally turn.

To get the base, observe that both sides of (1.28) are equal to 1 when $n = 0$. So assume the equation in (1.28) as it stands and compute as follows:

$$\begin{aligned} c_{n+2} &= \frac{4n+2}{n+2}\,c_{n+1}, \quad \text{by (1.30)}, \\ &= \frac{4n+2}{n+2} \cdot \frac{1}{n+1}\binom{2n}{n}, \quad \text{by the IH}, \\ &= \frac{2n+1}{n+2} \cdot \frac{2}{n+1} \cdot \frac{n+1}{n+1}\binom{2n}{n} \\ &= \frac{1}{n+2}\binom{2n+2}{n+1}, \end{aligned}$$

which completes the inductive step. □

EXERCISES

1.45 Given a sequence a_n, $n \geq 1$, formulate inductive definitions of the expressions

$$\sigma_n = \sum_{k=1}^{n} a_n, \qquad \pi_n = \prod_{k=1}^{n} a_n.$$

1.46 Define $x^{-1} = 1/x$ and $x^{-n} = (x^{-1})^n$ for $n \geq 2$. Prove that $(x^m)^{-1} = (x^{-1})^m$ for all integers m. Deduce that the rules of indices in Theorem 1.9 hold for all integers m and n, positive, negative and zero.

1.47 Use double induction (Strategy 2) to prove the symmetry of the binomial coefficients:

$$b(m,n) = b(n,m) \text{ for all } m, n \geq 0,$$

directly from the definition.

1.48 Similarly prove the closed form $b(m,n) = \frac{(m+n)!}{m!\,n!}$ for all $m, n \geq 0$.

1.49 Similarly, prove that $\binom{m+n}{m}$ is the number of ways of choosing m things from $m+n$, for all $m, n \geq 0$.

1.50 Prove that $\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2$ for all $n \geq 0$.

1.51 Deduce from Theorem 1.10 the more general form of the binomial theorem

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

for all $n \geq 0$ and any numbers a and b.

1.52 Prove that, for all $n \geq 0$, the consecutive Fibonacci numbers u_n, u_{n+1} have no common (positive integer) divisor greater than 1. Is this also true for u_n and u_{n+2}?

1.53 Consider the statements

$$P(n) : u_{2n} = u_{n-1}u_n + u_n u_{n+1}, \qquad Q(n) : u_{2n+1} = u_n^2 + u_{n+1}^2$$

about the Fibonacci numbers. Show that

$$P(n) \text{ and } Q(n) \Rightarrow P(n+1)\ ,\ Q(n) \text{ and } P(n+1) \Rightarrow Q(n+1).$$

Use a simultaneous induction to deduce that the statement $P(n)$ and $Q(n)$ is true for all $n \geq 1$.

1.54 Prove that the Fibonacci numbers are given by the formula

$$u_n = (\theta^n - \phi^n)/\sqrt{5} \text{ for all } n \geq 0,$$

where $\theta = (1+\sqrt{5})/2$ and $\phi = (1-\sqrt{5})/2$.

1.55 Prove that

$$2 \nmid c_n \Leftrightarrow n = 2^m$$

for some positive integer m, that is, the nth Catalan number c_n is odd when n is power of 2 and even otherwise.

1.6 The Well-ordering Principle

In this section we attempt to justify the PMI by appealing to an assertion that may be intuitively more reasonable. Recall your acceptance, at the age of six, and again about 21 pages ago ((1.17) in Section 1.2), of the idea of a fraction in lowest terms. This is an application of the following fact.

Well-ordering principle (WOP). If A is a property of the positive integers possessed by at least one of them, then there is a least positive integer, l say, with A, that is:

$$\text{(i)} \quad l \text{ has } A, \quad \text{and} \quad \text{(ii)} \quad \text{no } k \text{ with } 1 \leq k < l \text{ has } A.$$

Given a rational number r, let A stand for the property of "being a possible denominator of r", that is, a positive integer b has property A if br is an integer. By definition of rational number, there is at least one such b. The WOP then guarantees that there is a least such, and this is the denominator of r in lowest terms.

Theorem 1.13

The PMI is a consequence of the WOP.

Proof

We shall prove the contrapositive of the implication WOP $\Rightarrow$ PMI. So assume the PMI to be invalid. This means that there is a statement $P(n)$ about positive integers n such that

(a) $P(1)$ is true,

(b) $P(n-1) \Rightarrow P(n)$ for all $n > 1$, but

(c) for some positive integer m, $P(m)$ is false.

Let A be the property that $P(n)$ is false, so that m has property A because of (c). By the WOP, there is a least l with A, so that $P(l)$ is false. So $l \neq 1$ because of (a), whence $l \geq 2$ and $l - 1$ is a positive integer. By the minimality of l, $P(l-1)$ is true, and because of (b), $P(l)$ is true. Contradiction. □

It turns out (Exercise 1.56 below) that the converse of this theorem is also true. Thus the PMI and WOP are logically equivalent, and in some areas of application the WOP is easier to use. Some examples follow, of which the first

is the very reasonable assertion that, in the normal process of division, the remainder is less than what you divide by.

Theorem 1.14 (Euclid)

Given positive integers a and b, we can find integers q and r such that

$$a = bq + r \quad \text{and} \quad 0 \leq r < b. \tag{1.31}$$

Proof

If $b \mid a$, then we can take $q = a/b$ and $r = 0$. If $b \nmid a$, let A be the property: being of the form $a - bn$ with n an integer. By taking $n = 0$, we see that a has property A, and by the WOP there is a least positive integer l with A. Put $l = a - bm$ with m an integer. We prove by contradiction that $l < b$. If this is false, we can write $l = b+x$ with $x \geq 0$. Then $x = l-b = a-bm-b = a-b(m+1)$. Since $b \nmid a$, $x \neq 0$ and the last equation then asserts that x has property A. By the minimality of l, $x \geq l = b + x > x$, a contradiction. Thus, $l = a - bm$ and $l < b$. Taking q to be m and r to be l, we get (1.31). □

This theorem forms the basic method for calculating the following important quantity associated with two positive integers.

Definition 1.8

The **highest common factor** of two positive integers a and b is the largest positive integer h such that $h \mid a$ and $h \mid b$; we often write $h = (a, b)$. a and b are called **relatively prime** (or **coprime**) if $(a, b) = 1$.

Some discussion is in order here. Note first that (a, b) always exists, since the number of common divisors of a and b is non-zero and finite: 1 divides a and b, and if c divides a and b, then $1 \leq c \leq \min(a, b)$. Next, the definition applies equally well if a or b (or both) is negative. *One* of them can even be zero: since every integer divides zero, $(a, 0) = |a|$. The highest common factor is thus defined for any integers a, b except the pair $0, 0$.

The key to the problem of calculating (a, b) for a given a and b is provided by (1.31). Since $(a, b) = (b, a)$, we can assume that $a \geq b$. If c divides both b and r, then clearly $c \mid a$. On the other hand, as $r = a - bq$, if d divides both a and b, then $d \mid r$. The common divisors of a and b are thus the same as the common divisors of b and r:

$$c \mid a \text{ and } c \mid b \Leftrightarrow c \mid b \text{ and } c \mid r.$$

It follows that $(a,b) = (b,r)$. The point is that $r < b$, and the problem has been reduced.

If $r = 0$, we are finished, for

$$(a,b) = (b,r) = (b,0) = b.$$

If not, repeat the process with b, r in place of a, b respectively to get

$$b = rq_1 + r_1, \qquad 0 \leq r_1 < r$$

and $(b,r) = (r,r_1)$. If $r_1 = 0$, then $(a,b) = r$ and we are finished. If not, repeat with (r,r_1) in place of (b,r) to get a remainder $r_2 < r_1$, and so on. Proceeding in this way we get a strictly decreasing sequence $b, r, r_1, r_2, \ldots$ of non-negative integers which, after a finite number of steps (at most b), must reach zero: $r_n = 0$ say. Then, from what has been said, $(a,b) = r_{n-1}$. This process for calculating (a,b) is called **Euclid's algorithm**.

An important but rather unexpected property of the hcf is described in the following theorem.

Theorem 1.15

Given positive integers a and b with $(a,b) = h$, we can find integers s and t such that

$$h = sa + tb. \tag{1.32}$$

Proof

Note first that in general, one of s, t will be positive and the other negative. Let A be the property of positive integers that they can be written in the form of the right-hand side of (1.32): a positive integer c has A if $c = ma + nb$ for some integers m and n. Since at least one positive integer, $a + b$ for example, has A, there is a least such, call it $l : l = ma + nb$. We make the claim that $l \mid a$ and prove it by contradiction.

Assume that $l \nmid a$. Then by Theorem 1.14 we can write

$$a = ql + r, \qquad 0 < r < l.$$

Then

$$r = a - ql = a - q(ma + nb) = (1 - qm)a + (-qn)b,$$

so that r has A. The fact that $0 < r < l$ contradicts the minimality of l, and our claim that $l \mid a$ is established. The fact that $l \mid b$ is proved in the same way.

We have shown that l is a common factor of a and b, whence $l \leq h$. But h divides the right-hand side of the equation $l = ma + nb$. Thus $h \mid l$, so that $h \leq l$. Therefore $h = l = ma + nb$, and we can take $s = m$, $t = n$ to get (1.32).

□

This proof illustrates a weakness in the WOP: it is not constructive. We have shown that s and t satisfying (1.32) *exist*, but no indication is given of how to calculate them. Fortunately, Euclid's algorithm comes to the rescue: s and t can be found by substituting back through the equations that led to $r_{n-1} = (a, b)$.

Example 1.1

Calculate the hcf h of the numbers 89 and 55, and find integers s and t such that

$$h = 89s + 55t.$$

The first steps in the algorithm are as follows:

$$\begin{aligned} 89 &= 55 \,.\, 1 + 34, \\ 55 &= 34 \,.\, 1 + 21, \\ 34 &= 21 \,.\, 1 + 13, \\ 21 &= 13 \,.\, 1 + 8, \\ 13 &= 8 \,.\, 1 + 5, \\ 8 &= 5 \,.\, 1 + 3, \\ 5 &= 3 \,.\, 1 + 2, \\ 3 &= 2 \,.\, 1 + 1, \\ 2 &= 1 \,.\, 2 + 0, \end{aligned}$$

and so $h = 1$. Working backwards,

$$\begin{aligned} 1 = 3 - 2 &= 3 - (5 - 3) = 2 \,.\, 3 - 5 \\ &= 2(8 - 5) - 5 = 2 \,.\, 8 - 3 \,.\, 5 \\ &= 2 \,.\, 8 - 3(13 - 8) = 5 \,.\, 8 - 3 \,.\, 13 \\ &= 5(21 - 13) - 3 \,.\, 13 = 5 \,.\, 21 - 8 \,.\, 13 \\ &= 5 \,.\, 21 - 8(34 - 21) = 13 \,.\, 21 - 8 \,.\, 34 \\ &= 13(55 - 34) - 8 \,.\, 34 = 13.55 - 21 \,.\, 34 \\ &= 13 \,.\, 55 - 21(89 - 55) = 34 \,.\, 55 - 21 \,.\, 89, \end{aligned}$$

so that $s = -21$ and $t = 34$.

So the Fibonacci numbers turn up as the canonical worst case of Euclid's algorithm. This is atypical: the algorithm is in general very efficient.

As a consequence of Theorem 1.15 we shall now obtain an important property of prime numbers, which is a necessary preliminary for the last theorem in this chapter.

Theorem 1.16

If n is a prime dividing a product ab of positive integers, then $n \mid a$ or $n \mid b$.

Proof

The proof is by contradiction. So assume that n is a prime with $n \mid ab$, but that n divides nether a nor b. Since n is prime, (n, a) can only be n or 1. Our hypotheses rule out the first possibility, and so n and a are coprime: $(n, a) = 1$. Similarly $(n, b) = 1$. By Theorem 1.15, we can find integers s, t, u, v such that

$$1 = sn + ta, \qquad 1 = un + vb.$$

Then

$$1 = (sn + ta)(un + vb) = sun^2 + (uta + vbs)n + tv\,ab.$$

Since $n \mid ab$, n divides the right-hand side, whence also $n \mid 1$. Contradiction. □

We are now in the happy position of being able to supply a proof of the Fundamental Theorem of Arithmetic.

Theorem 1.17

Every positive integer n can be expressed as a product of primes. Writing

$$n = p_1^{r_1} p_2^{r_2} \cdots p_l^{r_l} \tag{1.33}$$

with $p_1, p_2, \ldots, p_l$ all primes subject to the conditions

$$\text{(i)} \quad p_1 < p_2 < \cdots < p_l, \qquad \text{(ii)} \quad r_1, r_2, \ldots, r_l \text{ all } \geq 1, \tag{1.34}$$

this expression is unique.

Proof

First the existence of such a decomposition is just the assertion of Theorem 1.7 (in the trivial case $n = 1$, the product is empty).

To prove the uniqueness (and this is typical of such proofs) assume that

$$n = q_1^{s_1} q_2^{s_2} \cdots q_m^{s_m} \tag{1.35}$$

is another such decomposition, that is, $q_1, q_2, \ldots, q_m$ are all prime and

$$\text{(i)} \quad q_1 < q_2 < \cdots < q_m, \qquad \text{(ii)} \quad s_1, s_2, \ldots, s_m \text{ all } \geq 1. \tag{1.36}$$

Then we have to prove that the decompositions (1.33) and (1.35) are *identical*, that is,

$$l = m, \quad \text{and} \quad p_k = q_k, \quad r_k = s_k \qquad \text{for all } k, \quad 1 \leq k \leq l. \tag{1.37}$$

Proceed by (cumulative) induction on n. When $n = 1$, both products must be empty, so that $l = m = 0$ and (1.37) holds. Now let $n \geq 2$ and assume uniqueness for all k with $1 \leq k < n$ as the IH.

Let p be the smallest prime dividing n. Then an easy induction (Exercise 1.63 below) based on Theorem 1.16 shows that p divides one of the p_k, $1 \leq k \leq l$, and condition (i) of (1.34) forces $p = p_1$. Similarly, $p = q_1$, so $p_1 = q_1$.

It follows that

$$\frac{n}{p} = p_1^{r_1 - 1} p_2^{r_2} \cdots p_l^{r_l} = q_1^{s_1 - 1} q_2^{s_2} \cdots q_m^{s_m}. \tag{1.38}$$

Since $n/p < n$, we can apply the IH, provided that the analogues of conditions (1.34) and (1.36) hold for (1.38), that is, with r_1, s_1 replaced by $r_1 - 1$, $s_1 - 1$ respectively.

This is so when r_1, s_1 are both at least 2 (the case $p \mid n/p$), and we can deduce (1.37) from the fact that the products in (1.38) are identical, as $r_1 - 1 = s_1 - 1 \Rightarrow r_1 = s_1$. In the other case, $p \nmid n/p$, we have $r_1 = s_1 = 1$, and the analogous conditions are those obtained from (1.34) and (1.36) by removing the terms p_1, r_1, q_1, s_1. The fact that the products in (1.38) are identical again guarantees (1.37), as $l - 1 = m - 1 \Rightarrow l = m$. Thus, (1.37) holds in both cases and the induction is complete. □

EXERCISES

1.56 Give a proof by contradiction of the converse of Theorem 1.13:

$$\text{PMI} \Rightarrow \text{WOP}.$$

1.57 If l is the least positive denominator of a rational number s, prove that the possible denominators are just the positive multiples of l.

1.58 Show that Theorem 1.14 remains true when a is allowed to be negative.

1.59 Let h be the highest common factor of the positive integers a and b. Prove that the common factors of a and b are just the divisors of h.

1.60 Given positive integers a, b their product is a multiple of both. By the WOP, they have a **least common multiple**, often written $[a, b]$. Prove that $(a, b)\ [a, b] = ab$.

1.61 Find the highest common factor of 582 and 285, and express it in terms of these numbers.

1.62 Find the highest common factor of the polynomials $a(x) = x^6 - 1$ and $b(x) = x^3 + x^2 + x + 1$, and express it in terms of them.

1.63 Let p be a prime and $n \geq 2$ an integer. Prove that if p divides a product $b_1 b_2 \cdots b_n$ of positive integers, then p already divides one of the b_k, $1 \leq k \leq n$.

1.64 Let m and n be positive integers and $p_1, p_2, \ldots, p_l$ a list of the primes that divide at least one of them. Write

$$m = \prod_{k=1}^{l} p_k^{r_k}, \qquad n = \prod_{k=1}^{l} p_k^{s_k},$$

where r_k, $s_k \geq 0$ for $1 \leq k \leq l$. Prove that

$$h = \prod_{k=1}^{l} p_k^{t_k}, \qquad t_k = \min(r_k, s_k), \quad 1 \leq k \leq l,$$

is equal to the highest common factor (m, n) of m and n.

1.65 With m, n as in the previous exercise, write down the prime factorization for the least common multiple $[m, n]$ of m and n. Give an alternative solution of Exercise 1.60.