Chapter 4

The fundamental theorem of arithmetic

We prove two important results in this chapter: the fact that every natural number greater than or equal to 2 can be written uniquely as a product of powers of primes — this is the fundamental theorem of arithmetic — and the proof that certain numbers are irrational.

4.1 Writing numbers down

We begin at the beginning by reviewing how numbers are dealt with graphically.

4.1.1 From tallies to the positional number system

I don't think our hunter-gatherer ancestors worried too much about writing numbers down because there wasn't any need: they didn't have to fill in tax-returns and so didn't need accountants. However, organizing cities does need accountants and so ways had to be found of writing numbers down. The simplest way of doing this is to use a mark like |, called a *tally*, for each thing being counted. So

||||||||||

means 10 things. This system has advantages and disadvantages. The advantage is that you don't have to go on a training course to learn it. The

disadvantage is that even quite small numbers need a lot of space like

||||||||||||||||||||||||||||||||||||

It's also hard to tell whether

||||||||||||||||||||||||||||||||||||

is the same number or not. (It's not.) It's inevitable that people will introduce abbreviations to make the system easier to use. Perhaps it was in this way that the next development occurred. Both the ancient Egyptians and Romans used similar systems but I'll describe the Roman system because it involves letters rather than pictures. First, you have a list of basic symbols:

number	1	5	10	50	100	500	1000
symbol	I	V	X	L	C	D	M

There are more symbols for bigger numbers. Numbers are then written according to the *additive principle*. Thus MMVIIII is 2009. Incidently, I understand that the custom of also using a *subtractive principle* so that, for example, IX means 9 rather than using VIIII, is a more modern innovation. This system is clearly a great improvement on the tally-system. Even quite big numbers are written compactly and it is easy to compare numbers. On the other hand, there is more to learn. The other disadvantage is that we need separate symbols for different powers of 10 and their multiples by 5. This was probably not too inconvenient in the ancient world where it is likely that the numbers needed on a day-to-day basis were never going to be that big. A common criticism of this system is that it is hard to do multiplication in. However, that turns out to be a non-problem because, like us, the Romans used pocket calculators or, more accurately, a device called an *abacus* that could easily be carried under a toga. The real evidence for the usefulness of this system of writing numbers is that it survived for hundreds and hundreds of years.

The system used throughout the world today is quite different and is called the *positional number system*. It seems to have been in place by the ninth century in India but it was hundreds of years in development and the result of ideas from many different cultures: the invention of zero on its own

is one of the great steps in human intellectual development. The genius of the system is that it requires only 10 symbols

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9$$

and every natural number can be written using a sequence of these symbols. The trick to making the system work is that we use the *position* on the page of a symbol to tell us what number it means. Thus 2009 means

10^3	10^2	10^1	10^0
2	0	0	9

In other words

$$2 \times 10^3 + 0 \times 10^2 + 0 \times 10^1 + 9 \times 10^0.$$

Notice the important rôle played by the symbol 0 which makes it clear which column a symbol belongs in otherwise we couldn't tell 29 from 209 from 2009. The disadvantage of this system is that you *do* have to go on a course to learn it because it is a highly sophisticated way of writing numbers. On the other hand, it has the enormous advantage that any number can be written down in a compact way. Once the basic system had been accepted it could be adapted to deal not only with positive whole numbers but also negative whole numbers, using the symbol $-$, and also fractions with the introduction of the decimal point. By the end of the sixteenth century, the full decimal system was in place.

Notation warning! In the UK, we use a raised decimal point like $0 \cdot 123$ and not a comma. Also we generally write the number 1 without a long hook at the top. If you do write it like that there is a danger that people will confuse it with the number 7 which is not always written in the UK with a line through it.

4.1.2 Number bases

We shall now look in more detail at the way in which numbers can be written down using a positional notation. In order not to be biased, we shall not just

work in base 10 but show how any base can be used. There is one result that we shall use throughout this section. You can take it as an axiom, but I shall set a proof as one of the exercises.

Lemma 4.1.1 (Remainder Theorem)**.** *Let a and b be integers where $b > 0$. Then there are unique integers q and r such that*

$$a = bq + r$$

where $0 \leq r < b$.

The number q is called the *quotient* and the number r is called the *remainder*. For example, if we consider the pair of natural numbers 14 and 3 then

$$14 = 3 \cdot 4 + 2$$

where 4 is the quotient and 2 is the remainder.

Let me say a little more about the Remainder Theorem. Your first reaction to it should probably be that it looks obvious. You might conclude from that that it is therefore uninteresting. But this would be wrong. It is certainly not hard to understand but despite that it is important. The reason is that whenever we have a question that involves divisibility, it is very likely going to require the use of this result. For example, it is this result that tells us that odd numbers are precisely those that leave remainder 1 when divided by 2.

Let a and b be integers. We say that a *divides* b or that b *is divisible by* a if there is a q such that $b = aq$. In other words, there is no remainder. We also say that a is a *divisor or factor* of b. We write $a \mid b$ to mean the same thing as 'a divides b'.

Warning! $a \mid b$ does not mean the same thing as $\frac{a}{b}$. The latter is a number, the former is a statement about two numbers.

Let's see how to represent numbers in *base* b where $b \geq 2$. If $d \leq 10$ then we represent numbers by sequences of symbols taken from the set

$$\mathbb{Z}_d = \{0, 1, 2, 3, \ldots d - 1\}$$

but if $d > 10$ then we need new symbols for 10, 11, 12 and so forth. It's convenient to use A,B,C, ... For example, if we want to write numbers in

base 12 we use the set of symbols

$$\{0, 1, \ldots, 9, A, B\}$$

whereas if we work in base 16 we use the set of symbols

$$\{0, 1, \ldots, 9, A, B, C, D, E, F\}.$$

If x is a sequence of symbols then we write x_d to make it clear that we are to interpret this sequence as a number in base d. Thus BAD_{16} is a number in base 16.

The symbols in a sequence x_d, reading from right to left, tell us the contribution each power of d such as d^0, d^1, d^2, etc makes to the number the sequence represents. Here are some examples.

Examples 4.1.2. Converting from base d to base 10.

1. $11A9_{12}$ is a number in base 12. This represents the following number in base 10:
$$\mathbf{1} \times 12^3 + \mathbf{1} \times 12^2 + \mathbf{A} \times 12^1 + \mathbf{9} \times 12^0,$$
which is just the number
$$12^3 + 12^2 + 10 \times 12 + 9 = 2001.$$

2. BAD_{16} represents a number in base 16. This represents the following number in base 10:
$$B \times 16^2 + A \times 16^1 + D \times 16^0,$$
which is just the number
$$11 \times 16^2 + 10 \times 16 + 13 = 2989.$$

3. 5556_7 represents a number in base 7. This represents the following number in base 10:
$$5 \times 7^3 + 5 \times 7^2 + 5 \times 7^1 + 6 \times 7^0 = 2001.$$

These examples show how easy it is to convert from base d to base 10.

There are two ways to convert from base 10 to base d.

1. The first runs in outline as follows. Let n be the number in base 10 that we wish to write in base d. Look for the largest power m of d such that $ad^m \leq n$ where $a < d$. Then repeat for $n - ad^m$. Continuing in this way, we write n as a sum of multiples of powers of d and so we can write n in base d.

2. The second makes use of the remainder theorem. The idea behind this method is as follows. Let

$$n = a_m \dots a_1 a_0$$

in base d. We may think of this as

$$n = (a_m \dots a_1)d + a_0$$

It follows that a_0 is the remainder when n is divided by d, and the quotient is $n' = a_m \dots a_1$. Thus we can generate the digits of n in base d from *right to left* by repeatedly finding the next quotient and next remainder by dividing the current quotient by d; the process starts with our input number as first quotient.

Examples 4.1.3. Converting from base 10 to base d.

1. Write 2001 in base 7. I'll solve this question in two different ways: the long but direct route and then the short but more thought-provoking route.

We see that $7^4 > 2001$. Thus we divide 2001 by 7^3. This goes 5 times plus a remainder. Thus $2001 = 5 \times 7^3 + 286$. We now repeat with 286. We divide it by 7^2. It goes 5 times again plus a remainder. Thus $286 = 5 \times 7^2 + 41$. We now repeat with 41. We get that $41 = 5 \times 7 + 6$. We have therefore shown that

$$2001 = 5 \times 7^3 + 5 \times 7^2 + 5 \times 7 + 6.$$

Thus 2001 in base 7 is just 5556.

Now for the short method.

	quotient	remainder
7	2001	
7	285	6
7	40	5
7	5	5
	0	5

Thus 2001 in base 7 is:

$$5556.$$

2. Write 2001 in base 12.

	quotient	remainder
12	2001	
12	166	9
12	13	$10 = A$
12	1	1
	0	1

Thus 2001 in base 12 is:

$$11A9.$$

3. Write 2001 in base 2.

	quotient	remainder
2	2001	
2	1000	1
2	500	0
2	250	0
2	125	0
2	62	1
2	31	0
2	15	1
2	7	1
2	3	1
2	1	1
	0	1

Thus 2001 in base 2 is (reading from bottom to top):

$$11111010001.$$

When converting from one base to another it is always wise to **check** your calculations by converting back.

Number bases have some special terminology associated with them which you might encounter:

Base 2 *binary*.

Base 8 *octal*.

Base 10 *decimal*.

Base 12 *duodecimal*.

Base 16 *hexadecimal*.

Base 20 *vigesimal*.

Base 60 *sexagesimal*.

Binary, octal and hexadecimal occur in computer science; there are remnants of a vigesimal system in French and the older Welsh system of counting; base 60 was used by astronomers in ancient Mesopotamia and is still the basis of time measurement (60 seconds = 1 minute, and 60 minutes = 1 hour) and angle measurement.

Exercises 4.1

1. Find the *quotients* and *remainders* for each of the following pair of numbers. Divide the smaller into the larger.

   (a) 30 and 6.

   (b) 100 and 24.

   (c) 364 and 12.

2. Write the number 2009 in

(a) Base 5.

(b) Base 12.

(c) Base 16.

3. Write the following numbers in base 10.

(a) DAB_{16}.

(b) $ABBA_{12}$.

(c) 44332211_5.

4. Prove the following properties of the division relation on $\mathbb{Z}$.

(a) If $a \neq 0$ then $a \mid a$.

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

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

(d) If $a \mid b$ and $a \mid c$ then $a \mid (b + c)$.

5. This question develops a proof of the remainder theorem. Let a and b be integers with $b > 0$. Then there exist a unique pair of integers q and r such that $a = qb + r$ where $0 \leq r < b$.

(a) Let
$$X = \{a - nb \colon n \in \mathbb{Z}\}.$$
Show that this set contains non-negative elements.

(b) Let X^+ be the subset of X consisting of non-negative elements. This subset if non-empty by the first step. Use the well-ordering principle to deduce that this set contains a minimum element r. Thus $r = a - qb \geq 0$ for some $q \in \mathbb{Z}$.

(c) Show that if $r \geq b$ then X^+ in fact contains a smaller element, which is a contradiction.

(d) We therefore have that $a = bq + r$ where $0 \leq r < b$. It remains to prove that q and r are unique with these propertries. Assume therefore that $a = bq' + r'$ where $0 \leq r' < b$. Deduce that $q = q'$ and $r = r'$.

4.2 Greatest common divisors

The ideas in this section are simple but their ramifications substantial. Let $a, b \in \mathbb{N}$. A number d which divides both a and b is called a *common divisor* of a and b. The largest number which divides both a and b is called the *greatest common divisor* of a and b and is denoted by $\gcd(a, b)$. A pair of natural numbers a and b is said to be *coprime* if $\gcd(a, b) = 1$.

Note that for us $\gcd(0, 0)$ is undefined but that if $a \neq 0$ then $\gcd(a, 0) = a$.

Example 4.2.1. Consider the numbers 12 and 16. The set of divisors of 12 is the set $\{1, 2, 3, 4, 6, 12\}$. The set of divisors of 16 is the set $\{1, 2, 4, 8, 16\}$. The set of common divisors is the set of numbers that belong to both of these two sets: namely, $\{1, 2, 4\}$. The greatest common divisor of 12 and 16 is therefore 4. Thus $\gcd(12, 16) = 4$.

One application of greatest common divisors is in simplifying fractions. For example, the fraction $\frac{12}{16}$ is equal to the fraction $\frac{3}{4}$ because we can divide out the common divisor of numerator and denominator. The fraction which results cannot be simplified further and is in its *lowest terms*.

Lemma 4.2.2. *Let* $d = \gcd(a, b)$. *Then* $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$.

Proof. Because d divides both a and b we may write $a = a'd$ and $b = b'd$ for some natural numbers a' and b'. We therefore need to prove that $\gcd(a', b') = 1$. Suppose that $e \mid a'$ and $e \mid b'$. Then $a' = ex$ and $b' = ey$ for some natural numbers x and y. Thus $a = exd$ and $b = eyd$. Observe that $ed \mid a$ and $ed \mid b$ and so ed is a common divisor or both a and b. But d is the *greatest* common divisor and so $e = 1$, as required. □

Let me paraphrase what the result above says since it is not surprising. If I divide two numbers by their greatest common divisor then the numbers that remain are coprime. This seems intuitively plausible and the proof ensures that our intuition is correct.

If the numbers a and b are large, then calculating their gcd in the way I did above would be time-consuming and error-prone. We want to find an *efficient* way of calculating the greatest common divisor. The following lemma is the basis of just such an efficient method.

Lemma 4.2.3. *Let $a, b \in \mathbb{N}$, where $b \neq 0$, and let $a = bq + r$ where $0 \leq r < b$. Then*

$$\gcd(a, b) = \gcd(b, r).$$

Proof. Let d be a common divisor of a and b. Since $a = bq + r$ we have that $a - bq = r$ so that d is also a divisor of r. It follows that any divisor of a and b is also a divisor of b and r.

Now let d be a common divisor of b and r. Since $a = bq + r$ we have that d divides a. Thus any divisor of b and r is a divisor of a and b.

It follows that the set of common divisors of a and b is the same as the set of common divisors of b and r. Thus $\gcd(a, b) = \gcd(b, r)$. □

The point of the above result is that $b < a$ and $r < b$. So calculating $\gcd(b, r)$ will be easier than calculating $\gcd(a, b)$ because the numbers involved are smaller. Compare

$$\overbrace{a = b}\, q + r$$

with

$$a = \underbrace{bq + r}\,.$$

The above result is the basis of an efficient algorithm for computing greatest common divisors. It was described in Propositions 1 and 2 of Book VII of the Elements.

Algorithm 4.2.4 (Euclid's algorithm)**.**
Input: $a, b \in \mathbb{N}$ such that $a \geq b$ and $b \neq 0$.
Output: $\gcd(a, b)$.
Procedure: write $a = bq + r$ where $0 \leq r < b$. Then $\gcd(a, b) = \gcd(b, r)$. If $r \neq 0$ then repeat this procedure with b and r and so on. The last non-zero remainder is $\gcd(a, b)$

Example 4.2.5. Let's calculate $\gcd(19, 7)$ using Euclid's algorithm. I have highlighted the numbers that are involved at each stage.

$$\begin{aligned} \mathbf{19} &= \mathbf{7} \cdot 2 + \mathbf{5} \\ \mathbf{7} &= \mathbf{5} \cdot 1 + \mathbf{2} \\ \mathbf{5} &= \mathbf{2} \cdot 2 + \mathbf{1} \ * \\ \mathbf{2} &= \mathbf{1} \cdot 2 + 0 \end{aligned}$$

By Lemma 1.3.3 we have that

$$\gcd(19, 7) = \gcd(7, 5) = \gcd(5, 2) = \gcd(2, 1) = \gcd(1, 0).$$

The last non-zero remainder is 1 and so $\gcd(19, 7) = 1$ and, in this case, the numbers are coprime.

There are occasions when we need to extract more information from Euclid's algorithm as we shall discover later when we come to deal with prime numbers.

Theorem 4.2.6 (Bézout's theorem)**.** *Let a and b be natural numbers. Then there are integers x and y such that*

$$\gcd(a, b) = xa + yb.$$

I shall prove this theorem by describing an algorithm that will compute the integers x and y above. This is achieved by running Euclid's algorithm in reverse and is called *the extended Euclidean algorithm*. The procedure for doing so is outlined below but the details are explained in the example that follows it.

Algorithm 4.2.7 (Extended Euclidean algorithm)**.**
Input: $a, b \in \mathbb{N}$ where $a \geq b$ and $b \neq 0$.
Output: numbers $x, y \in \mathbb{Z}$ such that $\gcd(a, b) = xa + yb$.
Procedure: apply Euclid's algorithm to a and b; working from bottom to top rewrite each remainder in turn.

Example 4.2.8. This is a little involved so I have split the process up into steps. I shall apply the extended Euclidean algorithm to the example I calculated above. I have highlighted the non-zero remainders wherever they occur, and I have discarded the last equality where the remainder was zero. I have also marked the last non-zero remainder.

$$\begin{aligned} 19 &= 7 \cdot 2 + \mathbf{5} \\ 7 &= \mathbf{5} \cdot 1 + \mathbf{2} \\ \mathbf{5} &= \mathbf{2} \cdot 2 + \mathbf{1} \ * \end{aligned}$$

The first step is to rearrange each equation so that the non-zero remainder is alone on the lefthand side.

$$\begin{aligned} \mathbf{5} &= 19 - 7 \cdot 2 \\ \mathbf{2} &= 7 - \mathbf{5} \cdot 1 \\ \mathbf{1} &= \mathbf{5} - \mathbf{2} \cdot 2 \end{aligned}$$

Next we reverse the order of the list

$$\begin{aligned} \mathbf{1} &= \mathbf{5} - \mathbf{2} \cdot 2 \\ \mathbf{2} &= 7 - \mathbf{5} \cdot 1 \\ \mathbf{5} &= 19 - 7 \cdot 2 \end{aligned}$$

We now start with the first equation. The lefthand side is the gcd we are interested in. We treat all other remainders as algebraic quantities and systematically substitute them in order. Thus we begin with the first equation

$$\mathbf{1} = \mathbf{5} - \mathbf{2} \cdot 2.$$

The next equation in our list is

$$\mathbf{2} = 7 - \mathbf{5} \cdot 1$$

so we replace **2** in our first equation by the expression on the right to get

$$\mathbf{1} = \mathbf{5} - (7 - \mathbf{5} \cdot 1) \cdot 2.$$

We now rearrange this equation by collecting up like terms treating the highlighted remainders as algebraic objects to get

$$\mathbf{1} = 3 \cdot \mathbf{5} - 2 \cdot 7.$$

We can of course make a check at this point to ensure that our arithmetic is correct. The next equation in our list is

$$\mathbf{5} = 19 - 7 \cdot 2$$

so we replace **5** in our new equation by the expression on the right to get

$$\mathbf{1} = 3 \cdot (19 - 7 \cdot 2) - 2 \cdot 7.$$

Again we rearrange to get

$$\mathbf{1} = 3 \cdot \mathit{19} - 8 \cdot \mathit{7}.$$

The algorithm now terminates and we can write

$$\gcd(19, 7) = 3 \cdot \mathit{19} + (-8) \cdot \mathit{7},$$

as required. We can also, of course, easily check the answer!

I shall describe a much more efficient algorithm for implementing the extended Euclidean algorithm later in this book when I have discussed matrices.

The key application for the material of this section is the following.

Lemma 4.2.9. *Let a and b be natural numbers. Then a and b are coprime if, and only if, we may find integers x and y such that*

$$1 = xa + yb.$$

Proof. Suppose first that a and b are coprime. By Bézout's theorem

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

for some integers a and b. But, by assumption, $\gcd(a, b) = 1$. Conversely, suppose that

$$1 = xa + yb.$$

Then any natural number that divides both a and b must divide 1. It follows that $\gcd(a, b) = 1$. □

The significance of the above lemma is that whenever you know that a and b are coprime, you can actually write down an expression $1 = xa + yb$ which actually means the same thing. This turns out to be enormously useful.

The greatest common divisor of two numbers a and b is the largest number that divides into both a and b. On the other hand, if $a \mid c$ and $b \mid c$ then we say that c is a *common multiple* of a and b. The smallest common multiple of a and b is called the *least common multiple* of a and b and is denoted by $\mathrm{lcm}(a, b)$. You might expect that to calculate the least common multiple we would need a new algorithm, but in fact we can use Euclid's algorithm as the following result shows. I shall prove the following result later once I have proved the fundamental theorem of arithmetic.

Proposition 4.2.10. *Let a and b be natural numbers. Then*

$$\gcd(a,b) \times \operatorname{lcm}(a,b) = ab.$$

The notions of gcd and lcm play a natural role in the arithmetic of fractions. The key property of fractions is that a fraction $\frac{a}{b}$ is unchanged when numerator and denominator are both multiplied by the same non-zero integer. Thus

$$\frac{a}{b} = \frac{ac}{bc}.$$

Given a fraction $\frac{a}{b}$ we often want to simplify it as much as possible and this is accomplished by calculating $\gcd(a,b) = d$. We have $a = a'd$ and $b = b'd$ and so

$$\frac{a}{b} = \frac{a'd}{b'd} = \frac{a'}{b'}.$$

We have proved above that $\gcd(a',b') = 1$ and so the fraction cannot be simplified any further. Thus $\frac{a'}{b'}$ is a fraction in its lowest terms. When we come to add fractions, the problem is the reverse of simplification. We cannot immediately add $\frac{a}{b} + \frac{c}{d}$ because the denominators b and d are different. To make progress, we have to rewrite each fraction so that their denominators are the same. The simplest way to do this is to rewrite each fraction as a fraction over bd by multiplying the first fraction by d and the second by b to get

$$\frac{ad}{bd} + \frac{bc}{bd} = \frac{ad+bc}{bd}.$$

However, the most efficient way is to write each fraction over $\operatorname{lcm}(b,d)$. Let $\operatorname{lcm}(b,d) = b'b = d'd$. Then

$$\frac{a}{b} + \frac{c}{d} = \frac{b'a}{b'b} + \frac{d'c}{d'd} = \frac{b'a + d'c}{\operatorname{lcm}(b,c)}.$$

Exercises 4.2

1. Use Euclid's algorithm to find the gcd's of the following pairs of numbers.

   (a) 35, 65.

   (b) 135, 144.

(c) 17017, 18900.

2. Use the extended Euclidean algorithm to find integers x and y such that $\gcd(a, b) = ax + by$ for each of the following pairs of numbers. You should ensure that your answers for x and y have the correct signs.

   (a) 112, 267.

   (b) 242, 1870.

3. Find the lowest common multiples of the following pairs of numbers.

   (a) 22, 121.

   (b) 48, 72.

   (c) 25, 116.

4. We know how to find the greatest natural number that divides two numbers. Define now $\gcd(a, b, c)$ to be the greatest common divisor of a and b and c jointly. Prove that

$$\gcd(a, b, c) = \gcd(\gcd(a, b), c).$$

   Deduce that

$$\gcd(\gcd(a, b), c) = \gcd(a, \gcd(b, c)).$$

   We may similarly define $\gcd(a, b, c, d)$ to be the greatest common divisor of a and b and c and d jointly. Calculate $\gcd(910, 780, 286, 195)$ and justify your calculations.

5. The following question is by Dubisch *Amer. Math. Mon.* **69**. Define $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$. A binary operation $\circ$ defined on $\mathbb{N}^*$ is known to have the following properties:

   (a) $a \circ b = b \circ a$.

   (b) $a \circ a = a$.

   (c) $a \circ (a + b) = a \circ b$.

   Prove that $a \circ b = \gcd(a, b)$. Hint: the question is *not* asking you to prove that $\gcd(a, b)$ has these properties.

6. You have an unlimited supply of 3 cent stamps and an unlimited supply of 5 cent stamps. By combining stamps of different values you can make up other values: for example, three 3 cent stamps and two 5 cent stamps make the value 19 cents. What is the largest value you *cannot* make? Hint: you need to show that the question makes sense.

7. Let $n \geq 1$. Define $\phi(n)$ to be the number of numbers less than n and coprime to n. This is the *Euler totient function*. Tabulate the values of $\phi(n)$ for $1 \leq n \leq 12$.

4.3 The fundamental theorem of arithmetic

The goal of this section is to state and prove the most basic result about the natural numbers: each natural number, excluding 0 and 1, can be written as a product of powers of primes in essentially one way. The primes are therefore the 'atoms' from which all natural numbers can be built.

4.3.1 Primes: the atoms of number

A *proper divisor* of a natural number n is a divisor that is neither 1 nor n. A natural number n is said to be *prime* if $n \geq 2$ and the only divisors of n are 1 and n itself. A number bigger than or equal to 2 which is not prime is said to be *composite*.

Warning! The number 1 is not a prime.

The properties of primes have exercised a great fascination ever since they were first studied and continue to pose questions that mathematicians have yet to solve. We shall just describe their basic properties in this section.

Lemma 4.3.1. *Let $n \geq 2$. Either n is prime or the smallest proper divisor of n is prime.*

Proof. Suppose n is not prime. Let d be the smallest proper divisor of n. If d were not prime then d would have a smallest proper divisor and this divisor would in turn divide n, but this would contradict the fact that d was the smallest proper divisor of n. Thus d must itself be prime. □

The following was also proved by Euclid: it is Proposition 20 of Book IX of Euclid.

Theorem 4.3.2. *There are infinitely many primes.*

Proof. Let $p_1, \ldots, p_n$ be the first n primes. Put

$$N = (p_1 \ldots p_n) + 1.$$

If N is a prime, then N is a prime bigger than p_n. If N is composite, then N has a prime divisor p by Lemma 4.3.1. But p cannot equal any of the primes $p_1, \ldots, p_n$ because N leaves remainder 1 when divided by p_i. It follows that p is a prime bigger than p_n. Thus we can always find a bigger prime. It follows that there must be an infinite number of primes. □

Algorithm 4.3.3. To decide whether a number n is prime or composite. Check to see if any prime $p \leq \sqrt{n}$ divides n. If none of them do, the number n is prime.

Let's think about why this works. If a divides n then we can write $n = ab$ for some number b. If $a < \sqrt{n}$ then $b > \sqrt{n}$ whilst if $a > \sqrt{n}$ then $b < \sqrt{n}$. Thus to decide if n is prime or not we need only carry out trial divisions by all numbers $a \leq \sqrt{n}$. However, this is inefficient because if a divides n and a is not prime then a is divisible by some prime p which must therefore also divide n. It follows that we need only carry out trial divisions by the primes $p \leq \sqrt{n}$.

Example 4.3.4. Determine whether 97 is prime using the above algorithm. We first calculate the largest whole number less than or equal to $\sqrt{97}$. This is 9. We now carry out trial divisions of 97 by each prime number p where $2 \leq p \leq 9$; by the way, if you aren't certain which of these numbers is prime: just try them all. You'll get the right answer although not as efficiently. You might also want to remember that if m doesn't divide a number neither can any multiple of m. In any event, in this case we carry out trial divisions by $2, 3, 5$ and 7. None of them divides 97 exactly and so 97 is prime.

The following is the key property of primes we shall need to prove the fundamental theorem of arithmetic. We use Bézout's Theorem to prove it. It is Proposition 30 of Book VII of Euclid.

Lemma 4.3.5 (Euclid's lemma)**.** *Let $p \mid ab$ where p is a prime. Then $p \mid a$ or $p \mid b$.*

Proof. Suppose that p does not divide a. We shall prove that p must then divide b. If p does not divide a, then a and p are coprime, and so there exist integers x and y such that $1 = px + ay$. Thus $b = bpx + bay$. Now $p \mid bp$ and $p \mid ba$, by assumption, and so $p \mid b$, as required. □

Example 4.3.6. The above result is not true if p is not a prime. For example, $6 \mid 9 \times 4$ but 6 divides neither 9 nor 4.

Lemma 4.3.5 is so important, I want to spell out in words what it says

> *If a prime divides a product of numbers it must divide at least one of them.*

Theorem 4.3.7 (Fundamental theorem of arithmetic)**.** *Every number $n \geq 2$ can be written as a product of primes in one way if we ignore the order in which the primes appear. By* product *we allow the possibility that there is only one prime.*

Proof. Let $n \geq 2$. If n is already a prime then there is nothing to prove, so we can suppose that n is composite. Let p_1 be the smallest prime divisor of n. Then we can write $n = p_1 n'$ where $n' < n$. Once again, n' is either prime or composite. Continuing in this way, we can write n as a product of primes.

We now prove uniqueness. Suppose that

$$n = p_1 \dots p_s = q_1 \dots q_t$$

are two ways of writing n as a product of primes. Now $p_1 \mid n$ and so $p_1 \mid q_1 \dots q_t$. By Euclid's Lemma, the prime p_1 must divide one of the q_i's and, since they are themselves prime, it must actually equal one of the q_i's. By relabelling if necessary, we can assume that $p_1 = q_1$. Cancel p_1 from both sides and repeat with p_2. Continuing in this way, we see that every prime occurring on the lefthand side occurs on the righthand side. Changing sides, we see that every prime occurring on the righthand side occurs on the lefthand side. We deduce that the two prime decompositions are identical. □

When we write a number as a product of primes we usually gather together the same primes into a prime power, and write the primes in increasing order which then gives a unique representation. This is illustrated in the example below.

Example 4.3.8. Let $n = 999,999$. Write n as a product of primes. There are a number of ways of doing this but in this case there is an obvious place to start. We have that

$$n = 3^2 \cdot 111,111 = 3^3 \cdot 37,037 = 3^3 \cdot 7 \cdot 5,291 = 3^3 \cdot 7 \cdot 11 \cdot 481 = 3^3 \cdot 7 \cdot 11 \cdot 13 \cdot 37.$$

Thus the prime factorisation of $999,999$ is

$$999,999 = 3^3 \cdot 7 \cdot 11 \cdot 13 \cdot 37.$$

We can use the prime factorizations of numbers to give a nice proof of Proposition 4.2.10. Let m and n be two integers. To keep things simple, we suppose that their prime factorizations are

$$m = p_1^{\alpha} p_2^{\beta} p_3^{\gamma} \text{ and } n = p_1^{\delta} p_2^{\epsilon} p_3^{\zeta}$$

where p_1, p_2, p_3 are primes. It will be obvious how to extend this argument to the general case. The prime factorizations of $\gcd(m,n)$ and $\text{lcm}(m,n)$ are

$$\gcd(m,n) = p_1^{\min(\alpha,\delta)} p_2^{\min(\beta,\epsilon)} p_3^{\min(\gamma,\zeta)}$$

and

$$\text{lcm}(m,n) = p_1^{\max(\alpha,\delta)} p_2^{\max(\beta,\epsilon)} p_3^{\max(\gamma,\zeta)}$$

respectively. I shall let you work out why and also work out how we can use these results to prove the above proposition.

4.3.2 Numerical partial fractions

This section is intended as motivation for the partial fraction representation of rational functions described in a later chapter, so it can be omitted at first reading. The idea is to show how a fraction be written as a sum of other fractions having a particular shape. Specfically, the goal of this section is to show how a proper fraction can be written as a sum of proper fractions over prime power denominators. This involves two steps which I shall describe by means of examples. The theory is an application of the fundamental theorem or arithmetic and the extended Euclidean algorithm.

In order to add two fractions together, we first have to ensure that both are expressed over the same denominator. For example, suppose we want to add $\frac{5}{7}$ and $\frac{8}{13}$. Since $7 \times 13 = 91$ we have the following

$$\frac{5}{7} + \frac{8}{13} = \frac{65+56}{91} = \frac{121}{91}.$$

We shall now consider the reverse process, using the fraction $\frac{810}{1003}$ as an example. Observe that $1003 = 17 \times 59$ where 17 and 59 are coprime. Our goal is to write

$$\frac{810}{1003} = \frac{a}{17} + \frac{b}{59}$$

for some natural numbers a and b. By the extended Euclidean algorithm, we can write

$$1 = 7 \cdot 17 - 2 \cdot 59.$$

It follows that

$$\frac{1}{1003} = \frac{7 \cdot 17 - 2 \cdot 59}{17 \cdot 59} = \frac{7}{59} - \frac{2}{17}.$$

Now multiply both sides by 810 to get

$$\frac{810}{1003} = \frac{7 \cdot 810}{59} - \frac{2 \cdot 810}{17} = 96\frac{6}{59} - 95\frac{5}{17} = 1 + \frac{6}{59} - \frac{5}{17}.$$

Simplifying we get

$$\frac{810}{1003} = \frac{6}{59} + \frac{12}{17}$$

as required.

We shall now do something different. Consider the fraction $\frac{10}{16}$. We have that $16 = 2^4$ and so we cannot write it as a product of coprime numbers. However, we can do something else. We can write $10 = 2+8 = 2^1+2^3$. Thus

$$\frac{10}{16} = \frac{2^1 + 2^3}{2^4} = \frac{2^1}{2^4} + \frac{2^3}{2^4} = \frac{1}{2^3} + \frac{1}{2}.$$

Thus

$$\frac{10}{16} = \frac{1}{2^1} + \frac{1}{2^3}.$$

Let's now combine these two steps. Consider the fraction $\frac{41}{90}$. The prime factorisation of 90 is $2 \cdot 3^2 \cdot 5$. Our first goal is to write

$$\frac{41}{90} = \frac{a}{2} + \frac{b}{3^2} + \frac{c}{5}.$$

Thus we have to find a, b, c such that

$$41 = 45a + 10b + 18c.$$

By trial and error, remembering that a, b, c have to be integers, we find that

$$41 = 45 \cdot 1 + 10 \cdot 5 + (-3) \cdot 18.$$

It follows that

$$\frac{41}{90} = \frac{1}{2} + \frac{5}{3^2} - \frac{3}{5}.$$

We now want to write

$$\frac{5}{3^2} = \frac{d}{3} + \frac{e}{3^2}$$

where $|d|, |e| < 3$. But $5 = 2 + 3$ and so

$$\frac{5}{3^2} = \frac{1}{3} + \frac{2}{3^2}.$$

It follows that

$$\frac{41}{90} = \frac{1}{2} + \frac{1}{3} + \frac{2}{9} - \frac{3}{5}.$$

We may summarise what we have found in the following theorem.

Theorem 4.3.9.

(i) *Let $\frac{a}{b}$ be a proper fraction, and let $b = p_1^{n_1} \dots p_r^{n_r}$ be the prime factorisation of b. Then*

$$\frac{a}{b} = \sum_{i=1}^{r} \frac{c_i}{p_i^{n_i}}$$

for some integers c_i, where each of the fractions is proper.

(ii) *Now let p be a prime and $\frac{c}{p^n}$ a proper fraction. Then*

$$\frac{c}{p^n} = \sum_{j=1}^{n} \frac{d_j}{p^j}$$

where each d_j is such that $|d_j| < p$.

4.3.3 The prime number theorem

There are no nice formulae to tell us what the nth prime is but there are still some interesting results in this direction. The polynomial

$$p(n) = n^2 - n + 41$$

has the property that its value for $n = 1, 2, 3, 4, \ldots, 40$ is always prime. Of course, for $n = 41$ it is clearly not prime. In 1971, the mathematician Yuri Matijasevic found a polynomial in 26 variables of degree 25 with the property that when non-negative integers are substituted for the variables the positive values it takes are all and only the primes. However, this polynomial does not generate the primes in any particular order.

If we adopt a *statistical approach* then we can obtain much more useful results. The idea is that for each natural number n we count the number of primes $\pi(n)$ less than or equal to n. If we are going to do this then our first problem is to compile a table of sufficiently many of them. The simplest way of doing this is to use the *Sieve of Eratosthenes*. Suppose we want to construct a table of all primes up to the number N. We begin by listing all numbers from 2 to N inclusive. Mark 2 as prime and then cross out from the table all numbers which are multiples of 2. The first number after 2 which we have not crossed out is 3. We mark this as prime and then cross out all multiples of 3. The first number after 3 not crossed out is 5. We mark this as prime and continue in the same way. We stop when we have crossed out all multiples of the largest prime less than or equal to $\sqrt{N}$. All marked numbers will be prime as well as those numbers which remain not crossed out.

If you compile tables of primes in this way, you can calculate the function $\pi(x)$. Its graph has a staircase shape — it certainly isn't smooth — but as you zoom away it begins to look smoother and smoother. This raises the question whether there is a smooth function that is a good approximation to $\pi(n)$. This seems to have been what Gauss did. He set up a table something like the following (this is taken from LeVeque's book *Fundamentals of number theory*, Dover, 1977) where

$$\Delta(x) = \frac{\pi(x) - \pi(x - 1000)}{1000}$$

represents an approximate slope of the *curve* $\pi(x)$.

x	$\pi(x)$	$\Delta(x)$	$\frac{1}{\ln(x)}$
1000	168	$0 \cdot 168$	$0 \cdot 145$
2000	303	$0 \cdot 135$	$0 \cdot 132$
3000	430	$0 \cdot 127$	$0 \cdot 125$
4000	550	$0 \cdot 120$	$0 \cdot 121$
5000	669	$0 \cdot 119$	$0 \cdot 117$
6000	783	$0 \cdot 114$	$0 \cdot 115$
7000	900	$0 \cdot 117$	$0 \cdot 113$
8000	1007	$0 \cdot 107$	$0 \cdot 111$
9000	1117	$0 \cdot 110$	$0 \cdot 110$
10000	1229	$0 \cdot 112$	$0 \cdot 109$

Gauss noticed, because that was the kind of person he was, that the *slope* of $\pi(x)$ looked very much like $\frac{1}{\ln(x)}$. This suggests that the function, defined by integrating these *slopes*, is given by

$$\mathrm{li}(x) = \int_{t=2}^{x} \frac{1}{\ln(t)} dt$$

should be an approximation to $\pi(x)$. It is called the *logarithmic integral*. Of course, this is not a theorem: it is a conjecture. It was proved in 1896 by two mathematicians: Hadamard in France and de la Vallée Poussin in Belgium.

Theorem 4.3.10 (The Prime Number Theorem (PNT): version 1)**.**

$$\lim_{x \to \infty} \frac{\pi(x)}{\mathrm{li}(x)} = 1.$$

This version of the PNT is not that easy for us to use. However by l'Hôpital's rule, we can show that

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

If we assume the first version of the PNT and use the above result, we obtain the second version of the PNT.

Theorem 4.3.11 (The Prime Number Theorem: version 2)**.**

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

The above theorem can be interpreted as saying that for large values of x the value of $\pi(x)$ is approximately given by $\frac{x}{\ln(x)}$. This result is a huge improvement on the theorem that there are infinitely many primes: it tells us not only that there are infinitely many of them but also how they are distributed.

Prime numbers also play an important role in computing: specifically, in exchanging secret information. In 1976, Whitfield Diffie and Martin Hellman wrote a paper on cryptography that can genuinely be called ground-breaking. In 'New directions in cryptography' *IEEE Transactions on Information Theory* **22** (1976), 644–654, they put forward the idea of a *public-key cryptosystem* which would enable

> . . . a private conversation . . . [to] be held between any two individuals regardless of whether they have ever communicated before.

With considerable farsightedness, Diffie and Hellman foresaw that such cryptosystems would be essential if communication between computers was to reach its full potential. However, their paper did not describe a concrete way of doing this. It was R. I. Rivest, A. Shamir and L. Adleman (RSA) who found just such a concrete method described in their paper, 'A method for obtaining digital signatures and public-key cryptosystems' *Communications of the ACM* **21** (1978), 120–126. Their method is based on the following observation. Given two prime numbers it takes very little time to multiply them together, but if I give you a number that is a product of two primes and ask you to factorize it then it takes a lot of time. After considerable experimentation, RSA showed how to use little more than undergraduate mathematics to put together a public-key cryptosystem that is an essential ingredient in e-commerce. Ironically, this secret code had in fact been invented in 1973 at GCHQ, who had kept it secret.

Exercises 4.3

1. List the primes less than 100.

2. For each of the following numbers use Algorithm 4.3.3 to determine whether they are prime or composite. When they are composite find a prime factorization. Show all working.

   (a) 131.

(b) 689.

(c) 5491.

3. Given $2^4 \cdot 3 \cdot 5^5 \cdot 11^2$ and $2^2 \cdot 5^6 \cdot 11^4$, calculate their greatest common divisor and least common multiple.

4. Use the fundamental theorem of arithmetic to show that we can always write $\sqrt{n}$, where n is a natural number, as a product of a natural number and a product of square roots of primes. Calculate the square roots of the following numbers exactly using the above method.

   (a) 10.

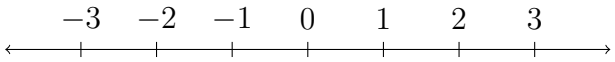
   (b) 42.

   (c) 54.

5. Let a and b be coprime. Prove that if $a \mid bc$ then $a \mid c$.

4.4 Modular arithmetic

From an early age, we are taught to think of numbers as being strung out along the *number line*



But that is not the only way we count. We count the seasons in a cyclic manner

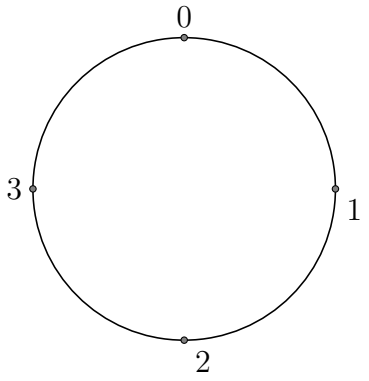
... autumn, winter, spring, summer ...

and likewise the days of the week

... Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday ...

Also the months of the year or the hours in a day, whether by means of the 12-hour clock or the 24-hour clock. The fact that we use words for these events obscures the fact that we really are counting. This is clearer in the names for the months since October, November and December were originally the eighth, ninth and tenth months, respectively, until Roman politics intervened and they were shifted. But the counting in all these cases is not *linear* but

cyclic. Rather than using a number *line* to represent this type of counting, we use instead number *circles*, and rather than using the words above, I shall use numbers. Here is the number circle for the seasons with numbers replacing words.



Adding in these systems of arithmetic means stepping around in a clockwise direction, whereas subtracting means stepping around in an anticlockwise direction. *Modular arithmetic* is the name given to these different systems of cyclic counting. It was Carl Friedrich Gauss (1777–1855) who realised that these different systems of counting were mathematically interesting.

4.4.1 Congruences

Let $n \geq 2$ be a fixed natural number which in this context we call the *modulus*. If $a, b \in \mathbb{Z}$ we write $a \equiv b$ if, and only if, a and b leave the same remainder when divided by n or, what amounts to the same thing, $n \mid a - b$.

Here are a couple of simple examples. If $n = 2$, then $a \equiv b$ if, and only if, a and b are either both odd or both even. On the other hand, if $n = 10$ then $a \equiv b$ if, and only if, a and b have the same units digit.

The symbol $\equiv$ is a modification of the equality symbol $=$. If $a \equiv b$ with respect to n we say that *a is congruent to b modulo n*. In fact, congruence behaves like a weakened form of equality as we now show.

Lemma 4.4.1. *Let $n \geq 2$ be a fixed modulus.*

1. $a \equiv a$.

2. *$a \equiv b$ implies $b \equiv a$.*

3. *$a \equiv b$ and $b \equiv c$ implies that $a \equiv c$.*

4. *$a \equiv b$ and $c \equiv d$ implies that $a + c \equiv b + d$.*

5. *$a \equiv b$ and $c \equiv d$ implies that $ac \equiv bd$.*

Here is a very simple application of modular arithmetic.

Lemma 4.4.2. *A natural number n is divisible by 9 if, and only if, the sum of the digits of n is divisible by 9.*

Proof. We shall work modulo 9. The proof hinges on the fact that $10 \equiv 1$ modulo 9. By using Lemma 4.4.1, we quickly find that $10^r \equiv 1$ for all natural numbers $r \geq 1$. We use this result now. Let

$$n = a_n 10^n + a_{n-1} 10^{n-1} + \ldots + a_1 10 + a_0.$$

Then $n \equiv a_n + \ldots + a_0$. Thus n and the sum of the digits of n leave the same remainder when divided by 9, and so n is divisible by 9 if, and only if, the sum of the digits of n are divisible by 9. □

Solving a linear equation such as $ax+by = c$ is very easy. For each possible real value of x we can compute the corresponding real value of y. But suppose now that a, b and c are integers and we only want to find solutions (x, y) whose co-ordinates are integers? This is an example of a *Diophantine equation*. We shall show how it may solved with the help of modular arithmetic. First, we shall show that the problem of finding integer solutions is equivalent to solving a simple kind of liner equation in one unknown in modular arithmetic.

Lemma 4.4.3. *Let a, b and c be integers. Then the following are equivalent.*

1. *The pair (x_1, y_1) is an integer solution to $ax + by = c$ for some y_1.*

2. *The integer x_1 is a solution to the equation $ax \equiv c \pmod{b}$.*

Proof. (1) $\Rightarrow$ (2). Suppose that $ax_1 + by_1 = c$. Then it is immediate that $ax_1 \equiv c \pmod{b}$.

(2) $\Rightarrow$ (1). Suppose that $ax_1 \equiv c \pmod{b}$. Then by definition, $ax_1 - c = bz_1$ for some integer z_1. Thus $ax_1 + b(-z_1) = c$. We may therefore put $y_1 = z_1$. □

We shall now describe how to solve all equations of the form

$$ax \equiv b \pmod{n}.$$

Lemma 4.4.4. *Consider the linear congruence* $ax \equiv b \pmod{n}$.

1. *The linear congruence has a solution if, and only if,* $d = \gcd(a, n)$ *is such that* $d \mid b$.

2. *If the condition in part (1) holds and* x_0 *is any solution, then all solutions have the form*
$$x = x_0 + t\frac{n}{d}$$
where $t \in \mathbb{Z}$.

Proof. (1). Suppose first that x_1 is a solution to our linear congruence. Then by definition, $ax_1 - b = nq$ for some integer q. It follows that $ax_1 + n(-q) = b$. By definition $d \mid a$ and $d \mid n$ and so $d \mid b$.

We now prove the converse. By Bézout's theorem, we may find integers u and v such that $au + nv = d$. By assumption, $d \mid b$ and so $b = dw$ for some integer w. It follows that $auw + nvw = dw = b$. Thus $a(uw) \equiv b \pmod{n}$, and we have found a solution.

(2) Let x_0 be any one solution to $ax \equiv b \pmod{n}$. It is routine to check that $x = x_0 + t\frac{n}{d}$ for any $t \in \mathbb{Z}$. Let x_1 be any solution to $ax \equiv b \pmod{n}$. Then $a(x_1 - x_0) \equiv 0 \pmod{n}$. Thus $a(x_1 - x_0) = tn$ for some integer t. The result now follows. □

There is a special case of the above result that is very important. Its proof is immediate.

Corollary 4.4.5. *Let* p *be a prime. Then the linear congruence* $ax \equiv b \pmod{p}$, *where* a *is not congruent to 0 modulo* p, *always has a solution, and all solutions are congruent modulo* p.

Example 4.4.6. Let's find all the points on the line $2x + 3y = 5$ that have integer co-ordinates. Observe first that $\gcd(2, 3) = 1$. Thus such points exist. In this case, by inspection, $1 = 2 \cdot 2 + (-1)3$. Thus $5 = 10 \cdot 2 + (-5)3$. It follows that $(10, -5)$ is one point on the line with integer co-ordinates. Thus the set of integer solutions is

$$\{(10 + 3t, -5 - 2t) \colon t \in \mathbb{Z}\}.$$

4.4.2 Wilson's theorem

I shall finish off this section with an application of congruences to primes. It is the first hint of hidden patterns in the primes. We need some notation first. For each natural number n define $n!$, pronounced n *factorial*, or if you are more extrovert n *shriek*, as follows: $0! = 1$ and for $n > 0$ define $n! = n \cdot (n-1)!$. In other words, $n!$ is what you get when you multiply together all the positive integers less than or equal to n. For each natural number n, we shall be interested in the value of $(n-1)!$ modulo n. Observe that there is no point in studying $n!$ (mod n) since the answer is always 0. It's worth doing some numerical calculations first to see if you can spot a pattern.

Theorem 4.4.7 (Wilson's Theorem)**.** *Let n be a natural number. Then n is a prime if, and only if,*

$$(n-1)! \equiv n-1 \pmod{n}$$

Since $n - 1 \equiv -1 \pmod{n}$ this is usually expressed in the form

$$(n-1)! \equiv -1 \pmod{n}.$$

Proof. The statement to be proved is an 'if, and only if' and so we have to prove two statements: (1) If n is prime then $(n-1)! \equiv n-1 \pmod{n}$. (2) If $(n-1)! \equiv n-1 \pmod{n}$ then n is prime.

We prove (1) first. Let n be a prime. The result is clearly true when $n = 2$ so we may assume n is an odd prime. For each $1 \leq a \leq n-1$ there is a unique number $1 \leq b \leq n-1$ such that $ab \equiv 1 \pmod{n}$. If $a = b$ then $a^1 \equiv 1 \pmod{n}$ which means that $n \mid (a-1)(a+1)$. Since n is a prime either $n \mid a-1$ or $a \mid a+1$. This can only occur if $a = 1$ or $a = n-1$. Thus $(n-1)! \equiv n-1 \pmod{n}$, as claimed.

We now prove (2). Suppose that $(n-1)! \equiv n-1 \pmod{n}$. We prove that n is a prime. Observe that when $n = 1$ we have that $(n-1)! = 1$ which is not congruent to 0 modulo 1. When $n = 4$, we get that $(4-1)! \equiv 2 \pmod{4}$. Suppose that $n > 4$ is not prime. Then $n = ab$ where $1 < a, b < n$. If $a \neq b$ then ab occurs as a factor of $(n-1)!$ and so this is congruent to 0 modulo n. If $a = b$ then a occurs in $(n-1)!$ and so does $2a$. Thus n is again a factor of $(n-1)!$. □

This theorem is interesting for another reason. To show that a number is prime, we would usually apply the algorithm we described earlier which

is just a systematic way of carrying out trial division. This theorem shows that a number is prime in a completely different way. Although it is not a pratical test for deciding whether a number is prime or composite, since $n!$ gets very big very quickly, it shows that there might be backdoor ways of showing that a number is prime. This is a very important question in the light of the rôle of prime numbers in cryptography.

4.5 Rational vs. irrational

In Chapter 3, I described the set of real numbers in quite abstract terms. In this section, I shall regard them from the point of view of decimal representations. This will lead to a new insight into the difference between rational and irrational numbers.

4.5.1 Irrationals

Real numbers are the actual values of quantities such as mass length and time. We cannot measure them exactly: the result of a measurement will always be a rational number. We begin by proving that there are real numbers that are not rationals. Recall the basic property of prime numbers: if a prime divides the product of two numbers then it must divide at least of the numbers. We use this property below. A real number which is not rational is said to be *irrational*. We may now generalize the result, obtained in Chapter 2, that $\sqrt{2}$ is irrational.

Theorem 4.5.1. *The square root of every prime number is irrational.*

Proof. We shall prove this by contradiction. Assume that we can write $\sqrt{p}$ as a rational. I shall show that this assumption leads to a contradiction and so must be false. We are assuming that $\sqrt{p} = \frac{a}{b}$. By cancelling the greatest common divisor of a and b we can in fact assume that $\gcd(a, b) = 1$. This will be crucial to our argument. Squaring both sides of the equation $\sqrt{p} = \frac{a}{b}$ and multiplying the resulting equation by b^2 we get that

$$pb^2 = a^2.$$

This says that a^2 is divisible by p. But if a prime divides a product of two numbers it must divide at least one of those numbers by Euclid's lemma.

Thus p divides a. Thus we can write $a = pc$ for some natural number c. Substituting this into our equation above we get that

$$pb^2 = p^2c^2.$$

Dividing both sides of this equation by p gives

$$b^2 = pc^2.$$

This tells us that b^2 is divisible by p and so in the same way as above p divides b. We have therefore shown that our assumption that $\sqrt{p}$ is rational leads to both a and b being divisible by p. But this contradicts the fact that $\gcd(a, b) = 1$. Our assumption is therefore wrong, and so $\sqrt{p}$ is not a rational number. □

Irrational numbers abound: both e and π can be proved to be irrational, for example. The discovery of irrational numbers is due to the Ancient Greeks and was one of the first great mathematical discoveries.

Although we cannot calculate irrational numbers exactly, we can calculate them to any degree of accuracy needed and it is by means of such approximations that irrational numbers are handled practically. For example, suppose we want to calculate $\sqrt{n}$ where n is not a perfect square. Make a first guess a to $\sqrt{n}$. Put $b = \frac{n}{a}$. Then their average $a' = \frac{a+b}{2}$ is in general a better guess. This process can be repeated, as the following example illustrates, and enables us to calculate square roots to any desired degree of accuracy.

Example 4.5.2. I shall calculate some approximations to $\sqrt{3}$ using the above method. We observe that $1^2 < 3 < 2^2$ so my first guess is 1. We have that $\frac{3}{1} = 3$ and the average of 1 and 3 is 2. I now start the process all over again with 2 as my guess. We have that $\frac{3}{2} = 1 \cdot 5$ and the average of 2 and $1 \cdot 5$ is $1 \cdot 75$. This is my new guess. The number 3 divided by $1 \cdot 75$ is $1 \cdot 714$ (approximately). The average of $1 \cdot 75$ and $1 \cdot 714$ is $1 \cdot 732$. My new guess is $1 \cdot 732$. 3 divided by $1 \cdot 732$ is $1 \cdot 732$ to 3 decimal places. Observe that $(1 \cdot 732)^2 = 2 \cdot 999\ldots$ which isn't bad.

4.5.2 Decimal fractions

I shall describe in this section the decimal fractions which correspond to rational numbers. To see what's involved, let's calculate some decimal fractions.

Examples 4.5.3.

(i) $\frac{1}{20} = 0 \cdot 05$. This fraction has a finite decimal representation.

(ii) $\frac{1}{7} = 0 \cdot 142857142857142857142857142857\ldots$. This fraction has an infinite decimal representation, which consists of the same sequence of numbers repeated. We abbreviate this decimal to $0 \cdot \overline{142857}$.

(iii) $\frac{37}{84} = 0 \cdot 44\overline{047619}$. This fraction has an infinite decimal representation, which consists of a non-repeating part followed by a part which repeats.

Case (ii) is said to be a *purely periodic* decimal whereas case (iii), which is more general, is said to be *ultimately periodic*.

Proposition 4.5.4. *A proper rational number $\frac{a}{b}$ in its lowest terms has a finite decimal expansion if and only if $b = 2^m 5^n$ for some natural numbers m and n.*

Proof. Let $\frac{a}{b}$ have the finite decimal representation $0 \cdot a_1 \ldots a_n$. This means

$$\frac{a}{b} = \frac{a_1}{10} + \frac{a_2}{10^2} + \ldots + \frac{a_n}{10^n}.$$

The righthand side is just the fraction

$$\frac{a_1 10^{n-1} + a_2 10^{n-2} + \ldots + a_n}{10^n}.$$

The denominator contains only the prime factors 2 and 5 and so the reduced form will also only contain at most the prime factors 2 and 5.

To prove the converse, consider the proper fraction

$$\frac{a}{2^\alpha 5^\beta}.$$

If $\alpha = \beta$ then the denominator is 10^α. If $\alpha \neq \beta$ then multiply the fraction by a suitable power of 2 or 5 as appropriate so that the resulting fraction has denominator a power of 10. But any fraction with denominator a power of 10 has a finite decimal expansion. □

Proposition 4.5.5. *An infinite decimal fraction represents a rational number if and only if it is ultimately periodic.*

Proof. Consider the ultimately periodic decimal number

$$r = 0 \cdot a_1 \dots a_s \overline{b_1 \dots b_t}.$$

We shall prove that r is rational. Observe that

$$10^s r = a_1 \dots a_s \cdot \overline{b_1 \dots b_t}$$

and

$$10^{s+t} = a_1 \dots a_s b_1 \dots b_t \cdot \overline{b_1 \dots b_t}.$$

From which we get that

$$10^{s+t} r - 10^s r = a_1 \dots a_s b_1 \dots b_t - a_1 \dots a_s$$

where the righthand side is the decimal form of some integer that we shall call a. It follows that

$$r = \frac{a}{10^{s+t} - 10^s}$$

is a rational number.

The proof of the converse is based on the method we use to compute the decimal expansion of $\frac{m}{n}$. We carry out repeated divisions by n and at each step of the computation we use the remainder obtained to calculate the next digit. But there are only a finite number of possible remainders and our expansion is assumed infinite. Thus at some point there must be repetition. □

Example 4.5.6. We shall write the ultimately periodic decimal $0 \cdot 9\bar{4}.$ as a proper fraction in its lowest terms. Put $r = 0 \cdot 9\bar{4}$. Then

- $r = 0 \cdot 9\bar{4}.$
- $10r = 9.444\dots$
- $100r = 94.444\dots.$

Thus $100r - 10r = 94 - 9 = 85$ and so $r = \frac{85}{90}$. We can simplify this to $r = \frac{17}{18}$. We can now easily check that this is correct.

Exercises 4.5

1. For each of the following fractions determine whether they have finite or infinite decimal representations. If they have infinite decimal representations determine whether they are purely periodic or ultimately periodic; in both cases determine the periodic block.

   (a) $\frac{1}{2}$.

   (b) $\frac{1}{3}$.

   (c) $\frac{1}{4}$.

   (d) $\frac{1}{5}$.

   (e) $\frac{1}{6}$.

   (f) $\frac{1}{7}$.

2. Write the following decimals as fractions in their lowest terms.

   (a) $0 \cdot 5\overline{34}$.

   (b) $0 \cdot 2\overline{106}$.

   (c) $0 \cdot \overline{076923}$.

4.6 Continued fractions

The goal of this section is to show how some of the ideas we have introduced so far can interact with each other. The material we cover is not needed elsewhere in this book.

4.6.1 Fractions of fractions

We return to an earlier calculation. We used Euclid's algorithm to calculate $\gcd(19, 7)$ as follows.

$$\begin{aligned} 19 &= 7 \cdot 2 + 5 \\ 7 &= 5 \cdot 1 + 2 \\ 5 &= 2 \cdot 2 + 1 \\ 2 &= 1 \cdot 2 + 0 \end{aligned}$$

We first rewrite each line, except the last, as follows

$$\begin{aligned} \frac{19}{7} &= 2+\frac{5}{2} \\ \frac{7}{5} &= 1+\frac{2}{5} \\ \frac{5}{2} &= 2+\frac{1}{2} \end{aligned}$$

Take the first equality

$$\frac{19}{7} = 2+\frac{5}{2}.$$

But $\frac{5}{7}$ is the reciprocal of $\frac{7}{5}$, and from the second equality

$$\frac{7}{5} = 1+\frac{2}{5}.$$

If we combine them, we get

$$\frac{19}{7} = 2+\frac{1}{1+\frac{2}{5}}$$

however strange this may look. We may repeat the process to get

$$\frac{19}{7} = 2+\frac{1}{1+\dfrac{1}{2+\frac{1}{2}}}$$

Fractions like this are called *continued fractions*. Suppose I just gave you

$$2+\frac{1}{1+\dfrac{1}{2+\frac{1}{2}}}$$

You could work out what the usual rational expression was by working from the bottom up. First compute the part in bold below

$$2+\frac{1}{1+\dfrac{1}{\mathbf{2+\frac{1}{2}}}}$$

to get

$$2 + \frac{1}{1 + \frac{1}{\frac{5}{2}}}$$

which simplifies to

$$2 + \frac{1}{1 + \frac{2}{5}}$$

This process can no be repeated and we shall eventually obtain a standard fraction.

I am not going to develop the theory of continued fractions, but I shall show you one more application. Let r be a real number. We may write r as $r = m_1 + r_1$ where $0 \leq r_1 < 1$. For example, π may be written as $\pi = 3 \cdot 14159265358\ldots$ where here $m = 3$ and $r_1 = 0 \cdot 14159265358\ldots$. Now since $r_1 < 1$ and assume that it is non-zero. Then $\frac{1}{r_1} > 1$. We may therefore repeat the above process and write $\frac{1}{r_1} = m_2 + r_2$ where once again $r_2 < 1$. This begin to feel an aweful lot like what we did above. In fact, we may write

$$r = m_1 + \frac{1}{m_2 + r_2},$$

and we can continue the above process with r_2. It looks like we would obtain a continued fraction representation of r with the big difference that it could be infinite. Here is a concrete example.

Example 4.6.1. We apply the above process to $\sqrt{3}$. Clearly, $1 < \sqrt{3} < 2$. Thus

$$\sqrt{3} = 1 + (\sqrt{3} - 1)$$

where $\sqrt{3} - 1 < 1$. We now focus on

$$\frac{1}{\sqrt{3} - 1}.$$

To convert this into a more usable form we multiple top and bottom by $\sqrt{3} + 1$. We therefore get that

$$\frac{1}{\sqrt{3} - 1} = \frac{1}{2}(\sqrt{3} + 1).$$

It is clear that $1 < \frac{1}{2}(\sqrt{3}+1) < 1\frac{1}{2}$. Thus

$$\frac{1}{\sqrt{3}-1} = 1 + \frac{\sqrt{3}-1}{2}.$$

We now focus on

$$\frac{2}{\sqrt{3}-1}$$

which simplifies to $\sqrt{3}+1$. Clearly

$$2 < \sqrt{3}+1 < 3.$$

Thus $\sqrt{3}+1 = 2+(\sqrt{3}-1)$. However, we have now gone full circle. Let's see what we have obtained. We have that

$$\sqrt{3} = 1 + \cfrac{1}{1+\cfrac{1}{2+(\sqrt{3}-1)}}.$$

However, we saw above that the pattern repeats as $\sqrt{3}-1$, so what we actually have is

$$\sqrt{3} = 1 + \cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{\cdots}}}}.$$

Let's see where we are by computing

$$1 + \cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1}}}$$

which simplifies to $\frac{7}{4}$. You can check that this is an approximation to $\sqrt{3}$.

4.6.2 Rabbits and pentagons

We now illustrate some of the ways that algebra and geometry may interact. We begin with an artificial looking question. In his book, *Liber Abaci*, Fibonacci raised the following little puzzle which I've taken from *MacTutor*:

> "A certain man put a pair of rabbits in a place surrounded on all sides by a wall. How many pairs of rabbits can be produced from that pair in a year if it is supposed that every month each pair begets a new pair which from the second month on becomes productive?"

These are obviously mathematical rabbits rather than real ones so let me spell out the rules more explicitly:

Rule 1 The problem begins with one pair of immature rabbits.[1]

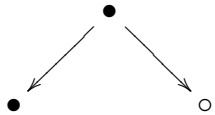
Rule 2 Each immature pair of rabbits takes one month to mature.

Rule 3 Each mature pair of rabbits produces a new immature pair at the end of a month.

Rule 4 The rabbits are immortal.

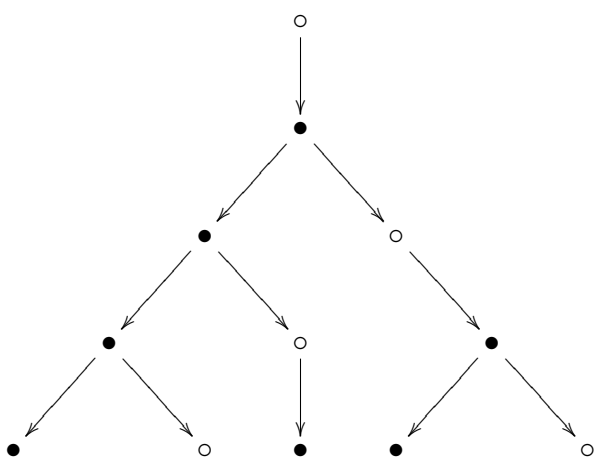
The important point is that we must solve the problem using the rules we have been given. To do this, I am going to draw a picture. I will represent an immature pair of rabbits by ∘ and a mature pair by •. Rule 2 will be represented by

and Rule 3 will be represented by



Rule 1 tells us that we start with ∘. Applying the rules we obtain the following picture for the first 4 months.

[1] Fibonacci himself seems to have assumed that the starting pair was already mature but we shan't.



We start with 1 pair and at the end of the first month we still have 1 pair, at the end of the second month 2 pairs, at the end of the third month 3 pairs, and at the end of the fourth month 5 pairs. I shall write this $F_0 = 1$, $F_1 = 1$, $F_2 = 2$, $F_3 = 3$, $F_4 = 5$, and so on. Thus the problem will be solved if we can compute F_{12}. There is an apparent pattern in the sequence of numbers $1, 1, 2, 3, 5, \ldots$ after the first two terms in the sequence each number is the sum of the previous two. Let's check that we are not just seeing things. Suppose that the number of immature pairs of rabbits at a given time t is I_t and the number of mature pairs is M_t. Then using our rules at time $t+1$ we have that $M_{t+1} = M_t + I_t$ and $I_{t+1} = M_t$. Thus

$$F_{t+1} = 2M_t + I_t.$$

Similarly

$$F_{t+2} = 3M_t + 2I_t.$$

It is now easy to check that

$$F_{t+2} = F_{t+1} + F_t.$$

The sequence of numbers such that $F_0 = 1$, $F_1 = 1$ and satisfying the rule $F_{t+2} = F_{t+1} + F_t$ is called the *Fibonacci sequence*. We have that

$$F_0 = 1, F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, F_5 = 8, F_6 = 13, F_7 = 21,$$

$$F_8 = 34, F_9 = 55, F_{10} = 89, F_{11} = 144, F_{12} = 233.$$

The solution to the original question is therefore 233 pairs of rabbits.

Fibonacci numbers arise in the most diverse situations: famously, in *phyllotaxis* which is the study of how leaves and petals are arranged on plants. We shall now look for a formula that will enable us to calculate F_n directly. To begin, we'll follow an idea due to the astronomer Jonannes Kepler, and look at the behaviour of the fractions $\frac{F_{n+1}}{F_n}$ as n gets bigger and bigger. I have tabulated some calculations below.

$\frac{F_1}{F_0}$	$\frac{F_2}{F_1}$	$\frac{F_3}{F_2}$	$\frac{F_4}{F_3}$	$\frac{F_5}{F_4}$	$\frac{F_6}{F_5}$	$\frac{F_7}{F_6}$	$\frac{F_{14}}{F_{13}}$
1	2	$1 \cdot 5$	$1 \cdot 6$	$1 \cdot 625$	$1 \cdot 615$	$1 \cdot 619$	$1 \cdot 6180$

These ratios seem to be going somewhere; the question is: where? Notice that

$$\frac{F_{n+1}}{F_n} = \frac{F_n + F_{n-1}}{F_n} = 1 + \frac{F_{n-1}}{F_n} = 1 + \frac{1}{\frac{F_n}{F_{n-1}}}.$$

But for very large n we suspect that $\frac{F_{n+1}}{F_n}$ and $\frac{F_n}{F_{n-1}}$ will be almost the same. This suggests, but doesn't prove, that we need to find the positive solution x to

$$x = 1 + \frac{1}{x}.$$

Thus x is a number that when you take its reciprocal and add 1 you get x back again. This problem is really a quadratic equation in disguise

$$x^2 = x + 1 \text{ or more usually } x^2 - x - 1 = 0.$$

This equation can be solved very simply to give us

$$x = \frac{1 \pm \sqrt{5}}{2}.$$

That is

$$\phi = \frac{1 + \sqrt{5}}{2} \text{ and } \bar{\phi} = \frac{1 - \sqrt{5}}{2}.$$

The number ϕ is called *the golden ratio*, about which a deal of nonsense has been written. Let's go back and see if this calculation makes sense. First we calculate ϕ and we get

$$\phi = 1 \cdot 618033988 \ldots$$

I compute

$$\frac{F_{19}}{F_{18}} = \frac{6765}{4181} = 1 \cdot 618033963$$

on my pocket calculator. This is pretty close.

We can now get our formula for the Fibonacci numbers. Define

$$f_n = \frac{1}{\sqrt{5}}\left(\phi^{n+1} - \bar{\phi}^{n+1}\right).$$

I'm going to show you that $F_n = f_n$. To do this, I'll use the following identities which are straightforward to check

$$\phi - \bar{\phi} = \sqrt{5} \quad \phi^2 = \phi + 1 \text{ and } \bar{\phi}^2 = \bar{\phi} + 1.$$

Let's start with f_0. We know that

$$\phi - \bar{\phi} = \sqrt{5}$$

and so we really do have that $f_0 = 1$. To calculate f_1 we use the other formulae and again we get $f_1 = 1$. We now calculate $f_n + f_{n+1}$ we get

$$\begin{aligned} f_n + f_{n+1} &= \frac{1}{\sqrt{5}}\left(\phi^{n+1} - \bar{\phi}^{n+1}\right) + \frac{1}{\sqrt{5}}\left(\phi^{n+2} - \bar{\phi}^{n+2}\right) \\ &= \frac{1}{\sqrt{5}}\left(\phi^{n+1} + \phi^{n+2} - (\bar{\phi}^{n+1} + \bar{\phi}^{n+2})\right) \\ &= \frac{1}{\sqrt{5}}\left(\phi^{n+1}(1+\phi) - \bar{\phi}^{n+1}(1+\bar{\phi})\right) \\ &= \frac{1}{\sqrt{5}}\left(\phi^{n+1}\phi^2 - \bar{\phi}^{n+1}\bar{\phi}^2\right) \\ &= \frac{1}{\sqrt{5}}\left(\phi^{n+3} - \bar{\phi}^{n+3}\right) = f_{n+2} \end{aligned}$$

Because f_n and F_n start in the same place and satisfy the same rules, we have therefore proved that

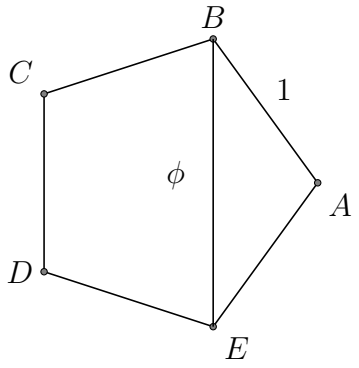
$$\boxed{F_n = \tfrac{1}{\sqrt{5}}\left(\phi^{n+1} - \bar{\phi}^{n+1}\right).}$$

At this point, we can go back and verify our original idea that the fractions $\frac{F_{n+1}}{F_n}$ seem to get closer and closer to ϕ as n gets larger and larger. We have that

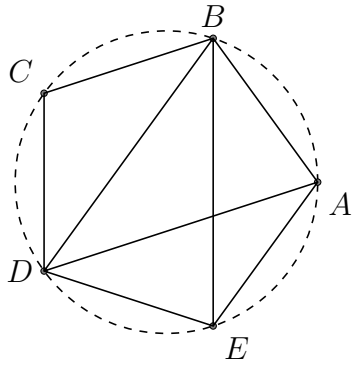
$$\begin{aligned} \frac{F_{n+1}}{F_n} &= \frac{\phi^{n+2} - \bar{\phi}^{n+2}}{\phi^{n+1} - \bar{\phi}^{n+1}} \\ &= \frac{\phi}{1 - (\frac{\bar{\phi}}{\phi})^{n+1}} - \frac{1}{\frac{1}{\phi}(\frac{\phi}{\bar{\phi}})^{n+1} - \frac{1}{\phi}} \end{aligned}$$

I have rewritten it like this so that we can see what happens as n gets larger and larger. Observe that the absolute value of $\frac{\bar{\phi}}{\phi}$ is less than 1. So as n gets larger and larger the first term above gets closer and closer to ϕ. Now look at the second term. The absolute value of the fraction $\frac{\phi}{\phi}$ is strictly greater than 1. Thus as n gets larger and larger the denominator of the second term gets larger and larger and so the fraction as a whole gets smaller and smaller. Thus we have proved that $\frac{F_{n+1}}{F_n}$ really is close to ϕ when n is large.

So far, what we have been doing is algebra. I shall now show that there is geometry here as well. Below is a picture of a regular pentagon. I have assumed that the length of the sides is 1. I claim that the length of a diagonal, such as BE, is equal to ϕ.



To prove this I am going to use Ptolomy's theorem. We shall concentrate on the cyclic quadrilateral formed by the vertices $ABDE$.

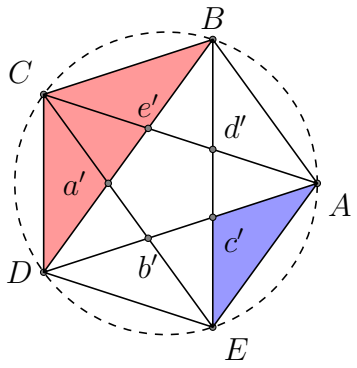


I'll let the side of a diagonal be x. Then by Ptolomy's theorem, we have that

$$x^2 = 1 + x.$$

But this is precisely the quadratic equation we solved above. Its positive solution is ϕ and so the length of a diagonal of a regular pentagon with side 1 is ϕ.

This raises the question of whether we can somehow see the Fibonacci numbers in the regular pentagon. The answer is: almost. Consider the diagram below.



I've drawn in all the diagonals. The shaded triangle BCD is similar to the shaded triangle $Ac'E$. This means that they have exactly the same shapes just different sizes. It follows that

$$\frac{Ac'}{AE} = \frac{BC}{BD}.$$

But AE is a side of the pentagon and so has unit length, and BD is of length ϕ. Thus

$$AC' = \frac{1}{\phi}.$$

Now, Dc' has the same length as BC which is a side of the pentagon. Thus $Dc' = 1$. We now have

$$\phi = DA = Dc' + c'A = 1 + \frac{1}{\phi}.$$

Thus, just from geometry, we get

$$\phi = 1 + \frac{1}{\phi}.$$

This is a very odd equation because ϕ is mentioned on both sides. Let's go with it and repeat:

$$\phi = 1 + \frac{1}{1 + \frac{1}{\phi}}$$

and

$$\phi = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{\phi}}}$$

and

$$\phi = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{\phi}}}}$$

and so on. We therefore obtain a continued fraction. For each of these fractions cover up the term $\frac{1}{\phi}$ and then calculate what you see to get

$$1, \quad 1 + \frac{1}{1} = 2, \quad 1 + \frac{1}{1 + \frac{1}{1}} = \frac{3}{2}, \quad 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1}}} = \frac{5}{3}, \ldots$$

and the Fibonacci sequence reappears.