1
Squares

I met a man once who told me that, far from believing in the square root of minus one, he didn't even believe in minus one.

E. C. Titchmarsh (1899–1963)

1.1 Square and Triangular Numbers

Everyone knows the whole numbers, also called the *positive integers* or the *natural numbers*,

$$1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12,\ 13,\ 14,\ \ldots\ .$$

Young children often ask, "What is the biggest number?" and we have to answer, "There is no biggest number," since the sequence of positive integers goes on for ever. We say that the sequence of positive integers is *infinite*; this is what is meant by the row of dots after the number 14 above. Every positive integer is either *even* or *odd*. The even integers are those that are divisible by 2,

$$2,\ 4,\ 6,\ 8,\ 10,\ 12,\ 14,\ 16,\ 18,\ 20,\ 22,\ 24,\ 26,\ \ldots\ ,$$

and the odd integers are those that are *not* divisible by 2,

$$1,\ 3,\ 5,\ 7,\ 9,\ 11,\ 13,\ 15,\ 17,\ 19,\ 21,\ 23,\ 25,\ \ldots\ .$$

The number 2 is the smallest of the *prime numbers*, which we now define.

Definition 1.1.1 A prime number is a positive integer greater than 1 that is divisible only by 1 and itself. A number greater than 1 that is not prime is called *composite*. ■

The sequence of prime numbers begins 2, 3, 5, 7, 11, and I will have more to say about the primes later in this book.

Perhaps you recognize one or both of the following infinite sequences:

$$1,\ 4,\ 9,\ 16,\ 25,\ 36,\ 49,\ 64,\ 81,\ 100,\ 121,\ 144,\ \ldots\ , \tag{1.1}$$

$$1,\ 3,\ 6,\ 10,\ 15,\ 21,\ 28,\ 36,\ 45,\ 55,\ 66,\ 78,\ 91,\ \ldots\ . \tag{1.2}$$

The sequence (1.1) is the sequence of *squares*. Each square number can be depicted by an array of nodes arranged in a square formation, as in Figure 1.1, which shows the first six squares, 1, 4, 9, 16, 25, and 36.

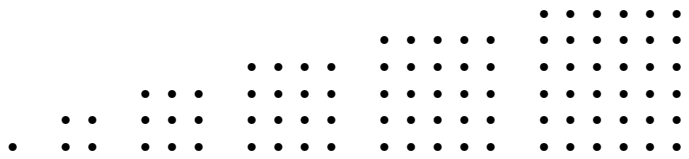


FIGURE 1.1. The first six squares, 1, 4, 9, 16, 25, and 36.

The sequence (1.2) is the sequence of *triangular* numbers. The first triangular number is 1, and the next three are

$$1 + 2 = 3, \quad 1 + 2 + 3 = 6, \quad 1 + 2 + 3 + 4 = 10.$$

The first six members of this sequence are displayed as sets of nodes in triangular formation in Figure 1.2.

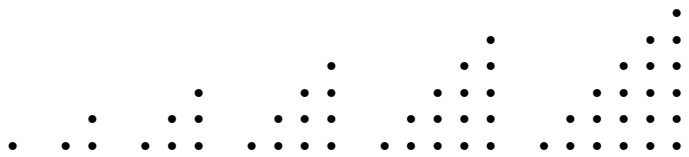


FIGURE 1.2. The first six triangular numbers, 1, 3, 6, 10, 15, and 21.

There is a famous story concerning triangular numbers and C. F. Gauss (1777–1855), who is generally regarded as one of the finest mathematicians of all time. When Gauss was in the early years of school his teacher decided to set the class an exercise that would keep them out of mischief for some time. The problem posed to the young children was to find the sum of the first hundred positive integers. To the teacher's astonishment, Carl Gauss almost immediately gave the correct answer, 5050. How did he do it? Gauss

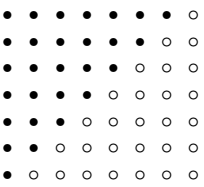


FIGURE 1.3. A geometrical interpretation of the equation $2T_n = n(n+1)$.

saw that the first and last numbers in the sum, 1 and 100, add up to 101, and also that the second and the second to last numbers, 2 and 99, add up to 101, and so on. It is easier to follow Gauss's argument if we write out the sum

$$T = 1 + 2 + 3 + \cdots + 98 + 99 + 100$$

and reverse the order of the numbers to give also

$$T = 100 + 99 + 98 + \cdots + 3 + 2 + 1.$$

I have called the sum T to remind us that it is a triangular number, and have used three dots in the latter two equations to denote the sum of the missing numbers, from 4 to 97. We can now add corresponding terms in the above two sums to give

$$2T = 101 + 101 + 101 + \cdots + 101 + 101 + 101.$$

Since there are a hundred terms in the latter sum, we see that

$$2T = 100 \times 101$$

and so

$$T = 50 \times 101 = 5050,$$

as Gauss obtained. Thus the 100th triangular number is 5050, and we can now follow Gauss's method to evaluate *any* triangular number. For if we write T_n to denote the nth triangular number, we have

$$T_n = 1 + 2 + 3 + \cdots + (n-2) + (n-1) + n,$$

and we can reverse the order of the terms in the above sum to give also

$$T_n = n + (n-1) + (n-2) + \cdots + 3 + 2 + 1.$$

Notice that if we replace n by 100 we obtain the two sums we had above for T. On adding corresponding terms in the above two sums for T_n, we obtain

$$2T_n = (n+1) + (n+1) + (n+1) + \cdots + (n+1) + (n+1) + (n+1),$$

and since there are n terms in the latter sum, we see that

$$2T_n = n(n+1),$$

meaning n multiplied by $n+1$. A geometrical interpretation of this equation is given in Figure 1.3, which shows that $2T_7 = 7 \times 8$. On dividing $2T_n$ by 2, we find that the triangular number T_n can be written in the form

$$T_n = \frac{1}{2}n(n+1). \tag{1.3}$$

As a check, we can evaluate this expression for some values of n, and we find, for example, that $T_3 = 6$, $T_5 = 15$, and $T_{100} = 5050$, in agreement with what we found above.

There is a special notation for handling sums. Suppose we have a sequence of numbers that begins with a_1, a_2, and a_3. We write the sum of the first n members of this sequence as

$$a_1 + a_2 + \cdots + a_n = \sum_{r=1}^{n} a_r. \tag{1.4}$$

The symbol $\sum$ is the uppercase letter *sigma*, the eighteenth letter of the classical Greek alphabet, and it stands for *sum*. We read the symbols on the right of the above equation as "sum a_r over all integer values of r from 1 to n." For example, we can express the nth triangular number as

$$T_n = \sum_{r=1}^{n} r.$$

There is a very simple connection between the triangular numbers and the squares, which follows immediately from Figure 1.4, where we see that every square is the sum of two *consecutive* triangular numbers. If we write S_n to denote n^2, the nth square, we have

$$S_n = T_{n-1} + T_n. \tag{1.5}$$

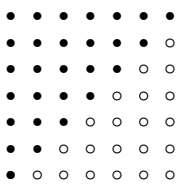


FIGURE 1.4. Every square is the sum of two consecutive triangular numbers.

This relation, which we have obtained by using a geometrical argument, can be verified algebraically. For we saw above that $T_n = \frac{1}{2}n(n+1)$, and so $T_{n-1} = \frac{1}{2}(n-1)n$. Thus we may write

$$T_{n-1} + T_n = \frac{1}{2}(n-1)n + \frac{1}{2}n(n+1) = \frac{1}{2}n\Big((n-1) + (n+1)\Big).$$

Since $(n-1) + (n+1) = 2n$, the above calculation shows that

$$T_{n-1} + T_n = n^2 = S_n,$$

which verifies the result displayed geometrically in Figure 1.4.

Having summed the first n positive integers, can we find the sum of the first n squares? This would settle the problem of determining how many oranges there are in a pyramid containing a square array of n^2 oranges on the bottom layer, an array of $(n-1)^2$ oranges on the second layer, and so on, with a single orange at the apex of the pyramid. Can you visualize how each layer of oranges nestles into the spaces between the oranges in the layer below? The similar problem of finding the sum of the first n triangular numbers, called the nth *tetrahedral number*, is equivalent to finding the number of oranges arranged in *triangular* layers of diminishing size. Let us imagine the oranges in each layer arranged in the form of an equilateral triangle (that is, with all three sides equal), rather than a right-angled triangle, as in Figure 1.2. The whole triangular configuration is called a triangular pyramid or a *tetrahedron*, which means "four faces" in Greek. Let us find the sum of the first n triangular numbers, and use that result to find the sum of the first n squares. We begin by writing

$$r(r+1)(r+2) - (r-1)r(r+1) = r(r+1)\Big((r+2) - (r-1)\Big).$$

If we divide the above equation throughout by 6 and simplify the right side, we find that

$$\frac{1}{6}r(r+1)(r+2) - \frac{1}{6}(r-1)r(r+1) = \frac{1}{2}r(r+1), \tag{1.6}$$

and we observe that the number on the right of (1.6) is the rth triangular number. Our next move is to sum each number in (1.6) over r, from $r = 1$ to n, giving

$$\sum_{r=1}^{n} \frac{1}{6}r(r+1)(r+2) - \sum_{r=1}^{n} \frac{1}{6}(r-1)r(r+1) = \sum_{r=1}^{n} \frac{1}{2}r(r+1). \tag{1.7}$$

At first sight, it may look as if we have made things worse, by expressing the sum of the first n triangular numbers in a more complicated way. But if we look more carefully at the left side of (1.7), we see that we are very close to the solution. For on putting $r = 1$ in the first sum on the left of (1.7),

and $r = 2$ in the second sum, we find that these terms cancel. Likewise, the second term of the first sum cancels with the third term of the second sum, and so on. All that remains is the last term of the first sum and the first term of the second sum. Since the first term of the second sum is zero, we can express the nth tetrahedral number as

$$\sum_{r=1}^{n} T_r = \sum_{r=1}^{n} \frac{1}{2} r(r+1) = \frac{1}{6} n(n+1)(n+2). \tag{1.8}$$

As we saw in Figure 1.4,

$$S_r = T_{r-1} + T_r, \tag{1.9}$$

and we can use (1.9) and (1.8) to derive an expression for the sum of the first n squares. We have defined S_r and T_r for all positive integers r. We now define $T_0 = 0$, and then (1.9) will hold for all positive integers r. If we sum each number in (1.9) over r, from $r = 1$ to n, we obtain

$$\sum_{r=1}^{n} S_r = \sum_{r=1}^{n} T_{r-1} + \sum_{r=1}^{n} T_r.$$

Since

$$\sum_{r=1}^{n} T_{r-1} = \sum_{r=1}^{n-1} T_r,$$

we can use (1.8) to give

$$\sum_{r=1}^{n} S_r = \sum_{r=1}^{n-1} T_r + \sum_{r=1}^{n} T_r = \frac{1}{6}(n-1)n(n+1) + \frac{1}{6} n(n+1)(n+2).$$

We simplify the right side of the last equation, writing

$$\frac{1}{6}(n-1)n(n+1) + \frac{1}{6} n(n+1)(n+2) = \frac{1}{6} n(n+1)\Big((n-1) + (n+2)\Big),$$

and so obtain an expression for the sum of the first n squares,

$$\sum_{r=1}^{n} S_r = \frac{1}{6} n(n+1)(2n+1). \tag{1.10}$$

Above, we used a geometrical argument to show how a square can be expressed as the sum of two consecutive triangular numbers. Figure 1.5 shows another way of expressing a square as a sum. We begin with an $n \times n$ array of nodes and remove the L-shape of $2n - 1$ nodes consisting of those in the first column and the last row, leaving an $(n-1) \times (n-1)$ square. This geometrical observation corresponds to the algebraic equation

$$n^2 - (n-1)^2 = 2n - 1.$$

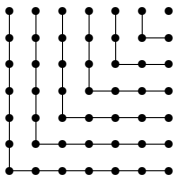


FIGURE 1.5. Every square is the sum of consecutive odd numbers.

We then remove an L-shape of $2n-3$ nodes from the $(n-1)\times(n-1)$ square to leave an $(n-2)\times(n-2)$ square. After removing $n-1$ L-shapes, only one node remains, showing that

$$n^2 = 1+3+5+\ldots+(2n-3)+(2n-1) = \sum_{r=1}^{n}(2r-1). \qquad (1.11)$$

This shows that n^2 is the sum of the first n odd numbers.

Mathematicians have been interested in square numbers at least since the development of Babylonian mathematics in the second millennium BC, between three and four thousand years ago. The number system used by the Babylonians was not very suitable for carrying out calculations, and perhaps this is why they devised the very clever method of multiplying two numbers that I will now describe.

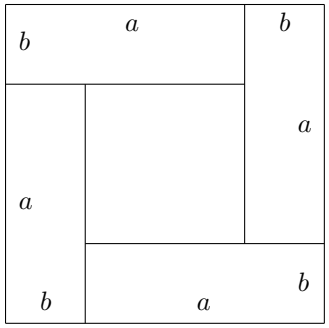


FIGURE 1.6. Geometrical interpretation of $(a+b)^2 = (a-b)^2 + 4ab$.

We write $a > b$ to mean a is greater than b, and write $a < b$ to mean a is less than b. We also write $a \geq b$ to mean a is greater than or equal to b, and then obviously $a \leq b$ means that a is less than or equal to b. Figure 1.6 consists of a large square of area $(a+b)^2$ that is subdivided into four

identical rectangles of area ab and a small square of area $(a-b)^2$, where $a > b$. Thus we have

$$(a+b)^2 = (a-b)^2 + 4ab. \tag{1.12}$$

It follows from its geometrical derivation that this equation holds for all positive real values of a and b, with $a \geq b$. If $a = b$, the smaller square shrinks to zero, and the equation becomes $(a+a)^2 = 4a^2$. Can you see what happens to Figure 1.6 when $a = b$? We now make use of (1.12) by writing

$$x = \frac{1}{2}(a+b) \qquad \text{and} \qquad y = \frac{1}{2}(a-b). \tag{1.13}$$

Then it follows from (1.13) and (1.12) that

$$x^2 - y^2 = \frac{1}{4}(a+b)^2 - \frac{1}{4}(a-b)^2 = ab,$$

so that

$$ab = x^2 - y^2. \tag{1.14}$$

Equations (1.13) and (1.14) show how Babylonian mathematicians were able to multiply two numbers by using a table of squares. Suppose we have a table of squares of all positive integers from 1 to 10,000. Then, to multiply any two numbers $a > b$ of up to 4 decimal digits, we need only compute x and y from (1.13), look up their squares in the table, and then find ab from (1.14). We would need to modify this process if $a + b$ happens to be an odd number, because then $\frac{1}{2}(a+b)$ is not an integer. This happens only when one of the numbers a and b is even and the other is odd, and then $a+b$ and $a-b$ are both odd. (Problems 1.1.6 and 1.1.7 show how we can deal with this case.) If a and b both lie between 1 and 10,000, then since $a \times b = b \times a$, there are $T_{10,000} = 50{,}005{,}000$ possible multiplications. Thus the Babylonian method of multiplying two numbers would make a table of 10,000 squares equivalent to a gigantic multiplication table with more than fifty million entries.

Example 1.1.1 Let us evaluate 53×89 using the Babylonian method. Using (1.13) we compute

$$x = \frac{1}{2}(89+53) = 71 \qquad \text{and} \qquad y = \frac{1}{2}(89-53) = 18.$$

Then, using Table 1.1, we obtain from (1.14) that

$$89 \times 53 = 71^2 - 18^2 = 5041 - 324 = 4717. \qquad \blacksquare$$

In earlier eras, the need to perform multiplications arose in agrarian societies with the requirement to measure land areas, estimate amounts of agricultural products, and so on. There was little practical need for calculations other than simple additions, subtractions, multiplications, and

1	1	26	676	51	2601	76	5776
2	4	27	729	52	2704	77	5929
3	9	28	784	53	2809	78	6084
4	16	29	841	54	2916	79	6241
5	25	30	900	55	3025	80	6400
6	36	31	961	56	3136	81	6561
7	49	32	1024	57	3249	82	6724
8	64	33	1089	58	3364	83	6889
9	81	34	1156	59	3481	84	7056
10	100	35	1225	60	3600	85	7225
11	121	36	1296	61	3721	86	7396
12	144	37	1369	62	3844	87	7569
13	169	38	1444	63	3969	88	7744
14	196	39	1521	64	4096	89	7921
15	225	40	1600	65	4225	90	8100
16	256	41	1681	66	4356	91	8281
17	289	42	1764	67	4489	92	8464
18	324	43	1849	68	4624	93	8649
19	361	44	1936	69	4761	94	8836
20	400	45	2025	70	4900	95	9025
21	441	46	2116	71	5041	96	9216
22	484	47	2209	72	5184	97	9409
23	529	48	2304	73	5329	98	9604
24	576	49	2401	74	5476	99	9801
25	625	50	2500	75	5625	100	10000

TABLE 1.1. A table of squares.

divisions. However, even from the time of ancient Babylon, Greece, and China, mathematicians showed a sophisticated interest in the concept of number that was driven by sheer mathematical curiosity rather than commercial utility, and we will discuss some of these things later in this book. The growth of scientific knowledge, which greatly accelerated from the seventeenth century onwards, created a quantum leap in mankind's need to calculate, and encouraged the discovery of more efficient ways of carrying out arithmetical operations. In particular, logarithms were invented in the early years of the seventeenth century, and much effort went into the construction of logarithm tables. These were in constant use for carrying out multiplications and other calculations until the latter decades of the twentieth century, when the widespread availability of cheap electronic calculators made logarithm tables obsolete.

Although we can appreciate its ingenuity, and enjoy using it, the Babylonian method of multiplication using a table of squares is not necessary in our familiar decimal system, because calculations like those in Example 1.1.1 are easily carried out using "long multiplication." Let us work through such a calculation, for the sake of clarity. For example, we set out the evaluation of the product 53×89 in the form

$$\begin{array}{r} 89 \\ \times 53 \\ \hline 267 \\ 445 \\ \hline 4717 \end{array}$$

In the above calculation we first multiply 89 by 3, the least-significant digit of 53, to give 267, and write it down. Then we multiply 89 by 5, the other digit of 53, to give 445, and write it down. We write 445 one place to the left of 267, because we are really multiplying 89 by 50 rather than 5. Finally, we add the last two numbers to give the result, which is 4717. To carry out such calculations, we need to know how to multiply two numbers a and b, where a and b are between 0 and 9, and add any carry digit. We also need to know how to write down the intermediate results, as in the above example, to take account of the decimal place.

The Roman numeral system is even more complicated than that of the Babylonians. Table 1.2 gives the Roman numerals that correspond to certain numbers in the system most familiar to us, which is called the Arabic or Hindu–Arabic system. Note the use of IV for 4, which is 5 minus 1, while VI denotes 6. Similarly, we write IX for 9, and XI for 11, XC for 90, and CX for 110, and so on. Table 1.2 gives us the basis for writing down all positive integers in Roman numerals up to a thousand, which is denoted by M. All other numbers are written down by combining those in the table in an obvious way. For example, we write XXIV for 24, CCCLV for 355, and MDCCCLXXXVIII for 1888. But if we contemplate how we would

Arabic	1	2	3	4	5	6	7	8	9
Roman	I	II	III	IV	V	VI	VII	VIII	IX
Arabic	10	20	30	40	50	60	70	80	90
Roman	X	XX	XXX	XL	L	LX	LXX	LXXX	XC
Arabic	100	200	300	400	500	600	700	800	900
Roman	C	CC	CCC	CD	D	DC	DCC	DCCC	CM

TABLE 1.2. The Arabic and Roman numeral systems.

multiply XLVII by LXXIII, for example, we realize that the Roman system is not convenient for doing arithmetic.

Before leaving the topic of multiplication, let us consider a method based on a table of triangular numbers instead of a table of squares. Consider Figure 1.7, which we can view as a rectangular array of 8×5 nodes together with a triangular array of $T_3 = 6$ nodes above it. The rectangular array of 40 nodes is split into a triangular array of $T_4 = 10$ nodes and a trapezoidal array of $30 = 4 + 5 + 6 + 7 + 8 = T_8 - T_3$ nodes, a trapezoidal figure being a quadrilateral (a four-sided figure) with only one pair of sides parallel.

In general, we can see that for any two positive integers $m \geq n$, we can split the rectangular array of nodes arranged in m columns and n rows into a triangular array of T_{n-1} nodes and a trapezoidal array containing $T_m - T_{m-n}$ nodes. Thus we have the remarkable result that

$$mn = T_m + T_{n-1} - T_{m-n}. \tag{1.15}$$

Notice how this equation agrees with (1.5) in the special case $m = n$.

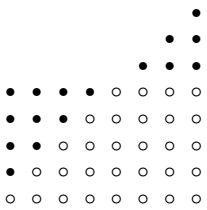


FIGURE 1.7. This diagram illustrates the equation $mn = T_m + T_{n-1} - T_{m-n}$, where $m = 8$ denotes the number of columns of nodes, and $n = 5$.

Example 1.1.2 Using (1.15) and Table 1.3 we find that

$$89 \times 53 = T_{89} + T_{52} - T_{36} = 4005 + 1378 - 666 = 4717$$

and

$$94 \times 38 = T_{94} + T_{37} - T_{56} = 4465 + 703 - 1596 = 3572. \qquad \blacksquare$$

1	1	26	351	51	1326	76	2926
2	3	27	378	52	1378	77	3003
3	6	28	406	53	1431	78	3081
4	10	29	435	54	1485	79	3160
5	15	30	465	55	1540	80	3240
6	21	31	496	56	1596	81	3321
7	28	32	528	57	1653	82	3403
8	36	33	561	58	1711	83	3486
9	45	34	595	59	1770	84	3570
10	55	35	630	60	1830	85	3655
11	66	36	666	61	1891	86	3741
12	78	37	703	62	1953	87	3828
13	91	38	741	63	2016	88	3916
14	105	39	780	64	2080	89	4005
15	120	40	820	65	2145	90	4095
16	136	41	861	66	2211	91	4186
17	153	42	903	67	2278	92	4278
18	171	43	946	68	2346	93	4371
19	190	44	990	69	2415	94	4465
20	210	45	1035	70	2485	95	4560
21	231	46	1081	71	2556	96	4656
22	253	47	1128	72	2628	97	4753
23	276	48	1176	73	2701	98	4851
24	300	49	1225	74	2775	99	4950
25	325	50	1275	75	2850	100	5050

TABLE 1.3. A table of triangular numbers.

The number 1 is both a square and a triangular number, and an inspection of Tables 1.1 and 1.3 shows that 36 and 1225 also have this property. Such numbers correspond to solutions of the equation

$$n(n+1) = 2m^2, \tag{1.16}$$

as we see on dividing by 2. Equations for which we seek solutions in integers are called Diophantine equations, named after Diophantus of Alexandria, who lived in the third century AD, in the latter part of the glorious era of ancient Greek mathematics, which flourished for about a thousand years. Diophantine equations can be very difficult to solve. Even when we believe that a given Diophantine equation has *no* solution, this is often difficult to prove. The most famous, or notorious, Diophantine equation is that associated with Pierre de Fermat (1601–1665), to which (see (1.44)) we refer in Section 1.3. Happily, equation (1.16) is one for which we *can* find solutions. Let

$$\alpha = \left(1+\sqrt{2}\right)^2 = 3 + 2\sqrt{2}, \tag{1.17}$$

where α is *alpha*, the first letter of the Greek alphabet. (The English word alphabet is derived from alpha and *beta*, the second letter of the Greek alphabet, which is written β.) We can show that equation (1.16) is satisfied by

$$m = m_k = \frac{1}{4\sqrt{2}}\left(\alpha^k - \alpha^{-k}\right), \tag{1.18}$$

where α^k means α multiplied by itself k times and $\alpha^{-k} = 1/\alpha^k$, and

$$n = n_k = \frac{1}{4}\left(\alpha^k + \alpha^{-k} - 2\right), \tag{1.19}$$

for every positive integer k. (See Problems 3.2.8 and 3.2.9 for a derivation of (1.18) and (1.19).) We will verify below that $m = m_k$ and $n = n_k$ are indeed positive integers, and that they satisfy (1.16). But first let us see why we might expect the presence of the quantity $\sqrt{2}$ in the above equations. We have from (1.16) that

$$2 = \frac{n(n+1)}{m^2} = \left(1 + \frac{1}{n}\right)\frac{n^2}{m^2}.$$

This shows that

$$\frac{2m^2}{n^2} = 1 + \frac{1}{n},$$

and we see that $2m^2/n^2$ tends to the limit 1 as n tends to infinity. Hence $n^2 \approx 2m^2$, where the symbol $\approx$ means "approximately equals," and so $n \approx \sqrt{2}\,m$ for large values of m and n.

Now let us replace n by $n+1$ in (1.19) to give

$$n+1 = \frac{1}{4}\left(\alpha^k + \alpha^{-k} + 2\right),$$

and so

$$n(n+1) = \frac{1}{16}\left(\alpha^k + \alpha^{-k} - 2\right)\left(\alpha^k + \alpha^{-k} + 2\right) = \frac{1}{16}\left(\left(\alpha^k + \alpha^{-k}\right)^2 - 4\right).$$

This simplifies to give

$$n(n+1) = \frac{1}{16}\left(\alpha^{2k} + \alpha^{-2k} - 2\right).$$

We find from (1.18) that

$$2m^2 = \frac{1}{16}\left(\alpha^{2k} + \alpha^{-2k} - 2\right),$$

which shows that m and n defined by (1.18) and (1.19) do indeed satisfy $2m^2 = n(n+1)$, for all positive integers k. Finally, we need to show that these values of m and n are positive integers, and so justify our claim that there is an infinite number of squares that are also triangular numbers. One way of doing this is to begin by evaluating (1.18) and (1.19) with $k = 1$ and $k = 2$ to show that

$$m_1 = 1,\ m_2 = 6 \qquad \text{and} \qquad n_1 = 1,\ n_2 = 8. \tag{1.20}$$

In verifying (1.20) it is helpful to observe that

$$\left(3 + 2\sqrt{2}\right)\left(3 - 2\sqrt{2}\right) = 9 - 8 = 1,$$

and hence

$$\frac{1}{\alpha} = 3 - 2\sqrt{2}. \tag{1.21}$$

Then we show (see Problem 1.1.10) that

$$m_{k+1} = 6m_k - m_{k-1}, \quad k \geq 2, \tag{1.22}$$

and

$$n_{k+1} = 6n_k - n_{k-1} + 2, \quad k \geq 2. \tag{1.23}$$

We can compute as many values of m_k and n_k as we wish from (1.22) and (1.23), which are called *recurrence relations*, and we can see that m_k and n_k are indeed all positive integers. It follows from the recurrence relations that $m_3 = 35$ and $n_3 = 49$, and

$$35^2 = \frac{1}{2}49 \cdot 50 = 1225,$$

which, as we found above, is the third number that is both a square and a triangular number. Although we have not proved it here, it can be shown that the only numbers that are both squares and triangular numbers are those obtained via the above recurrence relations. The next values of m_k and n_k are $m_4 = 204$, $n_4 = 288$, and $m_5 = 1189$, $n_5 = 1681$.

Problem 1.1.1 Show that the sum of the first n odd numbers can be expressed in the form

$$\sum_{r=1}^{n}(2r-1) = 2\sum_{r=1}^{n} r - \sum_{r=1}^{n} 1 = 2T_n - n,$$

where T_n is the nth triangular number, and so verify that

$$\sum_{r=1}^{n}(2r-1) = n^2,$$

as we found using a geometrical argument.

Problem 1.1.2 Deduce from the equation

$$1+3+5+\cdots+(2n-1) = (1+2+3+\cdots+2n) - (2+4+6+\cdots+2n)$$

that

$$\sum_{r=1}^{n}(2r-1) = \sum_{r=1}^{2n} r - \sum_{r=1}^{n} 2r = T_{2n} - 2T_n = n^2,$$

in agreement with the result obtained in Problem 1.1.1.

Problem 1.1.3 Imagine n^3 nodes arranged in a cube, just as we arranged n^2 nodes in a square in Figure 1.1. If we remove the nodes that lie on three faces of the cube that meet at a corner, we will be left with a cube of $(n-1)^3$ nodes. Verify that the number of nodes removed is $3n^2 - 3n + 1$, and deduce that

$$n^3 = \sum_{r=1}^{n} \left(3r^2 - 3r + 1\right).$$

Hence show that

$$\sum_{r=1}^{n} r^2 = \frac{1}{3}\left(n^3 - n\right) + \sum_{r=1}^{n} r = \frac{1}{3}\left(n^3 - n\right) + \frac{1}{2}n(n+1),$$

and so verify (1.10).

Problem 1.1.4 Verify that

$$n^2(n+1)^2 - (n-1)^2n^2 = 4n^3$$

and deduce that

$$\sum_{r=1}^{n} r^3 = \frac{1}{4}n^2(n+1)^2,$$

so that

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = (1+2+3+\cdots+n)^2.$$

Problem 1.1.5 Let a and b be positive integers with $a+b$ odd. By considering the difference between $a+b$ and $a-b$, show that $a-b$ is also odd.

Problem 1.1.6 Let a and b be positive integers, with $a > b$ and $a+b$ odd. Then we may write

$$\frac{1}{2}(a+b) = d + \frac{1}{2},$$

where d is a positive integer. Show that

$$\frac{1}{2}(a-b) = d - b + \frac{1}{2},$$

and deduce from (1.12) that

$$ab = \left(d + \frac{1}{2}\right)^2 - \left(d - b + \frac{1}{2}\right)^2 = d^2 - (d-b)^2 + b.$$

Use this result and Table 1.1 to evaluate 86×57.

Problem 1.1.7 With the notation and conditions of Problem 1.1.6, write

$$\frac{1}{2}(a+b) = c - \frac{1}{2},$$

so that $c = d+1$ is a positive integer, and deduce that

$$ab = c^2 - (c-b)^2 - b.$$

Problem 1.1.8 Verify (1.15) algebraically, using (1.3), and also show that

$$T_{m-1} + T_n - T_{m-n-1} = mn,$$

where $m > n$.

Problem 1.1.9 When he was only nineteen, Gauss proved that every positive integer can be expressed as the sum of three or fewer triangular numbers. Check that this property holds for the first 100 positive integers.

Problem 1.1.10 Deduce from (1.18) that

$$m_{k+1} + m_{k-1} = \frac{1}{4\sqrt{2}}\left(\left(\alpha + \frac{1}{\alpha}\right)\alpha^k - \left(\alpha + \frac{1}{\alpha}\right)\alpha^{-k}\right),$$

and use (1.17) and (1.21) to show that $\alpha + 1/\alpha = 6$, thus verifying the recurrence relation (1.22). Use the same method to verify (1.23).

1.2 Pythagoras's Theorem

One of the best known theorems in mathematics is Pythagoras's theorem, that in a right-angled triangle whose longest side has length c, and whose other two sides have lengths a and b, we have

$$a^2 + b^2 = c^2. \tag{1.24}$$

This is depicted in Figure 1.8. The longest side, which is opposite the right

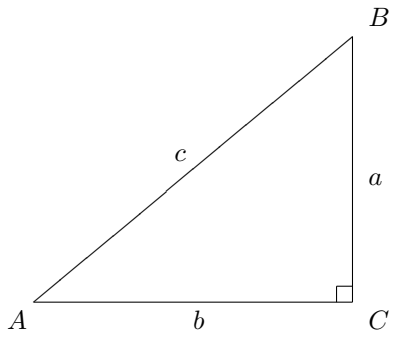


FIGURE 1.8. The sides of a right-angled triangle satisfy $a^2 + b^2 = c^2$.

angle, is called the *hypotenuse*. It is also true that if the sides of a triangle satisfy (1.24), then the angle opposite the longest side is a right angle. Although this theorem is named after the Greek mathematician Pythagoras, who lived in the sixth century BC, and the earliest known proof in the general case is due to the Greeks, the theorem was known much earlier to the Egyptians and even earlier to the Babylonians. The relation (1.24) holds for *all* right-angled triangles, for example the triangle whose sides are $a = b = 1$ and c, so that $c^2 = 2$. However, from the earliest times, there was an interest in finding right-angled triangles whose sides are positive integers. The simplest of these is the triangle whose sides are 3, 4, and 5. Eves [9] states that surveyors in ancient Egypt laid out right angles by constructing a 3, 4, 5 triangle using a rope divided into 12 equal parts by 11 knots, and that a proof that the 3, 4, 5 triangle is indeed right-angled was obtained in China, possibly as early as the second millennium BC.

To discuss this proof, we use Figure 1.9. It is clear by construction that there is a unique triangle, say T, that has sides of length 3 and 4 with a right angle contained between them. It also follows by construction that there is a unique triangle, say T', whose sides are of lengths 3, 4 and 5. We need to prove that the triangles T and T' are the same. We observe that two copies of the triangle T can be put together to form a rectangle with area 3×4, and so T has area 6. Figure 1.9 shows a quadrilateral $ABCD$ composed of a small square of unit length surrounded by four right-angled triangles that

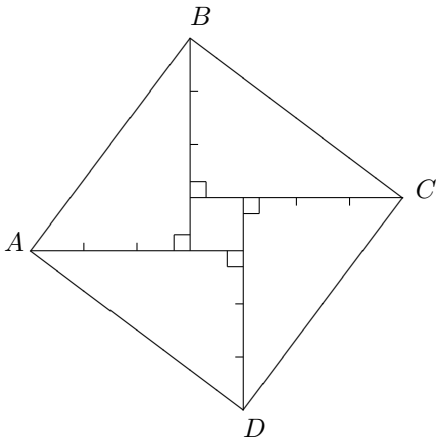


FIGURE 1.9. Chinese proof that the 3, 4, 5 triangle is right-angled.

are *congruent* to T. (Two or more geometrical figures that are identical are said to be congruent.) It follows that the area of the quadrilateral is $4 \times 6 + 1 = 25$. We next observe that because the four triangles are congruent, all four angles of the quadrilateral $ABCD$ are equal. Since, as we will prove in Problem 1.2.1, the sum of the angles of a quadrilateral is four right angles, $ABCD$ is a square. We then have

$$\text{area of the square } ABCD = 25,$$

and consequently AB has length 5. Thus the two triangles T and T' are congruent, which confirms that the 3, 4, 5 triangle is right-angled.

There are many proofs of Pythagoras's theorem. One proof, which is very easy to follow because of its geometric simplicity, is based on the two different dissections of a square shown in Figure 1.10. Both squares contain four right-angled triangles with sides a, b, and c. The square on the right, with area $(a + b)^2$, is dissected into the four triangles and two squares of areas a^2 and b^2. The square on the left, also of area $(a + b)^2$, is split into the four triangles and a quadrilateral whose four sides are all of length c. It is clear that the four angles of this quadrilateral are all equal, and so they

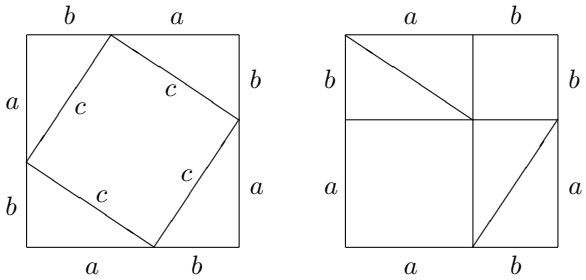


FIGURE 1.10. A pictorial proof of Pythagoras's theorem.

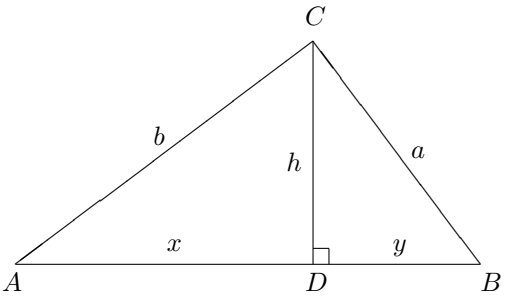


FIGURE 1.11. Proof of Pythagoras's theorem by similar triangles.

must all be right angles. Thus the quadrilateral is a square, with area c^2. We need only remove the four triangles from each of the two large squares of equal area to see that $a^2 + b^2 = c^2$, and this completes the proof.

We now give another simple proof of Pythagoras's theorem. This proof uses the notion of similar figures. Two figures drawn on the page are called *similar* if they have the same shape, apart from the *scale* of the figure. Thus, for example, all circles are similar, all squares are similar, and a 3, 4, 5 triangle is similar to a 9, 12, 15 triangle, but a 2×3 rectangle is not similar to a 3×5 rectangle.

Consider Figure 1.11, in which angle ACB is a right angle and CD is perpendicular to AB. Now, in any right-angled triangle, the sum of the two smaller angles is equal to a right angle. (We can see that this is true by putting two identical right-angled triangles together to make a rectangle.) Then we can deduce that the two smaller triangles in Figure 1.11, namely triangles ACD and CBD, have the same angles as the main triangle ABC, and so all three triangles are similar. In Figure 1.12, triangle ABC is the same as its namesake in Figure 1.11, and triangles $A'B'C$ and $A''B''C$ are obtained by cutting out triangles ACD and CBD from Figure 1.11, flipping them over, and pasting them into Figure 1.12. Thus the sum of the

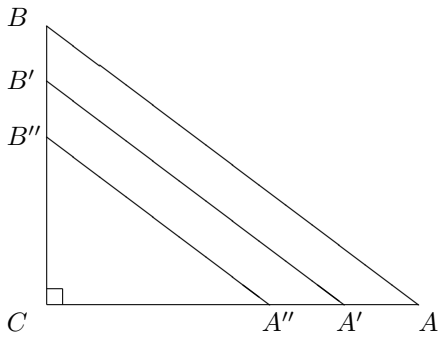


FIGURE 1.12. The three similar triangles in Figure 1.11.

areas of the two smaller triangles $A'B'C$ and $A''B''C$ is equal to the area of the largest triangle ABC. Now let

$$\frac{|BC|}{|AB|} = \frac{a}{c} = \lambda \qquad \text{and} \qquad \frac{|CA|}{|AB|} = \frac{b}{c} = \mu, \tag{1.25}$$

where $|BC|$ denotes the length of the line segment BC, and λ and μ (the Greek letters *lambda* and *mu*) are positive constants that apply to all triangles similar to triangle ABC. Thus $a = \lambda c$, $b = \mu c$, and we have

$$\Delta_c = \text{area of triangle } ABC = \frac{1}{2}ab = \frac{1}{2}\lambda\mu c^2, \tag{1.26}$$

where Δ is the uppercase Greek letter *delta*. Then, since triangles CBD and ABC are similar, we have

$$\frac{|BD|}{|BC|} = \frac{y}{a} = \lambda \qquad \text{and} \qquad \frac{|DC|}{|BC|} = \frac{h}{a} = \mu, \tag{1.27}$$

and it follows that

$$\Delta_a = \text{area of triangle } CBD = \frac{1}{2}yh = \frac{1}{2}\lambda\mu a^2. \tag{1.28}$$

We likewise derive from the similarity of triangles ACD and ABC that

$$\frac{|CD|}{|CA|} = \frac{h}{b} = \lambda \qquad \text{and} \qquad \frac{|DA|}{|CA|} = \frac{x}{b} = \mu, \tag{1.29}$$

so that

$$\Delta_b = \text{area of triangle } ACD = \frac{1}{2}hx = \frac{1}{2}\lambda\mu b^2. \tag{1.30}$$

Since $\Delta_a + \Delta_b = \Delta_c$, it follows from the above equations that

$$\frac{1}{2}\lambda\mu a^2 + \frac{1}{2}\lambda\mu b^2 = \frac{1}{2}\lambda\mu c^2,$$

and thus $a^2 + b^2 = c^2$. This completes our second proof of Pythagoras's theorem.

Among the very large number of proofs of Pythagoras's theorem, there is one that is attributed to James Garfield (1831–1881), who became the twentieth president of the United States in 1881. Garfield's proof depends on Figure 1.13, which consists of two right-angled triangles with sides a, b, and c, and half of a square of side c. Observe that we can put two copies of the trapezoidal shape $ABCD$ together to give the left-hand diagram in Figure 1.10, and so the quadrilateral $ABCD$ has area

$$\frac{1}{2}(a+b)^2 = \frac{1}{2}(a^2 + 2ab + b^2) = \frac{1}{2}(a^2 + b^2) + ab. \tag{1.31}$$

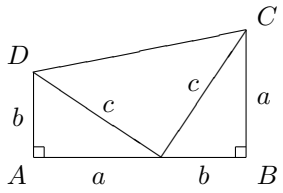


FIGURE 1.13. Garfield's proof of Pythagoras's theorem.

(The above expansion of $(a+b)^2$ can be verified by a geometrical argument, such as that given in Problem 1.2.3.) Continuing Garfield's proof, on adding the areas of the two triangles and the half square, we see that the area of the quadrilateral $ABCD$ can also be expressed as

$$\frac{1}{2}c^2 + ab. \tag{1.32}$$

We next subtract ab from each of the two expressions (1.31) and (1.32) for the area of the quadrilateral $ABCD$, to give $\frac{1}{2}(a^2+b^2) = \frac{1}{2}c^2$, and we need only multiply by 2 to complete the proof.

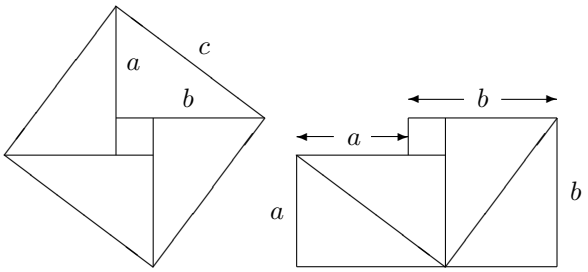


FIGURE 1.14. Chinese dissection method.

Many proofs of Pythagoras's theorem rely on dissecting a figure and rearranging it to give another shape with the same area, like the above proof based on Figure 1.10. Another such proof (see Figure 1.14) begins with four congruent right-angled triangles with sides a, b, and c placed around a square of side $b-a$, where a is the shortest side, to give a square of side c. This is rearranged to give a figure that can be viewed as two squares, one of side a and one of side b, nestling side by side. If $a = b$, the little square shrinks to zero.

Pythagoras's theorem appears as Proposition 47 in Book I of the famous set of books, the *Elements*, compiled by the Greek mathematician Euclid circa 300 BC. These books are a record of the finest achievements of Greek

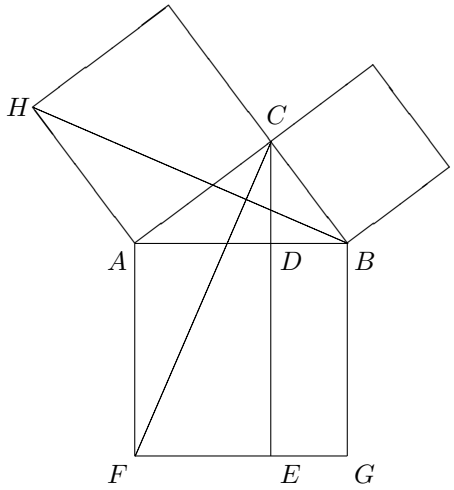


FIGURE 1.15. Euclid's diagram for Pythagoras's theorem.

mathematics up to that date. Euclid displays the right-angled triangle ABC (see Figure 1.15) and the squares constructed on each of its sides. The key step in Euclid's construction is to draw the line CE that splits the largest square into two rectangles. Euclid's proof depends on showing that each of these rectangles has the same area as the square that lies above it, so that the rectangle $ADEF$ has the same area as the square on the side AC, and the rectangle $DBGE$ has the same area as the square on the side BC. To do this, he first compares the triangles CAF and HAB. The *angles* CAF and HAB are equal, since each is a right angle plus the angle CAB. Since in each of the two triangles, the sides that enclose the equal angles CAF and HAB are also equal (that is, $|CA| = |HA|$ and $|AF| = |AB|$), it follows that the triangles CAF and HAB are congruent. Then, since the area of a triangle is half the length of its base times its height, the area of the triangle CAF is half the area of the rectangle $ADEF$, and also the area of triangle HAB is half the area of the square on AC. (See Problem 1.2.4.) Thus the the rectangle $ADEF$ has indeed the same area as the square on the side AC. Similarly, the rectangle $DBGE$ has the same area as the square on the side BC, and this completes Euclid's proof.

Problem 1.2.1 Put two congruent right-angled triangles together to form a rectangle, and hence show that the sum of the two smaller angles in the right-angled triangle add up to a right angle. Now, beginning with *any* triangle ABC, draw a perpendicular from a vertex to the opposite side, choosing the vertex so that the perpendicular lies within the triangle, thus

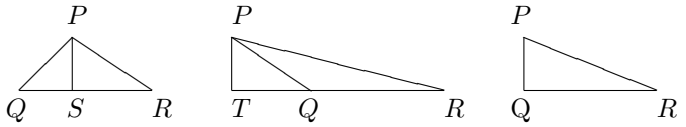
dissecting the triangle ABC into two right-angled triangles. Hence show that the sum of the angles in triangle ABC is two right angles. Deduce that the sum of the angles of a quadrilateral is four right angles by splitting the quadrilateral into two triangles.

Problem 1.2.2 Deduce from (1.25), (1.27), and (1.29) that $h = ab/c$, $x = b^2/c$, and $y = a^2/c$. Show also that $\lambda^2 + \mu^2 = 1$.

Problem 1.2.3 By dissecting a square of side $a+b$ into four pieces, namely a square of side a, a square of side b, and two equal rectangles of area ab, show that

$$(a+b)^2 = a^2 + 2ab + b^2.$$

Problem 1.2.4 Verify that the area of a triangle is half the length of its base times its height, by considering the three types of triangle PQR depicted in the figure that follows.



First show that this is true for the third figure, where the angle PQR is a right angle, since two copies of this triangle can be put together to form a rectangle. In the first figure, write $|QR| = |QS| + |SR|$ and express triangle PQR as the sum of two right-angled triangles, and in the second figure, write $|QR| = |TR| - |TQ|$ and express triangle PQR as the difference of two right-angled triangles.

Problem 1.2.5 Consider a triangle ABC. Let D denote the midpoint of BC, and let E denote the foot of the perpendicular from A to BC. (That is, the point where the line through A perpendicular to BC meets BC.) Consider the case in which E lies between B and D. Show that

$$|AB|^2 = |AE|^2 + |BE|^2 = |AD|^2 - |DE|^2 + |BE|^2,$$

so that

$$|AB|^2 = |AD|^2 + |BD|\big(|BE| - |DE|\big),$$

and similarly show that

$$|AC|^2 = |AD|^2 + |CD|\big(|EC| + |ED|\big).$$

By adding the latter two equations, show that

$$|AB|^2 + |AC|^2 = 2\left(|AD|^2 + |BD|^2\right). \tag{1.33}$$

Does (1.33) hold if E does not lie between B and D?

1.3 The Equation $a^2 + b^2 = c^2$

In this section we will write (a, b, c) to denote a triple of numbers. Let us consider the equation

$$a^2 + b^2 = c^2. \tag{1.34}$$

Obviously, if $c = \sqrt{a^2 + b^2}$, where a and b are any positive real numbers, the triples (a, b, c) and $(a, b, -c)$ are both solutions of (1.34). For example, $(1, 1, \sqrt{2})$ and $(1, 1, -\sqrt{2})$ satisfy (1.34). When we write $y = \sqrt{x}$, with x positive, y is defined as the *positive* number that satisfies $y^2 = x$. Thus $\sqrt{25} = 5$. It is more challenging to regard (1.34) as a Diophantine equation and seek solutions (a, b, c) in which a, b, and c are positive integers. Such a solution is called a *Pythagorean triple*, although much was known about such triples perhaps even a thousand years before the time of Pythagoras. We have already seen that $(3, 4, 5)$ is a Pythagorean triple. By trying some small values of a, b, and c, we find that $(5, 12, 13)$ is a Pythagorean triple, and you can find other solutions by such experimentation. However, we can do very much better than this because (1.34) is one of the rather rare Diophantine equations for which we can find *all* solutions. First we see that if (a, b, c) is a Pythagorean triple, then

$$(\lambda a)^2 + (\lambda b)^2 = \lambda^2(a^2 + b^2) = \lambda^2 c^2 = (\lambda c)^2,$$

and so if λ is any positive integer, $(\lambda a, \lambda b, \lambda c)$ is also a Pythagorean triple. For example, $(6, 8, 10)$, $(9, 12, 15)$, and all other multiples of $(3, 4, 5)$ are Pythagorean triples. Thus we can concentrate on finding solutions (a, b, c) such that a, b, and c are positive integers with no common factor. This means that no two of a, b, and c have a common factor. For if any two of the numbers a, b, and c have a common factor, say m, we can deduce from (1.34) that the third number would also have m as a factor. A Pythagorean triple (a, b, c) such that a, b, and c have no common factor is called a *primitive* Pythagorean triple.

Obviously, a and b cannot both be even, for this means they would have the common factor 2. If $a = 2m + 1$, so that a is odd, we have

$$a^2 = (2m + 1)^2 = 4m^2 + 4m + 1$$

and thus a^2 has a remainder of 1 when we divide by 4. Thus if a and b are both odd, $c^2 = a^2 + b^2$ has a remainder of 2 when we divide by 4. Thus c must be even and so c^2 must be a multiple of 4. This gives a contradiction, since c^2 cannot have a remainder of 0 *and* 2 when we divide by 4. Thus a and b cannot both be odd; the only possibility is that one of the pair a and b is even, and the other is odd, and then c must be odd. Since it does not matter which is which between a and b, we will take a to be even and b to be odd. It follows that $c + b$ and $c - b$ are both even. Thus $\frac{1}{2}(c + b)$ and $\frac{1}{2}(c - b)$ are both positive integers, and they can have no common factor.

For if they had, we could show by adding and subtracting $\frac{1}{2}(c+b)$ and $\frac{1}{2}(c-b)$ that c and b would share this same common factor. Suppose now that p is an odd prime number that is a factor of a. Thus p^2 is a factor of a^2 and so must be is a factor of $\frac{1}{2}(c+b)$ or $\frac{1}{2}(c-b)$. Note that p cannot be a factor of both $\frac{1}{2}(c+b)$ and $\frac{1}{2}(c-b)$. For then the argument we used above concerning divisibility by 2 shows that p would be a factor of both b and c, and then (a, b, c) would not be a primitive triple. This crucial observation concerning division by a prime number plus a little further thought shows us that $\frac{1}{2}(c+b)$ and $\frac{1}{2}(c-b)$ must both be squares, say

$$\frac{1}{2}(c+b) = u^2 \qquad \text{and} \qquad \frac{1}{2}(c-b) = v^2,$$

where u and v are positive integers. On adding and subtracting the last two equations, we find that

$$c = u^2 + v^2 \qquad \text{and} \qquad b = u^2 - v^2. \tag{1.35}$$

It then follows from (1.34) and (1.12) that

$$a^2 = c^2 - b^2 = (u^2+v^2)^2 - (u^2-v^2)^2 = 4u^2v^2,$$

so that

$$a = 2uv. \tag{1.36}$$

The values of a, b, and c are said to be given in *parametric form* in terms of u and v in (1.35) and (1.36). We need to choose $u > v > 0$ such that a, b, and c are all positive, and further restrictions on u and v are required so that (1.35) and (1.36) yield a primitive solution (a, b, c). Obviously, u and v must have no common factor $d > 1$, for then a, b, and c would have d^2 as a common factor. Also, u and v cannot both be odd, for otherwise, a, b, and c would all have the common factor 2. Since any positive multiple of a primitive solution of (1.34) is also a solution, we have the following theorem.

Theorem 1.3.1 All solutions of the Diophantine equation

$$a^2 + b^2 = c^2$$

in positive integers are of the form

$$a = 2\lambda uv, \quad b = \lambda(u^2 - v^2), \quad c = \lambda(u^2 + v^2), \tag{1.37}$$

where λ, u, and v are positive integers such that $u > v$, u and v have no common factor greater than 1, and $u + v$ is odd. ■

Given a solution of (a, b, c) of equation (1.34), we can determine unique values of λ, u, and v so that a, b, and c are given in the parametric form (1.37). The positive integer λ is just the greatest common divisor of a, b,

u	2	3	4	4	5	5	6	6	7	7	7
v	1	2	1	3	2	4	1	5	2	4	6
a	4	12	8	24	20	40	12	60	28	56	84
b	3	5	15	7	21	9	35	11	45	33	13
c	5	13	17	25	29	41	37	61	53	65	85

TABLE 1.4. The first few primitive Pythagorean triples (a, b, c).

and c. So it suffices to consider the case $\lambda = 1$. We then determine u and v uniquely as the positive numbers satisfying

$$u^2 = \frac{1}{2}(c + b) \qquad \text{and} \qquad v^2 = \frac{1}{2}(c - b),$$

where b is the smaller odd number, and c is the larger odd number. The first few primitive solutions of (1.34) are given in Table 1.4. We notice that each value of a in the table is a multiple of 4, and we can see that this holds for every a defined in (1.37), since either u or v is even.

Table 1.4 shows that the first Pythagorean triple $(3, 4, 5)$ is not the only one in which the sum of two consecutive squares is a square, since we also have $20^2 + 21^2 = 29^2$. It turns out that there is an infinite number of primitive Pythagorean triples that have this property. The key to the solution lies in the sequence (U_n) that we now define.

Consider the recurrence relation

$$U_{n+1} = 2U_n + U_{n-1}, \quad n \geq 1, \quad \text{with} \quad U_0 = 0,\ U_1 = 1. \tag{1.38}$$

This generates an infinite sequence whose next few values are $U_2 = 2$, $U_3 = 5$, $U_4 = 12$, $U_5 = 29$, and $U_6 = 70$. Let us now write

$$a = 2uv,\ \ b = u^2 - v^2,\ \ c = u^2 + v^2, \quad \text{where} \quad u = U_{n+1},\ \ v = U_n, \tag{1.39}$$

so that a, b, and c are as defined in (1.35) and (1.36). Obviously, (a, b, c) is a Pythagorean triple. We will show that for any $n \geq 1$, a, b, and c have no common factor and $a - b = \pm 1$. Since

$$b - a = u^2 - v^2 - 2uv,$$

it will follow that $a - b = \pm 1$ if

$$W_n = U_{n+1}^2 - U_n^2 - 2U_nU_{n+1} = \pm 1. \tag{1.40}$$

It is easily verified that $W_0 = 1$ and $W_1 = -1$. If you compute the next few values of W_n, you will be encouraged by the fact that they continue to take the values 1 and -1 alternately. To *prove* that $W_n + W_{n+1} = 0$ for all values of n, we write

$$W_n + W_{n+1} = (U_{n+1}^2 - U_n^2 - 2U_nU_{n+1}) + (U_{n+2}^2 - U_{n+1}^2 - 2U_{n+1}U_{n+2}),$$

and so obtain

$$W_n + W_{n+1} = U_{n+2}^2 - U_n^2 - 2U_nU_{n+1} - 2U_{n+1}U_{n+2}. \tag{1.41}$$

Then, since

$$U_{n+2}^2 - U_n^2 = (U_{n+2} - U_n)(U_{n+2} + U_n),$$

we can simplify (1.41) to give

$$W_n + W_{n+1} = (U_{n+2} + U_n)(U_{n+2} - 2U_{n+1} - U_n) = 0. \tag{1.42}$$

Note that it follows from the recurrence relation (1.38) that the second factor on the right of (1.42) is zero.

Let us review what we know about the sequence (W_n). The initial values are $W_0 = 1$ and $W_1 = -1$, and we see from (1.42) that $W_{n+1} = -W_n$ for *all* n. Then, by the *principle of mathematical induction*, all members of the sequence (W_n) take the values $+1$ and -1 alternately, beginning with $W_0 = 1$. Mathematical induction has been likened to climbing a ladder. We need to get onto the first rung of the ladder, and when we are standing on *any* rung of the ladder, we need to be able to step up to the next rung. If these two conditions are met, we can climb the ladder.

Thus there is an infinite number of Pythagorean triples

$$(a, b, c) = (2uv, u^2 - v^2, u^2 + v^2), \quad \text{with} \quad u = U_{n+1},\ v = U_n, \tag{1.43}$$

for which $a - b = \pm 1$. The Pythagorean triples (a, b, c) defined by (1.43) with values of n between 1 and 5 are

$$(4, 3, 5),\ \ (20, 21, 29),\ \ (120, 119, 169),\ \ (696, 697, 985),\ \ (4060, 4059, 5741).$$

The condition that $a - b = \pm 1$ implies that a and b have no common factor greater than 1, and thus all the Pythagorean triples defined by (1.43) are primitive.

Eves [9] writes that since the first half of the nineteenth century, about half a million clay tablets have been unearthed by archaeologists working in Mesopotamia, of which about 300 have shed light on ancient Babylonian mathematics. One of these tablets, housed at Columbia University, in New York, and known as Plimpton 322, is part of the collection named after G. A. Plimpton. From the style of the script used, it is thought to date from the period 1900 to 1600 BC. Plimpton 322 contains the b and c components of fifteen Pythagorean triples (a, b, c), of which all except two are primitive. The largest of the fifteen is the primitive triple that satisfies

$$13500^2 + 12709^2 = 18541^2.$$

Did the Babylonians discover such triples by a numerical search or through mathematical insight? I believe that these triples were found by people who

knew what they were doing. If so, this was not blind, mechanical arithmetic but serious mathematics.

Pierre de Fermat famously asserted that the equation

$$a^n + b^n = c^n \tag{1.44}$$

has no solutions in positive integers for $n > 2$, and even claimed he had a proof, although none was ever found. This came to be called Fermat's last theorem. A proof eluded mathematicians for more than 350 years until one was obtained by Andrew Wiles, and published in the *Annals of Mathematics* in 1995. Simon Singh [16] has given a fine nontechnical account of the story behind the solution of this problem. Even a proof that (1.44) has no solution in positive integers for exponent $n = 3$ is beyond the scope of this book. The usual proof given for the case $n = 4$ is very much simpler than that for $n = 3$. It relies on showing that the equation

$$x^4 + y^4 = z^2 \tag{1.45}$$

has no solution in positive integers and thus (1.44) with $n = 4$ has no solution in positive integers. (If $x^4 + y^4$ cannot be a square it certainly cannot be a fourth power.) We begin by assuming that (1.45) *has* solutions, and let (x, y, z) denote the solution with the smallest value of z such that x, y, and z have no common factor greater than 1. We use the fact that $\left(x^2, y^2, z\right)$ is a primitive Pythagorean triple. The proof is completed by deducing the existence of a *smaller* solution of (1.45), and this contradicts the above assumption that there is a solution of (1.45). This type of proof, which is called the *method of infinite descent*, was pioneered by Fermat. For a complete proof that the Diophantine equation (1.45) has no solution see Hardy and Wright [12] or Phillips [14].

Problem 1.3.1 If $u = 2m + 2n + 1$ and $v = 2m$, show that

$$u^2 - v^2 - 2uv - 1 = (u - v)^2 - 2v^2 - 1 = 4\left(n(n + 1) - 2m^2\right),$$

and show that with $a = 2uv$ and $b = u^2 - v^2$, we have $b - a = 1$ for all values of m and n satisfying the Diophantine equation $n(n + 1) = 2m^2$. Which of the Pythagorean triples in Table 1.4 are of this form?

Problem 1.3.2 Given the primitive Pythagorean triple

$$(a, b, c) = (13500, 12709, 18541)$$

that was mentioned in the text, find values of u and v such that a, b, and c are given by (1.37) with $\lambda = 1$.

Problem 1.3.3 Verify that

$$(2n^2 + 2n)^2 + (2n + 1)^2 = (2n^2 + 2n + 1)^2$$

for all positive integers n and express the Pythagorean triple that satisfies the above equation in the form (1.37). Note that this gives Pythagorean triples (a, b, c) for which $c - a = 1$.

Problem 1.3.4 Show that

$$x = 2uv, \quad y = 2u^2 - v^2, \quad z = 2u^2 + v^2,$$

where u and v are positive integers such that $2u^2 > v^2$, is a solution of the Diophantine equation $2x^2 + y^2 = z^2$.

Problem 1.3.5 Find solutions of the Diophantine equation $3x^2 + y^2 = z^2$.

Problem 1.3.6 Let S_n denote the sum of the squares of the first n positive integers. Verify that $S_1 = 1$. *Assume* that for some integer $n \geq 1$,

$$S_n = \frac{1}{6}n(n+1)(2n+1) \tag{1.46}$$

and use the fact that $S_{n+1} = S_n + (n+1)^2$ to deduce that

$$S_{n+1} = \frac{1}{6}(n+1)(n+2)(2n+3).$$

Thus conclude by mathematical induction that (1.46) holds for all $n \geq 1$.

1.4 Sum of Two Squares

In this section we will discuss the question of which positive integers can be expressed as the sum of two squares. For example, we have $13 = 2^2 + 3^2$, but 3 cannot be expressed as a sum of two squares. I think it is appropriate to include this topic, because it fits in so well with the rest of the material in this chapter. However, I will not prove all of the results that I state in this section, because that would take us deeper into the theory of numbers than seems appropriate in this book. The reader who wishes to gain a fuller understanding of the material will find the omitted proofs in Hardy and Wright [12], or in Phillips [14]. We begin by stating without proof that every positive integer n may be written uniquely in the form

$$n = p_1^{\alpha_1} p_2^{\alpha_2} p_3^{\alpha_3} \cdots p_N^{\alpha_N} \tag{1.47}$$

for some choice of N, where $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, and so on, are the prime numbers, each exponent α_1, α_2, $\alpha_3, \ldots \alpha_{N-1}$ is a nonnegative integer, and the last exponent, α_N, is positive. Thus if an exponent α_j is zero, it means that p_j is not present in the above representation of n. Given any n, we can derive the unique factorization (1.47) by testing whether n is divisible by $p_1 = 2$. We divide by 2 as many times as we can. Then, beginning with

the quotient that remains after repeated division by 2, we divide this by $p_2 = 3$ as many times as we can. We repeat this process with successive primes until we obtain the quotient 1. For example, we find that

$$56852 = 2^2 \times 61 \times 233 \quad \text{and} \quad 550368 = 2^5 \times 3^3 \times 7^2 \times 13.$$

Beginning with the representation of n as a product of powers of primes, as in (1.47), we can derive a second expression for n that is also unique, writing

$$n = \lambda^2 n_1, \qquad \text{with} \qquad n_1 = p_1^{\beta_1} p_2^{\beta_2} p_3^{\beta_3} \cdots p_N^{\beta_N},$$

where $\lambda \geq 1$ is a positive integer and each exponent β_j is either 0 or 1. We then say that the above number n_1 is *square-free*, since it is not divisible by any square greater than 1. Observe that a square-free number is simply a product of distinct primes. For example, we have

$$n = 550368 = 2^5 3^3 7^2 13 = (2^4 3^2 7^2) \times 2 \times 3 \times 13 = (2^2 \times 3 \times 7)^2 \times 2 \times 3 \times 13,$$

so that $\lambda = 2^2 \times 3 \times 7 = 84$, and the number $n_1 = 2 \times 3 \times 13 = 78$ is square-free. If we can express n_1 as the sum of two squares, say

$$n_1 = a^2 + b^2,$$

then we can write n as the sum of two squares, since

$$n = \lambda^2 n_1 = (\lambda a)^2 + (\lambda b)^2.$$

This greatly simplifies our task of finding all positive integers that can be expressed as the sum of two squares. For we have reduced our original problem to that of finding which square-free numbers can be expressed as the sum of two squares.

Apart from 2, the first prime, every prime is odd and must have one of the forms $4m + 1$ or $4m + 3$. Consider the following two statements concerning these two classes of odd primes.

1. Every prime number of the form $4m + 1$ can be expressed as the sum of two squares, and this can be done uniquely.

2. No prime number of the form $4m + 3$ can be expressed as the sum of two squares.

Both statements are true. The second statement is easily verified, as follows. We saw in the last section that a^2 is divisible by 4 if a is even, and leaves a remainder of 1 on division by 4 if a is odd. Thus $a^2 + b^2$ has a remainder of either 0, 1, or 2 on division by 4. It follows that if n is of the form $4m + 3$, it cannot be expressed as the sum of two squares. The first of the above statements is very much harder to prove than the second. One proof (see Hardy and Wright [12] or Phillips [14]) relies on the use of complex numbers, which we will meet in the next section.

We further state without proof that if a square-free n has any prime factor of the form $4m + 3$, then n *cannot* be expressed as the sum of two squares. (See Hardy and Wright [12].) Finally, we make use of the identity

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2. \tag{1.48}$$

This identity was used implicitly by Diophantus of Alexandria in his book *Arithmetica*, which was published in the third century AD. It was quoted by Leonardo of Pisa (circa 1175–1220), who is also known as Fibonacci, in his book *Liber Abaci*, published in 1202. By repeatedly applying (1.48), we can deduce that any square-free n whose prime factorization may or may not contain 2 and otherwise consists only of odd primes of the form $4m+1$ is expressible as the sum of two squares. Furthermore, it follows from what has been said above that these are the only square-free numbers that are expressible as the sum of two squares.

Observe that (1.48) reduces to

$$2(a^2 + b^2) = (a - b)^2 + (a + b)^2 \tag{1.49}$$

when we put $c = d = 1$. Note also that if $a \neq b$ (meaning a is not equal to b) and $c \neq d$, we can interchange c and d in (1.48) and so obtain two different expressions of $(a^2 + b^2)(c^2 + d^2)$ as the sum of two squares.

Example 1.4.1 We have

$$2 = 1^2 + 1^2, \qquad 13 = 3^2 + 2^2, \qquad 41 = 5^2 + 4^2,$$

and we have from (1.49) that $26 = 2(3^2 + 2^2) = 1^2 + 5^2$. We can now use (1.48) to give

$$26 \times 41 = (1^2 + 5^2)(5^2 + 4^2) = 15^2 + 29^2$$

and also

$$26 \times 41 = (1^2 + 5^2)(4^2 + 5^2) = 21^2 + 25^2.$$

Thus we can write 1066 as the sum of two squares in two ways,

$$1066 = 15^2 + 29^2 = 21^2 + 25^2. \qquad \blacksquare$$

Problem 1.4.1 Show that 1776 cannot be expressed as the sum of two squares.

Problem 1.4.2 Express 1314 as the sum of two squares.

Problem 1.4.3 C. F. Gauss was born in Braunschweig in 1777 and died in Göttingen in 1855. Express one of these numbers as the sum of two squares, and show that the other number cannot be expressed in this form.

Problem 1.4.4 Let n be an odd integer such that

$$2n^2 = u^2 + v^2,$$

where u and v are positive integers and $u > v$. Show that u and v must both be odd and that consequently, $\frac{1}{2}(u+v)$ and $\frac{1}{2}(u-v)$ are both positive integers. Deduce that n may be expressed as the sum of two squares, in the form

$$n^2 = \left(\frac{1}{2}(u+v)\right)^2 + \left(\frac{1}{2}(u-v)\right)^2.$$

Problem 1.4.5 Find the smallest integer that can be expressed as the sum of two distinct squares in two different ways.

1.5 Complex Numbers

Consider the *quadratic* equation

$$az^2 + bz + c = 0, \tag{1.50}$$

where a, b, and c are any real numbers, with a nonzero. It is called a quadratic equation because it involves the square (Latin *quadrum*) of the unknown quantity z. We will find all values of z that satisfy it. These are called *solutions* of the equation. We begin by dividing the quantities on each side of the equation by the nonzero number a, giving

$$z^2 + \left(\frac{b}{a}\right)z + \frac{c}{a} = 0. \tag{1.51}$$

It should be clear that the two equations (1.50) and (1.51) have the same solutions. If we subtract c/a from both sides of (1.51) we obtain

$$z^2 + \left(\frac{b}{a}\right)z = -\frac{c}{a}, \tag{1.52}$$

which also has the same solutions as the two earlier equations. We now "complete the square," adding a suitable constant to each side of (1.52) so that we can write the left side in the form $(z+\alpha)^2$, where α is real. Since

$$\left(z + \frac{b}{2a}\right)^2 = z^2 + \left(\frac{b}{a}\right)z + \frac{b^2}{4a^2},$$

the quantity we need to add to both sides of (1.52) is $b^2/4a^2$, giving

$$\left(z + \frac{b}{2a}\right)^2 = -\frac{c}{a} + \frac{b^2}{4a^2},$$

which can be written in the form

$$\left(z + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}. \tag{1.53}$$

This quadratic equation has the same solutions as each of the equations (1.50), (1.51), and (1.52). We now see that there are three possibilities, depending on whether $b^2 - 4ac$ is zero, positive, or negative.

1. If $b^2 - 4ac$ is zero, it follows from (1.53) that equation (1.50) has only one solution, namely

$$z = -\frac{b}{2a}. \tag{1.54}$$

2. If $b^2 - 4ac$ is positive, we see from (1.53) that

$$z + \frac{b}{2a} = +\frac{\sqrt{b^2 - 4ac}}{2a} \quad \text{or} \quad z + \frac{b}{2a} = -\frac{\sqrt{b^2 - 4ac}}{2a},$$

where $\sqrt{b^2 - 4ac}$ is the positive number whose square is $b^2 - 4ac$. In this case the quadratic equation (1.50) has the two solutions

$$z = -\frac{b}{2a} + \frac{\sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad z = -\frac{b}{2a} - \frac{\sqrt{b^2 - 4ac}}{2a}.$$

3. If $b^2 - 4ac$ is *negative*, we have the most interesting case! It was said before the invention of complex numbers that (1.53), and consequently (1.50), has *no solutions* when $b^2 - 4ac$ is negative, because a negative number has no square root. However, as we will see below, we *now* say that (1.53) has no solutions in real numbers, but has two solutions that are complex numbers.

Consider the quadratic equation $z^2 + 1 = 0$, which is equivalent to

$$z^2 = -1. \tag{1.55}$$

Since the square of any real number is nonnegative, it is clear that this equation has no solution in real numbers. Let us assume that there is some other system of "numbers" in which there is a solution of (1.55). Let us denote such a solution by i, so that $i^2 = -1$. We will also assume that this number system behaves algebraically like the real numbers so that, for example, any number α in this system has the property that $\alpha^2 = (-\alpha)^2$. This entails that (1.55) has the two solutions $z = i$ and $z = -i$. We then find that

$$(z - i)(z + i) = z(z + i) - i(z + i) = z^2 + iz - iz - i^2,$$

and since $i^2 = -1$, we obtain

$$(z - i)(z + i) = z^2 + 1.$$

Thus we have expressed $z^2 + 1$ as a product of the two factors $z - i$ and $z + i$. Then $z^2 + 1 = 0$ is equivalent to

$$(z - i)(z + i) = 0,$$

and so $z - i = 0$ or $z + i = 0$. Although this tells us only what we already know, that $z = i$ or $z = -i$, it gives us some confidence in the algebra of this system.

We can now pursue the solution of (1.53) for the case $b^2 - 4ac < 0$. For then

$$\left(z + \frac{b}{2a}\right)^2 = \frac{4ac - b^2}{4a^2} \cdot (-1),$$

and so

$$z + \frac{b}{2a} = \frac{\sqrt{4ac - b^2}}{2a}\, i \quad \text{or} \quad z + \frac{b}{2a} = -\frac{\sqrt{4ac - b^2}}{2a}\, i,$$

where $\sqrt{4ac - b^2}$ is the positive number whose square is the positive number $4ac - b^2$. Thus, when $b^2 - 4ac < 0$, the equation (1.50) has the two solutions

$$z = -\frac{b}{2a} + \frac{\sqrt{4ac - b^2}}{2a}\, i \quad \text{and} \quad z = -\frac{b}{2a} - \frac{\sqrt{4ac - b^2}}{2a}\, i. \tag{1.56}$$

These have the form $z = x + yi$ and $z = x - yi$, where the two numbers

$$x = -\frac{b}{2a} \quad \text{and} \quad y = \frac{\sqrt{4ac - b^2}}{2a}$$

are both real and $i^2 = -1$. Any number of the form $x + yi$, with x and y real, is called a *complex number*.

Definition 1.5.1 Given the complex number $z = x + yi$, we say that x is the *real* part of z and y is the *imaginary* part. We write the real and imaginary parts of z as

$$x = \mathbf{Re}(z) \quad \text{and} \quad y = \mathbf{Im}(z). \qquad \blacksquare \tag{1.57}$$

Remark 1.5.1 The terms "real part" and "imaginary part" are a little misleading, since mathematicians would agree that the imaginary part of a complex number is no more "imaginary," in the everyday sense of the word, than the real part. One might also say that a real number is no more or less "real," in the everyday sense, than any other kind of number. $\blacksquare$

We can evaluate sums and products of complex numbers in an obvious way. Thus, if $z_1 = a + bi$ and $z_2 = c + di$, their sum is

$$z_1 + z_2 = (a + c) + (b + d)i, \tag{1.58}$$

and their product is

$$z_1 z_2 = (a + bi)(c + di) = a(c + di) + bi(c + di) = ac + adi + bci + bdi^2.$$

Since $i^2 = -1$, we find that

$$z_1 z_2 = (a + bi)(c + di) = (ac - bd) + (ad + bc)i. \tag{1.59}$$

Note from (1.58) and (1.59) that the sum and product of two complex numbers are both complex numbers. If we put $a = c = x$ and $b = -d = y$ in (1.59), we obtain, as a special case,

$$(x + yi)(x - yi) = x^2 + y^2. \tag{1.60}$$

Definition 1.5.2 If $z = x + yi$, we denote $x - yi$ by $\bar{z}$ and call $\bar{z}$ the *conjugate* of z. It follows from this definition that the conjugate of $\bar{z}$ is z, and we refer to z and $\bar{z}$ as a complex conjugate pair. ■

It follows from (1.56) that when the solutions of the quadratic equation (1.50) are complex, the two solutions are a complex conjugate pair.

We see from Definitions 1.5.1 and 1.5.2 that

$$z + \bar{z} = 2\,\mathbf{Re}(z) \quad \text{and} \quad z - \bar{z} = 2i\,\mathbf{Im}(z). \tag{1.61}$$

Definition 1.5.3 If $z = x + yi$, we write $|z|$ to denote $\sqrt{x^2 + y^2}$ and call $|z|$ the *modulus* of the complex number z. ■

Since $|x| \le \sqrt{x^2 + y^2}$ and $|y| \le \sqrt{x^2 + y^2}$, we see from Definitions 1.5.1 and 1.5.3 that

$$|\mathbf{Re}(z)| \le |z| \quad \text{and} \quad |\mathbf{Im}(z)| \le |z|. \tag{1.62}$$

It follows from Definitions 1.5.2 and 1.5.3 that $|\bar{z}| = |z|$, and it is then clear from (1.60) that for any complex number z,

$$z\bar{z} = |z|^2. \tag{1.63}$$

With $z_1 = a + bi$ and $z_2 = c + di$, we see from (1.59) that

$$|z_1 z_2|^2 = (ac - bd)^2 + (ad + bc)^2,$$

and we deduce from the identity (1.48) that $|z_1 z_2|^2 = |z_1|^2 |z_2|^2$, and thus

$$|z_1 z_2| = |z_1|\,|z_2|. \tag{1.64}$$

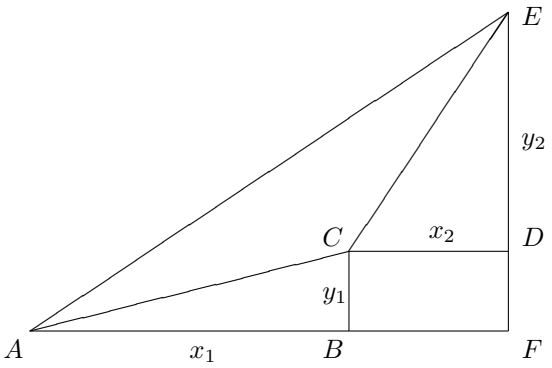


FIGURE 1.16. If $|AB| = x_1$, $|BC| = y_1$, $|CD| = x_2$, and $|DE| = y_2$, then $|AF| = x_1 + x_2$ and $|FE| = y_1 + y_2$. With $z_1 = x_1 + y_1 i$ and $z_2 = x_2 + y_2 i$, we have $|AC| = |z_1|$, $|CE| = |z_2|$, and $|AE| = |z_1 + z_2|$.

It is remarkable that the wonderful identity (1.48) was known in number theory fifteen hundred years before its application to complex numbers was appreciated. We will meet a generalization of (1.48) in Problem 7.1.9.

The modulus of a complex number behaves like a *length*. Indeed, if x and y are positive, the modulus of $z = x + yi$ is the length of the hypotenuse of a right-angled triangle whose shorter sides are x and y. This leads to the *triangle inequality*,

$$|z_1 + z_2| \le |z_1| + |z_2|, \tag{1.65}$$

as demonstrated geometrically by the inequality $|AC| + |CE| < |AE|$ in Figure 1.16. This would be an equality if A, C, and E were in a straight line. The triangle inequality can also be justified algebraically. Let us assume that z_1 and z_2 are both nonzero, for otherwise, (1.65) is trivial. We first verify that the conjugate of $z_1 + z_2$ is $\bar{z_1} + \bar{z_2}$ and that $z_1\bar{z_2}$ and $\bar{z_1}z_2$ are conjugates. Then we have from (1.63) that

$$|z_1 + z_2|^2 = (z_1 + z_2)(\bar{z_1} + \bar{z_2}) = z_1\bar{z_1} + z_1\bar{z_2} + \bar{z_1}z_2 + z_2\bar{z_2}. \tag{1.66}$$

Since $z_1\bar{z_2}$ and $\bar{z_1}z_2$ are conjugates, we see from (1.61) that

$$z_1\bar{z_2} + \bar{z_1}z_2 = 2\,\mathbf{Re}\,(z_1\bar{z_2})\,.$$

If we now use the latter equality and (1.63) in (1.66), we find that

$$|z_1 + z_2|^2 = |z_1|^2 + 2\,\mathbf{Re}\,(z_1 z_2) + |z_2|^2\,.$$

Then, on using the first inequality in (1.62), we obtain

$$|z_1 + z_2|^2 \le |z_1|^2 + 2\,|z_1\bar{z_2}| + |z_2|^2.$$

Finally, we use (1.64) and the property that $|\bar{z}| = |z|$ to give

$$|z_1 + z_2|^2 \le |z_1|^2 + 2\,|z_1|\;|z_2| + |z_2|^2 = (|z_1| + |z_2|)^2,$$

and hence (1.65) holds. This will be an equality if

$$\mathbf{Re}\,(z_1\bar{z_2}) = |z_1\bar{z_2}|\,,$$

which will hold if and only if

$$\mathbf{Re}\,(z_1\bar{z_2}) \geq 0 \quad \text{and} \quad \mathbf{Im}\,(z_1\bar{z_2}) = 0. \tag{1.67}$$

Let us write $z_1 = x_1 + y_1 i$ and $z_2 = x_2 + y_2 i$. Then

$$\mathbf{Re}\,(z_1\bar{z_2}) = x_1x_2 + y_1y_2 \quad \text{and} \quad \mathbf{Im}\,(z_1\bar{z_2}) = x_2y_1 - x_1y_2.$$

We can see that if z_2 is a positive multiple of z_1, so that

$$x_2 = \lambda x_1 \quad \text{and} \quad y_2 = \lambda y_1, \quad \text{with } \lambda > 0, \tag{1.68}$$

then *both* conditions in (1.67) will hold, and thus (1.65) will be an equality. The conditions in (1.68) are equivalent to the geometrical condition that the three points A, C, and E in Figure 1.16 lie in a straight line, with C lying between A and E. See also Problem 1.5.5.

We can generalize the quadratic equation, defined by (1.50), to give a *polynomial* equation of degree n,

$$a_0z^n + a_1z^{n-1} + \cdots + a_{n-1}z + a_n = 0, \tag{1.69}$$

where the coefficients $a_0, a_1, \ldots, a_n$ are all real and a_0 is nonzero. We had to introduce complex numbers to provide a system within which all quadratic equations have solutions. Having done that, it is pleasing to know that polynomial equations of *any* degree have solutions within the system of complex numbers. Its solutions satisfy the property that if z_1 is a solution, so is its complex conjugate. I will not pursue this further here.

The simplest equation of the form (1.69) for a general value of n is $z^n = 1$. We will see that this equation has n solutions, which are all complex numbers. Obviously, $z^2 = 1$ has the two solutions $z = \pm 1$, and it is easily verified that $z^4 = 1$ has the solutions $z = \pm 1$ and $z = \pm i$. The solutions of $z^n = 1$ for some other small values of n can be found by elementary methods. (See, for example, Problem 1.5.4.) However, we will now see how we can find the solutions of $z^n = 1$ for *any* positive integer n. First we need a formal definition of angle.

Definition 1.5.4 The *angle* BAC in Figure 1.17, denoted by θ, is defined as the ratio of the length of the circular arc BC to the length of the line segment AB, the radius of the circular arc. ■

The symbol θ denotes the Greek letter *theta*. Angles are sometimes measured in degrees, where 90 degrees corresponds to a right angle. This is not a fundamental way of measuring angles, since the number 90 is an arbitrary choice. Definition 1.5.4 gives the natural way of measuring angles, which

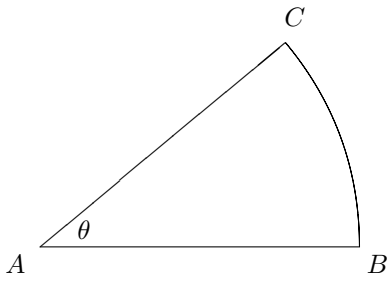


FIGURE 1.17. The angle BAC is defined as the ratio of the length of the circular arc BC to the length of the radius AB.

is called *radian* measure. In the case in which the circular arc BC has the same length as AB, the angle θ is one radian. The constant π is defined as the ratio of the circumference of a circle to its diameter, the symbol π being the Greek letter *pi*. It then follows from the definition of angle that the right angle has measure $\frac{\pi}{2}$ radians. The area bounded by AB, AC, and the circular arc BC is called a *sector*. Let $|AB| = |AC| = r$. Then the area of the sector ABC, which is obviously proportional to the angle θ, is equal to $\frac{1}{2}\theta r^2$. Thus, when $\theta = 2\pi$, we correctly obtain πr^2 for the area of the circle of radius r.

We will now define the sine and cosine functions, $\sin\theta$ and $\cos\theta$, with the aid of Figure 1.18, where P is a point on the circumference of the circle that has radius r, and whose center is the point O, with coordinates $(0, 0)$. The point P has coordinates (x, y), and the line OP makes an angle θ with the x-axis. For the point P displayed in Figure 1.18, the x-coordinate is negative and the y-coordinate is positive. (More on this topic is given in Section 6.4.)

Note that *positive* angles are measured in a counterclockwise direction from the x-axis, and *negative* angles are measured in a clockwise direction. Then, for any real value of θ, we define

$$\sin\theta = \frac{y}{r} \quad \text{and} \quad \cos\theta = \frac{x}{r}. \tag{1.70}$$

The functions $\sin\theta$ and $\cos\theta$ are called *circular* functions, after the way they are defined in (1.70). It follows from (1.70) that the values of $\sin\theta$ and $\cos\theta$ are unchanged if we replace θ by $\theta + 2k\pi$, where k is any integer. We say that the sine and cosine functions are *periodic*, with period 2π. It also follows from (1.70), on putting $x = 0$ and $x = \frac{\pi}{2}$, that

$$\sin 0 = 0, \quad \cos 0 = 1, \quad \sin\frac{\pi}{2} = 1, \quad \cos\frac{\pi}{2} = 0. \tag{1.71}$$

We also see from (1.70) that for all real values of θ,

$$\sin(-\theta) = -\sin\theta \quad \text{and} \quad \cos(-\theta) = \cos\theta. \tag{1.72}$$

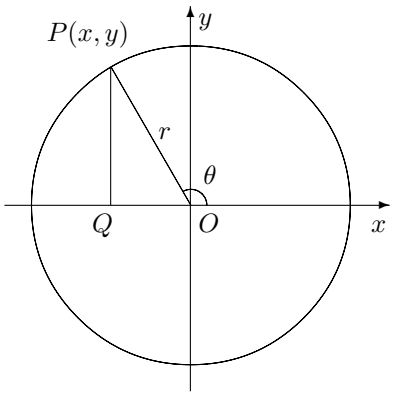


FIGURE 1.18. Circular functions: $\sin\theta = \frac{y}{r}$, $\cos\theta = \frac{x}{r}$.

In Figure 1.18 we note from Pythagoras's theorem that $x^2 + y^2 = r^2$. On dividing this last equation throughout by r^2, we see from (1.70) that

$$\cos^2\theta + \sin^2\theta = 1. \tag{1.73}$$

Following the usual custom, we have written $\cos^2\theta$ and $\sin^2\theta$ to denote $(\cos\theta)^2$ and $(\sin\theta)^2$, respectively.

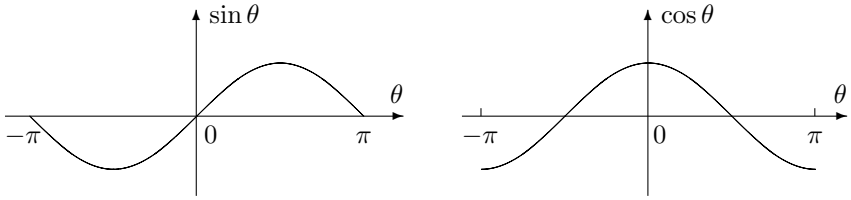


FIGURE 1.19. The functions $\sin\theta$ and $\cos\theta$ on the interval $[-\pi, \pi]$.

The graphs of the sine and cosine functions on the interval $[-\pi, \pi]$ are shown in Figure 1.19, and these shapes are copied endlessly to the left and to the right, following the periodic property of these functions. The graphs of the sine and cosine on $(-\infty, \infty)$ look the same. Indeed, we can see by considering how we defined the sine and cosine, using Figure 1.18, that

$$\sin\left(\theta + \frac{\pi}{2}\right) = \cos\theta. \tag{1.74}$$

We say that each graph is a *translation* of the other. The sine and cosine functions satisfy many identities, including the following two that we will

use in our further study of complex numbers:

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta, \tag{1.75}$$

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta. \tag{1.76}$$

The identities (1.75) and (1.76), which hold for all real values of α and β, can be verified by elementary geometrical arguments.

Given any nonzero complex number $z = x + yi$, we can write

$$z = r\left(\frac{x}{r} + \frac{y}{r}\,i\right), \quad \text{where} \quad r = \sqrt{x^2+y^2} = |z|.$$

Thus any nonzero complex number $z = x+yi$ can be expressed in the form

$$z = x + yi = r\left(\cos\theta + i\sin\theta\right), \tag{1.77}$$

where

$$\cos\theta = \frac{x}{r}, \quad \sin\theta = \frac{y}{r}, \quad \text{with} \quad r = \sqrt{x^2+y^2}. \tag{1.78}$$

It is the standard practice to write the symbol i before $\sin\theta$, as on the right side of (1.77), instead of after it, as I have done in writing the general complex number as $x + yi$. We call $r\left(\cos\theta + i\sin\theta\right)$ the *polar* form of a complex number. Because of the periodicity of the sine and cosine functions, the choice of θ in (1.77) is not unique. We can obtain a unique value for θ by requiring that it satisfy the inequalities $-\pi < \theta \le \pi$, and call this value of θ the *argument* of z. The following theorem is named after Abraham De Moivre (1667–1754).

Theorem 1.5.1 For any positive integer n we have

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta \tag{1.79}$$

for all real values of θ.

Proof. We will use mathematical induction. The result is obviously true for $n = 1$. Let us assume that (1.79) holds for some value of $n \ge 1$. Then

$$(\cos\theta + i\sin\theta)^{n+1} = (\cos\theta + i\sin\theta)^n \cdot (\cos\theta + i\sin\theta), \tag{1.80}$$

and using (1.79), we obtain

$$(\cos\theta + i\sin\theta)^{n+1} = (\cos n\theta + i\sin n\theta)\,(\cos\theta + i\sin\theta) = X + Y\,i,$$

say, where

$$X = \cos n\theta\cos\theta - \sin n\theta\sin\theta \quad \text{and} \quad Y = \sin n\theta\cos\theta + \cos n\theta\sin\theta.$$

On putting $\alpha = n\theta$ and $\beta = \theta$ in (1.75) and (1.76), we find that X and Y can be written more simply as $X = \cos(n+1)\theta$ and $Y = \sin(n+1)\theta$.

Thus (1.79) holds when n is replaced by $n+1$, and this completes the proof by mathematical induction. ■

Now let us find the solutions of the equation $z^n = 1$. It follows from (1.64) that $|z^n| = |z|^n$, so that $|z| = 1$. Thus $z = \cos\theta + i\sin\theta$, for some value of θ. Then, by Theorem 1.5.1, we obtain

$$z^n = \cos n\theta + i\sin n\theta = 1, \tag{1.81}$$

so that

$$\cos n\theta = 1 \quad \text{and} \quad \sin n\theta = 0. \tag{1.82}$$

In view of (1.71) and the periodicity of the sine and cosine functions, we deduce from (1.82) that $n\theta = 2k\pi$, where k is any integer. Consequently, $\theta = 2k\pi/n$, and all solutions of (1.81) are of the form

$$z = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}, \tag{1.83}$$

where k is an integer. The choice of $k = 0, 1, \ldots, n-1$ gives n distinct solutions, and because of the periodicity of the sine and cosine, no further choice of k yields any more solutions. We call the n solutions of $z^n = 1$ the nth roots of unity. They are equally spaced on the circumference of the circle $|z| = 1$, one root always being on the x-axis, corresponding to the root $z = 1$. The case $n = 7$ is illustrated in Figure 1.20.

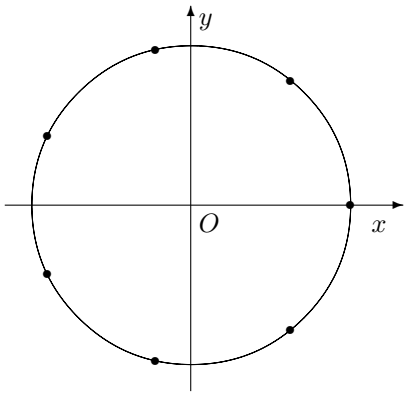


FIGURE 1.20. The seventh roots of unity.

Problem 1.5.1 Verify that the equation $z^2 - z + 1 = 0$ has solutions

$$z_1 = \frac{1}{2} + \frac{\sqrt{3}}{2}i \quad \text{and} \quad z_2 = \frac{1}{2} - \frac{\sqrt{3}}{2}i.$$

Evaluate $|z_1|$, $|z_2|$, $z_1 z_2$, and $|z_1 z_2|$.

Problem 1.5.2 With z_1 and z_2 as defined in Problem 1.5.1, show that

$$z_1^3 = z_2^3 = -1.$$

Problem 1.5.3 Show that the equation $z^2 + z + 1 = 0$ has solutions

$$z_3 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i \quad \text{and} \quad z_4 = -\frac{1}{2} - \frac{\sqrt{3}}{2}i.$$

Verify that $z_1^2 = z_3$ and $z_2^2 = z_4$, where z_1 and z_2 are defined in Problem 1.5.1, and that $z_3^3 = z_4^3 = 1$.

Problem 1.5.4 Show that

$$z^6 - 1 = (z^3 - 1)(z^3 + 1) = (z - 1)(z^2 + z + 1)(z + 1)(z^2 - z + 1).$$

Using the results obtained in the three problems above, show that the set of solutions of $z^6 - 1 = 0$ is $\{z_1, z_1^2, z_1^3, z_1^4, z_1^5, z_1^6\}$, where $z_1 = \frac{1}{2} + \frac{\sqrt{3}}{2}i$.

Problem 1.5.5 Deduce from the inequality $(x_1y_2 - x_2y_1)^2 \geq 0$ that

$$(x_1x_2 + y_1y_2)^2 \leq \left(x_1^2 + y_1^2\right)\left(x_2^2 + y_2^2\right)$$

and thus verify that

$$|x_1x_2 + y_1y_2| \leq |z_1|\,|z_2|,$$

so that

$$2x_1x_2 + 2y_1y_2 \leq 2|z_1|\,|z_2|,$$

where $z_1 = x_1 + y_1i$ and $z_2 = x_2 + y_2i$. By adding $x_1^2 + y_1^2 + x_2^2 + y_2^2$ to both sides of the latter inequality, show that

$$|z_1 + z_2|^2 \leq (|z_1| + |z_2|)^2.$$

Problem 1.5.6 Verify that

$$|z_1 - z_2|^2 = (z_1 - z_2)\,(\bar{z_1} - \bar{z_2}) = |z_1|^2 - 2\,\mathbf{Re}\,(z_1z_2) + |z_2|^2$$

and hence prove that

$$|z_1 - z_2|^2 \geq (|z_1| - |z_2|)^2.$$

Problem 1.5.7 Let $p(z)$ denote a polynomial of degree n in z with real coefficients. Show that the complex conjugate of $p(z)$ is $p(\bar{z})$. Deduce that $p(\alpha) = 0$ if and only if $p(\bar{\alpha}) = 0$, and that the complex zeros of $p(z)$ occur in conjugate pairs.