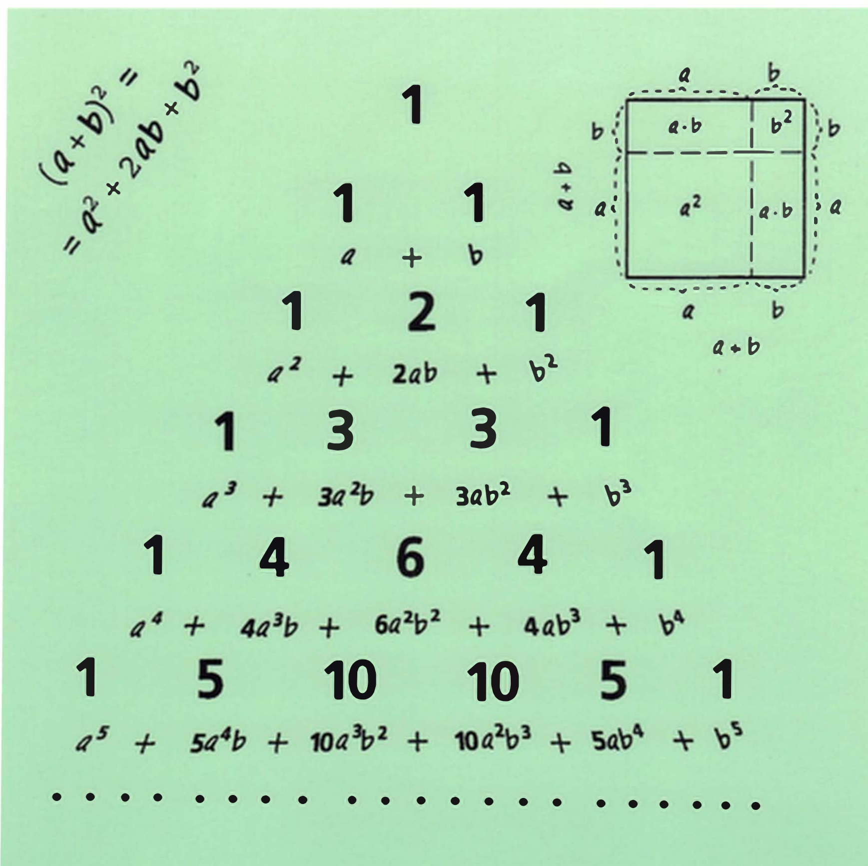
I.M. Gelfand A. Shen

Algebra



I. M. Gelfand A. Shen

Algebra

Myriad Mathematical Library

Birkhäuser
Boston · Basel · Berlin

Contents

1 Introduction

This book is about algebra. This is a very old science and its gems have lost their charm for us through everyday use. We have tried in this book to refresh them for you.

The main part of the book is made up of problems. The best way to deal with them is: Solve the problem by yourself – compare your solution with the solution in the book (if it exists) – go to the next problem. However, if you have difficulties solving a problem (and some of them are quite difficult), you may read the hint or start to read the solution. If there is no solution in the book for some problem, you may skip it (it is not heavily used in the sequel) and return to it later.

The book is divided into sections devoted to different topics. Some of them are very short, others are rather long.

Of course, you know arithmetic pretty well. However, we shall go through it once more, starting with easy things.

2 Exchange of terms in addition

Let's add 3 and 5:

$$3 + 5 = 8.$$

And now change the order:

$$5 + 3 = 8.$$

We get the same result. Adding three apples to five apples is the same as adding five apples to three – apples do not disappear and we get eight of them in both cases.

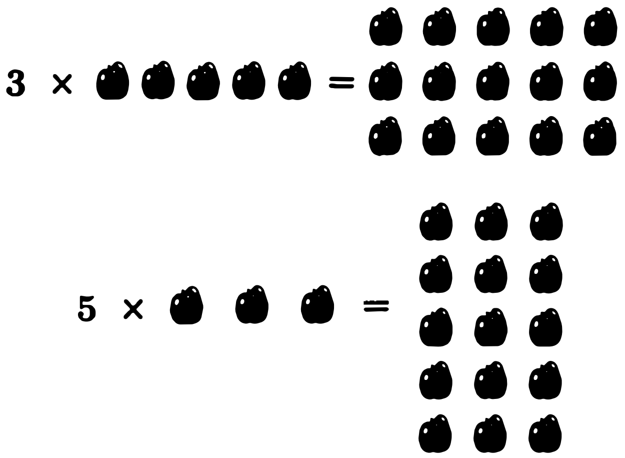
3 Exchange of terms in multiplication

Multiplication has a similar property. But let us first agree on notation. Usually in arithmetic, multiplication is denoted as "$\times$". In algebra this sign is usually replaced by a dot "$\cdot$". We follow this convention.

Let us compare $3 \cdot 5$ and $5 \cdot 3$. Both products are 15. But it is not so easy to explain why they are equal. To give each of three boys five apples is not the same as to give each of five boys three apples – the situations differ radically.

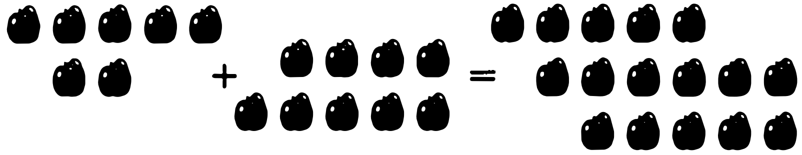
One of the authors of this book asked a seven-year-old girl, "How much is two times four?" "Eight", she answered immediately. "And four times two?" She started thinking, trying to add $2+2+2+2$. A year later she would know very well that the product remains the same when we exchange factors and she would forget that it was not so evident before.

The simplest way to explain why $5 \cdot 3 = 3 \cdot 5$ is to show a picture:



4 Addition in the decimal number system

If we want to know how much $7+9$ is, we may draw 7 apples and then 9 apples near them:



and then count all the apples together: one, two, three, four, ..., fifteen, sixteen. We get $7+9=16$. This method can be applied for any numbers; however, you need a lot of patience to try it on, say, 137 and 268. So mathematicians invented other methods. One of them is the standard addition method used in the positional number system.

In different countries and at different times, people used different notations for numbers, and entire books are written about them. We are so used to the familiar decimal number system using the digits $0, 1, 2, \ldots, 8, 9$ that we don't realize how unbelievably convenient this convention has proved to be. Even the possibility of writing down very big numbers quickly was not self-evident for ancient people. A great mathematician of ancient Greece, Archimedes, even wrote a book called *The Sand Reckoner*. The main point of the book was to show that it is possible to write down the number that is greater than the number of sand particles filling the sphere whose radius is the distance between Earth and the stars.

Now the decimal number system has no rivals – except the binary number system, which is popular among computers, not people. This binary system has only two digits, 0 and 1 – but numbers have more digits. The computer does not worry about the length of numbers, but still wishes to keep rules of operation as simple as possible.

We shall speak about the binary system in another section, but now we return to our ordinary decimal system and to the addition method. We shall not explain it to you once more – you know it without us. Let us solve some problems instead.

Problem 1. Several digits "8" are written and some "+" signs are inserted to get the sum 1000. Figure out how it is done. (For example, if we try $88 + 88 + 8 + 8 + 88$, we fail because we get only 280 instead of 1000.)

Solution. Assume that

$$\begin{array}{r} \ldots 8 \\ \ldots. \\ \underline{\ldots 8} \\ 1000 \end{array}$$

We do not know how many rows are here nor how many digits are used in each number. But we do know that each number ends with "8" and that the last digit of the sum is zero. How many numbers do we need to get this zero? If we use only one number, we get 8. If we use two numbers, we get 6 $(8 + 8 = 16)$, etc. To get zero we need at least five numbers:

$$\begin{array}{r} \ldots 8 \\ \ldots 8 \\ \ldots 8 \\ \ldots 8 \\ \underline{\ldots 8} \\ 1000 \end{array}$$

After we get this zero, we keep "4" in mind because $8+8+8+8+8 = 40$. To get the next zero in the "tens place" from this "4", we need to add at least two 8's since $4+8+8 = 20$.

$$\begin{array}{r} 8 \\ 8 \\ 8 \\ ..88 \\ \underline{..88} \\ 1000 \end{array}$$

We keep "2" in mind and we need only one more "8" to get 10:

$$\begin{array}{r} 8 \\ 8 \\ 8 \\ 88 \\ \underline{888} \\ 1000 \end{array}$$

The problem is solved: $8+8+8+88+888 = 1000$.

Problem 2. In the addition example

$$\begin{array}{r} \mathrm{AAA} \\ \underline{\mathrm{BBB}} \\ \mathrm{AAAC} \end{array}$$

all A's denote some digit, all B's denote another digit and C denotes a third digit. What are these digits?

Solution. First of all A denotes 1 because no other digit can appear as a carry in the thousands position of the result. To find what B is let us ask ourselves: Do we get a (nonzero) carry adding the rightmost A and B? If we had no carry, we would get the same digit in the other two places (tens and hundreds), but this is not so. Therefore, the carry digit is not zero, and this is possible only if $B = 9$. Therefore we get the answer:

$$\begin{array}{r} 111 \\ \underline{999} \\ 1110 \end{array}$$

5 The multiplication table and the multiplication algorithm

To compute the product of, say, 17 and 38, we may draw a picture of 17 rows, each containing 38 points, and then count all the points. But of course, nobody does this – we know an easier method of multiplying using the positional system.

This method (called the *multiplication algorithm*) is based on the multiplication table for digits and requires that you memorize the table. There is – sorry! – no way around it, and if, on being asked, "What is seven times eight?" in the middle of the night, you cannot answer "Fifty-six!" immediately, and instead try to add up seven eights half-asleep, we are unable to help you.

There is some good news, however. You don't need to memorize the product $17 \cdot 38$. Instead, you can compute it in two different ways:

$$\begin{array}{r} 17 \\ \underline{38} \\ 136 \\ \underline{51} \\ 646 \end{array} \qquad\qquad \begin{array}{r} 38 \\ \underline{17} \\ 266 \\ \underline{38} \\ 646 \end{array}$$

Both results are equal, though the intermediate results are different. A lucky coincidence, isn't it?

Here are some problems concerning multiplication.

Problem 3. A boy claims that he can multiply any three-digit number by 1001 instantly. If his classmate says to him "715" he gives the answer immediately. Compute this answer and explain the boy's secret.

Problem 4. Multiply 101010101 by 57.

Problem 5. Multiply 10001 by 1020304050.

Problem 6. Multiply 11111 by 1111.

Problem 7. A six-digit number having 1 as its leftmost digit becomes three times bigger if we take this digit off and put it at the end of the number. What is this number?

Solution. Look at the multiplication procedure:

$$\begin{array}{r} 1\,A\,B\,C\,D\,E \\ 3 \\ \hline A\,B\,C\,D\,E\,1 \end{array}$$

Here A,B,C,D and E denote some digits (we do not know whether all these digits are different or not). Digit E must be equal to 7, because among the products $3 \times 0 = 0$, $3 \times 1 = 3$, $3 \times 2 = 6$, $3 \times 3 = 9$, $3 \times 4 = 12$, $3 \times 5 = 15$, $3 \times 6 = 18$, $3 \times 7 = 21$, $3 \times 8 = 24$, $3 \times 9 = 27$ only $3 \times 7 = 21$ has the last digit 1. So we get:

$$\begin{array}{r} 1\,A\,B\,C\,D\,7 \\ 3 \\ \hline A\,B\,C\,D\,7\,1 \end{array}$$

When multiplying 7 by 3 we get a carry of 2, so $3 \times D$ must have its last digit equal to 5. This is possible only if $D = 5$:

$$\begin{array}{r} 1\,A\,B\,C\,5\,7 \\ 3 \\ \hline A\,B\,C\,5\,7\,1 \end{array}$$

In the same way, we find that $C = 8$, $B = 2$, $A = 4$. So we get the solution:

$$\begin{array}{r} 1\,4\,2\,8\,5\,7 \\ 3 \\ \hline 4\,2\,8\,5\,7\,1 \end{array}$$

6 The division algorithm

Division is the most complicated thing among all the four arithmetic operations. To make yourself confident, you may try the following problems.

Problem 8. Divide 123123123 by 123. (Check your answer by multiplication!)

Problem 9. Can you predict the remainder when $111\ldots1$ (100 ones) is divided by 1111111?

Problem 10. Divide $1000\ldots0$ (20 zeros) by 7.

Problem 11. While solving the two preceding problems you may have discovered that quotient digits (and remainders) became periodic:

$$
\begin{array}{l}
\overbrace{142857}14\ldots\\
7\overline{)100000000\ldots}\\
\underline{7}\\
30\\
\underline{28}\\
20\\
\underline{14}\\
60\\
\underline{56}\\
40\\
\underline{35}\\
50\\
\underline{49}\\
10\\
\underline{7}\\
30\\
\underline{28}\\
2\ldots
\end{array}
$$

Is it just a coincidence, or will this pattern repeat?

Problem 12. Divide $2000\ldots000$ (20 zeros), $3000\ldots000$ (20 zeros), $4000\ldots000$ (20 zeros), etc. by 7. Compare the answers you get and explain what you see.

A multiplication fan may enjoy the following problem:

Problem 13. Multiply 142857 by 1, 2, 3, 4, 5, 6, 7, and look at the results. (It is easy to memorize these results and become a famous number cruncher who is able to multiply a random number, for example, 142857, by almost any digit!)

Problem 14. Try to invent similar tricks based on the division of $1000\ldots0$ by other numbers instead of 7.

7 The binary system

Problem 15. Find a generating rule, and write five or ten more lines:

0
1
10
11
100
101
110
111
1000
1001
1010
1011
1100
...

Problem 16. You have weights of 1, 2, 4, 8, and 16 grams. Show that it is possible to get any weight from 0 to 31 grams using the following table ("+" means "the weight is used", "−" means "not used"):

	A					B	C
	16	8	4	2	1		
0	−	−	−	−	−	00000	0
1	−	−	−	−	+	00001	1
2	−	−	−	+	−	00010	10
3	−	−	−	+	+	00011	11
4	−	−	+	−	−	00100	100
5	−	−	+	−	+	00101	101
6	−	−	+	+	−	00110	110
7	−	−	+	+	+	00111	111
8	−	+	−	−	−	01000	1000
9	−	+	−	−	+	01001	1001
10	−	+	−	+	−	01010	1010
11	−	+	−	+	+	01011	1011
					...		

We can replace "−" by 0 and "+" by 1 (column B) and omit the leading zeros (column C). Then we get the same result as in the preceding problem.

This table is called a conversion table between decimal and binary number systems:

Decimal	Binary
0	0
1	1
2	10
3	11
4	100
5	101
6	110
7	111
8	1000
9	1001
10	1010
11	1011
12	1100
...	...

Problem 17. What corresponds to 14 in the right column? What corresponds to 10000 in the left column?

The binary system has an advantage: you don't need to memorize as many as 10 digits; two is enough. But it has a disadvantage also: numbers are too long. (For example, 1024 is 10000000000 in binary.)

Problem 18. How is 45 (decimal) written in the binary system?

Problem 19. What (decimal) number is written as 10101101 in binary?

Problem 20. Try the usual addition method in binary version:

$$\begin{aligned} 1010 &+ 101 = ? \\ 1111 &+ 1 = ? \\ 1011 &+ 1 = ? \\ 1111 &+ 1111 = ? \end{aligned}$$

Check your answers, converting all the numbers (the numbers being added and the sums) into the decimal system.

Problem 21. Try the usual subtraction algorithm in its binary version:

$$
\begin{aligned}
1101 &- 101 = ?\\
110 &- 1 = ?\\
1000 &- 1 = ?
\end{aligned}
$$

Check your answers, converting all the numbers into the decimal system.

Problem 22. Now try to multiply 1101 and 1010 (in binary):

$$
\begin{array}{r}
1101\\
\underline{1010}\\
????
\end{array}
$$

Check your result, converting the factors and the product into the decimal system.

Hint: Here are two patterns:

$$
\begin{array}{r}
1011\\
\underline{11}\\
1011\\
\underline{1011}\\
100001
\end{array}
\qquad
\begin{array}{r}
1011\\
\underline{101}\\
1011\\
\underline{1011}\\
110111
\end{array}
$$

Problem 23. Divide 11011 (binary) by 101 (binary) using the ordinary method. Check your result, converting all numbers into the decimal system.

Hint: Here is a pattern:

$$
\begin{array}{rl}
110 & \leftarrow \text{ the quotient}\\
100\overline{)11001} & \\
\underline{100} & \\
100 & \\
\underline{100} & \\
1 & \leftarrow \text{ the remainder}
\end{array}
$$

Problem 24. In the decimal system the fraction 1/3 is written as 0.333.... What happens with 1/3 in the binary system?

8 The commutative law

Let us return to the rule "exchange of terms in addition does not change the sum". It can be written as

$$\text{First term} + \text{Second term} = \text{Second term} + \text{First term}$$

or in short

$$\text{F.t.} + \text{S.t.} = \text{S.t.} + \text{F.t.}$$

But even this short form seems too long for mathematicians, and they use single letters such as a or b instead of "F.t." and "S.t.". So we get

$$\boxed{a + b = b + a}$$

The law "exchange of factors does not change the product" can be written now as

$$\boxed{a \cdot b = b \cdot a}$$

Here "$\cdot$" is a multiplication symbol. Often it is omitted:

$$ab = ba$$

The property $a + b = b + a$ is called the *commutative law for addition*; the property $ab = ba$ is called the *commutative law for multiplication*.

Remark. Sometimes it is impossible to omit the multiplication sign $(\cdot)$ in a formula; for example, $3 \cdot 7 = 21$ is not the same as $37 = 21$. By the way, multiplication had good luck in getting different symbols: the notations $a \times b$, $a \cdot b$, ab, and a*b (in computer programming) are all used.

9 The associative law

Now let us add three numbers instead of two:

$$3 + 5 + 11 = 8 + 11 = 19.$$

But there is another way:

$$3 + 5 + 11 = 3 + 16 = 19.$$

Usually parentheses are used to show the desired order of operations:

$$(3+5)+11$$

means that we have to add 3 and 5 first, and

$$3+(5+11)$$

means that we have to add 5 and 11 first.

The result does not depend on the order of the operations. This fact is called the *associative law* by mathematicians. In symbols:

$$\boxed{(a+b)+c=a+(b+c)}$$

If you would like to have a real-life example, here it is. You can get sweet coffee with milk if you add milk to the coffee with sugar or if you add sugar to the coffee with milk. You get the same result – and this is the associative law:

$$(\text{sugar}+\text{coffee})+\text{milk} = \text{sugar}+(\text{coffee}+\text{milk})$$

Problem 25. Try it.

Problem 26. Add $357 + 17999 + 1$ without paper and pencil.

Solution. It is not so easy to add 357 and 17999. But if you add $17999 + 1$, you get 18000 and now it is easy to add 357:

$$357 + (17999 + 1) = 357 + 18000 = 18357.$$

Problem 27. Add $357 + 17999$ without paper and pencil.

Solution. $357 + 17999 = (356 + 1) + 17999 = 356 + (1 + 17999) = 356 + 18000 = 18356$.

Problem 28. Add $899 + 1343 + 101$.

Hint. Remember the commutative law.

Multiplication is also associative:

$$(a \cdot b) \cdot c = a \cdot (b \cdot c)$$

or, in short,

$$(ab)c = a(bc).$$

Problem 29. Compute $37 \cdot 25 \cdot 4$.

Problem 30. Compute $125 \cdot 37 \cdot 8$.

10 The use of parentheses

A pedant is completely right saying that a notation like

$$2 \cdot 3 \cdot 4 \cdot 5$$

has no sense until we fix the order of operations. Even if we agree not to permute the factors, we have a lot of possibilities:

$$\begin{aligned}
((2 \cdot 3) \cdot 4) \cdot 5 &= (6 \cdot 4) \cdot 5 = 24 \cdot 5 = 120 \\
(2 \cdot (3 \cdot 4)) \cdot 5 &= (2 \cdot 12) \cdot 5 = 24 \cdot 5 = 120 \\
(2 \cdot 3) \cdot (4 \cdot 5) &= 6 \cdot 20 = 120 \\
2 \cdot ((3 \cdot 4) \cdot 5) &= 2 \cdot (12 \cdot 5) = 2 \cdot 60 = 120 \\
2 \cdot (3 \cdot (4 \cdot 5)) &= 2 \cdot (3 \cdot 20) = 2 \cdot 60 = 120
\end{aligned}$$

Problem 31. Find all possible ways to put parentheses in the product $2 \cdot 3 \cdot 4 \cdot 5 \cdot 6$ (not changing the order of factors; see the example just shown). Try to invent a systematic way of searching so as not to forget any possibilities.

Problem 32. How many "(" and ")" symbols do you need to specify completely the order of operations in the product

$$2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdots 99 \cdot 100\,?$$

The parentheses are often omitted because the result is independent of the order of the operations. The reader may reconstruct them as he or she wishes.

The following problem shows what can be achieved by clever permutation and grouping.

Problem 33. Compute $1 + 2 + 3 + 4 + \cdots + 98 + 99 + 100$.

Solution. Group the 100 terms in 50 pairs: $1+2+3+4+\cdots+98+99+100 = (1+100)+(2+99)+(3+98)+\cdots+(49+52)+(50+51)$. Each pair has the sum 101. We have 50 pairs, so the total sum is $50 \cdot 101 = 5050$.

A legend says that as a schoolboy Karl Gauss (later a great German mathematician) shocked his school teacher by solving this problem instantly (as the teacher was planning to relax while the children were busy adding the hundred numbers).

11 The distributive law

There is one more law for addition and multiplication, called the *distributive law.* If two boys and three girls get 7 apples each, then the boys get $2 \cdot 7 = 14$ apples, the girls get $3 \cdot 7 = 21$ apples – and together they get

$$2 \cdot 7 + 3 \cdot 7 = 14 + 21 = 35$$

apples. The same answer can be computed in another way: each of $2 + 3 = 5$ children gets 7 apples, so the total number of apples is

$$(2 + 3) \cdot 7 = 5 \cdot 7 = 35\,.$$

Therefore,

$$(2 + 3) \cdot 7 = 2 \cdot 7 + 3 \cdot 7$$

and, in general,

$$(a + b) \cdot c = a \cdot c + b \cdot c$$

This property is called the *distributive law.* Changing the order of factors we may also write

$$c \cdot (a + b) = c \cdot a + c \cdot b$$

Problem 34. Compute $1001 \cdot 20$ without pencil and paper.

Solution. $1001 \cdot 20 = (1000 + 1) \cdot 20 = 1000 \cdot 20 + 1 \cdot 20 = 20{,}000 + 20 = 20{,}020$.

Problem 35. Compute $1001 \cdot 102$ without pencil and paper.

Solution. $1001 \cdot 102 = 1001 \cdot (100 + 2) = 1001 \cdot 100 + 1001 \cdot 2 = (1000 + 1) \cdot 100 + (1000 + 1) \cdot 2 = 100{,}000 + 100 + 2000 + 2 = 102{,}102$.

The distributive law is a rule for removing brackets or parentheses. Let us see how it is used to transform the product of two sums

$$(a + b)(m + n)\,.$$

The number $(m + n)$ is the sum of the two numbers m and n and can replace c in the distributive law above:

$$(a + b) \cdot \boxed{c} = a \cdot \boxed{c} + b \cdot \boxed{c}$$

$$(a + b) \cdot \boxed{m + n} = a \cdot \boxed{m + n} + b \cdot \boxed{m + n}$$

Now we remember that $\boxed{m+n}$ is the sum of m and n and continue:

$$\ldots = a(m+n) + b(m+n) = am + an + bm + bn\,.$$

The general rule: To multiply two sums you need to multiply each term of the first sum by each term of the second one and then add all the products.

Problem 36. How many additive terms would be in

$$(a+b+c+d+e)(x+y+z)$$

after we use this rule?

12 Letters in algebra

In algebra we gradually make more and more use of letters (such as $a, b, c, \ldots, x, y, z$, etc.). Traditionally the use of letters (x's) is considered one of the most difficult topics in the school mathematics curriculum. Many years ago primary school pupils studied "arithmetic" (with no x's) and secondary school pupils started with "algebra" (with x's). Later "arithmetic" was renamed "mathematics" and x's were introduced (and created a mess, some people would say).

We hope that you, dear reader, never had difficulties understanding "what all these letters mean", but we still wish to give you some advice. If you ever want to explain the meaning of letters to your classmates, brothers and sisters, your parents, or your children (some day), just say that the letters are abbreviations for words. Let us explain what we mean.

In the equality

$$a+b=b+a$$

the letters a and b mean "the first term" and "the second term". When we write $a+b=b+a$ we mean that any numbers substituted instead of a and b give a true assertion. Therefore, $a+b=b+a$ can be considered as a unified short version of the equalities $1+7=7+1$ or $1028+17=17+1028$ as well as infinitely many other equalities of the same type.

Another example of the use of letters:

Problem 37. A small vessel and a big vessel contain (together) 5 liters. Two small and three big vessels contain together 13 liters. What are the volumes of the vessels?

Solution. (The "arithmetic" one.) The small and big vessels together contain 5 liters. Therefore, two small vessels and two big vessels together contain 10 liters ($10 = 2 \cdot 5$). As we know, two small vessels and three big vessels contain 13 liters. So we get 13 liters instead of 10 by adding one big vessel. Therefore the volume of a big vessel is 3 liters. Now it is easy to find the volume of a small vessel: together they contain 5 liters, so a small vessel contains $5 - 3 = 2$ liters. Answer: The volume of a small vessel is 2 liters, the volume of a big vessel is 3 liters.

This solution can be shortened if we use "Vol.SV" instead of "Volume of a Small Vessel" and "Vol.BV" instead of "Volume of a Big Vessel". Thus, according to the statement of the problem,

$$\text{Vol.SV} + \text{Vol.BV} = 5,$$

therefore

$$2 \cdot \text{Vol.SV} + 2 \cdot \text{Vol.BV} = 10\,.$$

We know also that

$$2 \cdot \text{Vol.SV} + 3 \cdot \text{Vol.BV} = 13\,.$$

If we subtract the preceding equality from the last one we find that Vol.BV = 3. Now the first equality implies that Vol.SV = $5 - 3 = 2$.

Now the only thing to do is to replace our "Vol.SV" and "Vol.BV" by standard unknowns x and y – and we get the standard "algebraic" solution of our problem. Here it is: Denote the volume of a small vessel by x and the volume of a big vessel by y. We get the following system of equations:

$$\begin{aligned} x + y &= 5 \\ 2x + 3y &= 13\,. \end{aligned}$$

Multiplying the first equation by 2 we get

$$2x + 2y = 10$$

and subtracting the last equation from the second equation of our system we get

$$y = 13 - 10 = 3.$$

Now the first equation gives

$$x = 5 - y = 5 - 3 = 2.$$

Answer: $x = 2$, $y = 3$.

Finally, one more example of the use of letters in algebra.

"**Magic trick**". Choose any number you wish. Add 3 to it. Multiply the result by 2. Subtract the chosen number. Subtract 4. Subtract the chosen number once more. You get 2, don't you?

Problem 38. Explain why this trick is successful.

Solution. Let us follow what happens with the chosen number (we denote it by x):

Choose the number you wish	x
add 3 to it	$x + 3$
multiply the result by 2	$2 \cdot (x + 3) = 2x + 6$
subtract the chosen number	$(2x + 6) - x = x + 6$
subtract 4	$(x + 6) - 4 = x + 2$
subtract the chosen number once more. You get 2.	$(x + 2) - x = 2$

13 The addition of negative numbers

It is easy to check that $3 + 5 = 8$: just take three apples, add five apples, and count all the apples together: "one, two, three, four, ..., seven, eight". But how can we check that $(-3) + (-5) = (-8)$ or that $3 + (-5) = (-2)$? Usually this is explained by examples like the following two:

$3+5=8$	Yesterday it was +3. Today the temperature is 5 degrees warmer and is 8 degrees.
$(-3)+5=2$	Yesterday it was −3 degrees. Today it is 5 degrees warmer, that is, +2.
$3+(-5)=-2$	Yesterday was +3, today it is 5 degrees colder, that is, −2.
$(-3)+(-5)=(-8)$	Yesterday was −3, today it is 5 degrees colder, that is, −8.

(Here all temperatures are measured in Celsius degrees.)

Here is another example:

$3+5=8$	Three protons + five protons = = eight protons.
$(-3)+5=2$	Three antiprotons + five protons = = two protons (ignoring γ-radiation).
$3+(-5)=-2$	Three protons + five antiprotons = = two antiprotons (ignoring γ-radiation).
$(-3)+(-5)=(-8)$	Three antiprotons + five antiprotons = = eight antiprotons.

(Protons and antiprotons are elementary particles. When a proton meets an antiproton they annihilate one another, producing gamma radiation.)

14 The multiplication of negative numbers

To find how much three times five is, you add three numbers equal to five:

$$5+5+5=15\,.$$

The same explanation may be used for the product $1\cdot5$ if we agree that a sum having only one term is equal to this term. But it is evidently not applicable to the product $0\cdot5$ or $(-3)\cdot5$: can you imagine a sum with zero or with minus three terms?

However, we may exchange the factors:

$$5 \cdot 0 = 0 + 0 + 0 + 0 + 0 = 0,$$

$$5 \cdot (-3) = (-3) + (-3) + (-3) + (-3) + (-3) = -15.$$

So if we want the product to be independent of the order of factors (as it was for positive numbers) we must agree that

$$0 \cdot 5 = 0, \quad (-3) \cdot 5 = -15.$$

Now let us consider the product $(-3) \cdot (-5)$. Is it equal to -15 or to $+15$? Both answers may have advocates. From one point of view, even one negative factor makes the product negative – so if both factors are negative the product has a very strong reason to be negative. From the other point of view, in the table

$3 \cdot 5 = +15$	$3 \cdot (-5) = -15$
$(-3) \cdot 5 = -15$	$(-3) \cdot (-5) = ?$

we already have two minuses and only one plus; so the "equal opportunities" policy requires one more plus. So what?

Of course, these "arguments" are not convincing to you. School education says very definitely that minus times minus is plus. But imagine that your small brother or sister asks you, "Why?" (Is it a caprice of the teacher, a law adopted by Congress, or a theorem that can be proved?) You may try to answer this question using the following example:

$3 \cdot 5 = 15$	Getting five dollars three times is getting fifteen dollars.
$3 \cdot (-5) = -15$	Paying a five-dollar penalty three times is a fifteen-dollar penalty.
$(-3) \cdot 5 = -15$	Not getting five dollars three times is not getting fifteen dollars.
$(-3) \cdot (-5) = 15$	Not paying a five-dollar penalty three times is getting fifteen dollars.

Another explanation. Let us write the numbers

$$1,\ 2,\ 3,\ 4,\ 5,\ldots$$

and the same numbers multiplied by three:

$$3,\ 6,\ 9,\ 12,\ 15,\ldots$$

Each number is bigger than the preceding one by three. Let us write the same numbers in the reverse order (starting, for example, with 5 and 15):

$$\begin{array}{rrrrr} 5, & 4, & 3, & 2, & 1 \\ 15, & 12, & 9, & 6, & 3 \end{array}$$

Now let us continue both sequences:

$$\begin{array}{rrrrrrrrrrr} 5, & 4, & 3, & 2, & 1, & 0, & -1, & -2, & -3, & -4, & -5,\ldots \\ 15, & 12, & 9, & 6, & 3, & 0, & -3, & -6, & -9, & -12, & -15,\ldots \end{array}$$

Here -15 is under -5, so $3\cdot(-5) = -15$; plus times minus is minus.

Now repeat the same procedure multiplying $1,2,3,4,5,\ldots$ by -3 (we know already that plus times minus is minus):

$$\begin{array}{rrrrr} 1, & 2, & 3, & 4, & 5 \\ -3, & -6, & -9, & -12, & -15 \end{array}$$

Each number is three units less than the preceding one. Now write the same numbers in the reverse order:

$$\begin{array}{rrrrr} 5, & 4, & 3, & 2, & 1 \\ -15, & -12, & -9, & -6, & -3 \end{array}$$

and continue:

$$\begin{array}{rrrrrrrrrrr} 5, & 4, & 3, & 2, & 1, & 0, & -1, & -2, & -3, & -4, & -5,\ldots \\ -15, & -12, & -9, & -6, & -3, & 0, & 3, & 6, & 9, & 12, & 15,\ldots \end{array}$$

Now 15 is under -5; therefore $(-3)\cdot(-5) = 15$.

Probably this argument would be convincing for your younger brother or sister. But you have the right to ask: So what? Is it possible to *prove* that $(-3)\cdot(-5) = 15$?

Let us tell the whole truth now. Yes, it is possible to prove that $(-3)\cdot(-5)$ *must be* 15 *if* we want the usual properties of addition, subtraction, and multiplication that are true for positive numbers to remain true for any integers (including negative ones).

Here is the outline of this proof: Let us prove first that $3 \cdot (-5) = -15$. What is -15? It is a number opposite to 15, that is, a number that produces zero when added to 15. So we must prove that

$$3 \cdot (-5) + 15 = 0 .$$

Indeed,

$$3 \cdot (-5) + 15 = 3 \cdot (-5) + 3 \cdot 5 = 3 \cdot (-5 + 5) = 3 \cdot 0 = 0 .$$

(When taking 3 out of the parentheses we use the law $ab+ac = a(b+c)$ for $a = 3$, $b = -5$, $c = 5$; we assume that it is true for all numbers, including negative ones.) So $3 \cdot (-5) = -15$. (The careful reader will ask why $3 \cdot 0 = 0$. To tell you the truth, this step of the proof is omitted – as well as the whole discussion of what zero is.)

Now we are ready to prove that $(-3) \cdot (-5) = 15$. Let us start with

$$(-3) + 3 = 0$$

and multiply both sides of this equality by -5:

$$((-3) + 3) \cdot (-5) = 0 \cdot (-5) = 0.$$

Now removing the parentheses in the left-hand side we get

$$(-3) \cdot (-5) + 3 \cdot (-5) = 0,$$

that is, $(-3) \cdot (-5) + (-15) = 0$. Therefore, the number $(-3) \cdot (-5)$ is opposite to -15, that is, is equal to 15. (This argument also has gaps. We should prove first that $0 \cdot (-5) = 0$ and that there is only one number opposite to -15.)

15 Dealing with fractions

If somebody asks you to compare the fractions

$$\frac{3}{5} \quad \text{and} \quad \frac{9}{15},$$

you would answer immediately that they are equal:

$$\frac{9}{15} = \frac{3 \cdot 3}{3 \cdot 5} = \frac{3}{5}.$$

But what would you say now: Are the fractions

$$\frac{221}{391} \quad \text{and} \quad \frac{403}{713}$$

equal or not?

If you remember the multiplication table for two-digit numbers, you would say immediately that they are equal:

$$\frac{221}{391} = \frac{17 \cdot 13}{17 \cdot 23} = \frac{13}{23} = \frac{31 \cdot 13}{31 \cdot 23} = \frac{403}{713}$$

But what are we to do if we do not remember this multiplication table? Then we should find the common denominator for the two fractions,

$$\frac{221}{391} = \frac{221 \cdot 713}{391 \cdot 713} \quad \text{and} \quad \frac{403}{713} = \frac{403 \cdot 391}{713 \cdot 391}$$

and compare numerators,

$$\begin{array}{r}
713\\
\underline{221}\\
713\\
1426\\
\underline{1426}\\
157573
\end{array}
\qquad\qquad
\begin{array}{r}
391\\
\underline{403}\\
1173\\
\underline{1564}\\
157573
\end{array}$$

After that we would know that the fractions are equal but would never discover that in fact they are equal to $13/23$.

Problem 39. Which is bigger, $1/3$ or $2/7$?

Solution. $1/3 = 7/21$, $2/7 = 6/21$, so $1/3 > 2/7$.

The real-life version of this problem says, "Which is better, one bottle for three or two bottles for seven?" It suggests another solution: One bottle for three is equivalent to getting two bottles for six (and not for seven), so $1/3 > 2/7$. In scientific language, we found the "common numerator" instead of the common denominator:

$$\frac{1}{3} = \frac{2}{6} > \frac{2}{7}.$$

Problem 40. Which of the fractions

$$\frac{10001}{10002} \quad \text{and} \quad \frac{100001}{100002}$$

is bigger?

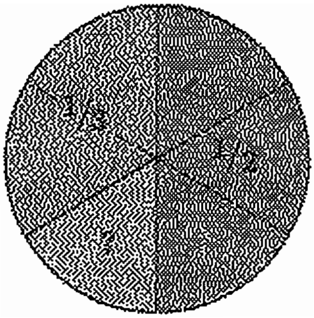
Hint. Both fractions are less than 1. What is the difference between them and 1?

Problem 41. Which of the fractions

$$\frac{12345}{54321} \quad \text{and} \quad \frac{12346}{54322}$$

is bigger?

Finding a common denominator is a traditional problem in teaching arithmetic. How much pie remains for you if your brother wants one-half and your sister wants one-third? The answer to this question is explained by the following picture:



Generally speaking, you need to find a common denominator when adding fractions. It is a horrible error (which, of course, you avoid) to add numerators and denominators separately:

$$\frac{2}{3} + \frac{5}{7} \longrightarrow \frac{2+5}{3+7} = \frac{7}{10}.$$

Instead of the sum this operation gives you something in between the two fractions you started with ($7/10 = 0.7$ is between $2/3 = 0.666\ldots$ and $5/7 = 0.714285\ldots$).

This is easy to understand in a real-life situation. Assume that one team has two bottles for three people ($2/3$ for each) and the other team has five bottles for seven people (5/7 for each). After they meet they have something in between ($2+5$ bottles for $3+7$ people).

Problem 42. Fractions $\frac{a}{b}$ and $\frac{c}{d}$ are called neighbor fractions if their difference $\frac{ad-bc}{bd}$ has numerator ± 1, that is, $ad-bc=\pm 1$. Prove that

(a) in this case neither fraction can be simplified (that is, neither has any common factors in numerator and denominator);

(b) if $\frac{a}{b}$ and $\frac{c}{d}$ are neighbor fractions, then $\frac{a+b}{c+d}$ is between them and is a neighbor fraction for both $\frac{a}{b}$ and $\frac{c}{d}$; moreover,

(c) no fraction $\frac{e}{f}$ with positive integer e and f such that $f<b+d$ is between $\frac{a}{b}$ and $\frac{c}{d}$.

Problem 43. A stick is divided by red marks into 7 equal segments and by green marks into 13 equal segments. Then it is cut into 20 equal pieces. Prove that any piece (except the two end pieces) contains exactly one mark (which may be red or green).

Solution. End pieces carry no marks because $\frac{1}{20}$ is smaller than $\frac{1}{7}$ and $\frac{1}{13}$. We have 18 other pieces – and it remains to prove that none of them can have more than one mark. (We have 18 marks – 6 red and 12 green – so no piece will be left without a mark.) Red marks correspond to numbers of the form $\frac{k}{7}$, green marks correspond to numbers of the form $\frac{l}{13}$. A fraction

$$\frac{k+l}{7+13}=\frac{k+l}{20}$$

is between them and is a cut point dividing these marks. Therefore, two marks of different colors cannot belong to the same piece. Two marks of the same color also cannot appear on one piece because the distance between them (either 1/7 or 1/13) is bigger than the piece length 1/20.)

Problem 44. What is better, to get five percent of seven billion or seven percent of five billion?

Problem 45. How can you cut from a 2/3-meter-long string a piece of length 1/2 meter, without having a meter stick?

Solution. A piece of length 1/2 m constitutes three-fourths of the whole string:

$$\frac{3}{4} \cdot \frac{2}{3} = \frac{2}{4} = \frac{1}{2}$$

and you need to cut off one-fourth of the string.

16 Powers

In the sequence of numbers

$$2,\ 4,\ 8,\ 16, \ldots$$

each number is twice as large as the preceding one:

$$\begin{aligned} 4 &= 2 \cdot 2 \\ 8 &= 4 \cdot 2 = 2 \cdot 2 \cdot 2 \text{ (3 factors)} \\ 16 &= 8 \cdot 2 = 2 \cdot 2 \cdot 2 \cdot 2 \text{ (4 factors)} \\ &\ldots \end{aligned}$$

Mathematicians use the following useful notation:

$$\begin{aligned} 2 \cdot 2 &= 2^2 \\ 2 \cdot 2 \cdot 2 &= 2^3 \\ 2 \cdot 2 \cdot 2 \cdot 2 &= 2^4 \\ &\ldots \end{aligned}$$

so, for example, $2^6 = 64$.

Now the sequence $2, 4, 8, 16, \ldots$ can be written as $2, 2^2, 2^3, 2^4, \ldots$. We read a^n as "a to the n-th power" or "the n-th power of a"; a is called the *base*, and n is called an *exponent*.

There are special names for a^2 and a^3. They are "a squared" and "a cubed", respectively. (A square with side a has area a^2; a cube with edge a has volume a^3.)

Problem 46. Compute: (a) 2^{10}; (b) 10^3; (c) 10^7.

Problem 47. How many decimal digits do you need to write down 10^{1000}?

Astronomers use powers of 10 to write big numbers in a short form. For example, the speed of light is about 300,000 kilometers per second $= 3 \cdot 10^5$ km/s $= 3 \cdot 10^8$ m/s $= 3 \cdot 10^{10}$ cm/s.

Problem 48. In astronomy the distance covered by light in one year is called a *light-year*. What is the distance (approximately) between the Sun and the closest star measured in meters if it is about 4 light-years?

17 Big numbers around us

The number of molecules in one gram of water	$\simeq 3 \cdot 10^{22}$
The radius of Earth	$\simeq 6 \cdot 10^6$ m
The distance between Earth and the Moon	$\simeq 4 \cdot 10^8$ m
The distance between Earth and the Sun (the "astronomical unit")	$\simeq 1.5 \cdot 10^{11}$ m
The radius of the part of the universe observed up to now	$\simeq 10^{26}$ m
The mass of Earth	$\simeq 6 \cdot 10^{24}$ kg
The age of Earth	$\simeq 5 \cdot 10^9$ years
The age of the universe	$\simeq 1.5 \cdot 10^{10}$ years
The number of people on Earth	$\simeq 5 \cdot 10^9$
The average duration of a human life	$\simeq 2 \cdot 10^9$ seconds

Remark. When speaking of big numbers you must keep in mind that the same quantity may be big or small, depending on the unit you choose. For example, the distance between Earth and the Sun, measured in light-years, is about 0.000015 lt-yr, or, in meters (as seen from the table above), $1.5 \cdot 10^{11}$ m.

We shall see later that not only big numbers but also small numbers can be written conveniently using powers.

Programmers prefer to deal with powers of 2 (and not of 10). It turns out that $2^{10} = 1024$ is rather close to $1000 = 10^3$. So the prefix *kilo*, which usually means 1000 (1 kilogram = 1000 grams, 1 kilometer

= 1000 meters, etc.), means "1024" in programming: 1 kilobyte is 1024 bytes.

Problem 49. (a) How many decimal digits do you need to write down 2^{20}? (b) How many for the number 2^{100}? (c) Draw the graph showing how the number of decimal digits in 2^n depends on n.

(To answer the last question, the number of decimal digits in 2^n is approximately $0.3\,n$: $2^{10} \simeq 10^{0.3 \cdot 10}$, $2^n \simeq 10^{0.3\,n}$. Remember this when studying logarithms.)

Many types of pocket calculators use powers of 10 to show the product of two big numbers. For example,

$$370{,}000 \cdot 2{,}100{,}000 = 7.77 \cdot 10^{11},$$

but on the screen you do not see the dot and the base 10, just

`7.77    11`   or   `7.77    E11`

because of the screen limitations. In the usual form

$$777000000000$$

the answer would overflow the calculator screen.

18 Negative powers

We have seen the sequence of powers of 2:

$$2,\ 4,\ 8,\ 16,\ 32,\ 64,\ 128,\ \ldots$$

Now let us start with some number of the sequence (for example, 128) and write it in the reverse order:

$$128,\ 64,\ 32,\ 16,\ 8,\ 4,\ 2.$$

In the first sequence each number was two times bigger than the preceding one; in the second each number is two times smaller than the preceding one. Let us continue this sequence:

$$128,\ 64,\ 32,\ 16,\ 8,\ 4,\ 2,\ 1,\ \frac{1}{2},\ \frac{1}{4},\ \frac{1}{8},\ \frac{1}{16},\ \ldots.$$

The sequence

$$2,\ 4,\ 8,\ 16,\ 32,\ 64,\ 128,\ \ldots$$

could be written as

$$2,\ 2^2,\ 2^3,\ 2^4,\ 2^5,\ 2^6,\ 2^7,\ \ldots.$$

In reverse order,

$$128,\ 64,\ 32,\ 16,\ 8,\ 4,\ 2$$

could be written as

$$2^7,\ 2^6,\ 2^5,\ 2^4,\ 2^3,\ 2^2,\ 2.$$

The analogy suggests the following continuation:

$$\begin{array}{ccccccccccccc} 128, & 64, & 32, & 16, & 8, & 4, & 2, & 1, & \frac{1}{2}, & \frac{1}{4}, & \frac{1}{8}, & \frac{1}{16}, & \ldots \\ 2^7, & 2^6, & 2^5, & 2^4, & 2^3, & 2^2, & 2^1. & 2^0, & 2^{-1}, & 2^{-2}, & 2^{-3}, & 2^{-4} & ,\ldots \end{array}$$

This notation is widely used. So, for example,

$$2^3 = 8, \quad 2^1 = 2, \quad 2^0 = 1, \quad 2^{-1} = \frac{1}{2}, \quad 2^{-2} = \frac{1}{4}, \quad 2^{-3} = \frac{1}{8}, \text{ etc.}$$

When we spoke about powers before we said that 2^3 is "2 used 3 times as a factor" and 2^5 is "2 used 5 times as a factor". We can even say that 2^1 is "2 used once as a factor", but for 2^0 or 2^{-1} such an explanation cannot be taken seriously. It is just an agreement between mathematicians to understand 2^{-n} (for positive integer n) as $\frac{1}{2^n}$.

We hope that this agreement seems rather natural to you. Later we shall see that it is convenient and – in a sense – unavoidable.

Problem 50. Write down (a) 10^{-1}; (b) 10^{-2}; (c) 10^{-3} as decimal fractions.

19 Small numbers around us

1 cm	$= 10^{-2}$ m
1 mm	$= 10^{-3}$ m
1 μm	$= 10^{-6}$ m
1 nanometer	$= 10^{-9}$ m
1 angstrom	$= 10^{-10}$ m
The mass of a water molecule	$\simeq 3 \cdot 10^{-23}$ g
The size of a living cell	$\simeq$ 15 to $350 \cdot 10^{-9}$ m
The size at which modern physical laws become inapplicable (the "elementary length", as physicists say)	$\simeq 10^{-31}$ cm
The wavelength of red light	$\simeq 7 \cdot 10^{-7}$ m

As we have said already, there is no difference, in principle, between "big" and "small" numbers. For example, Earth's radius is about $6 \cdot 10^3$ km and at the same time about $4 \cdot 10^{-5}$ astronomical units.

Now let us return to the general definition of powers.

Definition. For positive integers n,

$$
\begin{aligned}
a^n &= a \cdot a \cdots a \quad (n \text{ times}) \\
a^{-n} &= \frac{1}{a^n} \\
a^0 &= 1
\end{aligned}
$$

Problem 51. Is the equality $a^{-n} = \dfrac{1}{a^n}$ valid for negative n and for $n = 0$?

Is it possible to *prove* that $a^{-n} = \dfrac{1}{a^n}$? No, because the notation a^{-n} makes no sense without an agreement (called a *definition* by mathematicians). If suddenly all mathematicians change their mind and agree to understand a^{-n} in another way, then the equality $a^{-n} = \dfrac{1}{a^n}$ would be false. But you may be sure that this would never happen because nobody wants to violate such a convenient agreement. We would get into a mess if we did so.

Our notation allows us to write the long expression

$$2 \cdot a \cdot a \cdot a \cdot a \cdot b \cdot b \cdot b \cdot c \cdot c \cdot d$$

in the shorter form

$$2a^4b^3c^2d$$

and also rewrite

$$\frac{2 \cdot a \cdot a \cdot a \cdot a \cdot c \cdot c}{b \cdot b \cdot b \cdot d}$$

as

$$2a^4b^{-3}c^2d^{-1}.$$

Problem 52. Write the short form for the following expressions:

(a) $a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot b \cdot b \cdot b \cdot b$

(b) $\dfrac{2 \cdot a \cdot a \cdot a}{b \cdot b}$

Answer: (a) $a^{10}b^4$; (b) $2a^3b^{-2}$.

Problem 53. Rewrite using only positive powers:

(a) a^3b^{-5}; (b) $a^{-2}b^{-7}$.

Answer: (a) $\dfrac{a^3}{b^5}$; (b) $\dfrac{1}{a^2b^7}$.

20 How to multiply a^m by a^n, or why our definition is convenient

It is easy to multiply a^m by a^n if m and n are positive. For example,

$$a^5 \cdot a^3 = \underbrace{(a \cdot a \cdot a \cdot a \cdot a)}_{\text{5 times}} \cdot \underbrace{(a \cdot a \cdot a)}_{\text{3 times}} = a^8.$$

In general, $a^m \cdot a^n = a^{m+n}$ (indeed, a^m is a repeated m times and a^n is a repeated n times). Also

$$a^m \cdot a^1 = a^m \cdot a = a^{m+1}.$$

But the powers may also be negative. It turns out that our rule is valid in this case, too. For example, for $m = 5$, $n = -3$, it states that

$$a^5 \cdot a^{-3} = a^{5+(-3)} = a^2.$$

Let us check it: By definition, $a^5 \cdot a^{-3}$ is

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

More pedantic readers would ask us to check also that

$$a^{-5} \cdot a^3 = a^{-5+3} = a^{-2}.$$

O.K. By definition,

$$a^{-5} \cdot a^3 = \frac{1}{a^5} \cdot a^3 = \frac{a^3}{a^5} = \frac{1}{a^2} = a^{-2}.$$

Even more pedantic readers would remember that both numbers m and n may be negative and ask to check, for example, that

$$a^{-5} \cdot a^{-3} = a^{(-5)+(-3)} = a^{-8}.$$

Indeed,

$$a^{-5} \cdot a^{-3} = \frac{1}{a^5} \cdot \frac{1}{a^3} = \frac{1}{a^8} = a^{-8}.$$

Don't relax – there are still other cases. One of the exponents (or even both) may be equal to zero, and a^0 was defined by a special agreement. So let us check that

$$a^m \cdot a^0 = a^{m+0} = a^m.$$

Indeed, $a^0 = 1$ by definition, so

$$a^m \cdot a^0 = a^m \cdot 1 = a^m.$$

Question. Is it necessary to consider the cases $m < 0$, $m = 0$ and $m > 0$ in the last argument separately?

Problem 54. Find a formula for $\dfrac{a^m}{a^n}$. Is your answer valid for all integers m and n?

21 The rule of multiplication for powers

When multiplying powers with the same base, you need to add exponents:

$$\boxed{a^m \cdot a^n = a^{m+n}}$$

This rule can be used to multiply small and big numbers in a convenient way. For example, to multiply $2 \cdot 10^7$ and $3 \cdot 10^{-11}$ we multiply 2 and 3 and add 7 and -11:

$$(2 \cdot 10^7) \cdot (3 \cdot 10^{-11}) = (2 \cdot 3) \cdot (10^7 \cdot 10^{-11}) = 6 \cdot 10^{7+(-11)} = 6 \cdot 10^{-4}.$$

This method is used in computers (but with base 2 instead of 10).

Problem 55. (a) You know that $2^{1001} \cdot 2^n = 2^{2000}$. What is n?

(b) You know that $2^{1001} \cdot 2^n = 1/4$. What is n?

(c) Which is bigger: 10^{-3} or 2^{-10}?

(d) You know that $\dfrac{2^{1000}}{2^n} = 2^{501}$. What is n?

(e) You know that $\dfrac{2^{1000}}{2^n} = 1/16$. What is n?

(f) You know that $4^{100} = 2^n$. What is n?

(g) You know that $2^{100} \cdot 3^{100} = a^{100}$. What is a?

(h) You know that $(2^{10})^{15} = 2^n$. What is n?

We said earlier that the definition of negative powers is in a sense unavoidable. Now we shall explain what we mean. Assume that we want to define negative power in some way, but want to keep the rule $a^{m+n} = a^m \cdot a^n$ true for all m and n. It turns out that the only way to do so is to follow our definition. Indeed, for $n = 0$ we must have $a^m \cdot a^0 = a^{m+0}$, that is, $a^m \cdot a^0 = a^m$. Therefore, $a^0 = 1$. But then $a^n \cdot a^{-n} = a^{n+(-n)} = a^0 = 1$ implies that $a^{-n} = 1/a^n$.

What do we get if the power is used as a base for another power? For example,

$$(a^2)^3 = \underbrace{a^2 \cdot a^2 \cdot a^2}_{3 \text{ times}} = (a \cdot a) \cdot (a \cdot a) \cdot (a \cdot a) = a^6.$$

Similarly,

$$\boxed{(a^m)^n = a^{m \cdot n}}$$

for any positive m, n. And again our conventions "think for us": the same formula is also true for negative m and n. For example,

$$(a^{-2})^3 = \left(\frac{1}{a^2}\right)^3 = \frac{1}{a^2} \cdot \frac{1}{a^2} \cdot \frac{1}{a^2} = \frac{1}{a^6} = a^{-6} = a^{(-2)\cdot 3}.$$

Problem 56. Check this formula for other combinations of signs (if $m > 0$, $n < 0$; if both m and n are negative; if one of them is equal to zero).

The last formula about powers:

$$\boxed{(ab)^n = a^n \cdot b^n}$$

Problem 57. Check this formula for positive and negative integers n.

Problem 58. What is $(-a)^{775}$? Is it a^{775} or $-a^{775}$?

Problem 59. Invent a formula for $\left(\frac{a}{b}\right)^n$.

Now a^n is defined for any integer n (positive or not) and for any a. But that is not the end of the game, because n may be a number that is not an integer.

Problem 60. Give some suggestions: What might $4^{1/2}$ be? And $27^{1/3}$? Motivate your suggestions as well as you can.

The definition of $a^{m/n}$ will be given later. (But that also is not the last possible step.)

22 Formula for short multiplication: The square of a sum

As we have seen already,

$$(a + b)(m + n) = am + an + bm + bn$$

(to multiply two sums you must multiply each term of the first sum by each term of the second sum and then add all the products). Now consider the case when the letters inside the parentheses are the same:

$$(a + b)(a + b) = aa + ab + ba + bb.$$

Remember that $ab = ba$ and that aa and bb are usually denoted as a^2 and b^2; we get

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

or

$$\boxed{(a+b)^2 = a^2 + 2ab + b^2}$$

Problem 61. (a) Compute 101^2 without pencil and paper.

(b) Compute 1002^2 without pencil and paper.

Problem 62. Each of the two factors of a product becomes 10 percent bigger. How does the product change?

The rule in words: "The square of the sum of two terms is the sum of their squares plus two times the product of the terms".

Be careful here: "the square of the sum" and "the sum of the squares" sound very similar, but are different; the square of the sum is $(a+b)^2$ and the sum of the squares is $a^2 + b^2$.

Problem 63. Are the father of the son of NN and the son of the father of NN the same person?

23 How to explain the square of the sum formula to your younger brother or sister

A kind wizard liked to talk with children and to make them gifts. He was especially kind when many children came together; each of them got as many candies as the number of children. (So if you came alone, you got one, and if you came with a friend you got two and your friend got two.)

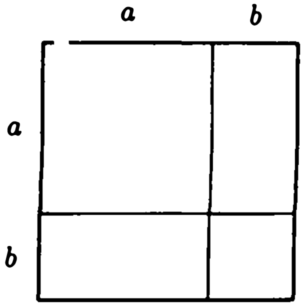
Once, a boys came together. Each of them got a candies – together they got a^2 candies. After they went away with the candies, b girls came and got b candies each – so the girls together got b^2 candies. So that day, the boys and girls got $a^2 + b^2$ candies together.

The next day, a boys and b girls decided to come together. Each of $a+b$ children got $a+b$ candies, so all the children together got $(a+b)^2$ candies. Did they get more or fewer candies than yesterday – and how big is the difference?

To answer this question we may use the following argument. The second time, each of the a boys got b more candies (because of the b girls), so all the boys together got ab more candies. Each girl got a more candies (because of the a boys), so all the girls got ba additional candies. So together, the boys and girls got $ab+ba = 2ab$ candies more than on the previous day. So $(a+b)^2$ is $2ab$ more than a^2+b^2, that is, $(a+b)^2 = a^2+b^2+2ab$.

Problem 64. Cut a square with edge $a+b$ into one square $a \times a$, one square $b \times b$ and two rectangles $a \times b$.

Solution.



The formula $(a+b)^2 = a^2+b^2+2ab$ may be considered as a generic formula for infinitely many equalities like $(5+7)^2 = 5^2+2\cdot5\cdot7+7^2$ or $(13+\frac{1}{3})^2 = 13^2+2\cdot13\cdot\frac{1}{3}+\left(\frac{1}{3}\right)^2$; we get these equalities by replacing a and b by specific numbers. These number may, of course, be negative. For example, for $a=7$, $b=-5$ we get

$$(7+(-5))^2 = 7^2+2\cdot7\cdot(-5)+(-5)^2.$$

Plus times minus is minus, and minus times minus is plus, so we get

$$(7-5)^2 = 7^2-2\cdot7\cdot5+5^2.$$

The same thing could be done for any other numbers, so the general rule is that

$$\boxed{(a-b)^2 = a^2-2ab+b^2}$$

Or in words: "The square of the difference is the sum of the squares minus two times the product of the terms".

Problem 65. Compute (a) 99^2; (b) 998^2 without pencil and paper.

Problem 66. What do the formulas $(a+b)^2 = a^2 + 2ab + b^2$ and $(a-b)^2 = a^2 - 2ab + b^2$ give when (a) $a = b$; (b) $a = 2b$?

24 The difference of squares

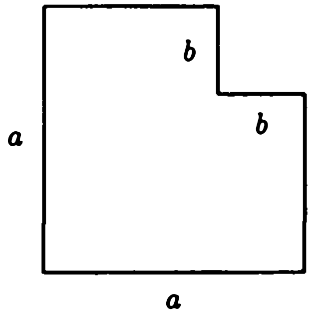
Problem 67. Multiply $a+b$ and $a-b$.

Solution. $(a+b)(a-b) = a(a-b) + b(a-b) = a^2 - ab + ba - b^2 = a^2 - b^2$ (here ab and ba compensate for each other). So we get the formula

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

Problem 68. Multiply $101 \cdot 99$ without pencil and paper.

Problem 69. A piece of size $b \times b$ was cut from an $a \times a$ square.



Cut the remaining part into pieces and combine the pieces into a rectangle with sides $a-b$ and $a+b$.

These three formulas – the square of a sum, the square of a difference, and the difference of squares – are called "short multiplication formulas".

Problem 70. Two integers differ by 2. Multiply them and add 1 to the product. Prove that the result is a perfect square (the square of an integer). For example,

$$\begin{aligned} 3 \cdot 5 + 1 &= 16 = 4^2, \\ 13 \cdot 15 + 1 &= 196 = 14^2. \end{aligned}$$

Solution. (First version.) Let n denote the smaller number. Then the other number is $n+2$. Their product is $n(n+2) = n^2+2n$. Adding 1, we get $n^2+2n+1 = (n+1)^2$ (the formula for the square of the sum).

(Second version.) Let n denote the bigger number. Then the smaller one is $n-2$. The product is $n(n-2) = n^2-2n$. Adding 1 we get $n^2-2n+1 = (n-1)^2$ (the square of the difference formula).

(Third version.) If we want to be fair and not choose between the bigger and the smaller number, let us denote by n the number halfway between the numbers. Then the smaller number is $n-1$, the bigger one is $n+1$, and the product is $(n+1)(n-1) = n^2-1$ (the difference of squares formula), that is, it is a perfect square minus one.

Problem 71. Write the sequence of squares of $1, 2, 3, \ldots$:

$$1,\ 4,\ 9,\ 16,\ 25,\ 36,\ 49,\ \ldots$$

and under any two consecutive numbers of this sequence write their difference:

$$\begin{array}{ccccccccccccccc} 1 & & 4 & & 9 & & 16 & & 25 & & 36 & & 49 & \ldots \\ & 3 & & 5 & & 7 & & 9 & & 11 & & 13 & & \ldots \end{array}$$

In the second sequence any two consecutive numbers differ by 2. Can you explain why?

Solution. The consecutive numbers n and $n+1$ have squares n^2 and $(n+1)^2 = n^2+2n+1$. The difference between these squares is $2n+1$, and it becomes greater by 2 if we add 1 to n.

Remark. A sequence where each term is greater than the preceding one by a fixed constant (as in $3, 5, 7, 9, \ldots$) is called an arithmetic (pronounced "arithmEtic", not "arIthmetic") progression. We shall meet progressions again later.

Problem 72. There is a rule that allows us to square any number with the last digit 5, namely, "Drop this last digit out and get some n; multiply n by $n+1$ and add the digits 2 and 5 to the end". For example, for 35^2, we delete 5 and get 3, multiplying 3 and 4 we get 12, adding "2" and "5" we get the answer: 1225. Explain why this rule works.

Problem 73. Compute $(a+b+c)^2$.

Solution. $(a+b+c)^2 = (a+b+c)(a+b+c) = a^2+ab+ac+ba+b^2+bc+ca+cb+c^2 = a^2+b^2+c^2+2ab+2ac+2bc$.

	a	b	c
a	a^2	ab	ac
b	ba	b^2	bc
c	ca	cb	c^2

Problem 74. Compute $(a+b-c)^2$.

Hint. Use the answer of the preceding problem.

Problem 75. Compute $(a+b+c)(a+b-c)$.

Hint. Use the difference-of-squares formula.

Problem 76. Compute $(a+b+c)(a-b-c)$.

Hint. The difference-of-squares formula is useful here also.

Problem 77. Compute $(a+b-c)(a-b+c)$.

Hint. Even here the difference-of-squares formula can be used!

Problem 78. Compute $(a^2-2ab+b^2)(a^2+2ab+b^2)$.

Solution. This is equal to

$$(a-b)^2(a+b)^2 = ((a-b)(a+b))^2 = (a^2-b^2)^2 = a^4-2a^2b^2+b^4.$$

Another solution:

$$\begin{aligned}
&(a^2-2ab+b^2)(a^2+2ab+b^2) = \\
&= ((a^2+b^2)+2ab)((a^2+b^2)-2ab) = (a^2+b^2)^2-(2ab)^2 = \\
&= a^4+2a^2b^2+b^4-4a^2b^2 = a^4+b^4-2a^2b^2.
\end{aligned}$$

25 The cube of the sum formula

Let us derive the formula for $(a+b)^3$. By definition,

$$(a+b)^3 = (a+b)(a+b)(a+b),$$

and we may start here. But part of the job is done already:

$$(a+b)^3 = (a+b)^2(a+b) = (a^2+2ab+b^2)(a+b).$$

Now we have to multiply each term of the first sum by each term of the second one and take the sum of all products:

$$\begin{aligned}(a^2+2ab+b^2)(a+b) &= \\ = \ & a^2\cdot a + 2ab\cdot a + b^2\cdot a\ + \\ + \ & a^2\cdot b + 2ab\cdot b + b^2\cdot b.\end{aligned}$$

Remembering how to multiply powers with a common base (that is, that $a^m\cdot a^n = a^{m+n}$) and putting a-factors first, we get

$$\begin{aligned}a^3 + 2a^2b + \ & ab^2\ + \\ + \ & a^2b + 2ab^2 + b^3.\end{aligned}$$

Here some terms are similar (only the numerical factors are different); they are written one under another. Collecting them, we get

$$\boxed{(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3}$$

Problem 79. Compute 11^3 without pencil and paper.

Hint. $11 = 10 + 1$.

Problem 80. Compute 101^3 without pencil and paper.

Problem 81. Compute $(a-b)^3$.

Solution. We may compute it in the same way as before, writing $(a-b)^3 = (a-b)^2(a-b) = (a^2-2ab+b^2)(a-b)$ etc. But an easier way is to substitute $(-b)$ for b in the formula for $(a+b)^3$:

$$(a+(-b))^3 = a^3 + 3a^2\cdot(-b) + 3a(-b)^2 + (-b)^3$$

or

$$\boxed{(a-b)^3 = a^3 - 3a^2b + 3ab^2 - b^3}$$

(recall that minus times minus is plus and plus times minus is minus).

26 The formula for $(a+b)^4$

Before computing $(a+b)^4$ let us try to guess the answer. To do so, look at the formulas we already have:

$$\begin{aligned}(a+b)^2 &= a^2+2ab+b^2\\(a+b)^3 &= a^3+3a^2b+3ab^2+b^3.\end{aligned}$$

To get more "experimental data" we can add the formula

$$(a+b)^1 = a+b\,.$$

So we have:

$$\begin{aligned}(a+b)^1 &= a+b\\(a+b)^2 &= a^2+2ab+b^2\\(a+b)^3 &= a^3+3a^2b+3ab^2+b^3\\(a+b)^4 &= ???\end{aligned}$$

How many additive terms do you expect in $(a+b)^4$? Five, of course. What is the first term? Definitely, a^4. The next term is a more difficult puzzle. (To tell you the truth, it will be $4a^3b$.) To explain how it can be guessed let us divide our question into two parts:

(1) What powers of a and b will appear?

(2) What numeric coefficients will appear?

Part (1) is simpler. If the formula for

$(a+b)^1$ uses a and b,
$(a+b)^2$ uses a^2, ab and b^2,
$(a+b)^3$ uses a^3, a^2b, ab^2 and b^3,

we may expect that

$$(a+b)^4 \text{ uses } a^4,\ a^3b,\ a^2b^2,\ ab^3, \text{ and } b^4.$$

Now look at the coefficients (we write the factor "1" to make our formulas more uniform):

$$\begin{aligned}(a+b)^1 &= 1a+1b\\(a+b)^2 &= 1a^2+2ab+1b^2\\(a+b)^3 &= 1a^3+3a^2b+3ab^2+1b^3\end{aligned}$$

or, without terms (only the coefficients):

$$\begin{array}{ccccc} 1 & 1 & & & \\ 1 & 2 & 1 & & \\ 1 & 3 & 3 & 1 & \\ ? & ? & ? & ? & ? \end{array}$$

(we have already said that we expect five terms in the $(a+b)^4$ formula). The first coefficient is, of course, 1. It seems that the second is 4 (because in the second column we have 1, 2 and 3). So we get

$$\begin{array}{ccccc} 1 & 1 & & & \\ 1 & 2 & 1 & & \\ 1 & 3 & 3 & 1 & \\ 1 & 4 & ? & ? & ? \end{array}$$

Two more coefficients can be guessed. In $(a+b)^4$, the letters a and b have equal rights, so b^4 must have the same coefficient as a^4, and ab^3 must have the same coefficient as a^3b – to avoid discrimination:

$$\begin{array}{ccccc} 1 & 1 & & & \\ 1 & 2 & 1 & & \\ 1 & 3 & 3 & 1 & \\ 1 & 4 & ? & 4 & 1. \end{array}$$

Now only a^2b^2 remains, and if we cannot guess it, we must compute it by brute force:

$$(a+b)^4 = (a+b)^3(a+b) = (a^3+3a^2b+3ab^2+b^3)(a+b) =$$

$$\begin{array}{llllll} = & a^3 \cdot a & + \ 3a^2b \cdot a & + \ 3ab^2 \cdot a & + \ b^3 \cdot a & + \\ & & + \ a^3 \cdot b & + \ 3a^2b \cdot b & + \ 3ab^2 \cdot b & + \ b^3 \cdot b = \\ = & a^4 & + \ 3a^3b & + \ 3a^2b^2 & + \ ab^3 & + \\ & & + \ a^3b & + \ 3a^2b^2 & + \ 3ab^3 & + \ b^4 \end{array}$$

(again the similar terms are written one under another). Collecting them, we get

$$\boxed{(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4}$$

All our guesses turn out to be true and we find the remaining coefficient of a^2b^2, which turns out to be 6.

27 Formulas for $(a+b)^5$, $(a+b)^6$, ... and Pascal's triangle

In $(a+b)^5$ we expect terms

$$a^5 \quad a^4b \quad a^3b^2 \quad a^2b^3 \quad ab^4 \quad b^5$$

with coefficients

$$1 \quad 5 \quad ? \quad ? \quad 5 \quad 1$$

To find the two remaining coefficients (they are expected to be equal, of course) let us proceed as usual:

$$(a+b)^5 = (a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4)(a+b) =$$

$$\begin{aligned} = a^4\cdot a &+ 4a^3b\cdot a + 6a^2b^2\cdot a + 4ab^3\cdot a + b^4\cdot a \;+ \\ &+ a^4\cdot b + 4a^3\cdot b + 6a^2b^2\cdot b + 4ab^3\cdot b + b^4\cdot b = \\ = a^5 &+ 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5. \end{aligned}$$

So our table of coefficients has one more row:

1	1				
1	2	1			
1	3	3	1		
1	4	6	4	1	
1	5	10	10	5	1

Probably you have already figured out the rule: Each coefficient is equal to the sum of the coefficient above it and the one to the left of it: $1+4=5$, $4+6=10$, $6+4=10$, $4+1=5$.

The reason this is so becomes clear if we look at our computation ignoring everything except coefficients:

$$\begin{aligned} 1\ldots &+ 4\ldots + 6\ldots + 4\ldots + 1\ldots \;+ \\ &+ 1\ldots + 4\ldots + 6\ldots + 4\ldots + 1\ldots = \\ 1\ldots &+ 5\ldots + 10\ldots + 10\ldots + 5\ldots + 1\ldots \end{aligned}$$

They are added exactly as the rule says.

For aesthetic reasons, we may write the table in a more symmetric way and add "1" on the top (because $(a+b)^0 = 1$). We get a triangle

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

which can be continued using the rule that each number is the sum of the two numbers immediately above it (except for the first and the last numbers, which are equal to 1). For example, the next row will be

$$1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1$$

and it corresponds to the formula

$$\boxed{(a+b)^6 = a^6 + 6a^5b + 15a^4b^2 + 20a^3b^3 + 15a^2b^4 + 6ab^5 + b^6}$$

This triangle is called *Pascal's triangle* (Blaise Pascal [1623–1662] was a French mathematician and philosopher.)

Problem 82. Compute 11^3, 11^4, 11^5 and 11^6.

Problem 83. Write a formula for $(a+b)^7$.

Problem 84. Find formulas for $(a-b)^4$, $(a-b)^5$ and $(a-b)^6$.

Problem 85. Compute the sums of all the numbers in the first, second, third, etc., rows of Pascal's triangle. Can you see the rule? Can you explain the rule?

Problem 86. What do the formulas for $(a+b)^2$, $(a+b)^3$, $(a+b)^4$, etc., give when $a = b$?

Problem 87. Do you see the connection between the two preceding problems?

Problem 88. What do the formulas for $(a+b)^2$, $(a+b)^3$, $(a+b)^4$, etc., give when $a = -b$?

28 Polynomials

By a *polynomial* we mean an expression containing letters (called *variables*), numbers, addition, subtraction and multiplication. Here are some examples:

$$a^4 + a^3b + ab^3 + b^4$$

$$(5 - 7x)(x - 1)(x - 3) + 11$$

$$(a + b)(a^3 + b^3)$$

$$(a + b)(a + 2b) + ab$$

$$(x + y)(x - y) + (y - x)(y + x)$$

$$0$$

$$(x + y)^{100}$$

These examples contain not only addition, subtraction and multiplication, but also positive integer constants as powers. These are legal because they can be considered as shortcuts (for example, a^4 may be considered as short notation for $a \cdot a \cdot a \cdot a$, which is perfectly legal). But a^{-7} or x^y are *not* polynomials.

A *monomial* is a polynomial that does not use addition or subtraction, that is, a product of letters and numbers. Here are some examples of monomials:

$$5 \cdot a \cdot 7 \cdot b \cdot a$$

$$127a^{15}$$

$$(-2)a^2b$$

(in the last example the minus sign is not subtraction but a part of the number "-2").

Usually numbers and identical letters are collected: for example, $5 \cdot a \cdot 7 \cdot b \cdot a$ is written as $35a^2b$.

Please keep in mind that a monomial is a polynomial, so sometimes for a mathematician one ("mono") is many ("poly").

Each polynomial can be converted into the sum of monomials if we remove parentheses. For example,

$$(a + b)(a^3 + b^3) = aa^3 + ab^3 + ba^3 + bb^3 = a^4 + ab^3 + ba^3 + b^4,$$

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

When doing so we can get similar monomials (having the same letters with the same powers and differing only in the coefficients). For example, in the second polynomial above, the terms $2ab$ and ba are similar. They can be collected into $3ab$ and we get

$$(a+b)(a+2b)=a^2+2ab+ba+2b^2=a^2+3ab+2b^2.$$

Problem 89. Convert $(1+x-y)(12-zx-y)$ into a sum of monomials and collect the similar terms.

Solution.

$$\begin{aligned}(1+x-y)&(12-zx-y)=\\ &= 12-zx\underline{-y}+12x-xzx-xy\underline{-12y}+yzx+y^2=\\ &= 12-zx-13y+12x-zx^2-yx+yzx+y^2.\end{aligned}$$

(The similar terms are underlined.)

Strictly speaking, this is not enough, because we need a sum of monomials and now we have subtraction. Therefore we need to do one more step to get

$$12+(-1)zx+(-13)y+12x+(-1)zx^2+(-1)yx+1yzx+1y^2$$

(to make the terms more uniform we added the factor "1" before xyz and before y^2).

A *standard form* of a polynomial is a sum of monomials, where each monomial is a product of a number (called a *coefficient*) and of powers of different letters, and where all similar monomials are collected.

To add two polynomials in standard form we must add the coefficients of similar terms. If we get a zero coefficient, the corresponding term vanishes:

$$(1x+(-1)y)+(1y+(-2)x+1z)=$$

$$(1+(-2))x+((-1)+1)y+1z=(-1)x+0y+1z=(-1)x+1z.$$

To multiply two polynomials in standard form we need to multiply each term of the first polynomial by each term of the second polynomial. When multiplying monomials, we add powers of each variable:

$$(a^5b^7c)\cdot(a^3bd^4)=a^{5+3}b^{7+1}cd^4=a^8b^8cd^4.$$

After this is done, we have to collect similar terms. For example,

$$(x-y)(x^2+xy+y^2) = x^3 + \underline{x^2y} + \underline{\underline{xy^2}} - \underline{yx^2} - \underline{\underline{xy^2}} - y^3 = x^3 - y^3.$$

(The pedantic reader may find that we have violated the rules adopted for the standard form of a polynomial, because the coefficients -1 and 1 are omitted.)

Problem 90.

(a) Multiply $(1+x)(1+x^2)$.

(b) Multiply $(1+x)(1+x^2)(1+x^4)(1+x^8)$.

(c) Compute $(1+x+x^2+x^3)^2$.

(d) Compute $(1+x+x^2+x^3+\cdots+x^9+x^{10})^2$.

(e) Find the coefficients of x^{30} and x^{29} in $(1+x+x^2+x^3+\cdots+x^9+x^{10})^3$.

(f) Multiply $(1-x)(1+x+x^2+x^3+\cdots+x^9+x^{10})$.

(g) Multiply $(a+b)(a^2-ab+b^2)$.

(h) Multiply $(1-x+x^2-x^3+x^4-x^5+x^6-x^7+x^8-x^9+x^{10})$ by $(1+x+x^2+x^3+x^4+x^5+x^6+x^7+x^8+x^9+x^{10})$.

29 A digression: When are polynomials equal?

The word "equal" for polynomials may be understood in many different ways. The first possibility: Polynomials are equal if they can be transformed into one another by using algebraic rules (removing parentheses, collecting similar terms, finding common factors, and so on). Another possibility: Two polynomials are considered to be equal if after substituting any numbers for the variables they have the same numeric value. It turns out that these two definitions are equivalent; if two polynomials are equal in the sense of one of these definitions they are also equal in the sense of the other one. Indeed, if one polynomial can be converted into another using algebraic transformations, these transformations are still valid when variables are replaced by numbers.

So these polynomials have the same numeric value after replacement. It is not easy to prove the reverse statement: *If two polynomials are equal for any values of variables, they can be converted into each other by algebraic transformations.* So we shall use it – sorry! – without proof.

If we want to convince somebody that two given polynomials are equal, the first version of the definition is preferable; it is enough to show the sequence of algebraic transformations needed to get the second polynomial from the first one. On the other hand, if we want to convince somebody that two polynomials are not equal, the second definition is better; it is enough to find numbers that lead to the different values of the polynomials.

Problem 91. Prove that

$$(x-1)(x-2)(x-3)(x-4) \neq (x+1)(x+2)(x+3)(x+4)$$

without computations.

Solution. When $x = 1$ the left-hand side is equal to zero and the right-hand side is not, therefore these polynomials are not equal according to the second definition.

Problem 92. In the (true) equality

$$(x^2-1)(x+\cdots) = (x-1)(x+3)(x+\cdots)$$

some numbers are replaced by dots. What are these numbers?

Hint. Substitute -1 and -3 for x.

Now assume that somebody gives us two polynomials, not saying whether they are different or equal. How can we check this? We can try to substitute different numbers for the variables. If at least once these polynomials have different numeric values we can be sure that they are different. Otherwise we may suspect that these polynomials are in fact equal.

Problem 93. George tries to check whether the polynomials $(x+1)^2-(x-1)^2$ and x^2+4x-1 are equal or not by substituting 1 and -1 for x. Is it a good idea?

Solution. No. These polynomials have equal values for $x = -1$ (both values are -4) and for $x = 1$ (both give 4). However, they are not equal; for example, they have different values for $x = 0$.

To check whether two polynomials are equal or not in a more regular way, we may convert them to a standard form. If after this they differ only in the order of the monomials (or in the order of the factors inside the monomials), then the polynomials are equal. If not, it is possible to prove that the polynomials are different.

Sometimes equal polynomials are called "identically equal", meaning that they are equal for all values of variables. So, for example, $a^2 - b^2$ is identically equal to $(a - b)(a + b)$.

Remark. Later we shall see that sometimes a finite number of tests is enough to decide whether two polynomials are equal or not.

30 How many monomials do we get?

Problem 94. Each of two polynomials contains four monomials. What is the maximal possible number of monomials in their product?

Remark. Of course, any polynomial can be extended by monomials with zero coefficients like this:

$$x^3 + 4 = x^3 + 0x^2 + 0x + 4$$

Such monomials are ignored.

Solution. Multiply $(a + b + c + d)$ by $(x + y + z + u)$:

$$\begin{aligned}(a + b + c + d)(x + y + z + u) &= \\ = ax + ay + az + au &+ \\ bx + by + bz + bu &+ \\ cx + cy + cz + cu &+ \\ dx + dy + dz + du &.\end{aligned}$$

We get 16 terms. It is clear that 16 is the maximum possible number (because each of 4 monomials of the first polynomial is multiplied by each of 4 monomials of the second one).

Problem 95. Each of two polynomials contains four monomials. Is it possible that their product contains fewer than 16 monomials?

Solution. Yes, if there are similar monomials among the products. For example,

$$(1+x+x^2+x^3)(1+x+x^2+x^3) = 1+2x+3x^2+4x^3+3x^4+2x^5+x^6,$$

that is, after collecting similar terms we get 7 monomials instead of 16.

Problem 96. Is it possible when multiplying two polynomials that, after collecting similar terms, all terms vanish (have zero coefficients)?

Answer. No.

Remark. Probably this problem seems silly; it is clear that it cannot happen. If you think so, please reconsider the problem several years from now.

Problem 97. Is it possible when multiplying two polynomials that after the collection of similar terms all terms vanish (have zero coefficients) except one? (Do not count the case when each of the polynomials has only one monomial.)

Problem 98. Is it possible that the product of two polynomials contains fewer monomials than each of the factors?

Solution. Yes:

$$\begin{aligned}
(x^2 + 2xy + 2y^2)(x^2 - 2xy + 2y^2) &= \\
= \quad & ((x^2 + 2y^2) + 2xy)((x^2 + 2y^2) - 2xy) = \\
= \quad & (x^2 + 2y^2)^2 - (2xy)^2 = \\
= \quad & x^4 + 4x^2y^2 + 4y^4 - 4x^2y^2 = \\
= \quad & x^4 + 4y^4.
\end{aligned}$$

31 Coefficients and values

Recall Pascal's triangle and the formulas for $(a+b)^n$ for different n:

$$\begin{array}{ccccc}
1 & & (a+b)^0 & = & 1 \\
1\ 1 & & (a+b)^1 & = & 1a + 1b \\
1\ 2\ 1 & & (a+b)^2 & = & 1a^2 + 2ab + 1b^2 \\
1\ 3\ 3\ 1 & & (a+b)^3 & = & 1a^3 + 3a^2b + 3ab^2 + 1b^3 \\
1\ 4\ 6\ 4\ 1 & & (a+b)^4 & = & 1a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1b^4
\end{array}$$

etc. Each of these formulas is an equality between two polynomials.

Problem 99. What do we get for $a = 1$, $b = 1$?

Solution.

$$\begin{aligned}
(1+1)^0 &= 1\\
(1+1)^1 &= 1+1\\
(1+1)^2 &= 1+2+1\\
(1+1)^3 &= 1+3+3+1\\
(1+1)^4 &= 1+4+6+4+1
\end{aligned}$$

etc. Recall that $1 + 1 = 2$; so we proved that the sum of any row of Pascal's triangle is a power of 2. For example, the sum of the 25th row of Pascal's triangle is equal to 2^{24}.

Problem 100. Add the numbers of some row of Pascal's triangle with alternating signs. You get 0:

$$1-1=0$$
$$1-2+1=0$$
$$1-3+3-1=0$$
$$1-4+6-4+1=0$$

etc. Why does this happen?

Hint. Try $a = 1$, $b = -1$.

Problem 101. Imagine that the polynomial $(1 + 2x)^{200}$ is converted to the standard form (the sum of powers of x with numerical coefficients). What is the sum of all the coefficients?

Hint. Try $x = 1$.

Problem 102. The same question for the polynomial $(1 - 2x)^{200}$ instead of $(1 + 2x)^{200}$.

Problem 103. Imagine that the polynomial $(1 + x - y)^3$ is converted to the standard form. What is the sum of its coefficients?

Problem 104. (*continued*) What is the sum of the coefficients of the terms not containing y?

Problem 105. (*continued*) What is the sum of the coefficients of the terms containing x?

32 Factoring

To multiply polynomials you may need a lot of patience, but you do not need to think; just follow the rules carefully. But to reconstruct factors if you know only their product you sometimes need a lot of ingenuity. And some polynomials cannot be decomposed into a product of nontrivial (nonconstant) factors at all. The decomposition process is called *factoring*, and there are many tricks that may help. We'll show some tricks now.

Problem 106. Factor the polynomial $ac + ad + bc + bd$.

Solution. $ac + ad + bc + bd = a(c+d) + b(c+d) = (a+b)(c+d)$.

Problem 107. Factor the following polynomials:

(a) $ac + bc - ad - bd$;

(b) $1 + a + a^2 + a^3$;

(c) $1 + a + a^2 + a^3 + \cdots + a^{13} + a^{14}$;

(d) $x^4 - x^3 + 2x - 2$.

Sometimes we first need to cut one term into two pieces before it is possible to proceed.

Problem 108. Factor $a^2 + 3ab + 2b^2$.

Solution. $a^2 + 3ab + 2b^2 = a^2 + ab + 2ab + 2b^2 = a(a+b) + 2b(a+b)$ $= (a+2b)(a+b)$.

Remark. When multiplying two polynomials we collect the similar terms into one term. So it is natural to expect that when going in the other direction we may have to split a term into a sum of several terms.

Problem 109. Factor:

(a) $a^2 - 3ab + 2b^2$;

(b) $a^2 + 3a + 2$.

The formula for the square of the sum can be read "from right to left" as an example of factoring: the polynomial $a^2 + 2ab + b^2$ is factored into $(a+b)(a+b)$. The same factorization can also be obtained as follows:

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

Problem 110. Factor:

(a) $a^2 + 4ab + 4b^2$;

(b) $a^4 + 2a^2b^2 + b^4$;

(c) $a^2 - 2a + 1$.

Sometimes it is necessary to add and subtract some monomial (reconstructing the annihilated terms). We show this trick factoring $a^2 - b^2$ (though we know the factorization in advance: it is the difference-of-squares formula):

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

Problem 111. Factor $x^5 + x + 1$.

Solution. $x^5 + x + 1 = x^5 + x^4 + x^3 - x^4 - x^3 - x^2 + x^2 + x + 1 = x^3(x^2+x+1) - x^2(x^2+x+1) + (x^2+x+1) = (x^3 - x^2 + 1)(x^2+x+1)$.

Probably you are discouraged by this solution because it seems impossible to invent it. The authors share your feeling.

Let us look at the factorization $a^2 - b^2 = (a+b)(a-b)$ once more from another viewpoint. If $a = b$, then the right-hand side is equal to zero (one of the factors is zero). Therefore the left-hand side must be zero, too. Indeed, $a^2 = b^2$ when $a = b$. Similarly, if $a + b = 0$ then $a^2 = b^2$ (in this case $a = -b$ and $a^2 = b^2$ because in changing the sign we do not change the square).

Problem 112. Prove that if $a^2 = b^2$ then $a = b$ or $a = -b$.

The moral of this story: When trying to factor a polynomial it is wise to see when it has a zero value. This may give you an idea what the factors might be.

Problem 113. Factor $a^3 - b^3$.

Solution. The expression $a^3 - b^3$ has a zero value when $a = b$. So it is reasonable to expect a factor $a - b$. Let us try: $a^3 - b^3 = a^3 - a^2b + a^2b - ab^2 + ab^2 - b^3 = a^2(a-b) + ab(a-b) + b^2(a-b) = (a^2 + ab + b^2)(a-b)$.

Problem 114. Factor $a^3 + b^3$.

Solution. $a^3+b^3 = a^3+a^2b-a^2b-ab^2+ab^2+b^3 = a^2(a+b)-ab(a+b)+b^2(a+b) = (a^2-ab+b^2)(a+b)$.

The same factorization can be obtained from the solution of the preceding problem by substituting $(-b)$ for b.

Problem 115. Factor a^4-b^4.

Solution. $a^4-b^4 = a^4-a^3b+a^3b-a^2b^2+a^2b^2-ab^3+ab^3-b^4 = a^3(a-b)+a^2b(a-b)+ab^2(a-b)+b^3(a-b) = (a-b)(a^3+a^2b+ab^2+b^3)$.

Problem 116. Factor:

(a) a^5-b^5;

(b) $a^{10}-b^{10}$;

(c) a^7-1.

Another factorization of a^4-b^4:

$$a^4-b^4=(a^2-b^2)(a^2+b^2).$$

These two factorizations are in fact related; both can be obtained from

$$(a^4-b^4)=(a-b)(a+b)(a^2+b^2)$$

by a suitable grouping of factors.

Problem 117. Factor a^2-4b^2.

Solution. Using that $4=2^2$ we write:

$$a^2-4b^2=a^2-2^2b^2=a^2-(2b)^2=(a-2b)(a+2b)$$

Let us try to apply the same trick to a^2-2b^2. Here we need a number called "the square root of two" and denoted by $\sqrt{2}$. It is approximately equal to $1.4142\ldots$; its main property is that its square is equal to 2: $(\sqrt{2})^2=2$. (Generally speaking, a square root of a nonnegative number a is defined as a nonnegative number whose square is equal to a. It is denoted by $\sqrt{a}$. Such a number always exists and is defined uniquely; see below.)

Using the square root of two we may write:

$$a^2-2b^2=a^2-(\sqrt{2}b)^2=(a-\sqrt{2}b)(a+\sqrt{2}b).$$

So we are able to factor $a^2 - 2b^2$, though we are forced to use $\sqrt{2}$ as a coefficient.

Remark. Look at the equality

$$a - b = (\sqrt{a})^2 - (\sqrt{b})^2 = (\sqrt{a} - \sqrt{b})(\sqrt{a} + \sqrt{b}).$$

So we have factored $a-b$, haven't we? No, we haven't, because $\sqrt{a}-\sqrt{b}$ is not a polynomial; taking the square root is not a legal operation for polynomials – only addition, subtraction and multiplication are allowed. But how about $a-\sqrt{2}b$? Why do we consider it as a polynomial? Because our definition of a polynomial allows it to be constructed from letters and numbers using addition, subtraction, and multiplication. And $\sqrt{2}$ is a perfectly legal number (though it is defined as a square root of another number). So in this case everything is O.K.

Problem 118. Factor: (a) a^2-2; (b) a^2-3b^2; (c) $a^2+2ab+b^2-c^2$; (d) $a^2 + 4ab + 3b^2$.

Problem 119. Factor a^4+b^4. (The known factorization of a^4-b^4 seems useless because substituting $(-b)$ for b we get nothing new.)

Solution. A trick: add and subtract $2a^2b^2$. It helps:

$$\begin{aligned} a^4 + b^4 &= a^4 + 2a^2b^2 + b^4 - 2a^2b^2 = \\ &= (a^2 + b^2)^2 - (\sqrt{2}ab)^2 = (a^2 + b^2 + \sqrt{2}ab)(a^2 + b^2 - \sqrt{2}ab). \end{aligned}$$

Let us see what we now know. We can factor a^n-b^n for any positive integer n (one of the factors is $a-b$). If n is odd, the substitution of $-b$ for b gives a factorization of $a^n + b^n$ (one of the factors is $a+b$). But what about a^2+b^2, a^4+b^4, a^6+b^6, etc.? We have just factored the second one.

Problem 120. Can you factor any other polynomial of the form $a^{2n}+b^{2n}$?

Hint. $a^6+b^6 = (a^2)^3 + (b^2)^3$. The same trick may be used if n has an odd divisor greater than 1.

But the simplest case, a^2+b^2, remains unsolved. It would be possible to write

$$a^2 + b^2 = a^2 - (\sqrt{-1}\cdot b)^2 = (a - \sqrt{-1}\cdot b)(a + \sqrt{-1}\cdot b)$$

if a square root of -1 exists. But – alas – it is not the case (the square of any nonzero number is positive and therefore not equal to -1). But mathematicians are tricky; if such a number does not exist, it must be invented. So they invented it, and got new numbers called *complex numbers*. But this is another story.

Problem 121. What would you suggest as the product of two complex numbers $(2+3\sqrt{-1})$ and $(2-3\sqrt{-1})$?

Let us finish this section with more difficult problems.

Problem 122. Factor:

(a) x^4+1;

(b) $x(y^2-z^2)+y(z^2-x^2)+z(x^2-y^2)$;

(c) $a^{10}+a^5+1$;

(d) $a^3+b^3+c^3-3abc$;

(e) $(a+b+c)^3-a^3-b^3-c^3$;

(f) $(a-b)^3+(b-c)^3+(c-a)^3$.

Problem 123. Prove that if $a,b>1$ then $a+b<1+ab$.

Hint. Factor $(1+ab)-(a+b)$.

Problem 124. Prove that if $a^2+ab+b^2=0$ then $a=0$ and $b=0$.

Hint. Recall the factorization of a^3-b^3. (Another solution will be discussed later when speaking about quadratic equations.)

Problem 125. Prove that if $a+b+c=0$ then $a^3+b^3+c^3=3abc$.

Problem 126. Prove that if

$$\frac{1}{a+b+c}=\frac{1}{a}+\frac{1}{b}+\frac{1}{c}$$

then there are two opposite numbers among a,b,c (i.e. $a=-b$, $a=-c$ or $b=-c$).

33 Rational expressions

One is not allowed to use division in a polynomial (only addition, subtraction, and multiplication). If we allow division too, we get what are called "rational expressions". (The only restriction is that the divisor must not be identically equal to zero.)

Examples:

(a) $\dfrac{ab}{c}$; (b) $\dfrac{a/b}{b/c}$; (c) $\dfrac{1}{1+\dfrac{1}{x}}$; (d) $\dfrac{1}{1+\dfrac{1}{1+\dfrac{1}{1+\dfrac{1}{x}}}}$;

(e) $\dfrac{\dfrac{x}{y}+\dfrac{y}{z}+\dfrac{z}{x}}{\dfrac{y}{x}+\dfrac{z}{y}+\dfrac{x}{z}}+1$; (f) $\dfrac{x^3+x^2+x+1}{x+1}$; (g) $\dfrac{1}{\left(\dfrac{1}{a}+\dfrac{1}{b}\right)/2}$

Let us mention that, for example, $\dfrac{x^2-x^2}{x-x}$ is not a rational expression because the denominator is identically equal to 0.

Let us mention as well that the permisson to use division is not an obligation to use it; therefore, any polynomial is a rational expression.

34 Converting a rational expression into the quotient of two polynomials

A rational expression may include several divisions (as in examples (b), (c), (d), or (g)). But it can be converted into a form in which only one division is used and the division operation is the last one. In other words, any rational expression may be converted into the quotient of two polynomials.

The following transformations are used to do the conversion:

1. Addition: Assume that we want to add $\dfrac{P}{Q}$ and $\dfrac{R}{S}$ where P, Q, R, S are polynomials. Find the common denominator for $\dfrac{P}{Q}$ and $\dfrac{R}{S}$ (if we have no better idea, just multiply P and Q by S and multiply R and S by Q):

$$\frac{P}{Q}+\frac{R}{S}=\frac{PS}{QS}+\frac{QR}{QS}=\frac{PS+QR}{QS}.$$

We've got a quotient of two polynomials.

2. The subtraction case is similar:

$$\frac{P}{Q}-\frac{R}{S}=\frac{PS}{QS}-\frac{QR}{QS}=\frac{PS-QR}{QS}.$$

3. Multiplication:

$$\frac{P}{Q}\cdot\frac{R}{S}=\frac{PR}{QS}.$$

4. Division:

$$\frac{P}{Q}\Big/\frac{R}{S}=\frac{PS}{QR}.$$

Sometimes during the transformation we are able to simplify the expression, eliminating a common factor in the numerator and the denominator:

$$\frac{PX}{QX}=\frac{P}{Q}.$$

Problem 127. Convert the expressions from the examples (b), (c), (d), (e), and (g) to this form (expressions (a) and (f) already are in this form).

Answers and solutions.

(b) $\frac{ac}{b^2}$; (c) $\frac{x}{x+1}$;

(d) $\dfrac{1}{1+\frac{1}{x}}=\dfrac{x}{x+1}$; $\quad 1+\dfrac{1}{1+\frac{1}{x}}=1+\dfrac{x}{x+1}=\dfrac{2x+1}{x+1}$;

$$\frac{1}{1+\dfrac{1}{1+\dfrac{1}{x}}}=\frac{x+1}{2x+1};\quad 1+\frac{1}{1+\dfrac{1}{1+\dfrac{1}{x}}}=\frac{3x+2}{2x+1};$$

thus, the answer is $\dfrac{2x+1}{3x+2}$.

(e) $$\frac{\frac{x}{y}+\frac{y}{z}+\frac{z}{x}}{\frac{y}{x}+\frac{z}{y}+\frac{x}{z}}+1=\frac{(x^2z+y^2x+z^2y)/xyz}{(y^2z+z^2x+x^2y)/xyz}+1=$$

$$=\frac{xyz\cdot(x^2z+y^2x+z^2y)/xyz}{xyz\cdot(y^2z+z^2x+x^2y)/xyz}+1=\frac{(x^2z+y^2x+z^2y)}{(y^2z+z^2x+x^2y)}+1=$$

$$=\frac{x^2z+y^2x+z^2y+y^2z+z^2x+x^2y}{y^2z+z^2x+x^2y};$$

(g) $\dfrac{2ab}{a+b}$.

Let us mention that in such problems the answer is not defined uniquely. For example, the expression

$$\frac{x^3+x^2+x+1}{x^2-1}$$

may be left as is, but may also be transformed as follows:

$$\frac{x^3+x^2+x+1}{(x+1)(x-1)} = \frac{(x+1)(x^2+1)}{(x+1)(x-1)} = \frac{x^2+1}{x-1}$$

Remark. Strictly speaking, the cancellation of common factors is not a perfectly legal operation, because sometimes the factor being cancelled may be equal to zero. For example, $\dfrac{x^3+x^2+x+1}{x^2-1}$ is undefined when $x=-1$; it is equal to $\dfrac{x^2+1}{x-1}$ where both are defined. Usually this effect is ignored but sometimes it may become important.

Sometimes the statement of a problem requires us to "simplify the expression" – to convert it to the simplest possible form. Though simplicity is a matter of taste, usually it is clear what the author of the problem meant.

Problem 128. Simplify the expression

$$\frac{(x-a)(x-b)}{(c-a)(c-b)} + \frac{(x-a)(x-c)}{(b-a)(b-c)} + \frac{(x-b)(x-c)}{(a-b)(a-c)}.$$

Solution. Let us first add two fractions. The common denominator is $(c-a)(c-b)(b-a)$. Additional factors are $b-a$ for the first fraction and $c-a$ for the second. We use the fact that $b-c=-(c-b)$, so

$$\begin{aligned}
&\frac{(x-a)(x-b)}{(c-a)(c-b)} + \frac{(x-a)(x-c)}{(b-a)(b-c)} = \\
&= \frac{(x-a)(x-b)(b-a)-(x-a)(x-c)(c-a)}{(c-a)(c-b)(b-a)} = \\
&= \frac{(x-a)[(x-b)(b-a)-(x-c)(c-a)]}{(c-a)(c-b)(b-a)} = \\
&= \frac{(x-a)[xb-\underline{xa}-b^2+ab-xc+\underline{xa}+c^2-ac]}{(c-a)(c-b)(b-a)} =
\end{aligned}$$

$$= \frac{(x-a)[x(b-c)+a(b-c)-(b-c)(b+c)]}{(c-a)(c-b)(b-a)} =$$

$$= \frac{(x-a)(b-c)(x+a-b-c)}{(c-a)(c-b)(b-a)}$$

Reducing the common factors $(c-b) = -(b-c)$ we get

$$\frac{(x-a)(b+c-a-x)}{(c-a)(b-a)}$$

Now we can add the third fraction (it has the same denominator, because minus times minus is plus):

$$\frac{(x-a)(b+c-a-x)}{(c-a)(b-a)} + \frac{(x-b)(x-c)}{(a-b)(a-c)} =$$

$$= \frac{xb+xc-xa-x^2-ab-ac+a^2+ax+x^2-xb-xc+bc}{(c-a)(b-a)} =$$

$$= \frac{a^2+bc-ab-ac}{(c-a)(b-a)} = \frac{a(a-b)-c(a-b)}{(c-a)(b-a)} = \frac{(a-c)(a-b)}{(c-a)(b-a)} = 1$$

So we have proved the identity

$$\frac{(x-a)(x-b)}{(c-a)(c-b)} + \frac{(x-a)(x-c)}{(b-a)(b-c)} + \frac{(x-b)(x-c)}{(a-b)(a-c)} = 1.$$

Problem 129. Check this identity in the special cases $x = a$, $x = b$, and $x = c$.

We shall see later that in fact these three cases are sufficient to be sure that the identity is true in the general case. (So the long computations we have done could be avoided.) But we need more theory to realize this.

To conclude this section we state some problems involving rational expressions.

The expression

$$\frac{1}{\left(\frac{1}{a}+\frac{1}{b}\right)/2}$$

(the inverse of the arithmetic mean of the inverses of a and b; see below) is called the *harmonic mean* of a and b. You may meet it in some situations.

Problem 130. A swimming pool is divided into two equal sections. Each section has its own water supply pipe. To fill one section (using its pipe) you need a hours. To fill the other section you need b hours. How many hours would you need if you turn on both pipes and remove the wall dividing the pool into sections?

Problem 131. A motor boat needs a hours to go from A to B down the river and needs b hours to go from B to A (up the river). How many hours would it need to go from A to B if there were no current in the river?

Problem 132. For the first half of a trip a car has velocity v_1; for the second half of a trip it has the velocity v_2. What is the mean velocity of the car?

Problem 133. You know that $x + \dfrac{1}{x} = 7$. Compute (a) $x^2 + \dfrac{1}{x^2}$; (b) $x^3 + \dfrac{1}{x^3}$.

Problem 134. You know that $x + \dfrac{1}{x}$ is an integer. Prove that $x^n + \dfrac{1}{x^n}$ is an integer for any $n = 1, 2, 3$, etc.

Problem 135. Solving problem (d) on pages 56–57 we have seen that

$$\frac{1}{1+\dfrac{1}{x}} = \frac{x}{x+1}, \qquad \frac{1}{1+\dfrac{1}{1+\dfrac{1}{x}}} = \frac{x+1}{2x+1}, \qquad \frac{1}{1+\dfrac{1}{1+\dfrac{1}{1+\dfrac{1}{x}}}} = \frac{2x+1}{3x+2}.$$

Represent the fractions

$$\frac{1}{1+\dfrac{1}{1+\dfrac{1}{1+\dfrac{1}{1+\dfrac{1}{x}}}}}, \qquad \frac{1}{1+\dfrac{1}{1+\dfrac{1}{1+\dfrac{1}{1+\dfrac{1}{1+\dfrac{1}{x}}}}}}, \ldots$$

as quotients of two polynomials and try to find a law governing the coefficients of these polynomials. (These fractions are examples of so-called *continued fractions.* The coefficients of the polynomials in question turn out to be the so-called *Fibonacci numbers*; see page 87)

35 Polynomial and rational fractions in one variable

If a polynomial contains only one variable, its standard form consists of its monomials written in the order of decreasing degrees. The monomial having the highest degree is called the *first* monomial. Its degree is called the degree of the polynomial. (Of course, monomials with zero coefficients must be ignored. For a zero polynomial the degree is undefined.) For example, the polynomial $7x^2 + 3x + 1$ has the first monomial $7x^2$ and degree 2. Constants (not equal to zero) are considered as polynomials of degree 0.

Problem 136. What is the first term of the polynomial $(2x+1)^5$?

Problem 137. Assume that a polynomial P has degree m and the polynomial Q has degree n. Find the degree of their product $P \cdot Q$.

Solution. When multiplying the first monomials we get a monomial of degree $m+n$ (because $x^m \cdot x^n = x^{m+n}$); its coefficient is the product of the coefficients of x^m and x^n in P and Q. This monomial is the only one having degree $m+n$; all the others have smaller degree. So there is nothing to cancel it and thus it will remain after reducing similar terms.

Problem 138. (a) What can be said about the degree of the sum of two polynomials having degrees 7 and 9? (b) What can be said about the degree of the sum of two polynomials both having degree 7?

Answer. (a) It is 9; (b) any degree not exceeding 7 is possible.

Problem 139. Consider a polynomial in one variable x having degree 10. Substitute $y^7 + 5y^2 - y - 4$ for x in this polynomial and get a polynomial in y. What can be said about its degree?

36 Division of polynomials in one variable; the remainder

Common fractions are either *proper* or *improper*. A proper fraction is a fraction where the numerator is smaller than the denominator, such as $\frac{3}{7}$ or $\frac{1}{15}$. An improper fraction is a fraction where the numerator is not less than the denominator, such as $\frac{7}{5}$, $\frac{11}{11}$, or $\frac{37}{7}$.

Any improper fraction has an integer part, which is obtained when we divide the numerator by the denominator, plus a proper fraction. For example:

$7 = 1 \cdot 5 + 2$ (quotient 1, remainder 2) $\qquad \frac{7}{5} = 1 + \frac{2}{5} \qquad 5\overline{)7}$ 1 r. 2

Another example:

$37 = 5 \cdot 7 + 2$ (quotient 5, remainder 2) $\qquad \frac{37}{7} = 5 + \frac{2}{7} \qquad 7\overline{)37}$ 5 r. 2

Now an example where the remainder is zero:

$11 = 1 \cdot 11 + 0$ (quotient 1, remainder 0) $\qquad \frac{11}{11} = 1 \qquad 11\overline{)11}$ 1 r. 0

Now we shall learn to do similar transformations for fractions whose numerators and denominators are polynomials in one variable. Such a fraction is considered proper if the degree of its numerator is less than the degree of its denominator. For example, the fractions

$$\frac{10x}{x^2}, \quad \frac{1}{x^3 - 1}$$

are proper, while the fractions

$$\frac{x^4}{x-2}, \quad \frac{x+1}{x+2}, \quad \frac{x^3}{5x}, \quad \frac{x^3+1}{x+1}$$

are improper.

Any improper fraction can be converted into the sum of a polynomial and a proper fraction. Several examples:

(a) $\dfrac{x+3}{x+1} = \dfrac{(x+1)+2}{x+1} = 1 + \dfrac{2}{x+1}.$

(b) $\dfrac{x}{x+2} = \dfrac{(x+2)-2}{x+2} = 1 - \dfrac{2}{x+2}$.

(c) $\dfrac{x}{2x+1} = \dfrac{x+(1/2)}{2x+1} - \dfrac{1/2}{2x+1} = \dfrac{1}{2} - \dfrac{1/2}{2x+1}$.

(When we say that a polynomial must not contain division it does not mean that all its coefficients must be integers; they may be any numbers, including fractions. So, for example, $\dfrac{1}{2}$ is a perfectly legal polynomial of degree 0.)

(d) $\dfrac{x^2}{x-2} = \dfrac{(x^2-4)+4}{x-2} = \dfrac{(x+2)(x-2)+4}{x-2} = (x+2) + \dfrac{4}{x-2}$.

(e) $\dfrac{x^4}{x-2} = \dfrac{(x^4-16)+16}{x-2} = \dfrac{(x^2+4)(x+2)(x-2)+16}{x-2} =$

$$= (x^2+4)(x+2) + \frac{16}{x-2}.$$

There is a standard way of converting an improper fraction (where the numerator and the denominator are polynomials) into a sum of a polynomial and a proper fraction. It is similar to the division of numbers. Let us illustrate it by examples.

Example. Converting the improper fraction $\dfrac{x^4}{x-2}$:

$$
\begin{array}{r|l}
 & x^3 + 2x^2 + 4x + 8 \quad \leftarrow \quad \text{the quotient} \\
\cline{2-2}
x-2 & x^4 \\
 & \underline{x^4 - 2x^3} \\
 & \quad 2x^3 \\
 & \quad \underline{2x^3 - 4x^2} \\
 & \qquad 4x^2 \\
 & \qquad \underline{4x^2 - 8x} \\
 & \qquad\quad 8x \\
 & \qquad\quad \underline{8x - 16} \\
 & \qquad\qquad 16 \quad \leftarrow \quad \text{the remainder}
\end{array}
$$

The same procedure can be written in another, less readable, way:

$$
\begin{aligned}
\frac{x^4}{x-2} &= \frac{x^4-2x^3}{x-2} + \frac{2x^3}{x-2} = x^3 + \frac{2x^3}{x-2} = x^3 + \frac{2x^3-4x^2}{x-2} + \frac{4x^2}{x-2} = \\
&= x^3 + 2x^2 + \frac{4x^2}{x-2} = x^3 + 2x^2 + \frac{4x^2-8x}{x-2} + \frac{8x}{x-2} = \\
&= x^3 + 2x^2 + 4x + \frac{8x-16}{x-2} + \frac{16}{x-2} = x^3 + 2x^2 + 4x + 8 + \frac{16}{x-2}.
\end{aligned}
$$

So we get

$$x^4 = (x^3 + 2x^2 + 4x + 8)(x - 2) + 16.$$

Example. Now let us convert the fraction $\dfrac{x^3 + 2x}{x^2 - x + 1}$:

$$\begin{array}{r|l} & \;x + 1 \\ \cline{2-2} x^2 - x + 1 & x^3 + 2x \\ & \underline{x^3 - x^2 + x} \\ & \qquad x^2 + x \\ & \qquad \underline{x^2 - x + 1} \\ & \qquad\quad 2x - 1 \end{array}$$

The same conversion written in another way:

$$\begin{aligned} \frac{x^3 + 2x}{x^2 - x + 1} &= \frac{x^3 - x^2 + x}{x^2 - x + 1} + \frac{x^2 + x}{x^2 - x + 1} = x + \frac{x^2 + x}{x^2 - x + 1} = \\ &= x + \frac{x^2 - x + 1}{x^2 - x + 1} + \frac{2x - 1}{x^2 - x + 1} = (x + 1) + \frac{2x - 1}{x^2 - x + 1}. \end{aligned}$$

So we get

$$x^3 + 2x = (x + 1)(x^2 - x + 1) + (2x - 1).$$

Example. The last example of fraction conversion: $\dfrac{x^3}{2x - 3}$.

$$\begin{array}{r|l} & (1/2)x^2 + (3/4)x + (9/8) \\ \cline{2-2} 2x - 3 & x^3 \\ & \underline{x^3 - (3/2)x^2} \\ & \quad (3/2)x^2 \\ & \quad \underline{(3/2)x^2 - (9/4)x} \\ & \qquad (9/4)x \\ & \qquad \underline{(9/4)x - 27/8} \\ & \qquad\qquad 27/8 \end{array}$$

The same conversion:

$$\begin{aligned} \frac{x^3}{2x - 3} &= \frac{x^3 - (3/2)x^2}{2x - 3} + \frac{(3/2)x^2}{2x - 3} = \\ &= \frac{1}{2}x^2 + \frac{(3/2)x^2 - (9/4)x}{2x - 3} + \frac{(9/4)x}{2x - 3} \end{aligned}$$

$$\begin{aligned} &= \frac{1}{2}x^2 + \frac{3}{4}x + \frac{(9/4)x - (27/8)}{2x-3} + \frac{27/8}{2x-3} = \\ &= \left(\frac{1}{2}x^2 + \frac{3}{4}x + \frac{9}{8}\right) + \frac{27/8}{2x-3}. \end{aligned}$$

So we get

$$x^3 = \left(\frac{1}{2}x^2 + \frac{3}{4}x + \frac{9}{8}\right)(2x-3) + \frac{27}{8}$$

Now it is time for the exact definition of polynomial division.

Definition. Assume that we have two polynomials (in one variable), called the dividend and the divisor. To perform a division means to find two other polynomials, called the quotient and the remainder, such that

$$\boxed{\text{(dividend)} = \text{(quotient)}\cdot\text{(divisor)} + \text{(remainder)}}$$

where the degree of the remainder is less than the degree of the divisor (or the remainder is zero).

Problem 140. What can you say about the degrees of the remainder and the quotient if a polynomial of degree 7 is divided by a polynomial of degree 3?

Answer. The degree of the quotient is 4; the degree of the remainder may be 0, 1, 2, or 3 or undefined (the remainder may be absent or, rather, equal to zero).

Problem 141. Prove that the quotient and the remainder with the desired properties do exist and are unique.

Solution. In the examples above we have seen a method of finding the quotient and the remainder with the desired properties, so they do exist. Their uniqueness can be proved as follows. Assume that we divide P by S and have two possible quotients Q_1 and Q_2 and two corresponding remainders R_1 and R_2. So we have

$$\begin{aligned} P &= Q_1S + R_1 \\ P &= Q_2S + R_2 \end{aligned}$$

and both R_1 and R_2 have degree less than the degree of S. Then

$$Q_1S + R_1 = Q_2S + R_2$$

and, therefore,

$$R_1 - R_2 = Q_2S - Q_1S = (Q_1 - Q_2)S.$$

Here $R_1 - R_2$ is a difference of two polynomials of degree smaller than the degree of S, so their difference has degree smaller than the degree of S and cannot be a multiple of S unless it is equal to 0. Therefore, $Q_1 - Q_2 = 0$, that is, $Q_1 = Q_2$, hence, $R_1 = R_2$.

Problem 142. What happens if the degree of the dividend is smaller than the degree of the divisor?

Answer. In this case the fraction is already proper, so the quotient is equal to 0 and the remainder is equal to the dividend.

Polynomial division is similar to ordinary division:

$$\begin{array}{r|l} & \underline{\;\;112} \\ 11 & 1234 \\ & \underline{11} \\ & 13 \\ & \underline{11} \\ & 24 \\ & \underline{22} \\ & 2 \end{array} \qquad\qquad \begin{array}{r|l} & x^2 + x + 2 \\ x + 1 & x^3 + 2x^2 + 3x + 4 \\ & \underline{x^3 + x^2} \\ & \phantom{x^3 + {}} x^2 + 3x \\ & \phantom{x^3 + {}} \underline{x^2 + \;x} \\ & \phantom{x^3 + x^2 + {}} 2x + 4 \\ & \phantom{x^3 + x^2 + {}} \underline{2x + 2} \\ & \phantom{x^3 + x^2 + 2x + {}} 2 \end{array}$$

$$1234 = 112 \cdot 11 + 2 \qquad x^3 + 2x^2 + 3x + 4 = (x^2 + x + 2)(x + 1) + 2$$

In this example we have a perfect analogy; to see it, substitute 10 for x in the polynomial division. In other cases such as

$$\begin{array}{r|l} & x^2 + 3x + 6 \\ x - 1 & x^3 + 2x^2 + 3x + 4 \\ & \underline{x^3 - x^2} \\ & \phantom{x^3 + {}} 3x^2 + 3x \\ & \phantom{x^3 + {}} \underline{3x^2 - 3x} \\ & \phantom{x^3 + 3x^2 + {}} 6x + 4 \\ & \phantom{x^3 + 3x^2 + {}} \underline{6x - 6} \\ & \phantom{x^3 + 3x^2 + 6x + {}} 10 \end{array}$$

$$x^3 + 2x^2 + 3x + 4 = (x^2 + 3x + 6)(x - 1) + 10$$

the analogy is incomplete; if we substitute 10 for x, we get the equality $1234 = 136 \cdot 9 + 10$, which is true but does not mean that dividing 124 by 9 we get quotient 136 and remainder 10 (in fact, 137 is the quotient and 1 is the remainder).

Problem 143.

(a) Divide $x^3 - 1$ by $x - 1$;

(b) Divide $x^4 - 1$ by $x - 1$;

(c) Divide $x^{10} - 1$ by $x - 1$;

(d) Divide $x^3 + 1$ by $x + 1$;

(e) Divide $x^4 + 1$ by $x + 1$.

Problems (a)–(c) are special cases of the formula

$$\frac{x^n - 1}{x - 1} = x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1$$

which can be easily checked by division (as described) or just by multiplication of $x - 1$ and $x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1$. This formula can also be considered as a way to compute the sum of consecutive powers of a number x:

$$1 + x + x^2 + \cdots + x^{n-1} = \frac{x^n - 1}{x - 1}$$

(it is valid for all x except 1). See below about the sum of a geometric progression.

Problem 144. The powers of 2,

$$1, 2, 4, 8, 16, 32, 64, \ldots$$

have the following property: The sum of all members of this sequence up to any term is 1 less than the next term; for example

$$\begin{aligned} 1 + 2 &= 3 = 4 - 1 \\ 1 + 2 + 4 &= 7 = 8 - 1 \\ 1 + 2 + 4 + 8 &= 15 = 16 - 1 \end{aligned}$$

and so on. Explain why.

Solution. Look at the equation

$$1 + x + x^2 + \cdots + x^{n-1} = \frac{x^n - 1}{x - 1}$$

when $x = 2$. We get

$$1 + 2 + 2^2 + \cdots + 2^{n-1} = \frac{2^n - 1}{2 - 1} = 2^n - 1.$$

Another solution. To compute the sum $1 + 2 + 4 + 8 + 16$, let us add and subtract 1:

$$\begin{aligned}
&1 + 2 + 4 + 8 + 16 = \\
&= \quad (1 + 1 + 2 + 4 + 8 + 16) - 1 = \\
&= \quad (2 + 2 + 4 + 8 + 16) - 1 = \\
&= \quad (4 + 4 + 8 + 16) - 1 = \\
&= \quad (8 + 8 + 16) - 1 = \\
&= \quad (16 + 16) - 1 = \\
&= \quad 32 - 1.
\end{aligned}$$

37 The remainder when dividing by $x - a$

There is a method that allows us to find the remainder of an arbitrary polynomial divided by $x - a$ without actually performing the division.

Assume that we want to find the remainder when x^4 is divided by $x - 2$. We can be sure that the remainder is a number (its degree must be less than the degree of $x - 2$). To find this number, look at the equality

$$x^4 = (\text{quotient})(x - 2) + (\text{remainder})$$

and substitute $x = 2$. We get

$$2^4 = (\ldots) \cdot 0 + (\text{remainder});$$

so the remainder is equal to $2^4 = 16$.

In general, if P is an arbitrary polynomial that we want to divide by $x - a$ (where a is some number), we write

$$P(x) = (\text{quotient})(x - a) + (\text{remainder})$$

and substitute a for x. Therefore,

> To find the remainder when P is divided by $x-a$, substitute a for x in P.

This rule is called the *remainder theorem*, or Bezout's theorem. It allows us to find the remainder without the actual division. However, if you want to know the quotient, you need to perform the division.

Here is a consequence of Bezout's theorem:

> To find whether a polynomial P is divisible by $x-a$ (without remainder), test whether it becomes zero after substitution of a for x.

If a polynomial P becomes zero when some number a is substituted for x, then this number a is called a *root* of the polynomial P. Therefore we may say

$$P \text{ is divisible by } x-a \iff a \text{ is a root of } P.$$

Problem 145. (a) For which n is the polynomial x^n-1 divisible by $x-1$? (b) For which n is the polynomial x^n+1 divisible by $x+1$?

After we find a root of a polynomial we may factor it; $x-a$ is one of the factors. Then we may try to apply the same method to the quotient.

Problem 146. Factor these polynomials:

(a) x^4+5x-6;

(b) x^4+3x^2+5x+1;

(c) x^3-3x-2.

Problem 147. The numbers 1 and 2 are roots of a polynomial P. Prove that P is divisible by $(x-1)(x-2)$.

Solution. P is divisible by $x-1$ because 1 is a root of P. Therefore $P=(x-1)\cdot Q$ for some polynomial Q. Substituting 2 for x in this equality we find that 2 is a root of Q, so Q is divisible by $x-2$, that is, $Q=(x-2)\cdot R$ for some polynomial R. So $P=(x-1)(x-2)R$.

Remark. A typical wrong solution goes as follows: P is divisible by $x-1$ (because 1 is a root) and by $x-2$ (because 2 is a root), therefore P is divisible by $(x-1)(x-2)$. The error: "therefore" is not justified. For example, 12 is divisible by 6 and by 4, but we may not say "therefore, 12 is divisible by $6 \cdot 4 = 24$".

A similar argument shows that

> If different numbers $a_1, a_2, \ldots, a_n$ are roots of a polynomial P, then P is divisible by $(x - a_1)(x - a_2)\cdots(x - a_n)$.

Problem 148. What is the maximal possible number of roots for a polynomial having degree 5?

Solution. The answer is 5. For example, the polynomial

$$(x-1)(x-2)(x-3)(x-4)(x-5)$$

has 5 roots. More than 5 roots is impossible; if a polynomial P had 6 roots $a_1, a_2, a_3, a_4, a_5, a_6$, then it would be divisible by

$$(x - a_1)(x - a_2)\cdots(x - a_6),$$

that is,

$$P = (x - a_1)(x - a_2)\cdots(x - a_6)Q$$

for some Q. That is impossible because the degree of the right-hand side is at least 6.

In general, a polynomial of degree n may have at most n different roots.

Remark. We used here the expression "different roots" because the words "the number of roots" may be used in a different sense. For example, what is the number of roots of the polynomial $x^2 - 2x + 1$? The polynomial is equal to $(x-1)^2$, so $x = 1$ is a root and all $x \neq 1$ are not roots. So we may say that it has exactly one root. On the other hand, the general formula for a polynomial with two roots a and b is

$$c(x-a)(x-b)$$

and our polynomial

$$x^2 - 2x + 1 = (x-1)^2 = (x-1)(x-1)$$

is a special case of this formula when $a = b = 1$ (and $c = 1$); so mathematicians often say that this polynomial has "two equal roots".

We shall not use this terminology in this book but you may see it, for example, in the statement of the so-called "fundamental theorem of algebra" claiming that "any polynomial of degree n has exactly n complex roots"

Problem 149. How should you check whether a given polynomial P is divisible by $x^2 - 1$?

Answer. Check whether 1 and -1 are roots of P.

Problem 150. For which n is the polynomial $x^n - 1$ divisible by $x^2 - 1$?

Now let us return to the identity

$$\frac{(x-a)(x-b)}{(c-a)(c-b)} + \frac{(x-a)(x-c)}{(b-a)(b-c)} + \frac{(x-b)(x-c)}{(a-b)(a-c)} - 1 = 0$$

which we discussed on page 59 (we have moved 1 to the left-hand side of the equation). Assume that a, b, c are different numbers. Consider the left-hand side as a polynomial with one variable x. The degree of this polynomial does not exceed 2. Therefore it may have at most two roots (if it is not equal to zero). But a, b, and c are its roots! Therefore, it is equal to zero.

A careful reader would say that we made a big mistake: we confuse the equality of rational expressions for all numerical values of a, b, c, x strictly speaking, not even for all, because, for example, the left-hand side is undefined when $a = b$) with the possibility of transforming the left-hand side to zero according to algebraic rules. What can be said about this? Bad news: this really is a problem. Good news: this problem is not fatal (but to justify this transition you need some theory).

Problem 151. The remainder of a polynomial P (in one variable x) when divided by $x^2 - 1$ is a polynomial of degree at most 1, that is, it has the form $ax + b$ for some numbers a and b. How can you find a and b if you know the values of P when $x = -1$ and $x = 1$?

Hint. Look at the equality

$$P = (x^2 - 1)(\text{quotient}) + (ax + b)$$

and substitute 1 and -1 for x.

Problem 152. The polynomial P gives a remainder of $5x-7$ when divided by x^2-1. Find the remainder when P is divided by $x-1$.

Problem 153. The polynomial $P = x^3+x^2-10x+1$ has three different roots (the authors guarantee it) denoted by x_1, x_2, x_3. Write a polynomial with integer coefficients having roots
(a) x_1+1, x_2+1, x_3+1; (b) $2x_1, 2x_2, 2x_3$; (c) $\dfrac{1}{x_1}, \dfrac{1}{x_2}, \dfrac{1}{x_3}$.

Problem 154. Assume that x^3+ax^2+x+b (where a and b are some numbers) is divisible by x^2-3x+2. Find a and b.

38 Values of polynomials, and interpolation

Assume that a polynomial P includes only one letter (variable) x. To stress this we denote this polynomial by $P(x)$ ("P of x"). Substitute some number, say 6, for x in P and perform all computations. We get a number. This number is called the value of the polynomial P for $x=6$ and is denoted by $P(6)$ ("P of 6").

For example, if $P(x) = x^2-x-4$ then $P(0) = 0^2-0-4 = -4$. Other values are $P(1) = -4$, $P(2) = -2$, $P(3) = 2$, $P(4) = 8$, $P(5) = 16$, $P(6) = 26$, etc.

Problem 155. Calculate a table of values $P(0), \ldots, P(6)$ for the polynomial $P(x) = x^3-2$.

Problem 156. Let us write the values $P(0)$, $P(1)$, $P(2), \ldots$ for $P(x) = x^2-x-4$:

$$-4,\ -4,\ -2,\ \ 2,\ \ 8,\ 16,\ 26, \ldots$$

Under any two adjacent numbers write their difference:

$$\begin{array}{ccccccccccccccc}
-4 & & -4 & & -2 & & 2 & & 8 & & 16 & & 26 & \ldots \\
 & 0 & & 2 & & 4 & & 6 & & 8 & & 10 & \ldots &
\end{array}$$

and repeat the same operation with this sequence of "first differences":

$$\begin{array}{ccccccccccccccc}
-4 & & -4 & & -2 & & 2 & & 8 & & 16 & & 26 & \ldots \\
 & 0 & & 2 & & 4 & & 6 & & 8 & & 10 & \ldots & \\
 & & 2 & & 2 & & 2 & & 2 & & 2 & \ldots & &
\end{array}$$

Now all numbers are 2s. Prove that it is not a coincidence and that all subsequent numbers (called "second differences") are also 2s.

Problem 157. Prove that for any polynomial of degree 2 all second differences are equal.

Problem 158. What can be said about polynomials having degree 3?

Problem 159. (L. Euler) Compute the values $P(x) = x^2 + x + 41$ for $x = 1, 2, 3, \ldots, 10$. Check that all these values are prime numbers (having no divisors except 1 and themselves). Might it be that all of $P(1), P(2), P(3), \ldots$ are prime numbers for this polynomial P?

Now we address the following topic: What can be said about a polynomial if we have some information about its values?

By a polynomial of degree *not exceeding* n we mean any polynomial of degree n, $n-1, \ldots, 2$, 1, 0, or the zero polynomial (whose degree is undefined).

For example, the general form of a polynomial of degree not exceeding 1 is $ax+b$. When $a \neq 0$ it has degree 1. When $a = 0, b \neq 0$ it has degree 0. When $a = b = 0$ we get the zero polynomial whose degree is undefined.

In a similar way the general form of a polynomial of degree not exceeding 2 is $ax^2 + bx + c$, etc.

Problem 160. You know that $P(x)$ is a polynomial of degree not exceeding 1, that $P(1) = 7$, and that $P(2) = 5$. Find $P(x)$.

Solution. By definition, $P(x) = ax + b$, where a and b are some numbers. Let us substitute $x = 1$ and $x = 2$. We get:

$$\begin{array}{rcccl} P(1) & = & a+b & = & 7 \\ P(2) & = & 2a+b & = & 5 \end{array}$$

Comparing this equations we see that after adding one more a (in the second one), 7 becomes 5, so $a = -2$. Therefore $b = 9$. Answer: $P(x) = -2x + 9$.

The same method can be applied to find a polynomial of degree not exceeding 1 if we know its values for any two different values of x.

If you know that the graph of a function $y = ax+b$ is a straight line you can easily explain this fact geometrically; for any two points there is exactly one straight line going through these points. (Two given values for two values of x correspond to two points in the plane.)

Problem 161. A polynomial $P(x)$ of degree not exceeding 1 satisfies the conditions $P(1) = 0$, $P(2) = 0$. Prove that $P(x) = 0$ for any x.

Now we consider polynomials of degree not exceeding 2. How many values do we need to reconstruct such a polynomial? We shall see that two is not enough.

Problem 162. A polynomial $P(x)$ of degree not exceeding 2 satisfies the conditions $P(1) = 0$, $P(2) = 0$. Can we conclude that $P(x) = 0$?

Solution. No; look at the polynomial $P(x) = (x-1)(x-2) = x^2 - 3x + 2$.

We already know that any polynomial $P(x)$ such that $P(1) = P(2) = 0$ has the form $P(x) = (x-1)(x-2)Q(x)$ where $Q(x)$ is some polynomial. If we also know that $P(x)$ has degree not exceeding 2, then $Q(x)$ must be a number (otherwise the degree of P will be too big).

Problem 163. A polynomial $P(x)$ of degree not exceeding 2 satisfies the conditions $P(1) = 0$, $P(2) = 0$, $P(3) = 4$. Find $P(x)$.

Solution. As we have seen, $P(x) = a(x-1)(x-2)$ where a is some constant. To find a, substitute $x = 3$:

$$P(3) = a(3-1)(3-2) = 2a = 4;$$

therefore $a = 2$. Answer: $P(x) = 2(x-1)(x-2) = 2x^2 - 6x + 4$.

Another solution. Any polynomial of degree not exceeding 2 has the form $ax^2 + bx + c$. Substituting $x = 1$, $x = 2$ and $x = 3$, we get

$$\begin{array}{rcccl} P(1) & = & a+b+c & = & 0 \\ P(2) & = & 4a+2b+c & = & 0 \\ P(3) & = & 9a+3b+c & = & 4. \end{array}$$

Therefore,

$$\begin{array}{rcccl} P(2)-P(1) & = & 3a+b & = & 0 \\ P(3)-P(2) & = & 5a+b & = & 4. \end{array}$$

Additional $2a$ make 4 from 0, therefore $a = 2$. Now we can find $b = -6$ and then $c = 4$. Answer: $2x^2 - 6x + 4$.

Problem 164. Prove that a polynomial of degree not exceeding 2 is defined uniquely by three of its values.

This means that if $P(x)$ and $Q(x)$ are polynomials of degree not exceeding 2 and $P(x_1) = Q(x_1)$, $P(x_2) = Q(x_2)$, $P(x_3) = Q(x_3)$ for three different numbers x_1, x_2, and x_3, then the polynomials $P(x)$ and $Q(x)$ are equal.

Solution. Consider a polynomial $R(x) = P(x) - Q(x)$. Its degree does not exceed 2. On the other hand, we know that

$$R(x_1) = R(x_2) = R(x_3) = 0;$$

in other words, x_1, x_2, and x_3 are roots of the polynomial $R(x)$. But a polynomial of degree not exceeding 2, as we know, cannot have more than 2 roots, unless it is equal to zero. Therefore $R(x)$ is equal to zero and $P(x) = Q(x)$.

Problem 165. Assume that

$$\begin{aligned} 16a + 4b + c &= 0 \\ 49a + 7b + c &= 0 \\ 100a + 10b + c &= 0. \end{aligned}$$

Prove that $a = b = c = 0$.

Problem 166. Prove that a polynomial of degree not exceeding n is defined uniquely by its $n + 1$ values. (We have already solved this problem for $n = 2$.)

Problem 167. Find a polynomial $P(x)$ of degree not exceeding 2 such that

(a) $P(1) = 0$, $P(2) = 0$, $P(3) = 4$;

(b) $P(1) = 0$, $P(2) = 2$, $P(3) = 0$;

(c) $P(1) = 6$, $P(2) = 0$, $P(3) = 0$;

(d) $P(1) = 6$, $P(2) = 2$, $P(3) = 4$.

Solution. Problem (a) was already solved, and the answer was $2(x-1)(x-2)$. Problems (b) and (c) may be solved by the same method; the answers are $-2(x-1)(x-3)$ for (b) and $3(x-2)(x-3)$ for (c). Now we are able to solve (d) by just adding the three polynomials from (a), (b), and (c). We get the following answer:

$$\begin{aligned} &2(x-1)(x-2)-2(x-1)(x-3)+3(x-2)(x-3)= \\ &= \quad 2x^2-6x+4-2x^2+8x-6+3x^2-15x+18= \\ &= \quad 3x^2-13x+16. \end{aligned}$$

Another solution for (d). Let us find any polynomial Q of degree not exceeding 2 such that $Q(1)=6$ and $Q(2)=2$. For example, the polynomial $Q(x)=10-4x$ (having degree 1) will work. It has two desired values $Q(1)$ and $Q(2)$, but unfortunately the value $Q(3)$ is not what we want: $Q(3)=-2$ (and we want 4). The remedy: consider a polynomial

$$P(x)=Q(x)+a(x-1)(x-2).$$

Any a would not damage the values $P(1)=Q(1)=6$, $P(2)=Q(2)=2$. And a suitable a will make $P(3)$ correct:

$$P(3)=Q(3)+2a.$$

To get $P(3)=4$ we use $a=3$. So the answer is

$$\begin{aligned} P(x) &= 10-4x+3(x-1)(x-2)= \\ &= \quad 10-4x+3x^2-9x+6=3x^2-13x+16. \end{aligned}$$

Problem 168. Find a polynomial $P(x)$ of degree not exceeding 3 such that $P(-1)=2$, $P(0)=1$, $P(1)=2$, $P(2)=7$.

Problem 169. Assume that $x_1,\dots,x_{10}$ are different numbers, and $y_1,\dots,y_{10}$ are arbitrary numbers. Prove that there is one and only one polynomial $P(x)$ of degree not exceeding 9 such that $P(x_1)=y_1$, $P(x_2)=y_2,\dots,P(x_{10})=y_{10}$.

Problem 170. Without any computations prove that there exist numbers a, b, and c such that

$$\begin{aligned} 100a+10b+c &= 18.37 \\ 36a+6b+c &= 0.05 \\ 4a+2b+c &= -3 \end{aligned}$$

(You don't need to find these a, b, and c; it is enough to prove that they exist.)

Problem 171. The highest coefficient of $P(x)$ is 1, and we know that $P(1) = 0$, $P(2) = 0$, $P(3) = 0, \ldots$, $P(9) = 0$, $P(10) = 0$. What is the minimal possible degree of $P(x)$? Find $P(11)$ for this case.

Answer. The minimal degree is 10 and $P(11)$ is 3628800 in this case.

39 Arithmetic progressions

In the sequence of numbers

$$3,\ 5,\ 7,\ 9,\ 11,\ \ldots$$

each term is greater than the preceding one by two units. In the sequence

$$10,\ 9,\ 8,\ 7,\ 6,\ \ldots$$

each term is one unit smaller than the preceding one. Such sequences are called *arithmetic progressions.* (Here "e" is stressed: arithmEtic, not arIthmetic!)

Definition. An *arithmetic progression* is a sequence of numbers where each term is a sum of the preceding one and a fixed number. This fixed number is called the *common difference*, or simply *difference*, of the arithmetic progression.

Problem 172. What are the differences in the examples above?

Answer. 2 and -1.

Problem 173. Find the third term of an arithmetic progression

$$5,\ -2,\ \ldots$$

Answer. -9.

Problem 174. Find the 1000th term of an arithmetic progression

$$2,\ 3,\ 4,\ 5,\ 6,\ \ldots$$

Solution. If the progression were

$$1,\ 2,\ 3,\ 4,\ 5,\ \ldots$$

the first term would be 1, the second term would be 2, ..., the 1000th term would be 1000. In our progression, each term is one unit bigger. So the answer is 1001.

Problem 175. Find the 1000th term of the progression

$$2,\ 4,\ 6,\ 8,\ \ldots$$

Problem 176. Find the 1000th term of the progression

$$1,\ 3,\ 5,\ 7,\ \ldots$$

Problem 177. The first term of a progression is a, its difference is d. Find the 1000th term of the progression. Find its nth term.

Solution.

1st term	a
2nd term	$a+d$
3rd term	$a+2d$
4th term	$a+3d$
5th term	$a+4d$
...	...
1000th term	$a+999d$
...	...
nth term	$a+(n-1)d$

Problem 178. An arithmetic progression with difference d is rewritten in the reverse order, from right to left. Do we get an arithmetic progression? If so, what is its difference?

Problem 179. In an arithmetic progression whose difference is d every second term is deleted. Do we get an arithmetic progression? If so, what is its difference?

Problem 180. The same question if every *third* term is deleted.

Problem 181. The first term of an arithmetic progression is 5, the third term is 8. Find the second term.

Answer. 6.5.

Problem 182. The first term of an arithmetic progression is a, the third term is b. Find the second term.

Answer. $(a+b)/2$.

Problem 183. The first term of an arithmetic progression is a, the 4th term is b. Find the second and the third terms.

Problem 184. Consider the progression

$$1,\ 3,\ 5,\ 7,\ \ldots,\ 993,\ 995,\ 997,\ 999\,.$$

How many terms does it have?

Hint. The nth term is equal to $2n-1$. (Another way is to compare it with the progression 2, 4, 6, ..., 1000.)

40 The sum of an arithmetic progression

Problem 185. Compute the sum

$$1+3+5+7+\cdots+999\,.$$

Solution. First of all we have to find out how many terms are in this sum (see the problem above). The nth term of this progresson is equal to $1+(n-1)\cdot 2 = 2n-2+1 = 2n-1$. It is equal to 999 when $n = 500$. So this progression contains 500 terms. Let us combine them into 250 pairs:

$$(1+999)+(3+997)+\cdots+(499+501)\,.$$

The sum of each pair is equal to 1000. Thus, the answer is 250,000.

Problem 186. The first term of a progression containing n terms is a, its last (nth) term is b. Find the sum of its terms.

Solution. Grouping terms into pairs (as in the preceding problem) we get $n/2$ pairs, and the sum of each pair is $a+b$. Thus, the answer is $\dfrac{n(a+b)}{2}$.

Problem 187. There is an error in the solution of the preceding problem (however, the answer is valid). Find and correct this error.

Solution. All is O.K. if n is even. But when n is odd, the middle term remains unpaired. To avoid the distinction between odd and even numbers of terms, we apply the following trick. Assume that the sum in question is

$$S = 3+5+7+9+11\,.$$

Rewrite it in the reverse order:

$$S = 11 + 9 + 7 + 5 + 3\,.$$

Now we add these two equalities:

$$\begin{array}{rrrrrrr} 2S = & 3 & +5 & +7 & +9 & +11 & + \\ & +11 & +9 & +7 & +5 & +\ 3 & \end{array}$$

and find out that in each column we have two numbers whose sum is 14:

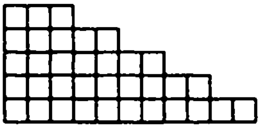
$$3 + 11 = 5 + 9 = 7 + 7 = 9 + 5 = 11 + 3 = 14\,.$$

So $2S = 5 \cdot 14 = 70$, $S = \dfrac{5 \cdot 14}{2} = 35$.

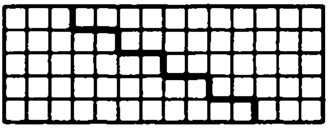
In the general case we have n columns with the same sum equal to the sum of the first and the last terms, that is, $a + b$. Therefore,

$$S = \frac{n(a+b)}{2}.$$

This argument can be illustrated by a picture. The sum $3+5+7+9+11$ can be drawn as

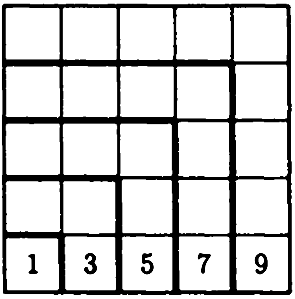


Two such pieces form a rectangle 5×14:



Problem 188. Prove that the sum of n first odd numbers is a perfect square ($1 = 1^2$, $1 + 3 = 2^2$, $1 + 3 + 5 = 3^2$ etc.)

Hint. You may use the preceding problem or the following picture:



41 Geometric progressions

In the sequence of numbers

$$3,\ 6,\ 12,\ 24,\ \ldots$$

each term is two times bigger than the preceding one. In the sequence

$$6,\ 2,\ \frac{2}{3},\ \frac{2}{9},\ \frac{2}{27},\ldots$$

each term is three times smaller than the preceding one. Such sequences are called *geometric progressions*.

Definition. A *geometric progression* is a sequence of numbers where each term is a product of the preceding one and a fixed number. This fixed number is called the *common ratio* (or *ratio*) of the geometric progression.

Problem 189. Find the common ratios of the progressions shown above.

Answer. 2, 1/3.

Problem 190. Find the third term of the geometric progression

$$2, 3, \ldots$$

Answer. 9/2.

Problem 191. Find the 1000th term of the geometric progression $3, 6, 12, \ldots$

Solution.

$$\begin{array}{rrcl}
\text{1st term} & 3 & = & 3\cdot 2^0 \\
\text{2nd term} & 6 & = & 3\cdot 2^1 \\
\text{3rd term} & 12 & = & 3\cdot 2^2 \\
\ldots & & & \ldots \\
\text{1000th term} & & = & 3\cdot 2^{999}
\end{array}$$

Problem 192. Find the 1000th term of a geometric progression whose first term is a and whose common ratio is q.

Solution.

$$\begin{array}{rcl} \text{1st term} & = & a = a \cdot q^0 \\ \text{2nd term} & = & a \cdot q = a \cdot q^1 \\ \text{3rd term} & = & a \cdot q^2 \\ \text{4th term} & = & a \cdot q^3 \\ \dots & & \dots \\ n\text{th term} & = & a \cdot q^{n-1} \end{array}$$

Problem 193. The first term of a geometric progression is 1, the third term is 4. Find the second term. Is your answer the only possible one?

Answer. There are two possibilities: 2 and -2.

Problem 194. A bacterium dividing one a minute fills a vessel in 30 minutes. How much time is necessary for two bacteria to fill the same vessel?

Let us look at the sequence

$$1, 0, 0, 0, \dots$$

Is it a geometric progression or not? According to our definition it is – each term is equal to the preceding one multiplied by zero (and there is no requirement for the common ratio to be nonzero). Though this sequence looks strange, we do consider it as a geometric progression. (But in some cases we would require the common ratio of a progression to be nonzero.)

Problem 195. A geometric progression whose common ratio is $q \neq 0$ is rewritten in the reverse order, from right to left. Do we get a geometric progression? If so, what is its common ratio?

Answer. $1/q$.

Problem 196. In a geometric progression whose common ratio is q every second term is deleted. Do we get a geometric progression? If so, what is its common ratio?

Problem 197. The same question if every *third* term is deleted.

Problem 198. The first term of a geometric progression is a and the third term is b. Find the second term.

Solution. Assume that x is the second term. Then the common ratio is equal to x/a and at the same time to b/x. Therefore, $x/a = b/x$; multiplying this equality by ax, we get $x^2 = ab$. Therefore, if $ab < 0$ the problem has no solutions (such a progression does not exist); if $ab = 0$ then $x = 0$; if $ab > 0$ there are two possibilities: $x = \sqrt{ab}$ and $x = -\sqrt{ab}$ (see below about square roots).

Remark. Our solution is not applicable when $x = 0$ or $a = 0$. But our formula turns out to be more clever than we may expect. For example, if $a = 1$, $b = 0$ then our formula gives the correct answer $x = \sqrt{1 \cdot 0} = 0$.

Problem 199. The first term of a geometric progression is 1, and its fourth term is $a > 0$. Find the second and the third terms of this progression.

Hint. See below about cube roots.

Answer. $\sqrt[3]{a}$, $\sqrt[3]{a^2}$.

42 The sum of a geometric progression

Problem 200. Compute the sum $1 + 2 + 4 + 8 + \cdots + 512 + 1024$ (each term is twice the preceding one).

Solution. Let us add 1 to this sum:

$$\begin{aligned}
1 + 1 + 2 + 4 + 8 + \cdots + 1024 &= \\
= 2 + 2 + 4 + 8 + \cdots + 1024 &= \\
= 4 + 4 + 8 + \cdots + 1024 &= \\
= 8 + 8 + \cdots + 1024 &= \\
= 16 + \cdots + 1024 &= \\
\cdots & \\
= 256 + 256 + 512 + 1024 &= \\
= 512 + 512 + 1024 &= \\
= 1024 + 1024 &= \\
= 2048 &
\end{aligned}$$

So the answer is $2048 - 1 = 2047$.

Another solution. Let us denote this sum by S. Then

$$S = 1 + 2 + 4 + 8 + \cdots + 512 + 1024$$

and

$$2S = 2 + 4 + 8 + 16 + \cdots + 1024 + 2048.$$

The latter sum (compared with the first one) contains an extra term 2048 but does not contain the term 1. So the difference is

$$2S - S = 2048 - 1.$$

Therefore, $S = 2048 - 1 = 2047$.

Problem 201. Compute the sums $1 + \frac{1}{2}$, $1 + \frac{1}{2} + \frac{1}{4}$, $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8}$, ..., $1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{1024}$.

Problem 202. The first term of a geometric progression is a and its common ratio is q. Find the sum of the first n terms of this progression.

Solution. The sum in question is equal to

$$a + aq + aq^2 + \cdots + aq^{n-1} = a(1 + q + q^2 + \cdots + q^{n-1}).$$

Recalling the factorization

$$q^n - 1 = (q - 1)(q^{n-1} + q^{n-2} + \cdots + q + 1)$$

we find out that

$$1 + q + \cdots + q^{n-1} = \frac{q^n - 1}{q - 1},$$

so the sum in question is equal to

$$a\frac{q^n - 1}{q - 1}.$$

Another solution. Let us denote the sum in question by S:

$$S = a + aq + \cdots + aq^{n-2} + aq^{n-1}.$$

Multiply it by q:

$$qS = aq + aq^2 + \cdots + aq^{n-1} + aq^n.$$

A new term aq^n appeared and term a disappeared, so

$$\begin{aligned} qS - S &= aq^n - a \\ (q-1)S &= a(q^n - 1) \\ S &= a\frac{q^n - 1}{q - 1}. \end{aligned}$$

Problem 203. The solution of the preceding problem has a gap; find it.

Solution. When $q = 1$ the answer given above is absurd; the quotient

$$\frac{1^n - 1}{1 - 1}$$

is undefined. In this case all terms of the progression are equal and the sum is equal to na. So one could say that in a sense

$$\frac{q^n - 1}{q - 1} = n, \quad \text{if } q = 1.$$

(This is a joke, of course – but it is also the computation of the derivative of the function $f(x) = x^n$ from the calculus textbooks!)

43 Different problems about progressions

Problem 204. Is it possible that numbers $1/2$, $1/3$, and $1/5$ are (not necessarily adjacent) terms of the same arithmetic progression?

Hint. Yes. Try $1/30$ as a difference.

Problem 205. Is it possible that the numbers 2, 3, and 5 are (not necessarily adjacent) terms of a geometric progression?

Solution. No, it is impossible. Assume that the common ratio of this progression is equal to q. Then

$$3 = 2q^n, \quad 5 = 3q^m$$

for some m and n. So we get

$$q^n = \frac{3}{2}, \quad q^m = \frac{5}{3}$$

and

$$\left(\frac{3}{2}\right)^m = q^{mn} = \left(\frac{5}{3}\right)^n.$$

Therefore,

$$\frac{3^m}{2^m} = \frac{5^n}{3^n}$$

and $3^{m+n} = 2^m \cdot 5^n$. The left-hand side is an odd number, and the right-hand side is an even number if $m \neq 0$. Hence, m must be equal to 0. But this is also impossible because in this case we would get

$$5 = 3q^m = 3 \cdot 1 = 3\,.$$

So we get a contradiction showing that the requirements $3 = 2q^n$ and $5 = 3q^m$ are inconsistent. Hence 2, 3, and 5 could not be terms of the same progression.

Problem 206. In this argument we assumed that the numbers 2, 3, and 5 occur in the progression in this order (because we assume implicitly that m and n are positive integers). What should we do in other cases?

Problem 207. Is it possible that two first terms of an arithmetic progression are integers, but all succeeding terms are not?

Solution. This is impossible; if two adjacent terms are integers, then the difference of the progression is an integer, and all the other terms are also integers.

Problem 208. Is it possible that the first 10 terms of a geometric progression are integers, but all succeeding terms are not?

Solution. Yes, it is possible:

$$512,\ 256,\ 128,\ 64,\ 32,\ 16,\ 8,\ 4,\ 2,\ 1,\ \frac{1}{2},\ \frac{1}{4},\ \ldots$$

Problem 209. Is it possible that the second term of an arithmetic progression is less than its first term and also less than its third term?

Solution. No, in this case the difference of the progression would be positive and negative at the same time.

Problem 210. The same question for a geometric progression.

Solution. Yes, for example, in the progression $1, -1, 1$.

Problem 211. Is it possible that an infinite arithmetic progression contains exactly one integer term?

Hint. Consider the progression with first term 0 and difference $\sqrt{2}$ and use the fact that $\sqrt{2}$ is an irrational number (see below).

Problem 212. Is it possible that an infinite arithmetic progression contains exactly two integer terms?

Answer. No.

Problem 213. In the sequence

$$1,\ 3,\ 7,\ 15,\ 31,\ \ldots$$

each term is equal to $2 \times$ (the preceding term) $+ 1$. Find the 100th term of this sequence.

Answer. $2^{100} - 1$.

Problem 214. In a geometric progression each term is equal to the sum of two preceding terms. What can be said about the common ratio of this progression?

Hint. See below about quadratic equations.

Answer. There are two possible common ratios:

$$\frac{1+\sqrt{5}}{2} \quad \text{and} \quad \frac{1-\sqrt{5}}{2}.$$

Problem 215. The *Fibonacci sequence*

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ \ldots$$

is defined as follows: The first two terms are equal to 1, and each subsequent term is equal to the sum of the two preceding terms. Find numbers A and B such that (for all n), the nth term of the Fibonacci sequence is equal to

$$A\left(\frac{1+\sqrt{5}}{2}\right)^n + B\left(\frac{1-\sqrt{5}}{2}\right)^n$$

44 The well-tempered clavier

A musical sound (a tone) consists of air oscillations (produced by string oscillations if we have a string instrument such as a violin or a piano).

The number of oscillations per second is called the *frequency* of oscillations. For example, the note A (*la*) of the fourth octave of a piano has a frequency of 440 oscillations per second (according to the modern standard; the frequency was lower in the past). The higher the tone is, the greater is its frequency.

When we hear two tones together, they form an *interval* (as musicians say). This interval may be consonant (harmonious, nice to hear) or dissonant (not so nice) – or something in between. It turns out that this depends on the ratio of frequencies of the two tones forming the interval. The rule is as follows: A consonant interval appears when the frequency ratio is equal (or very close) to the ratio of two small integers.

For example, an *octave* interval appears when the ratio is equal to $2 = 2/1$. The notes forming an octave interval have the same name. For example, all notes with frequencies 440, 880, 1760, and so on – as well as notes with frequencies 220, 110, and so on – share the name A (*la*), but belong to "different octaves", as musicians say. The octave interval opens the "Campanella" (the final part of the second violin concerto by Paganini; both tones are F-sharp).

A *fifth* is an interval whose frequencies ratio is equal to $3/2$ (or very close to $3/2$; see below). The adjacent strings on a violin form such an interval (G–D, D–A, or A–E). The final part of the Brahms concerto for violin and cello (A minor) starts with a fifth (A–E) played by the cello.

The interval with frequency ratio $4/3$ is called a *fourth*, the interval with ratio $5/4$ is called a *major third*, and the interval with ratio $6/5$ is called a *minor third*.

Problem 216. In a three-tone melody the first two tones form a fifth (the second tone is lower), and the next two tones form a fourth (the last tone is lower). What is the interval between the first and the third tones?

Solution. $\frac{3}{2} \times \frac{4}{3} = \frac{2}{1}$, so we get an octave.

Problem 217. A *minor sixth* complements a *major third* to form an octave; a *major sixth* complements a *minor third* to form an octave. What are the frequency ratios for minor and major sixths?

The same melody can be played in different keys; transposing it, we change the key. From the mathematical point of view transposition means that all frequencies are multiplied by a fixed number. So the frequency ratios remain unchanged, and consonant intervals remain consonant. You can observe this when you try your $33\frac{1}{3}$ record using a 45 rpm player. (The side effect is that music becomes not only higher in tone but also faster.)

Problem 218. How are the frequencies changed in this case?

Now we shall explain the connection between the well-tempered clavier and geometric progressions. It turns out that the following statement is true:

> *If the clavier (piano, harpsichord) is well tempered, that is, any melody can be transposed to start from any given tone, then the frequencies of the tones form a geometric progression.*

Problem 219. Prove this statement.

Solution. Consider the *chromatic scale*, that is, the sequence of tones starting from a certain tone and going in increasing order (without gaps). Let's transpose it; we still get a chromatic scale. (If this new melody did not use some specific tone, then adding this tone would give us a melody that could not be transposed back.) If the initial tone of the transposed chromatic scale and the original scale are neighbors, then each tone is mapped to its neighbor tone after the transposition. In other words, we get the frequency of a neighbor tone when multiplying the frequency of the original tone by some constant. This is the definition of a geometric progression.

Now let us denote the frequency of tone A by a, and the common ratio of the tone progression by q. Then the chromatic scale starting with A has frequencies

$$a,\ aq,\ aq^2,\ aq^3,\ \ldots$$

This scale must include the A tone of the next octave, whose frequency is $2a$. So $2a = a \times q^n$, when n is the number of tones per octave in the chromatic scale. If you have access to a piano (or to a synthesizer, if you cannot afford a piano or prefer "pop music"), you can easily find that $n = 12$ (do not forget to count black keys). Therefore $q^{12} = 2$ and $q = \sqrt[12]{2}$ (see the section below about roots).

Now we can understand the inherent difficulty in tuning a piano: the fifth (and other intervals, too) are not really true intervals. Indeed, between the tones A and E there are 7 steps:

$$\begin{array}{cccccccc} \mathrm{A} & \mathrm{A}\sharp = \mathrm{B}\flat & \mathrm{B} & \mathrm{C} & \mathrm{C}\sharp = \mathrm{D}\flat & \mathrm{D} & \mathrm{D}\sharp = \mathrm{E}\flat & \mathrm{E} \\ a & aq & aq^2 & aq^3 & aq^4 & aq^5 & aq^6 & aq^7 \end{array}$$

and to get a true fifth we need $q^7 = \frac{3}{2}$. But the requirements $q^{12} = 2$ and $q^7 = \frac{3}{2}$ are inconsistent; if both are fulfilled then

$$2^7 = (q^{12})^7 = (q^7)^{12} = \left(\frac{3}{2}\right)^{12}$$

which is false. Using a pocket calculator we may find that when $q^{12} = 2$ we get $q^7 = 1.498307\ldots$, which is close but not equal to 1.5. For other intervals the differences are even bigger:

Interval	should be	is about
		1.000000
		1.059463
		1.122462
minor third	1.2	1.189207
major third	1.25	1.259921
fourth	1.333...	1.334839
		1.414213
fifth	1.5	1.498307
minor sixth	1.6	1.587401
major sixth	1.666...	1.681792
		1.781797
		1.887748
octave	2.0	2.000000

In this table, the right column is a geometric progression accurate to 6 digits corresponding to a well-tempered clavier; the middle column shows the "true" intervals, which are ratios of small integers.

Problem 220. We assumed as a given fact that an octave contains 12 tones and found that in this case the well-tempered clavier cannot provide true fifths. What happens if we allow another number of tones in an octave? Is it possible to get true fifths or not?

Let us return to the history of music. In ancient times (before the eighteenth century), people tuned claviers (harpsichords at that time) trying to make at least some intervals harmonic (that is, corresponding to ratios of small integers). So the melodies sound nice in one key but become horrible when transposed into another key. Therefore some keys were avoided. A man named Andreas Werkmeister decided to go the other way and to make all intervals (that is, frequency ratios) the same. In this case, as we have seen, all intervals (except the octaves) are not exact but are close to the exact ones for all keys. It turns out that this is an acceptable solution. The great Bach honored this invention by writing his *Well-Tempered Clavier*. It contains two parts. Each part contains 24 preludes and fugues – one for each minor and major key.

Problem 221. Find a recording of Bach's *Well-Tempered Clavier* and enjoy it.

45 The sum of an infinite geometric progression

One of the famous "Paradoxes of Zeno" (Zeno was an ancient Greek philosopher) can be explained as follows. Assume that Achilles, who runs ten times faster than the turtle, starts to run after it. (The turtle runs away at the same time.) When Achilles comes to the place where the turtle was, it is not there but has moved on a distance equal to one tenth of the initial distance (between Achilles and the turtle). Achilles runs to that point – but at that time the turtle is again not there but has moved on a distance of one hundredth the initial distance, etc. This process has infinitely many stages; therefore Achilles will never meet the turtle. O.K?

We included this story in this section because the distances covered by Achilles form a geometric progression

$$1, \frac{1}{10}, \frac{1}{100}, \frac{1}{1000}, \ldots$$

whose common ratio is equal to $1/10$ (we assume that at the beginning the distance between Achilles and the turtle was equal to 1). So the total distance covered by Achilles is equal to the "sum of the infinite series"

$$1 + \frac{1}{10} + \frac{1}{100} + \frac{1}{1000} + \cdots$$

The pedantic view is that this infinite series has no sum (unless it is defined by a special definition) because when adding the numbers of an infinite series we never stop. And, of course, this is true. However, we shall not discuss this definition. Instead, we shall compute this undefined sum in different ways.

The first method is to denote this sum by S:

$$S = 1 + \frac{1}{10} + \frac{1}{100} + \frac{1}{1000} + \cdots$$

Then

$$10S = 10 + 1 + \frac{1}{10} + \frac{1}{100} + \cdots = 10 + S,$$

so

$$9S = 10, \quad S = \frac{10}{9}.$$

The second method is to add terms one by one:

$$\begin{array}{rcl} 1 + \frac{1}{10} & = & 1.1 \\ 1 + \frac{1}{10} + \frac{1}{100} & = & 1.11 \\ 1 + \frac{1}{10} + \frac{1}{100} + \frac{1}{1000} & = & 1.111 \\ & \cdots & \end{array}$$

After we add all terms, we get a periodic fraction $1.111\ldots$ equal to $1\frac{1}{9}$ (because $1/9 = 0.111\ldots$).

The third method is to apply the formula for the sum of the geometric progression:

$$1 + q + q^2 + q^3 + \cdots + q^{n-1} = \frac{q^n - 1}{q - 1}.$$

In our case $q = 1/10$ and n is infinitely big (so to speak). Then q^n is infinitely small (the bigger n is, the smaller $(1/10)^n$ is). Discarding it, we get the formula for the sum of an infinite geometric progression:

$$1 + q + q^2 + q^3 + \cdots = \frac{1}{1 - q}$$

(we have changed the signs of the numerator and the denominator). Recalling that $q = 1/10$, we get the answer $\dfrac{1}{0.9} = \dfrac{10}{9}$.

The fourth method. Let us return to Achilles and the turtle. Our common sense says that Achilles will meet the turtle after some distance S. During the race the turtle's speed is ten times less than that of Archilles, so the turtle covers the distance $S/10$. The initial distance (as we assume) was 1, so we get the equation

$$S - \frac{1}{10}S = 1.$$

Therefore, $(9/10)S = 1$ and $S = 10/9$.

Imagine now that Achilles is running ten times more slowly than the turtle. When he comes to the place where the turtle was, it is at the distance ten times further than the initial one. When Achilles comes to that place, the turtle is far away – at the distance that is one hundred times further than the initial one, etc. So we come to the sum

$$1 + 10 + 100 + \cdots$$

Of course, Achilles will never meet the turtle. But nevertheless we can substitute 10 for q in the formula

$$1 + q + q^2 + q^3 + \cdots = \frac{1}{1-q}.$$

and get an (absurd) answer

$$1 + 10 + 100 + 1000 + \cdots = \frac{1}{1-10} = -\frac{1}{9}.$$

Problem 222. Is it possible to give a reasonable interpretation of the (absurd) statement "Achilles will meet the turtle after running $-1/9$ meters"?

Hint. Yes, it is.

46 Equations

When we write, for instance, the equality

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

it has the following meaning: For any numbers a and b, the left-hand side and the right-hand side are equal. Such equalities are called identities. An identity may be proved (if we are lucky enough to transform the left-hand side to be equal to the right-hand side using algebraic rules). An identity may be refuted (if we managed to find values of variables such that the left-hand side is not equal to the right-hand side).

An equation also consists of a left-hand side and right-hand side connected by the equality sign, but the goal is different: it must be *solved*. To solve an equation means to find values of the variable(s) for which the left-hand side is equal to the right-hand side.

For example, the equation

$$5x + 3 = 2x + 7$$

may be solved as follows: Subtract $2x+3$ from both sides; you get the equivalent equation

$$3x = 4$$

(the equivalence means that if one of the equations is true for some x then the other one is also true for this x). Now dividing both sides by 3 we get

$$x = \frac{4}{3}.$$

So we say: "the equation $5x + 3 = 2x + 7$ has the unique solution $x = 4/3$".

Remark. The equation

$$\frac{x+1}{x+2} = 1$$

has no solutions. (Proof: if $\frac{x+1}{x+2} = 1$ then $x + 1 = x + 2$, which is impossible.) However, mathematicians do not say that this equation is unsolvable. On the contrary, they say that the equation is solved after they proved that it has no solutions. So "to solve an equation" means to find all solutions or to prove that there are no solutions.

47 A short glossary

unknowns	letters used in an equation
to *solve* an equation	to find all values of unknowns such that the left-hand side is equal to the right-hand side; to find all solutions
a *solution* of an equation	a set of values for the unknown for which the left-hand side is equal to the right-hand side (sometimes solutions are called *roots* when speaking about an equation with only one unknown)
equivalent equations	equations having the same solutions; equations that are true or false simultaneously, for the same values of the unknowns

48 Quadratic equations

By a quadratic equation, we mean an equation of the form

$$ax^2 + bx + c = 0,$$

where a, b, c are some fixed numbers and x is an unknown.

Problem 223. Solve the quadratic equation $x^2 - 3x + 2 = 0$.

Solution. Factor the left-hand side: $x^2 - 3x + 2 = (x - 1)(x - 2)$. Therefore, the equation may be rewritten as $(x - 1)(x - 2) = 0$. This equality is true in two cases: either $x - 1 = 0$ (so $x = 1$) or $x - 2 = 0$ (so $x = 2$). Thus, this equation has two roots, $x = 1$ and $x = 2$.

Problem 224. Solve the equations:

(a) $x^2 - 4 = 0$; (b) $x^2 + 2 = 0$;

(c) $x^2 - 2x + 1 = 0$; (d) $x^2 - 2x + 1 = 9$;

(e) $x^2 - 2x - 8 = 0$; (f) $x^2 - 2x - 3 = 0$;

(g) $x^2 - 5x + 6 = 0$; (h) $x^2 - x - 2 = 0$.

If in the equation

$$ax^2 + bx + c = 0$$

the coefficient a is equal to zero then the equation takes the form

$$bx + c = 0$$

and has the unique solution

$$x = -\frac{c}{b}.$$

Problem 225. Strictly speaking, the last sentence is wrong; when $b = 0$ the quotient c/b in undefined. How are we to correct this error?

If in the equation

$$ax^2 + bx + c = 0$$

the coefficient a is nonzero then we may divide by a and get an equivalent equation

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

So if we are able to solve a *reduced* quadratic equation (where x^2 has a coefficient 1) we can solve any quadratic equation. Usually the reduced quadratic equation is written as

$$x^2 + px + q = 0.$$

49 The case $p = 0$. Square roots

Let us start with the equation $x^2 + q = 0$. Three cases are possible:

(a) $q = 0$. The equation $x^2 = 0$ has a unique solution $x = 0$.

(b) $q > 0$. The equation has no solutions because the nonnegative number x^2 added to a positive number q cannot be equal to 0.

(c) $q < 0$. The equation may be rewritten as $x^2 = -q$ and we have to look for numbers whose square is a (positive) number $-q$.

Fact. For any positive number c there is a positive number whose square is c. It is called the square root of c; its notation is $\sqrt{c}$.

We met with $\sqrt{2}$ in factoring $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$. Now we use $\sqrt{c}$ for a similar purpose.

How to solve the equation $x^2 = c$:

$$\begin{aligned} x^2 - c &= 0; \\ x^2 - (\sqrt{c})^2 &= 0; \\ (x - \sqrt{c})(x + \sqrt{c}) &= 0; \end{aligned}$$

the last equation has two solutions, $x = \sqrt{c}$ and $x = -\sqrt{c}$ (and no other solutions).

The reader may ask now, why are we considering this? When $x = \sqrt{c}$ then $x^2 = c$ by definition (and when $x = -\sqrt{c}$, too). Yes, this is true. But we have proved also that *there is no other solution* (because if $x \neq \pm\sqrt{c}$ then both factors are nonzero).

Now let us return to the fact claimed above, the existence of a square root. Assume that we start with $x = 0$ and then x increases gradually. Its square x^2 also increases (greater values of x correspond to greater values of x^2). At the beginning, $x^2 = 0$ and x^2 is less than c. When x is very big, x^2 is even bigger and therefore $x^2 > c$ for x big enough. So x^2 was *smaller* than c and becomes *greater* than c. Therefore, it must cross this boundary sometime – for some x the value of x^2 must be *equal* to c.

In the last sentence the word "therefore" stands for several chapters of a good calculus textbook, where the existence of such an x is proved, based on considerations of continuity.

These days, when square roots can be found on almost any calculator, it is almost impossible to imagine the shock caused by square roots for ancient Greeks. They found that the square root of 2 cannot be written as a quotient of two integers – and they did not know any other numbers, so it was a crash of their foundations.

Problem 226. Prove that $\sqrt{2} \neq \dfrac{m}{n}$ for any integer m and n. In other words, $\sqrt{2}$ is irrational (rational numbers are fractions with integers as numerator and denominator).

Solution. Assume that $\sqrt{2} = \dfrac{m}{n}$. Three cases are possible:

(a) both m and n are odd;

(b) m is even and n is odd;

(c) m is odd and n is even.

(The fourth case "m and n are even" may be ignored, because we could divide m and n by 2 several times until at least one of them would be odd and we would get one of the cases (a)–(c).)

Let us show that cases (a)–(c) are all impossible. Recall that any even number can be represented as $2k$ for some integer n and any odd number can be represented as $2k+1$ for some integer k. So let us go through all three cases.

(a) Assume that $\sqrt{2} = \dfrac{2k+1}{2l+1}$; then

$$\left(\frac{2k+1}{2l+1}\right)^2 = 2,$$

$$\frac{(2k+1)^2}{(2l+1)^2} = 2,$$

$$(2k+1)^2 = 2 \cdot (2l+1)^2,$$

$$4k^2 + 4k + 1 = 2 \cdot (2l+1)^2.$$

Contradiction: (even number) + 1 = (even number).

(b) Assume that $\sqrt{2} = \dfrac{2k}{2l+1}$; then

$$\left(\frac{2k}{2l+1}\right)^2 = 2,$$

$$(2k)^2 = 2 \cdot (2l+1)^2,$$

$$4k^2 = 2 \cdot (4l^2 + 4l + 1),$$

$$2k^2 = 4l^2 + 4l + 1.$$

Contradiction: (even number) = (even number) + 1.

(c) Asssume that $\sqrt{2} = \dfrac{2k+1}{2l}$; then

$$(2k+1)^2 = 2 \cdot (2l)^2,$$

$$4k^2 + 4k + 1 = 2 \cdot (2l)^2.$$

Contradiction: (even number) + 1 = (even number).

So all three cases are impossible.

Problem 227. Prove that $\sqrt{3}$ is irrational.

Hint. Any integer has one of the forms $3k$, $3k+1$, $3k+2$.

When we claim that we have solved the equation $x^2-2=0$ and the answer is "$x=\sqrt{2}$ or $x=-\sqrt{2}$", we are in fact cheating. To tell the truth, we have not solved this equation but confessed our inability to solve it; $\sqrt{2}$ means nothing except "the positive solution of the equation $x^2-2=0$".

50 Rules for square roots

Problem 228. Prove that (for $a, b \geq 0$)

$$\sqrt{ab} = \sqrt{a} \cdot \sqrt{b}.$$

Solution. To show that $\sqrt{a} \cdot \sqrt{b}$ is the square root of ab we must (according to the definition of square roots) prove that it is a nonnegative number whose square is ab:

$$\left(\sqrt{a} \cdot \sqrt{b}\right)^2 = \left(\sqrt{a}\right)^2 \cdot \left(\sqrt{b}\right)^2 = a \cdot b.$$

Problem 229. Prove that for $a \geq 0$, $b > 0$

$$\sqrt{\frac{a}{b}} = \frac{\sqrt{a}}{\sqrt{b}}.$$

The following question is a traditional trap used by examiners to catch innocent pupils.

Problem 230. Is the equality $\sqrt{a^2} = a$ true for all a?

Solution. No. When a is negative, $\sqrt{a^2}$ is equal to $-a$. The correct statement is $\sqrt{a^2} = |a|$ where

$$|a| = \begin{cases} a, & \text{if } a \geq 0 \\ -a, & \text{if } a < 0 \end{cases}$$

Problem 231. Prove that

(a) $\dfrac{1}{2+\sqrt{3}} = 2-\sqrt{3}$;

(b) $\dfrac{1}{\sqrt{7}-\sqrt{5}} = \dfrac{\sqrt{5}+\sqrt{7}}{2}$.

Problem 232. Which is bigger: $\sqrt{1001}-\sqrt{1000}$, or $1/10$?

Problem 233. Simplify the expression $\sqrt{3+2\sqrt{2}}$

Solution. $3+2\sqrt{2} = 1+2+2\sqrt{2} = 1+(\sqrt{2})^2+2\sqrt{2} = (1+\sqrt{2})^2$. So we get the answer, $1+\sqrt{2}$.

Problem 234. Dan simplified an expression as follows:

$$\begin{aligned}\sqrt{3-2\sqrt{2}} &= \sqrt{1+2-2\sqrt{2}} = \\ &= \sqrt{1+(\sqrt{2})^2-2\sqrt{2}} = \sqrt{(1-\sqrt{2})^2} = 1-\sqrt{2}.\end{aligned}$$

Do you approve of his simplification?

Solution. The correct answer is $\sqrt{2}-1$ because $1-\sqrt{2}<0$.

51 *The equation $x^2+px+q=0$*

Problem 235. Solve the equation

$$x^2+2x-6=0\,.$$

Solution. The equation $x^2+2x-6=0$ may be rewritten as follows:

$$\begin{aligned}(x^2+2x+1)-7 &= 0;\\ (x+1)^2-7 &= 0;\\ (x+1)^2 &= 7;\\ x+1=\sqrt{7} \quad &\text{or} \quad x+1=-\sqrt{7};\\ x=-1+\sqrt{7} \quad &\text{or} \quad x=-1-\sqrt{7}.\end{aligned}$$

The same method can be applied to other equations.

Problem 236. Solve the equation

$$x^2+2x-8=0.$$

Problem 237. Solve the equation

$$x^2+3x+1=0.$$

Solution. Transform the left-hand side:

$$\begin{aligned}x^2+3x+1 &= x^2+2\cdot\frac{3}{2}x+\left(\frac{3}{2}\right)^2-\left(\frac{3}{2}\right)^2+1 = \\ &= \left(x+\frac{3}{2}\right)^2-\frac{9}{4}+1 = \left(x+\frac{3}{2}\right)^2-\frac{5}{4}.\end{aligned}$$

Now the equation can be written as follows:

$$\left(x+\frac{3}{2}\right)^2 = \frac{5}{4},$$

$$x+\frac{3}{2}=\sqrt{\frac{5}{4}} \quad \text{or} \quad x+\frac{3}{2}=-\sqrt{\frac{5}{4}},$$

$$x=-\frac{3}{2}+\sqrt{\frac{5}{4}} \quad \text{or} \quad x=-\frac{3}{2}-\sqrt{\frac{5}{4}}.$$

Remark. The answer to the preceding problem is usually written as

$$x=-\frac{3}{2}\pm\sqrt{\frac{5}{4}}.$$

Problem 238. Solve the equation $x^2-2x+2=0$.

Solution. $x^2-2x+2 = (x^2-2x+1)+1 = (x-1)^2+1$. The equation $(x-1)^2+1=0$ has no roots because its left-hand side is never less than 1 (a square is always nonnegative).

The method shown above is called "completing the square". In the general case it looks as follows:

$$\begin{aligned} x^2+px+q &= 0 \\ \left(x^2+2\cdot\frac{p}{2}\cdot x+\left(\frac{p}{2}\right)^2\right)-\left(\frac{p}{2}\right)^2+q &= 0 \\ \left(x+\frac{p}{2}\right)^2=\left(\frac{p}{2}\right)^2-q &= \frac{p^2}{4}-q \end{aligned}$$

Now three cases are possible:

- If $\frac{p^2}{4}-q>0$ then two solutions exist:

$$x+\frac{p}{2}=\pm\sqrt{\frac{p^2}{4}-q}.$$

Thus,

$$x=-\frac{p}{2}\pm\sqrt{\frac{p^2}{4}-q}.$$

- If $\frac{p^2}{4}-q=0$ then there is one solution:

$$x=-\frac{p}{2}.$$

- If $\frac{p^2}{4} - q < 0$ then there are no solutions.

Often all three cases are included in a single formula:

$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}$$

and people say that when $\frac{p^2}{4} - q = 0$ the solutions x_1 and x_2 coincide (because the square root of 0 is 0) and when $\frac{p^2}{4} - q < 0$ this formula gives no solutions, because the square root of a negative number is undefined. (To tell you the whole truth, in the latter case mathematicians agree that the square root of a negative number exists but is imaginary and there are two so-called complex roots. But this is another topic.)

We see that the sign of $D = \frac{p^2}{4} - q$ plays a crucial role (it determines how many solutions the equation has).

52 Vieta's theorem

Theorem. If a quadratic equation $x^2 + px + q$ has two (different) roots α and β then

$$\begin{aligned} \alpha + \beta &= -p \\ \alpha \cdot \beta &= q . \end{aligned}$$

Corollary. If a quadratic equation $x^2 + px + q$ has two different roots α and β then

$$x^2 + px + q = (x - \alpha)(x - \beta) .$$

This is another form of the same assertion because

$$(x - \alpha)(x - \beta) = x^2 - (\alpha + \beta)x + \alpha\beta$$

and two polynomials are equal if they have equal coefficients.

First proof. According to the formula for roots we have

$$\alpha = -\frac{p}{2} - \sqrt{D}, \quad \beta = -\frac{p}{2} + \sqrt{D}$$

where $D = \frac{p^2}{4} - q$. (Or vice versa:

$$\alpha = -\frac{p}{2} + \sqrt{D}, \quad \beta = -\frac{p}{2} - \sqrt{D},$$

but this makes no difference.) Then

$$\alpha + \beta = -\frac{p}{2} - \sqrt{D} - \frac{p}{2} + \sqrt{D} = -p$$

and

$$\alpha\beta = \left(-\frac{p}{2} - \sqrt{D}\right)\left(-\frac{p}{2} + \sqrt{D}\right) = \left(\frac{p}{2}\right)^2 - \left(\sqrt{D}\right)^2 =$$

$$= \frac{p^2}{4} - D = \frac{p^2}{4} - \frac{p^2}{4} + q = q.$$

That's what we want.

Second proof. Let us try to prove Vieta's theorem in the form stated in the corollary. We know that if a polynomial $P(x)$ has different roots α and β then it can be factored:

$$P(x) = (x - \alpha)(x - \beta)R(x)$$

where $R(x)$ is some polynomial. In our case (when P has degree 2) the polynomial R must be a constant (otherwise the degree of the right-hand side would be too big), and this constant is equal to 1, because the x^2-coefficients in $x^2 + px + q$ and $(x - \alpha)(x - \beta)$ are the same. Therefore

$$x^2 + px + q = (x - \alpha)(x - \beta).$$

The theorem is proved.

Problem 239. Can you generalize Vieta's theorem to the case of a quadratic equation having only one root? Are both proofs still valid for this case?

Problem 240. (Vieta's theorem for a cubic equation) Assume that a cubic equation $x^3 + px^2 + qx + r = 0$ has three different roots α, β, γ. Prove that

$$\begin{aligned} \alpha + \beta + \gamma &= -p \\ \alpha\beta + \alpha\gamma + \beta\gamma &= q \\ \alpha\beta\gamma &= -r \end{aligned}$$

Problem 241. The equation $x^2 + px + q = 0$ has roots x_1 and x_2. Find $x_1^2 + x_2^2$ (as an expression containing p and q).

Solution. $x_1^2+x_2^2 = x_1^2+2x_1x_2+x_2^2-2x_1x_2 = (x_1+x_2)^2-2x_1x_2 = p^2 - 2q$.

Problem 242. The equation $x^2 + px + q = 0$ has roots x_1 and x_2. Find $(x_1 - x_2)^2$ (as an expression containing p and q).

Solution. $(x_1-x_2)^2 = x_1^2-2x_1x_2+x_2^2 = x_1^2+2x_1x_2+x_2^2-4x_1x_2 = (x_1 + x_2)^2 - 4x_1x_2 = p^2 - 4q$.

Another solution. $x_1 - x_2$ is the difference between the roots; looking at the formula for the roots, we see that it is equal to $2\sqrt{D}$, so $(x_1 - x_2)^2 = 4D = 4(\frac{p^2}{4} - q) = p^2 - 4q$.

Problem 243. A cubic equation $x^3 + px^2 + qx + r = 0$ has three different roots x_1, x_2, x_3. Find

$$(x_1 - x_2)^2(x_2 - x_3)^2(x_1 - x_3)^2$$

as an expression containing p, q, r. This polynomial in p, q, r is called the *discriminant* of the cubic equation. As in the case of a quadratic equation (see page 107), it is small when two roots are close to each other.

Problem 244. The equation $x^2 + px + q = 0$ has roots x_1, x_2; the equation $y^2 + ry + s = 0$ has roots y_1, y_2. Find

$$(y_1 - x_1)(y_2 - x_1)(y_1 - x_2)(y_2 - x_2)$$

as a polynomial of p, q, r, s. (This polynomial is called the resultant of two quadratic polynomials; it is equal to zero if these two polynomials have a common root.)

Vieta's theorem allows us to construct a quadratic equation with given roots. More precisely, we should not say "Vieta's theorem" but "the converse to Vieta's theorem"; here it is:

Theorem. If α and β are any numbers, $p = -(\alpha + \beta)$, $q = \alpha\beta$, then the equation $x^2 + px + q = 0$ has roots α and β.

The proof is trivial: The equation $(x - \alpha)(x - \beta) = 0$ evidently has roots α and β. Multiplying the terms in parentheses we see that it is the equation $x^2 + px + q = 0$.

Problem 245. Find a quadratic equation with integer coefficients having $4 - \sqrt{7}$ as one of the roots.

Hint. The second root is $4 + \sqrt{7}$.

Problem 246. The integers p, q are coefficients of the quadratic equation $x^2 + px + q = 0$, which has two roots. Prove that

(a) the sum of squares of its roots is an integer;

(b) the sum of cubes of its roots is an integer;

(c) the sum of nth powers of its roots is an integer (for any natural number n)

Problem 247. (a) Prove that the square of any number of the form $a + b\sqrt{2}$ (where a, b are integers) also has this form (that is, is equal to $k + l\sqrt{2}$ for some integer k, l).

(b) Prove the same for $(a + b\sqrt{2})^n$ for any integer $n > 1$.

(c) The number $(a + b\sqrt{2})^n$ is equal to $k + l\sqrt{2}$ (here a, b, k, l are integers). What can be said about $(a - b\sqrt{2})^n$?

(d) Prove that there are infinitely many integers a, b such that $a^2 - 2b^2 = 1$.

Solution of (d). Let us start from the solution $3^2 - 2 \cdot 2^2 = 1$, and rewrite this equality as $(3 + 2\sqrt{2})(3 - 2\sqrt{2}) = 1$. Consider the nth powers of both sides: $(3 + 2\sqrt{2})^n(3 - 2\sqrt{2})^n = 1$. The number $(3+2\sqrt{2})^n$ is equal to $k + l\sqrt{2}$ for some integers k, l. Thus $(3 - 2\sqrt{2})^n$ is equal to $k - l\sqrt{2}$ and we get the equality

$$(k + l\sqrt{2})(k - l\sqrt{2}) = k^2 - 2l^2 = 1.$$

Therefore, k, l satisfy the equation.

For example, $(3 + 2\sqrt{2})^2 = 9 + 8 + 12\sqrt{2} = 17 + 12\sqrt{2}$. So $17, 12$ must satisfy the equation. Is this true? $17^2 - 2 \cdot 12^2 = 289 - 2 \cdot 144 = 289 - 288 = 1$. Our theory works!

Problem 248. Prove that the equation $x^2 + px + q = 0$ has two solutions having different signs if and only if $q < 0$

Solution. If the roots have opposite signs, then (recall Vieta's theorem) the coefficient q, being equal to their product, is negative. In the opposite direction, if the product of two roots is negative, they have opposite sign. (But we must be sure that the roots do exist; to check this, we look at $D = \frac{p^2}{4} - q$; if $q < 0$, then $D > 0$.)

Another explanation can be given as follows. Assume that $q < 0$. Then the value of the expression $x^2 + px + q$ is negative when $x = 0$. When x increases and becomes very big, $x^2 + px + q$ becomes positive (x^2 "outweighs" $px + q$). So $x^2 + px + q$ must cross the zero boundary somewhere in between – and the equation has a positive root. A similar argument shows that it also has a negative root.

53 Factoring $ax^2 + bx + c$

Problem 249. Factor $2x^2 + 5x - 3$.

Solution. Taking the factor 2 out of the parentheses, we get

$$2x^2 + 5x - 3 = 2\left(x^2 + \frac{5}{2}x - \frac{3}{2}\right).$$

Solving the equation $x^2 + \frac{5}{2}x - \frac{3}{2} = 0$ we get

$$x_{1,2} = -\frac{5}{4} \pm \sqrt{\frac{25}{16} + \frac{3}{2}} = -\frac{5}{4} \pm \sqrt{\frac{49}{16}} = -\frac{5}{4} \pm \frac{7}{4}$$

so $x_1 = -3$, $x_2 = \frac{1}{2}$. According to the corollary of Vieta's theorem, we get

$$x^2 + \frac{5}{2}x - \frac{3}{2} = (x - (-3))\left(x - \frac{1}{2}\right) = (x + 3)\left(x - \frac{1}{2}\right)$$

and

$$2x^2 + 5x - 3 = (x + 3)(2x - 1).$$

Problem 250. Factor $2x^2 + 2x + \frac{1}{2}$.

Problem 251. Factor $2a^2 + 5ab - 3b^2$.

Solution.

$$2a^2 + 5ab - 3b^2 = b^2\left(2\frac{a^2}{b^2} + 5\frac{a}{b} - 3\right).$$

Denote $\frac{a}{b}$ by x and use the factorization $2x^2 + 5x - 3 = (x+3)(2x-1)$. Then you can continue the equality:

$$\cdots = b^2\left(\frac{a}{b} + 3\right)\left(2\frac{a}{b} - 1\right) = (a + 3b)(2a - b).$$

54 A formula for $ax^2 + bx + c = 0$ (where $a \neq 0$)

Dividing by a, we get

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

and we can apply the formula for the equation

$$x^2 + px + q = 0$$

with $p = \frac{b}{a}$, $q = \frac{c}{a}$. We get

$$x_{1,2} = -\frac{b}{2a} \pm \sqrt{\left(\frac{b}{2a}\right)^2 - 4\frac{c}{a}} = -\frac{b}{2a} \pm \sqrt{\frac{b^2 - 4ac}{4a^2}} =$$

$$= -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The expression $D = b^2 - 4ac$ is called the *discriminant* of the equation $ax^2 + bx + c = 0$. If it is positive, the equation has two roots. If $D = 0$ the equation has one root. If $D < 0$ the equation has no roots.

Problem 252. We replaced

$$\sqrt{\frac{b^2 - 4ac}{4a^2}} = \frac{\sqrt{b^2 - 4ac}}{\sqrt{4a^2}}$$

by

$$\frac{\sqrt{b^2 - 4ac}}{2a}$$

but as we mentioned above, $\sqrt{4a^2}$ is equal not to $2a$ but to $|2a|$. Why does it not matter here?

Problem 253. Assume that the equation $ax^2 + bx + c = 0$ has roots x_1 and x_2. What are the roots of the equation $cx^2 + bx + a = 0$?

Solution. If $ax^2 + bx + c = 0$ then (divide by x^2)

$$a + \frac{b}{x} + \frac{c}{x^2} = 0, \text{ that is, } c \cdot \left(\frac{1}{x}\right)^2 + b \cdot \left(\frac{1}{x}\right) + a = 0.$$

So $\frac{1}{x_1}$ and $\frac{1}{x_2}$ will be the roots of the equation $cx^2 + bx + a = 0$.

Remark. We assumed implicitly that $x_1 \neq 0$, $x_2 \neq 0$. If one of the roots x_1 and x_2 is equal to 0 then (according to Vieta's theorem) c is equal to 0 and the equation $cx^2 + bx + a = 0$ has at most one root.

55 One more formula concerning quadratic equations

The formula

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

is well known to millions of school pupils all over the world. But there is another formula that has an equal right to be studied but is less known. Here it is:

$$x_{1,2} = \frac{2c}{-b \pm \sqrt{b^2 - 4ac}}$$

Let us prove it. If x is a root of the equation $ax^2 + bx + c = 0$ then $y = 1/x$ is a root of the equation $cy^2 + by + a = 0$, therefore

$$y_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}$$

and

$$x_{1,2} = \frac{1}{y_{1,2}} = \frac{2c}{-b \pm \sqrt{b^2 - 4ac}}.$$

Problem 254. Check by a direct calculation that

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{2c}{-b \mp \sqrt{b^2 - 4ac}}.$$

(we used $\mp$ and $\pm$ having in mind that plus in the left-hand side corresponds to the minus in the right-hand side and vice versa).

56 A quadratic equation becomes linear

Look at the quadratic equation $ax^2 - x + 1 = 0$. According to our general rule, it has two roots if and only if its discriminant $D = 1 - 4a$ is positive, that is, when $a < 1/4$.

Problem 255. Is this true?

Solution. No; when $a = 0$ the equation is not a quadratic one, it becomes $-x + 1 = 0$ and has only one root $x = 1$.

A pedant will describe what happens saying "our general rule is not applicable, because the equation is not quadratic". And he is right. But how can it be? We had an equation with two roots and were changing the coefficient. Suddenly one root disappeared, when a became zero. What happened to it?

To answer this question, let us look at the second formula for the roots of a quadratic equation:

$$x_{1,2} = \frac{2}{1 \pm \sqrt{1-4a}}.$$

If a is close to zero then $\sqrt{1-4a} \approx 1$, so

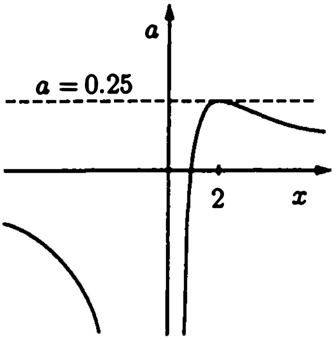
$$x_1 = \frac{2}{1+\sqrt{1-4a}} \approx \frac{2}{1+1} = 1,$$

but

$$x_2 = \frac{2}{1-\sqrt{1-4a}} = \frac{2}{\text{number close to zero}},$$

that is, x_2 is very big. So while a tends to zero, the root x_1 tends to 1 and x_2 goes to infinity (and returns from the other side of infinity).

One can see in detail how this happens looking at the following picture:



This picture shows points $\langle x, a \rangle$ such that $ax^2 - x + 1 = 0$. In other words, it shows the graph of the function $a = (x-1)/x^2$. To find the solutions of the equation $ax^2 - x + 1 = 0$ for a given a on the picture, we must intersect a horizontal straight line having height a with our graph. Assume that this horizontal line is moving downwards. At the beginning (when $a > 0.25$) it has no intersections (and the equation has no solutions). When $a = 0.25$ there is one intersection point, which splits immediately into two points when a becomes less than 0.25. One of the points is moving left, the other is moving right. The point moving right goes to plus infinity when a tends to zero, then disappears (when $a = 0$) and then returns from minus infinity. Then (when a becomes more and more negative), both roots go to zero from opposite sides.

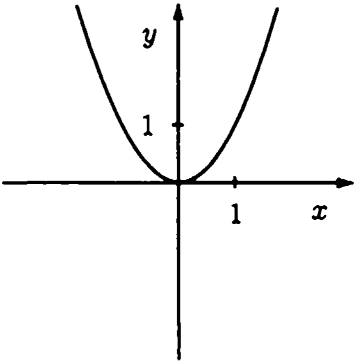
Problem 256. What happens with the roots of equations

(a) $x^2 - x - a = 0$; (b) $x^2 - ax + 1 = 0$

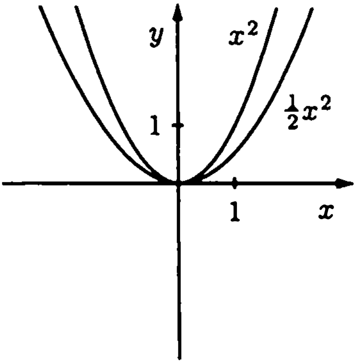
as a changes?

57 The graph of the quadratic polynomial

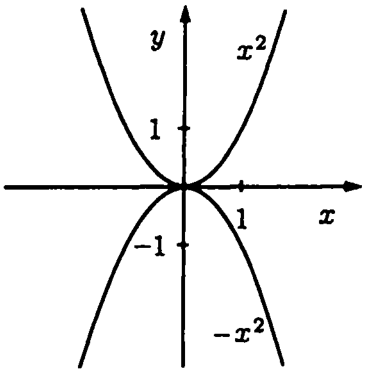
The graph of $y = x^2$ looks as follows:



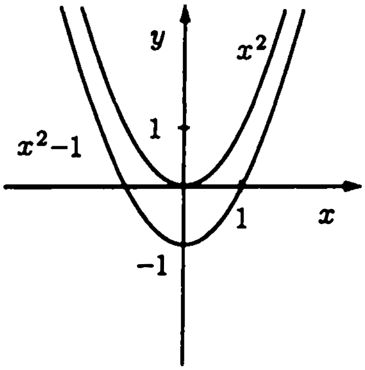
Using this graph we may draw graphs of other polynomials of degree 2. The graph of $y = ax^2$ (where a is a constant) can be obtained from $y = x^2$ by stretching (when $a > 1$) or shrinking (when $0 < a < 1$) in a vertical direction:



When a is negative the graph is turned upside down:

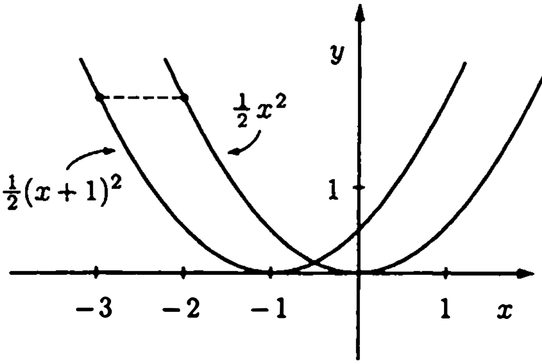


The graph $y = x^2 + c$ can be obtained from the graph of $y = x^2$ by a vertical translation by c (up if $c > 0$, down if $c < 0$).



In the same way, $y = ax^2 + c$ can be obtained from $y = ax^2$.

It is more difficult to understand what corresponds to a horizontal translation of the graph. Let us consider an example and compare the graphs $y = \frac{1}{2}x^2$ and $y = \frac{1}{2}(x+1)^2$. Let us start with one specific value of x; assume that $x = -3$. For this x the expression $\frac{1}{2}(x+1)^2$ is equal to $\frac{1}{2}(-2)^2$, that is, has the same value as $\frac{1}{2}x^2$ when $x = -2$. In general, the value of $\frac{1}{2}(x+1)^2$ for any value x coincides with the value of $\frac{1}{2}x^2$ for some other value x (greater by 1).



In terms of our graph this means that any point of the graph $y = \frac{1}{2}(x+1)^2$, when moved 1 unit to the right, becomes a point of the graph $y = \frac{1}{2}x^2$. Therefore, we get the graph $y = \frac{1}{2}x^2$ by translating the graph $y = \frac{1}{2}(x+1)^2$ one unit to the right, and vice versa, we get the graph $y = \frac{1}{2}(x+1)^2$ by translating the graph $y = \frac{1}{2}x^2$ one unit to the left.

The general rule is as follows: a graph $y = a(x+m)^2$ can be obtained from the graph of $y = ax^2$ by an m-unit shift to the left (when $m > 0$; when $m < 0$ we use a right shift).

Now we can get any graphs of the form

$$y = a(x+m)^2 + n$$

from the graph of $y = x^2$ in three stages:

(a) Stretch it vertically a times and you get $y = ax^2$.

(b) Move it m units to the left and you get $y = a(x+m)^2$.

(c) Move it n units up and you get $y = a(x+m)^2 + n$.

Problem 257. Find the coordinates of the top point (or the bottom point – it depends on the sign of a) of a graph $y = a(x+m)^2 + n$.

Answer. Its coordinates are $\langle -m, n\rangle$.

Problem 258. Is the ordering of operations (a), (b), and (c) important? Do we get the same graph applying, for example, (c), then (b), and then (a) to the graph $y = x^2$?

Answer. The ordering of operations is important. We get x^2+n after (c), then $(x+m)^2+n$ after (b) and finally $a(x+m)^2+an$ after (a). So we get an instead of n.

Problem 259. There are six possible orderings of operations (a), (b), and (c). Do we get six different graphs or do some of the graphs coincide?

Now we are able to draw the graph of any quadratic polynomial, because any quadratic polynomial may be converted to the form $a(x+m)^2+n$ by completing the square (as we did for the formula for the roots):

$$\begin{aligned} ax^2+bx+c &= a\left(x^2+\frac{b}{a}x\right)+c= \\ &= a\left(x^2+2\cdot\frac{b}{2a}\cdot x+\left(\frac{b}{2a}\right)^2-\left(\frac{b}{2a}\right)^2\right)+c= \\ &= a\left(x+\frac{b}{2a}\right)^2-\frac{b^2}{4a}+c. \end{aligned}$$

Denote $\dfrac{b}{2a}$ by m and $-\dfrac{b^2}{4a}+c$ by n and you get the desired result.

Problem 260. How can you determine the signs of a, b, c by looking at the graph of $y=ax^2+bx+c$?

Answer. If water can be kept in this graph then $a>0$, otherwise $a<0$. The sign of b/a is determined by the x-coordinate of the vertex of the graph (the left half of the plane corresponds to positive b/a). The sign of c can be found by looking at the intersection of the graph and y-axis (because $ax^2+bx+c=c$ when $x=0$).

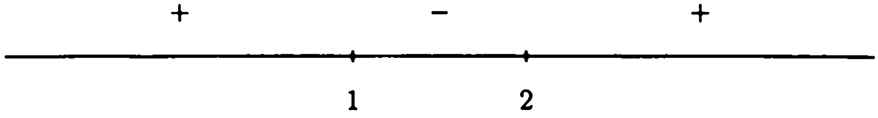
Remark. Another rule for finding the sign of b: If the graph intersects the y-axis going upwards, then b is positive; if the graph intersects it going downwards, then b is negative. This rule can be explained by means of calculus. When the function $f(x)=ax^2+bx+c$ is increasing near $x=0$, its derivative $f'(x)=2ax+b$ (which is equal to b when $x=0$) is positive.

58 Quadratic inequalities

Problem 261. Solve the inequality $x^2 - 3x + 2 < 0$. (To "solve an inequality" means to find all values of the variables for which it is true.)

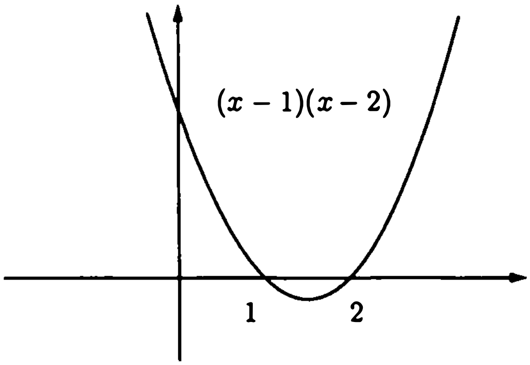
Solution. Factor the left-hand side:

$$x^2 - 3x + 2 = (x - 1)(x - 2)$$



The left-hand side is zero when $x = 1$ or $x = 2$. When $x > 2$, both factors are positive (and the product is positive). When we go through the point $x = 2$ into the interval $(1, 2)$, the second factor becomes negative (and the product is negative). When we go through the point $x = 1$, both factors become negative and the product is positive again. Therefore, we get an answer that the inequality is true for $1 < x < 2$.

You can get the same answer looking at the graph $y = x^2 - 3x + 2$ ($x = 1$ and $x = 2$ are intersection points with x-axis).



59 Maximum and minimum values of a quadratic polynomial

Problem 262. The sum of two numbers is equal to 1. What is the maximal possible value of its product?

Solution. Denote one of the numbers by x. Then the second number is $1-x$, and their product is $x \cdot (1-x) = x - x^2$. The graph of the quadratic polynomial $-x^2 + x$ is turned down and its roots are $x = 0$ and $x = 1$. Therefore its vertex, being in the middle, has $x = \frac{1}{2}$. Its value for $x = \frac{1}{2}$ (its maximal value) is $\frac{1}{2} \cdot (1 - \frac{1}{2}) = \frac{1}{4}$. So we get the answer that the maximal value is $\frac{1}{4}$.

Another solution. Assume that one of the numbers is $\frac{1}{2} + x$. Then the other number is $\frac{1}{2} - x$ and their product is

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

so the maximal value is obtained when $x = 0$ (and both numbers are equal to $\frac{1}{2}$).

Problem 263. Prove that a square has the maximum area of all rectangles having the same perimeter.

Problem 264. Prove that a square has the minimum perimeter of all rectangles having the same area.

Hint. Use the result of the preceding problem.

Problem 265. Find the minimal value of the expression $x + \dfrac{2}{x}$ for positive x.

Solution. Let us see what numbers $c > 0$ may be values of the expression $x + \dfrac{2}{x}$. In other words we want to know for which c the equation

$$x + \frac{2}{x} = c$$

has solutions. We may multiply this equation by x and ask for which c the resulting equation

$$x^2 + 2 = cx$$

has nonzero solutions. But no solutions of this equation are equal to zero ($x = 0$ is not a solution, $0^2 + 2 \neq c \cdot 0$). Therefore, the word "nonzero" may be omitted.

The equation $x^2 + 2 = cx$ may be rewritten as $x^2 - cx + 2 = 0$. It has solutions if and only if its discriminant

$$D = \left(\frac{c}{2}\right)^2 - 2$$

is nonnegative, that is, when $\left(\frac{c}{2}\right)^2 \geq 2$. The latter condition is satisfied when

$$\frac{c}{2} \geq \sqrt{2} \text{ or } \frac{c}{2} \leq -\sqrt{2}.$$

So the equation $x + \frac{2}{c} = c$ has solutions when $c \geq 2\sqrt{2}$ or $c \leq -2\sqrt{2}$. Therefore, the minimum value of $x + \frac{2}{x}$ for positive x is $2\sqrt{2}$.

Another solution. The numbers x and $\frac{2}{x}$ may be considered as edges of a rectangle having area 2, and $x + \frac{2}{x}$ is its semiperimeter. It will be minimal when the rectangle is a square (see the preceding problem), that is, when $x = \frac{2}{x}$, $x^2 = 2$, $x = \sqrt{2}$. For such an x the value of $x + \frac{2}{x}$ is $2\sqrt{2}$.

60 Biquadratic equations

Problem 266. Solve the equation $x^4 - 3x^2 + 2 = 0$.

Solution. If x is a root of this equation, then $y = x^2$ is a root of the equation $y^2 - 3y + 2 = 0$, and vice versa. This quadratic equation (where y is considered an unknown) has roots

$$y_{1,2} = \frac{3 \pm \sqrt{9-8}}{2} = \frac{3 \pm 1}{2};$$

hence, $y_1 = 1$, $y_2 = 2$.

Therefore the solutions of the initial equation are all x such that $x^2 = 1$ or $x^2 = 2$. So it has four solutions:

$$x = 1,\ x = -1,\ x = \sqrt{2},\ x = -\sqrt{2}.$$

The same method can be applied to any equation of the form

$$ax^4 + bx^2 + c = 0$$

(such equations are called biquadratic)

Problem 267. Construct a biquadratic equation

(a) having no solution;

(b) having exactly one solution;

(c) having exactly two solutions;

(d) having exactly three solutions;

(e) having exactly four solutions;

(f) having exactly five solutions.

Hint. One of the cases (a)–(f) is impossible.

Problem 268. What is the possible number of solutions of the equation

$$ax^6 + bx^3 + c = 0\,?$$

Hint. Remember that a, b or c may be equal to 0.

Answer. 0, 1, 2, or infinitely many.

Problem 269. The same question for the equation

$$ax^8 + bx^4 + c = 0.$$

61 Symmetric equations

Problem 270. Solve the equation

$$2x^4 + 7x^3 + 4x^2 + 7x + 2 = 0.$$

Solution. First of all, $x = 0$ is not a solution of this equation. Therefore we lose nothing dividing by x:

$$2x^2 + 7x + 4 + \frac{7}{x} + \frac{2}{x^2} = 0.$$

Now we group terms with equal coefficients and opposite powers of x:

$$2\left(x^2 + \frac{1}{x^2}\right) + 7\left(x + \frac{1}{x}\right) + 4 = 0.$$

Now we use that $x^2 + \dfrac{1}{x^2}$ may be expressed in terms of $x + \dfrac{1}{x}$:

$$\left(x + \frac{1}{x}\right)^2 \;=\; x^2 + 2 \cdot x \cdot \frac{1}{x} + \left(\frac{1}{x}\right)^2 \;=\; x^2 + \frac{1}{x^2} + 2,$$

and

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

Therefore, if x is a solution of the given equation, then $y = x + \frac{1}{x}$ is a solution of the equation $2(y^2 - 2) + 7y + 4 = 0$, or $2y^2 - 4 + 7y + 4 = 0$. or $2y^2 + 7y = 0$, or $y(2y + 7) = 0$, whose solutions are $y = 0$ and $y = -7/2$. Therefore the solutions of the initial equation are all x such that

$$x + \frac{1}{x} = 0 \quad \text{or} \quad x + \frac{1}{x} = -\frac{7}{2}.$$

Let us solve these two equations. The first one:

$$x + \frac{1}{x} = 0 \quad \Longrightarrow \quad x^2 + 1 = 0 \quad \Longrightarrow \quad x^2 = -1. \text{ No solutions.}$$

The second equation: $x + \frac{1}{x} = -\frac{7}{2}$ means (we know that $x \neq 0$) that $x^2 + 1 = -\frac{7}{2}x$, or $x^2 + \frac{7}{2}x + 1 = 0$; the roots are

$$x_{1,2} = \frac{-\frac{7}{2} \pm \sqrt{\frac{49}{4} - 4}}{2} = \frac{\frac{-7}{2} \pm \frac{\sqrt{33}}{2}}{2}$$

Answer. The given equation has two solutions:

$$x_1 = \frac{-7 - \sqrt{33}}{4}, \quad x_2 = \frac{-7 + \sqrt{33}}{4}.$$

62 How to confuse students on an exam

As usual, there are many ways to make evil use of knowledge. Here are the instructions for one of them, namely, how to invent a practically unsolvable equation.

1. Take a quadratic equation – preferably with non-integer roots, for example,

$$3x^2 + 2x - 10 = 0.$$

whose roots are $x_{1,2} = \frac{-2 \pm \sqrt{4 + 120}}{6} = \frac{-2 \pm \sqrt{124}}{6} = \frac{-1 \pm \sqrt{31}}{3}$.

2. Substitute some polynomial of degree 2 instead of x, for example, take $x = y^2 + y - 1$. You get

$$x^2 = (y^2 + y - 1)(y^2 + y - 1) = y^4 + 2y^3 - y^2 - 2y + 1,$$

$$\begin{array}{rlllll} 3x^2+2x-10 = & 3y^4 & +6y^3 & -3y^2 & -6y & +3 \\ & & & +2y^2 & +2y & -2 \\ & & & & & -10 \\ = & 3y^4 & +6y^3 & -y^2 & -4y & -9. \end{array}$$

3. Ask the students to solve the equation

$$3y^4+6y^3-y^2-4y-9=0.$$

4. Wait 10 to 15 minutes.

5. Tell the students that their time is up and they failed.

6. If somebody complains that the problem is too difficult and could not be solved by standard methods, you can explain that in fact this equation can be easily reduced to a quadratic:

$$\begin{array}{rllllll} & 3y^4 & +6\,y^3 & -\ y^2 & -\ 4y & -\ 9 & = \\ = & 3y^4 & +3\,y^3 & -\ 3y^2 & & & \\ & & +3\,y^3 & +\ 3y^2 & -\ 3y & & \\ & & & -\ y^2 & -\ y & +\ 1 & \\ & & & & & -\ 10 & = \end{array}$$

$$= 3y^2(y^2+y-1)+3y(y^2+y-1)-(y^2+y-1)-10 =$$

$$= 3(y^2+y)(y^2+y-1)-(y^2+y-1)-10.$$

If now we denote y^2+y-1 by x we get an equation

$$3(x+1)x-x-10=0,$$

$$3x^2+3x-x-10=0,$$

$$x_{1,2}=\frac{-1\pm\sqrt{31}}{3}$$

and it remains to solve the two equations

$$y^2+y-1=\frac{-1-\sqrt{31}}{3} \quad \text{and} \quad y^2+y-1=\frac{-1-\sqrt{31}}{3}.$$

That's all, isn't it?

Another efficient method is to choose two quadratic equations with non-integer roots, for example,

$$x^2+x-3=0 \quad \text{and} \quad x^2+2x-1=0$$

and multiply them:

$$(x^2 + x - 3)(x^2 + 2x - 1) = x^4 + 3x^3 - 2x^2 - 7x + 3 = 0.$$

The resulting equation can be given to students without a big risk of seeing it solved. But don't lose the sheet of paper with the factoring; otherwise you will be caught by your own trap when somebody asks you to show the solution!

63 Roots

A square root of a is defined as a number whose square is equal to a. (To be exact, a square root of a *nonnegative* number a is a *nonnegative* number whose square is equal to a.) In the same way we can define other roots: a *cube root* of $A \geq 0$ is a number $x \geq 0$ such that $x^3 = a$, a *fourth root* of $a \geq 0$ is a number $x \geq 0$ such that $x^4 = a$, etc. The notation for the nth root of a is $\sqrt[n]{a}$.

Definition. An nth root of a nonnegative number a is a nonnegative number x such that $x^n = a$. (We assume that n is a positive integer.)

This definition raises several questions.

Question. What happens if there are many numbers x having this property?

Answer. This cannot happen. The greater a nonnegative number x, the greater is x^n (if in a product of nonnegative factors all factors increase, the product increases also). So different nonnegative values of x have different nth powers.

Question. Is it possible that there is no x with the required property?

Answer. The same question was discussed for the square root. Those arguments are still valid, and we have no other (more convincing) ones.

Question. If the degree n is even, then the number $-\sqrt[n]{a}$ also has its nth power equal to a. Why do we prefer the positive x such that $x^n = a$ and reject the negative one?

Answer. This is a generally accepted convention.

Question. If the degree n is odd, then for negative a we can also find an x such that $x^n = a$. For example, $(-2)^3 = -8$. So why do we do not say that the cube root of -8 is -2?

Answer. It is possible to extend our definition to this case (and sometimes people do so), but for simplicity we will consider only nonnegative roots of nonnegative numbers. (Otherwise we should consider two cases – odd and even n – all the time.)

Problem 271. Which number is bigger: $\sqrt[10]{2}$ or 1.2?

Problem 272. Compute $\sqrt[7]{0.999}$ to three decimal digits.

Problem 273. Which number is bigger: $\sqrt{2}$ or $\sqrt[3]{3}$?

Problem 274. Which number is bigger: $\sqrt[3]{3}$ or $\sqrt[4]{4}$?

Problem 275. Which number is bigger: $\sqrt{\sqrt{2}}$ or $\sqrt[4]{2}$?

Problem 276. What is $\sqrt[1]{a}$ according to our definition?

Answer. $\sqrt[1]{a} = a$ (for $a \geq 0$).

Now we shall prove some properties of roots.

Problem 277. Prove that (for $a \geq 0$, $b \geq 0$)

$$\sqrt[n]{ab} = \sqrt[n]{a} \cdot \sqrt[n]{b}.$$

Solution. According to the definition of $\sqrt[n]{ab}$ we have to prove that

$$\left(\sqrt[n]{a} \cdot \sqrt[n]{b}\right)^n = ab.$$

Using that

$$(xy)^n = x^n \cdot y^n$$

we get (let $x = \sqrt[n]{a}$, $y = \sqrt[n]{b}$)

$$\left(\sqrt[n]{a} \cdot \sqrt[n]{b}\right)^n = \left(\sqrt[n]{a}\right)^n \cdot \left(\sqrt[n]{b}\right)^n = ab.$$

Problem 278. Prove that (for nonnegative a and b)

$$\sqrt[n]{\frac{a}{b}} = \frac{\sqrt[n]{a}}{\sqrt[n]{b}}.$$

Hint. You may use the equation

$$\left(\frac{x}{y}\right)^n = \frac{x^n}{y^n}$$

or the preceding problem.

Problem 279. Prove that for positive a

$$\sqrt[n]{\frac{1}{a}} = \frac{1}{\sqrt[n]{a}}.$$

Problem 280. Prove that for three nonnegative numbers a, b, and c

$$\sqrt[n]{abc} = \sqrt[n]{a} \cdot \sqrt[n]{b} \cdot \sqrt[n]{c}.$$

Solution.

$$\sqrt[n]{abc} = \sqrt[n]{(ab)c} = \sqrt[n]{ab} \cdot \sqrt[n]{c} = \sqrt[n]{a} \cdot \sqrt[n]{b} \cdot \sqrt[n]{c}.$$

The same statement is true for four, five, etc. numbers.

Problem 281. Prove that for nonnegative a

$$\sqrt[n]{a^m} = \left(\sqrt[n]{a}\right)^m.$$

Solution.

$$\sqrt[n]{a^m} = \sqrt[n]{\underbrace{a \cdot a \cdots a}_{m \text{ times}}} = \underbrace{\sqrt[n]{a} \cdot \sqrt[n]{a} \cdots \sqrt[n]{a}}_{m \text{ times}} = \left(\sqrt[n]{a}\right)^m.$$

(We used the statement of the preceding problem.)

Problem 282. There is a flaw in the solution of the preceding problem; find and correct it.

Solution. We assumed that $m \geq 2$; however, the statement makes sense for all integers m (and positive integers n). The cases $m = 0$ and $m = 1$ are trivial. Let us prove it for negative values of m. For example, assume that $m = -3$. Then

$$\sqrt[n]{a^{-3}} = \sqrt[n]{\frac{1}{a^3}} = \frac{1}{\sqrt[n]{a^3}} = \frac{1}{\left(\sqrt[n]{a}\right)^3} = \left(\sqrt[n]{a}\right)^{-3}.$$

Problem 283. Prove that

$$\sqrt[mn]{a} = \sqrt[m]{\sqrt[n]{a}}$$

for any positive integers m, n and for any nonnegative a.

Solution. According to the definition of mnth root we have to prove that

$$\left(\sqrt[m]{\sqrt[n]{a}}\right)^{mn} = a.$$

Indeed,

$$\left(\sqrt[m]{\sqrt[n]{a}}\right)^{mn} = \left(\left(\sqrt[m]{\sqrt[n]{a}}\right)^{m}\right)^{n} = \left(\sqrt[n]{a}\right)^{n} = a.$$

Problem 284. Prove that

$$\sqrt[mn]{a^n} = \sqrt[m]{a}$$

(here m, n are positive integers, $a \geq 0$).

Problem 285. Prove that

$$\sqrt[n]{a^n b} = a\sqrt[n]{b}$$

(n is a positive integer, $a \geq 0$, $b \geq 0$).

64 Non-integer powers

Different properties of roots are hard to remember. The following mnemonic rule may be useful: All of them can be obtained from the known properties of powers if we agree that

$$\sqrt{a} = a^{1/2}, \quad \sqrt[3]{a} = a^{1/3}, \quad \sqrt[4]{a} = a^{1/4} \qquad \text{etc.}$$

For example, the main property of roots (in fact, the definition)

$$\left(\sqrt[n]{a}\right)^n = a$$

now may be rewritten as

$$\left(a^{1/n}\right)^n = a$$

and becomes a special case of the general rule

$$(a^p)^q = a^{pq}$$

where $p = 1/n$, $q = n$.

The property

$$\sqrt[n]{\sqrt[m]{a}} = \sqrt[mn]{a}$$

now may be rewritten as

$$\left(a^{1/m}\right)^{1/n} = a^{1/mn}$$

and can be obtained if we let $p = 1/m$, $q = 1/n$.

Problem 286. Do the same thing for all properties of roots mentioned above (using appropriate properties of powers).

Mnemonic rules are always disappointing, so let us make the status of our rule higher and call it a *definition* of the $1/n$-th power (we may do so because before, we had only integer powers).

Definition. For any integer $n \geq 1$ let

$$a^{1/n} = \sqrt[n]{a}.$$

We immediately observe that this definition does not make us completely happy. For example, we would like to write that

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

(as a special case of the rule $a^m \cdot a^n = a^{m+n}$ where $m = n = \frac{1}{3}$). But we do not know what $a^{2/3}$ is. To fill this gap we *define* $a^{2/3}$ as $(a^{1/3})^2$ and, in general, $a^{m/n}$ as $(a^{1/n})^m$ or, in other words, as $(\sqrt[n]{a})^m$. So we come to the following

Definition. For any integer m and for any positive integer n the expression $a^{m/n}$ is defined as follows:

$$a^{\frac{m}{n}} = \left(\sqrt[n]{a}\right)^m.$$

The careful reader would mention that there is some cheating in this definition. Indeed,

$$a^{\frac{10}{15}} \text{ is defined as } \left(\sqrt[15]{a}\right)^{10}$$

and at the same time

$$a^{\frac{2}{3}} \text{ is defined as } \left(\sqrt[3]{a}\right)^2$$

At the same time $\frac{10}{15} = \frac{2}{3}$ so $a^{10/15}$ must be equal to $a^{2/3}$. So the correctness of our definition requires that

$$\left(\sqrt[15]{a}\right)^{10} = \left(\sqrt[3]{a}\right)^{2}.$$

Problem 287. Prove this fact.

Solution.

$$\left(\sqrt[3\cdot 5]{a}\right)^{2\cdot 5} = \left(\left(\sqrt[5]{\sqrt[3]{a}}\right)^{5}\right)^{2} = \left(\sqrt[3]{a}\right)^{2}.$$

Problem 288. Prove that reducing common factors in the fraction $\frac{m}{n}$ does not change the value of the expression $a^{\frac{m}{n}}$ (see the definition above).

Hint. In the preceding problem the common factor 5 was reduced in the fraction $\frac{10}{15}$.

Now the properties of powers that we know for integer powers should be checked for arbitary rational powers (where the exponent is a ratio of any integers).

Problem 289. Prove that

$$a^{p} \cdot a^{q} = a^{p+q}$$

for any rational p and q.

Solution. For example, let $p = 2/5$, $q = 3/7$. We have to check that

$$a^{\frac{2}{5}} \cdot a^{\frac{3}{7}} = a^{\frac{2}{5}+\frac{3}{7}}.$$

Let us find a common denominator for $2/5$ and $3/7$:

$$\frac{2}{5} = \frac{14}{35}, \quad \frac{3}{7} = \frac{15}{35}.$$

As we know already,

$$a^{\frac{2}{5}} = a^{\frac{14}{35}}, \quad a^{\frac{3}{7}} = a^{\frac{15}{35}},$$

therefore

$$\begin{aligned} a^{\frac{2}{5}} \cdot a^{\frac{3}{7}} &= a^{\frac{14}{35}} \cdot a^{\frac{15}{35}} = \left(\sqrt[35]{a}\right)^{14} \cdot \left(\sqrt[35]{a}\right)^{15} = \\ &= \left(\sqrt[35]{a}\right)^{14+15} = a^{\frac{14+15}{35}} = a^{\frac{14}{35}+\frac{15}{35}} = a^{\frac{2}{5}+\frac{3}{7}}. \end{aligned}$$

Problem 290. Prove that

$$(ab)^{m/n} = a^{m/n} \cdot b^{m/n}.$$

Problem 291. Prove that

$$(a^p)^q = a^{pq}$$

for any rational p and q.

Solution. Let us start with the case of integer q and arbitrary rational $p = m/n$. In this case

$$(a^p)^q = \left(a^{\frac{m}{n}}\right)^q = \left(\left(\sqrt[n]{a}\right)^m\right)^q = \left(\sqrt[n]{a}\right)^{mq} = a^{\frac{mq}{n}} = a^{pq}.$$

Assume now that $q = 1/k$ for some integer k and that $p = m/n$. Then

$$(a^p)^q = \left(a^{\frac{m}{n}}\right)^{\frac{1}{k}} = \sqrt[k]{a^{\frac{m}{n}}} = \sqrt[k]{\left(\sqrt[n]{a}\right)^m}$$

Let us denote $\sqrt[n]{a}$ as b and continue this chain of equalities:

$$\cdots = \sqrt[k]{b^m} = \left(\sqrt[k]{b}\right)^m = \left(\sqrt[k]{\sqrt[n]{a}}\right)^m = \left(\sqrt[kn]{a}\right)^m = a^{\frac{m}{kn}} = a^{\frac{m}{n}\cdot\frac{1}{k}} = a^{pq}.$$

Finally, for an arbitrary $q = l/k$ we have

$$(a^p)^q = (a^p)^{\frac{l}{k}} = \left(\sqrt[k]{a^p}\right)^l = \left((a^p)^{\frac{1}{k}}\right)^l = \left(a^{\frac{p}{k}}\right)^l = a^{\frac{pl}{k}} = a^{pq}.$$

We used that

$$\left(a^p\right)^{\frac{1}{k}} = a^{p\cdot\frac{1}{k}}$$

and then we used that

$$\left(a^{\frac{p}{k}}\right)^l = a^{\frac{p}{k}\cdot l}.$$

These two special cases of the statement of the problem are considered already.

Problem 292. Prove that for $a > 1$ the value of a^p increases when p increases. Prove that for $0 < a < 1$ the value of a^p *decreases* when p increases.

Hint. When comparing two values of p, find the common denominator. Do not forget that p may be negative (and the statement of the problem remains true).

This problem shows a possible way to extend the definition of a^x to the irrational values of x. For example, we may try to define

$$2^{\sqrt{2}}$$

as a number that is bigger than any of the numbers $2^{p/q}$ when $p/q < \sqrt{2}$ but smaller than any of the numbers $2^{p/q}$ when $p/q > \sqrt{2}$. Of course, to make this definition correct we must prove that such a number exists and is unique, but these topics belong to the scope of calculus.

Problem 293. How do you think one should define $\sqrt[1/2]{a}$ or $\sqrt[-1/2]{a}$?

Answer. As a^2 and a^{-2}.

65 Proving inequalities

Almost all the inequalities in this section in principle could be proved by "brute force" if we computed the values of all the expressions. But we shall look for a better way.

Problem 294. Prove that

$$\frac{1}{2} < \frac{1}{101} + \frac{1}{102} + \cdots + \frac{1}{200} < 1.$$

Solution. Each of 100 terms of the sum is between $\dfrac{1}{200}$ and $\dfrac{1}{100}$. If all terms were equal to $\dfrac{1}{200}$, the sum would be equal to $\dfrac{1}{2}$; if all terms were equal to $\dfrac{1}{100}$, the sum would be equal to 1.

Problem 295. Prove that

$$\frac{1}{2} < 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{1}{199} - \frac{1}{200} < 1.$$

Solution. The left inequality can be proved by grouping the terms with parentheses as

$$\left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{199} - \frac{1}{200}\right).$$

Here the first parenthesized grouping is equal to $1/2$, and all the others are positive.

To get the right inequality we rewrite the expression as

$$1-\left(\frac{1}{2}-\frac{1}{3}\right)-\left(\frac{1}{4}-\frac{1}{5}\right)-\cdots-\left(\frac{1}{198}-\frac{1}{199}\right)-\frac{1}{200}.$$

Here all parenthesized groupings are positive.

Remark. In fact the preceding two problems coincide in a sense:

$$\frac{1}{101}+\frac{1}{102}+\cdots+\frac{1}{200}=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{1}{199}-\frac{1}{200}.$$

Problem 296. Prove this coincidence.

Solution. Indeed,

$$\begin{aligned}
&\frac{1}{101}+\cdots+\frac{1}{200}=\\
&=\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{200}\right)-\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{100}\right)=\\
&=\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{200}\right)-2\cdot\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\frac{1}{8}+\cdots+\frac{1}{200}\right)=\\
&=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{1}{199}-\frac{1}{200}.
\end{aligned}$$

Problem 297. Prove that $(1.01)^{100}\geq 2$.

Solution. By definition

$$(1.01)^{100}=\underbrace{(1+0.01)(1+0.01)\cdots(1+0.01)}_{\text{100 factors}}$$

What happens if we remove the parentheses? We get a sum of many products (each term of this sum is a product of 100 numbers – one for each parenthesized expression). One of the terms is 1 (a product of all the ones). Among other terms there are terms being products of 99 ones and only one 0.01. We have 100 terms of this type (because the 0.01 term could be taken from any of the parenthesized expressions). The value of such a term is 0.01. There are other terms also (equal to 0.01^2, 0.01^3, etc.) but even if we omit them we get the sum

$$1+100\cdot 0.01=2.$$

Another solution to the same problem goes as follows:

$$\begin{aligned}
1.01^2 &= 1.0201 > 1.02 \\
1.01^3 &= 1.01^2 \cdot 1.01 > 1.02 \cdot 1.01 = (1+0.02)(1+0.01) = \\
&= 1 + 0.02 + 0.01 + 0.02 \cdot 0.01 > 1.03 \\
1.01^4 &= 1.01^3 \cdot 1.01 > 1.03 \cdot 1.01 = (1+0.03)(1+0.01) = \\
&= 1 + 0.03 + 0.01 + 0.03 \cdot 0.01 > 1.04 \\
1.01^5 &> 1.05 \\
1.01^6 &> 1.06 \\
&\dots \\
1.01^{99} &> 1.99 \\
1.01^{100} &> 2\,.
\end{aligned}$$

Problem 298. Prove that

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots + \frac{1}{100^2} < 2\,.$$

Solution.

$$\begin{aligned}
\frac{1}{4} = \frac{1}{2^2} &< \frac{1}{1 \cdot 2} = \frac{1}{1} - \frac{1}{2} \\
\frac{1}{9} = \frac{1}{3^2} &< \frac{1}{2 \cdot 3} = \frac{1}{2} - \frac{1}{3} \\
\frac{1}{16} = \frac{1}{4^2} &< \frac{1}{3 \cdot 4} = \frac{1}{3} - \frac{1}{4} \\
&\dots \\
\frac{1}{100^2} &< \frac{1}{99 \cdot 100} = \frac{1}{99} - \frac{1}{100},
\end{aligned}$$

hence (adding all the inequalities),

$$\begin{aligned}
&1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots + \frac{1}{100^2} < \\
&< 1 + \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{99} - \frac{1}{100}\right) = \\
&= 1 + 1 - \frac{1}{100} < 2\,.
\end{aligned}$$

Here the terms $-\frac{1}{2}$ and $\frac{1}{2}$, $-\frac{1}{3}$ and $\frac{1}{3}$, etc. cancel out.

Problem 299. Which is bigger: 1000^{2000} or 2000^{1000}?

Problem 300. Prove that

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{1{,}000{,}000}<20.$$

Prove that

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}>20.$$

for some n.

Hint. In the expression

$$\left(\frac{1}{2}\right)+\left(\frac{1}{3}+\frac{1}{4}\right)+\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}\right)+\left(\frac{1}{9}+\frac{1}{10}+\cdots+\frac{1}{16}\right)+\cdots$$

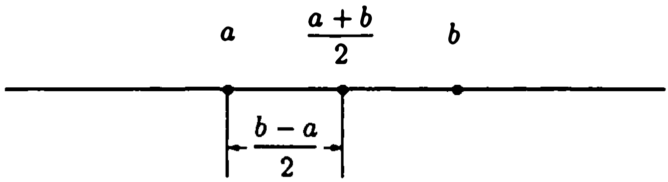
each expression in parentheses is between $1/2$ and 1 (compare with the first problem of this section).

66 Arithmetic and geometric means

The *arithmetic mean* (pronounced "arithmEtic", not "arIthmetic") of two numbers a and b is defined as $\frac{a+b}{2}$, that is, as half of their sum. The corresponding point on the real line is the midpoint of the segment with endpoints a and b.

Problem 301. Prove this fact.

Solution. Without loss of generality we may assume that $a<b$. In this case point a is on the left of point b.



The distance between these points is equal to $b-a$; if we add to a one-half of this distance we get

$$a+\frac{b-a}{2}=\frac{2a+b-a}{2}=\frac{a+b}{2}.$$

Problem 302. The arithmetic mean of two numbers 1 and a is equal to 7. Find a.

The *geometric mean* of two nonnegative numbers a and b is defined as the square root of their product, $\sqrt{ab}$. We restrict ourselves to nonnegative a and b; if a and b have different signs, their product is negative and the square root is undefined. If both numbers are negative, then $\sqrt{ab}$ is defined, but it would be strange to call the positive number $\sqrt{ab}$ a geometric mean of two negative numbers!

Problem 303. The geometric mean of two numbers 1 and a is equal to 7. Find a.

Problem 304. (a) Find the side of a square having the same perimeter as a rectangle with sides a and b. (b) Find the side of a square having the same area as a rectangle with sides a and b.

Problem 305. We have already heard about arithmetic and geometric progressions, and now we learn the terms "arithmetic mean" and "geometric mean". Can you explain this coincidence of terms?

Solution. The sequence

$$a,\quad (\text{the arithmetic mean of } a \text{ and } b),\quad b$$

is an arithmetic progression while the sequence

$$a,\quad (\text{the geometric mean of } a \text{ and } b),\quad b$$

is a geometric progression.

One more way to define the arithmetic and geometric mean:

- The arithmetic mean of a and b is a number x such that

$$x-a=b-x;$$

- The geometric mean of a and b is a number x such that

$$\frac{x}{a}=\frac{b}{x}\qquad (\text{for } a,b>0).$$

67 The geometric mean does not exceed the arithmetic mean

Problem 306. Prove that for nonnegative a and b

$$\sqrt{ab} \le \frac{a+b}{2}$$

Solution. To compare nonnegative numbers $\sqrt{ab}$ and $\frac{a+b}{2}$ let us compare their squares and prove that

$$ab \le \left(\frac{a+b}{2}\right)^2.$$

Taking into account that

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

we have to prove that

$$ab \le \frac{(a+b)^2}{4}$$

or, in other words, that $4ab \le (a+b)^2$, or $4ab \le a^2 + 2ab + b^2$, or $0 \le a^2 - 2ab + b^2$.

It is easy to recognize $(a-b)^2$ as the right-hand side of this inequality, therefore it is proved (a square is always nonnegative).

Problem 307. When is the arithmetic mean of two numbers equal to their geometric mean?

Solution. As we see from the solution of the preceeding problem, this happens if and only if $(a-b)^2 = 0$, that is, if $a = b$.

68 Problems about maximum and minimum

Problem 308. (a) What is the maximum value of the product of two nonnegative numbers whose sum is a fixed positive number c ? (b) What is its minimum value?

Solution. (a) The arithmetic mean of these numbers is $c/2$, so their geometric mean cannot exceed $c/2$, and its square (that is, the product of the numbers) never exceeds $c^2/4$. This maximum value is achieved when the numbers are equal.

(b) The minimum value is zero (one of the numbers is zero, the other one is equal to c).

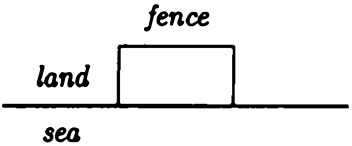
Problem 309. What are the maximum and minimum values of the sum of two nonnegative numbers whose product is a fixed $c > 0$?

Solution. The geometric mean of these numbers is $\sqrt{c}$. Therefore their arithmetic mean is not less than $\sqrt{c}$ and their sum (which is two times bigger) is not less than $2\sqrt{c}$. This value is achieved if the numbers are equal. The maximum value does not exist (the sum may be arbitrarily large if one of the numbers is close to zero and the other one is very large).

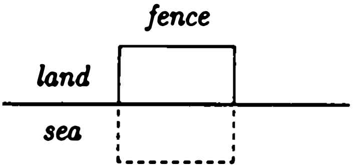
Remark. As you may remember, we have met the two last problems earlier when speaking about maximum and minimum values of quadratic polynomials.

Problem 310. What is the maximum possible area of a rectangular piece of land if you may enclose it with only 120 m of fence?

Problem 311. What is the maximum possible area of a rectangular piece of land near the (straight) sea shore if you may enclose it with only 120 m of fence? (You don't need the fence on the shore or in the water.)



Solution. Imagine the symmetric fence in the sea:



We get a rectangle (half in the water) with perimeter equal to 240 m. Its area will be maximal if it is a square with side 60 m. In this case the area is equal to $3600\,\mathrm{m}^2$. The real area (on the shore) is half of this and equals $1800\,\mathrm{m}^2$; the real fence contains of segments of length 30, 60, and 30 meters.

Problem 312. What is the maximum value of the product ab if a and b are nonnegative numbers such that $a + 2b = 3$?

Solution. It is easier to say when the product of two nonnegative numbers a and $2b$ (whose sum equals 3) is maximal. It is maximal when these numbers are equal, that is, $a = 2b = 3/2$. The product of a and b is half the product of a and $2b$; its maximum value is

$$\frac{3}{2}\cdot\frac{3}{4} = \frac{9}{8}.$$

69 Geometric illustrations

The inequality

$$\sqrt{ab} \le \frac{a+b}{2}$$

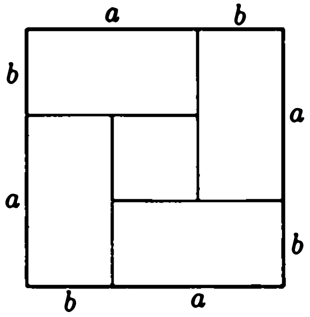
can be rewritten as

$$2\sqrt{ab} \le a + b$$

and then, after squaring, as

$$4ab \le (a+b)^2$$

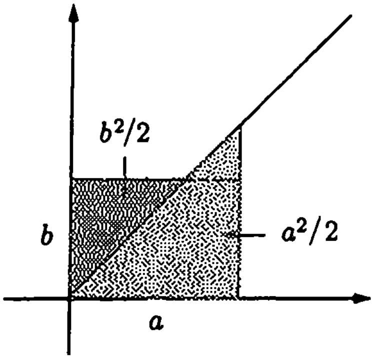
The last inequality can be illustrated as follows: Four rectangles $a \times b$ can be put into the square with side $a + b$ (and some space in the middle of the square remains, if $a \neq b$).



Problem 313. How much free space remains? Compare the result

with the algebraic proof of the inequality given above.

Another illustration is as follows. Consider the bisector of a right angle, and two triangles with sides a and b parallel to the sides of the angle:



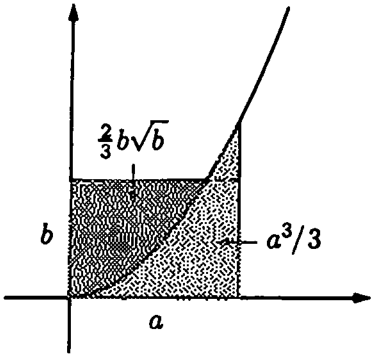
Their areas are $a^2/2$ and $b^2/2$. Together these triangles cover a rectangle with sides a and b; therefore

$$ab \leq \frac{a^2 + b^2}{2}.$$

To see that this illustrates the inequality between the arithmetic and the geometric mean, substitute $\sqrt{c}$ and $\sqrt{d}$ for a and b; you get

$$\sqrt{c} \cdot \sqrt{d} \leq \frac{c + d}{2}.$$

Remark. You may use almost any curve instead of the bisector – and obtain many other inequalities, if you are able to compute the areas of triangles formed by curves.



For example, for the curve $y = x^2$ you get (as calculus experts say) two "triangles" having areas $a^3/3$ and $\frac{2}{3}b\sqrt{b}$. So the inequality obtained is

$$ab \le \frac{a^3}{3} + \frac{2}{3}b\sqrt{b}.$$

This is true for any nonnegative a and b.

70 The arithmetic and geometric means of several numbers

The arithmetic mean of three numbers is defined as $\dfrac{a+b+c}{3}$; the geometric mean is defined as $\sqrt[3]{abc}$ (we assume that $a, b, c \ge 0$). Similar definitions are given for four, five, etc. numbers; the arithmetic mean of $a_1, \ldots, a_n$ is

$$\frac{a_1 + \cdots + a_n}{n};$$

the geometric mean is

$$\sqrt[n]{a_1 \cdot a_2 \cdots a_n}.$$

The inequality between the arithmetic and the geometric means can be generalized for the case of n numbers:

$$\sqrt[n]{a_1 \cdots a_n} \le \frac{a_1 + \cdots + a_n}{n}.$$

As for the case of two numbers (see above), equality is possible only if all numbers are equal.

Before proving this inequality we shall derive some of its consequences.

Problem 314. Using this inequality, prove that if $a_1, \ldots, a_n$ are nonnegative numbers and $a_1 + a_2 + \cdots + a_n \le n$, then $a_1 \cdot a_2 \cdots a_n \le 1$.

Solution.

$$a_1 + \cdots + a_n \le n \quad \Longrightarrow \quad \frac{a_1 + \cdots + a_n}{n} \le 1 \quad \Longrightarrow$$

$$\Longrightarrow \quad \sqrt[n]{a_1 \cdot a_2 \cdots a_n} \le \frac{a_1 + \cdots + a_n}{n} \le 1 \quad \Longrightarrow \quad a_1 \cdot a_2 \cdots a_n \le 1.$$

In the following two problems you may also use the inequality between the arithmetic and the geometric means without proof.

Problem 315. Prove that the product of n nonnegative numbers with a fixed sum is at a maximum when all the numbers are equal.

Problem 316. Prove that the sum of n nonnegative numbers with a given product is at a minimum when all the numbers are equal.

There are different proofs of the inequality between the arithmetic mean and the geometric mean of n numbers. Unfortunately, the most natural of them uses calculus (the notion of a derivative or something else). We shall avoid that, but our proofs will be tricky.

Problem 317. Prove the inequality between arithmetic and geometric means for $n = 4$.

Solution. We have four nonnegative numbers. During the proof we will change them but keep their sum unchanged. (Therefore, their arithmetic mean will be unchanged.) Their product will change and we'll keep track of how.

1. Replace a and b by two numbers equal to $\frac{a+b}{2}$; so we make a transition

$$a,\ b,\ c,\ d \longrightarrow \frac{a+b}{2},\ \frac{a+b}{2},\ c,\ d.$$

The sum remains unchanged while the product increases (when $a \neq b$) or remains the same (if $a = b$); two factors c and d do not change and the product of two numbers with fixed sum $a+b$ is maximal when numbers are equal (see above).

2. Do the same with c and d:

$$\frac{a+b}{2},\ \frac{a+b}{2},\ c,\ d \longrightarrow \frac{a+b}{2},\ \frac{a+b}{2},\ \frac{c+d}{2},\ \frac{c+d}{2}.$$

The sum remains unchanged, the product increases or remains the same (if $c = d$).

3. We have balanced the first and the second pair; now we balance numbers of different pairs:

$$\frac{a+b}{2},\ \frac{a+b}{2},\ \frac{c+d}{2},\ \frac{c+d}{2} \longrightarrow$$
$$\longrightarrow \frac{a+b+c+d}{4},\ \frac{a+b}{2},\ \frac{a+b+c+d}{4},\ \frac{c+d}{2}$$

and, finally,

$$\frac{a+b+c+d}{4}, \frac{a+b}{2}, \frac{a+b+c+d}{4}, \frac{c+d}{2} \longrightarrow$$
$$\longrightarrow \quad \frac{a+b+c+d}{4}, \frac{a+b+c+d}{4}, \frac{a+b+c+d}{4}, \frac{a+b+c+d}{4}.$$

So ultimately we replaced numbers

$$a, b, c, d$$

by numbers

$$S, S, S, S$$

where

$$S = \frac{a+b+c+d}{4}$$

is the arithmetic mean, and their product increased (or at least did not decrease), so

$$a \cdot b \cdot c \cdot d \leq S \cdot S \cdot S \cdot S$$

or

$$\sqrt[4]{abcd} \leq S.$$

The inequality is proved!

Problem 318. Prove that the inequality between the arithmetic and the geometric means of four numbers (see the preceding problem) becomes an equality only if all numbers are equal.

Hint. Look at the solution of the preceding problem; the final equality is possible only if at all stages numbers being balanced are equal.

Problem 319. Prove the inequality between arithmetic and geometric means for $n = 8$.

Solution. Do the same trick as in the preceding proof: balance numbers in four pairs, then (between pairs) in two quadruples, and then all eight.

Problem 320. Prove the inequality between arithmetic and geometric means for $n = 3$.

Solution. We reduce this problem to the case $n = 4$ by the following method: besides three given numbers a, b, c consider the fourth number, namely, their geometric mean. So we get four numbers

$$a, b, c, \sqrt[3]{abc}$$

and then use the inequality for $n = 4$; we get

$$\sqrt[4]{abc\sqrt[3]{abc}} \leq \frac{a+b+c+\sqrt[3]{abc}}{4}.$$

The left-hand side expression turns out to be equal to $\sqrt[3]{abc}$. To verify this, compute the fourth powers of both (nonnegative) numbers; we get

$$\left(\sqrt[4]{abc\sqrt[3]{abc}}\right)^4 = abc\sqrt[3]{abc}$$

and

$$(\sqrt[3]{abc})^4 = (\sqrt[3]{abc})^3\sqrt[3]{abc} = abc\sqrt[3]{abc},$$

which is the same. So we can rewrite the inequality we have as

$$\sqrt[3]{abc} \leq \frac{a+b+c+\sqrt[3]{abc}}{4}$$

and then

$$\begin{aligned} 4\sqrt[3]{abc} &\leq a+b+c+\sqrt[3]{abc}, \\ 3\sqrt[3]{abc} &\leq a+b+c, \\ \sqrt[3]{abc} &\leq \frac{a+b+c}{3}. \end{aligned}$$

That is what we want.

Problem 321. Using the inequality between arithmetic and geometric means for $n = 8$, prove it for $n = 7$.

Problem 322. Prove the inequality between arithmetic and geometric means for $n = 6$.

Hint. Recall the solution of the preceding problem.

Problem 323. Prove the inequality between arithmetic and geometric means for all integer $n \geq 2$.

Hint. Prove it for $n = 2, 4, 8, 16, 32, \ldots$ and then all integers in between them.

Problem 324. Prove that the inequality between arithmetic and geometric means becomes an equality only if all numbers are equal.

Another proof of the inequality between arithmetic and geometric means goes as follows. First of all we mention that if all numbers $a_1, \ldots, a_n$ are multiplied by the same constant (for example, if all numbers become three times bigger) then both the arithmetic and geometric means are multiplied by the same constant and the relation between them remains unchanged. Therefore, proving the inequality between them, we may multiply all numbers by some constant and assume without loss of generality that their arithmetic mean is equal to 1. Thus, it is enough to prove

$$a_1, \ldots, a_n \geq 0,\ a_1 + \cdots + a_n = n \quad \Rightarrow \quad a_1 \cdots a_n \leq 1.$$

Let's try.

A. For the case of two numbers: If the sum of two numbers is equal to 2, then these numbers can be represented as $1 + h$ and $1 - h$ and their product is $(1+h)(1-h) = 1 - h^2 \leq 1$.

B. Let us consider now the case of three numbers. Assume that the sum of three nonnegative numbers a, b, c is 3. If not all of a, b, c are equal to 1 (the latter case is trivial) some of them must be greater than 1 and some must be smaller. Assume, for instance, that $a < 1$ and $b > 1$. Then $a - 1 < 0$, $b - 1 > 0$ and the product

$$(a-1)(b-1) = ab - a - b + 1$$

is negative, so

$$ab + 1 < a + b.$$

Because

$$(a + b) + c = 3$$

we have

$$ab + 1 + c < (a + b) + c = 3$$

and

$$ab + c < 2.$$

Look – we now have two numbers ab and c, their sum is less than 2 and we have to prove that their product does not exceed 1. For two numbers we already know this fact from part **A**.

The careful reader may ask why we refer to part **A** where we proved that if the sum of two numbers is *equal* to 2 then the product does not exceed 1, and now the sum of two numbers is *smaller* than 2, not *equal* to 2. But this is not a big problem; if the sum is smaller than 2 we may increase one of the numbers and make the sum equal to 2; if the increased product does not exceed 1 then the original product also does not exceed 1.

C. Now assume that $n = 4$; we have to prove that

$$a,\ b,\ c,\ d \geq 0,\ a+b+c+d = 4 \implies abcd \leq 1\,.$$

Again we may assume without loss of generality that one of the numbers, say a, is less than 1 and the other, say b, is greater than 1. Then

$$ab + 1 < a + b, \quad (a+b) + c + d = 4,$$

therefore

$$ab + 1 + c + d < 4, \quad ab + c + d < 3\,.$$

And again it remains to prove that if a sum of three nonnegative numbers ab, c, and d is less than 3 then their product does not exceed 1 – and this is already proved.

The same argument can be applied for $n = 5, 6$, etc.

The next, third proof of the inequality between the arithmetic and the geometric means of three numbers is probably the shortest – but it looks mysterious.

We start from the identity

$$a^3 + b^3 + c^3 - 3abc = \frac{1}{2}(a+b+c)((a-b)^2 + (b-c)^2 + (a-c)^2)$$

which can be checked by a direct computation (perform the operations on the right-hand side). You see that if a, b, c are nonnegative then the right-hand side (and therefore left-hand side) is nonnegative, that is,

$$abc \leq \frac{a^3 + b^3 + c^3}{3}.$$

It remains to substitute $\sqrt[3]{p}$, $\sqrt[3]{q}$, $\sqrt[3]{r}$, for a, b, and c and you get

$$\sqrt[3]{pqr} \le \frac{p+q+r}{3}$$

which concludes the third proof.

Here is one more proof of the inequality between the arithmetic and geometric means. Let us prove that the product of n nonnegative numbers is the maximum when all the numbers are equal. As we have seen, it is easy to prove this fact for $n = 2$. Assume that for some n the product of n equal numbers with a given sum S is *not* the maximum, and that some other numbers $a_1, a_2, \ldots, a_n$ – not all of them equal – provide this maximum. Assume, for example, that $a_1 \neq a_2$. Then replacing both of a_1 and a_2 by their arithmetic mean, we do not change the sum, but the product increases. So we get a contradiction with our assumption that the product was the maximum.

Problem 325. There is a gap in this argument – find it.

Solution. We assumed that numbers $a_1, a_2, \ldots, a_n$ providing the maximum value of the product (for nonnegative numbers with fixed sum) do exist. This fact needs to be proved. In fact it can be proved using calculus methods, but this goes beyond the scope of the book.

Problem 326. Assume that $a_1, \ldots, a_n$ are positive numbers. Prove that

$$\frac{a_1}{a_2} + \frac{a_2}{a_3} + \cdots + \frac{a_{n-1}}{a_n} + \frac{a_n}{a_1} \ge n.$$

Problem 327. Prove that

$$\sqrt[3]{ab^2} \le \frac{a+2b}{3}.$$

Problem 328. Find the minimal value of $a + b$ if a and b are nonnegative numbers and $ab^2 = 1$.

Problem 329. Prove the inequality

$$\sqrt[6]{a} \cdot \sqrt[3]{b} \cdot \sqrt{c} \le \frac{a+2b+3c}{6}$$

for any nonnegative a, b, and c.

Problem 330. Prove the inequality

$$\sqrt[3]{abc} \le \frac{a+2b+3c}{3\sqrt[3]{6}}$$

Problem 331. Prove that

$$\left(1+\frac{1}{10}\right)^{10} < \left(1+\frac{1}{11}\right)^{11}.$$

Solution. The left-hand side $\left(1+\frac{1}{10}\right)^{10}$ is a product of 10 factors each equal to $\left(1+\frac{1}{10}\right)$. We may consider it also as a product of 11 factors, one of them equal to 1 and ten others equal to $\left(1+\frac{1}{10}\right)$. Comparing this product with the product in the right-hand side where we also have 11 factors but all of them are equal to $\left(1+\frac{1}{11}\right)$, we see that the sum of all factors are the same in both cases (namely 12). But in the right-hand side all factors are equal, so the product is bigger.

Problem 332. Prove that

$$\left(1+\frac{1}{10}\right)^{11} > \left(1+\frac{1}{11}\right)^{12}.$$

Hint. The right-hand side may be considered as a product of 11 factors – one equal to

$$\left(1+\frac{1}{11}\right)^2 = \left(1+\frac{2}{11}+\frac{1}{11^2}\right)$$

and the others equal to $\left(1+\frac{1}{11}\right)$. The left-hand side is also a product of 11 factors (but the factors are equal). It is enough to show that the sum of all factors in the left-hand side is bigger than the sum of all factors in the right-hand side and then use the inequality between arithmetic and geometric means.

Problem 333. Write down the four numbers mentioned in the two preceding problems in ascending order.

71 The quadratic mean

The *quadratic mean* of two nonnegative numbers a and b is defined as a nonnegative number whose square is the arithmetic mean of a^2 and b^2, that is, as

$$\sqrt{\frac{a^2+b^2}{2}}.$$

Problem 334. This definition uses the arithmetic mean. What happens if the arithmetic mean is replaced by the geometric mean?

Problem 335. Prove that the quadratic mean of two nonnegative numbers a and b is not less than their arithmetic mean:

$$\sqrt{\frac{a^2+b^2}{2}} \geq \frac{a+b}{2}.$$

(For example, the quadratic mean of 0 and a is $a/\sqrt{2}$ and their arithmetic mean is $a/2$.

Solution. Comparing the squares, we need to prove that

$$\frac{a^2+b^2}{2} \geq \frac{(a+b)^2}{4}.$$

Multiplying by 4 and using the square-of-the-sum formula, we get

$$2(a^2+b^2) \geq a^2+b^2+2ab$$

or

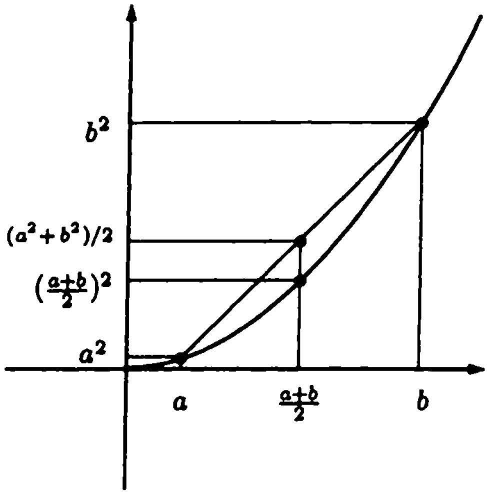
$$a^2+b^2 \geq 2ab, \quad a^2+b^2-2ab \geq 0.$$

Here the left-hand side is a square of $(a-b)$ and, therefore, is always nonnegative.

Problem 336. For which a and b is the arithmetic mean equal to the quadratic mean?

Problem 337. Prove that the geometric mean does not exceed the quadratic mean.

The geometric illustration of the inequality between the arithmetic mean and the quadratic mean can be given as follows.



Draw the graph $y = x^2$ and consider two points $\langle a, a^2 \rangle$ and $\langle b, b^2 \rangle$ on this graph. Connect these points with a segment. The middle point of this segment has coordinates that are the arithmetic means of the coordinates of the endpoints, that is,

$$\left(\frac{a+b}{2}, \quad \frac{a^2+b^2}{2}\right).$$

Look at the picture; you see that this point is higher than the graph point with the same x-coordinate

$$\left(\frac{a+b}{2}, \quad \left(\frac{a+b}{2}\right)^2\right)$$

so the y-coordinate of the first point is bigger than the y-coordinate of the second one:

$$\left(\frac{a+b}{2}\right)^2 \quad \leq \quad \frac{a^2+b^2}{2},$$

$$\frac{a+b}{2} \quad \leq \quad \sqrt{\frac{a^2+b^2}{2}}.$$

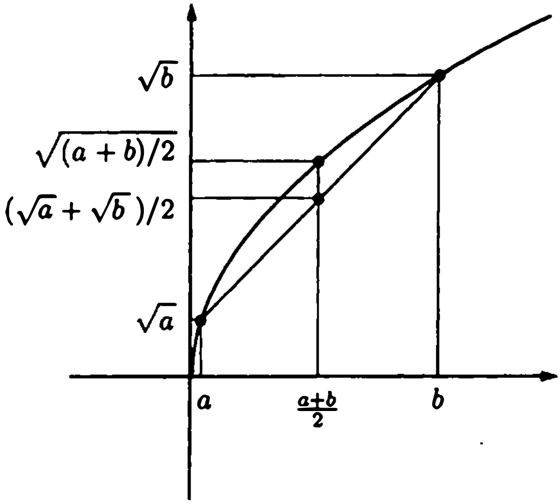
This argument may be considered as a proof of the inequality between arithmetic and quadratic means if we believe that the graph of $y = x^2$

is "concave" (that is, the curve goes under the chord connecting any two points).

Problem 338. Turning the graph $y = x^2$ around (that is, exchanging x- and y-axes), we get the graph of $y = \sqrt{x}$, which goes *above* any of its chords. What inequality corresponds to this fact?

Answer.

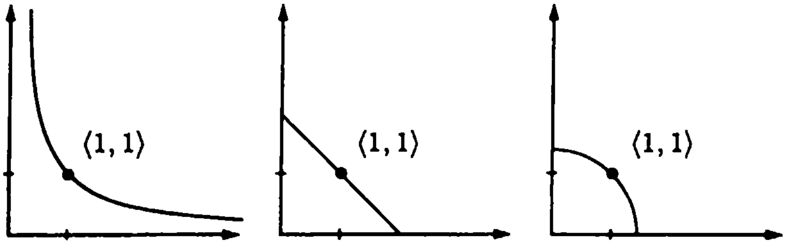
$$\sqrt{\frac{a+b}{2}} \geq \frac{\sqrt{a}+\sqrt{b}}{2}.$$



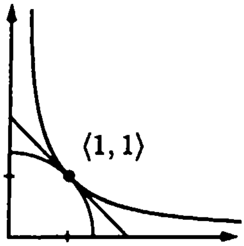
Now we know that for any nonnegative a and b

$$\sqrt{ab} \quad \leq \quad \frac{a+b}{2} \quad \leq \quad \sqrt{\frac{a^2+b^2}{2}}.$$

For any of these three expressions, let us draw in the coordinate plane the set of all points $\langle a, b\rangle$ where this type of mean value does not exceed 1:



If we put all of them in one picture, we see that the bigger expression corresponds to the smaller set (as it should).



Problem 339. Prove the inequality between the arithmetic mean and the quadratic mean for three numbers:

$$\frac{a+b+c}{3} \le \sqrt{\frac{a^2+b^2+c^2}{3}}.$$

Problem 340. (a) The sum of two nonnegative numbers is 2. What is the minimum value of the sum of their squares?

(b) The same question for three numbers.

72 The harmonic mean

The *harmonic mean* of two positive numbers a and b is defined (see above) as the number whose inverse is the arithmetic mean of the inverses of a and b, that is, as

$$\frac{1}{\left(\frac{1}{a}+\frac{1}{b}\right)/2}.$$

Problem 341. Prove that the harmonic mean does not exceed the geometric mean.

Solution. The inverse of the harmonic mean is the arithmetic mean of $1/a$ and $1/b$; the inverse of the geometric mean is the geometric mean of $1/a$ and $1/b$; so it is enough to recall the inequality between the arithmetic and geometric means (the inverse of the bigger number is smaller).

Problem 342. The numbers $a_1, \ldots, a_n$ are positive. Prove that

$$(a_1 + \cdots + a_n)\left(\frac{1}{a_1} + \cdots + \frac{1}{a_n}\right) \geq n^2.$$

Solution. The desired inequality may be rewritten as

$$\frac{a_1 + \cdots + a_n}{n} \geq \frac{1}{\left(\frac{1}{a_1} + \cdots + \frac{1}{a_n}\right) / n}$$

that is, we have to prove that the arithmetic mean of n numbers is greater than or equal to its harmonic mean. This becomes clear if we put the geometric mean between them:

$$\begin{aligned}\frac{a_1 + \cdots + a_n}{n} &\geq \sqrt[n]{a_1 \cdots a_n} = \\ = \frac{1}{\sqrt[n]{\frac{1}{a_1} \cdots \frac{1}{a_n}}} &\geq \frac{1}{\left(\frac{1}{a_1} + \cdots + \frac{1}{a_n}\right) / n};\end{aligned}$$

the last inequality follows from the inequality between the arithmetic and geometric mean of the numbers $1/a_1, \ldots, 1/a_n$.

Another solution uses the following trick. Our inequality becomes a consequence of the so-called Cauchy–Schwarz inequality

$$(p_1 q_1 + \cdots + p_n q_n)^2 \leq (p_1^2 + \cdots + p_n^2) \cdot (q_1^2 + \cdots + q_n^2)$$

if we substitute $\sqrt{a_i}$ for p_i and $\dfrac{1}{\sqrt{a_i}}$ for q_i.

Therefore, it remains to prove the Cauchy–Schwarz inequality. Consider the following quadratic polynomial (where x is considered to be a variable and p_i and q_i are constants):

$$(p_1 + q_1 x)^2 + (p_2 + q_2 x)^2 + \cdots + (p_n + q_n x)^2.$$

If we remove the parentheses and collect terms with x^2, with x, and without x, we get the polynomial

$$Ax^2 + Bx + C$$

where

$$\begin{aligned} A &= q_1^2 + q_2^2 + \cdots + q_n^2\,, \\ B &= 2(p_1q_1 + p_2q_2 + \cdots + p_nq_n)\,, \\ C &= p_1^2 + p_2^2 + \cdots + p_n^2\,. \end{aligned}$$

This polynomial is nonnegative for all x (because it was a sum of squares). Therefore its discriminant $B^2 - 4AC$ must be negative or zero, that is, $B^2 \le 4AC$, or $(B/2)^2 \le AC$, which is to say,

$$(p_1q_1 + \cdots + p_nq_n)^2 \le (p_1^2 + \cdots + p_n^2) \cdot (q_1^2 + \cdots + q_n^2)\,.$$

How do you like this trick?

OTHER BOOKS IN THE SERIES

Algebra is the third book in this series of books for high school students. The first two, published in 1990, are *Functions and Graphs* and *The Method of Coordinates.* Future books will include:
Pre-Geometry
Geometry
Trigonometry
Calculus

As organized and directed by I. M. Gelfand for a Mathematical School by Correspondence, the books are intended to cover the basics in mathematics. *Functions and Graphs* and *The Method of Coordinates* were written more than 25 years ago for the Mathematical School by Correspondence in the former Soviet Union. Still under the guidance of I. M. Gelfand, the School continues to thrive at such places as Rutgers University, New Brunswick, NJ and Bures-sur-Yvette, France.

As Gelfand himself has stated:
"It was not our intention that all of the students who study from these books or even completed the School by Correspondence should choose mathematics as their future career. Nevertheless, no matter what they would later choose, the results of this mathematical training remain with them. For many, this is a first experience in being able to do something completely independently of a teacher."

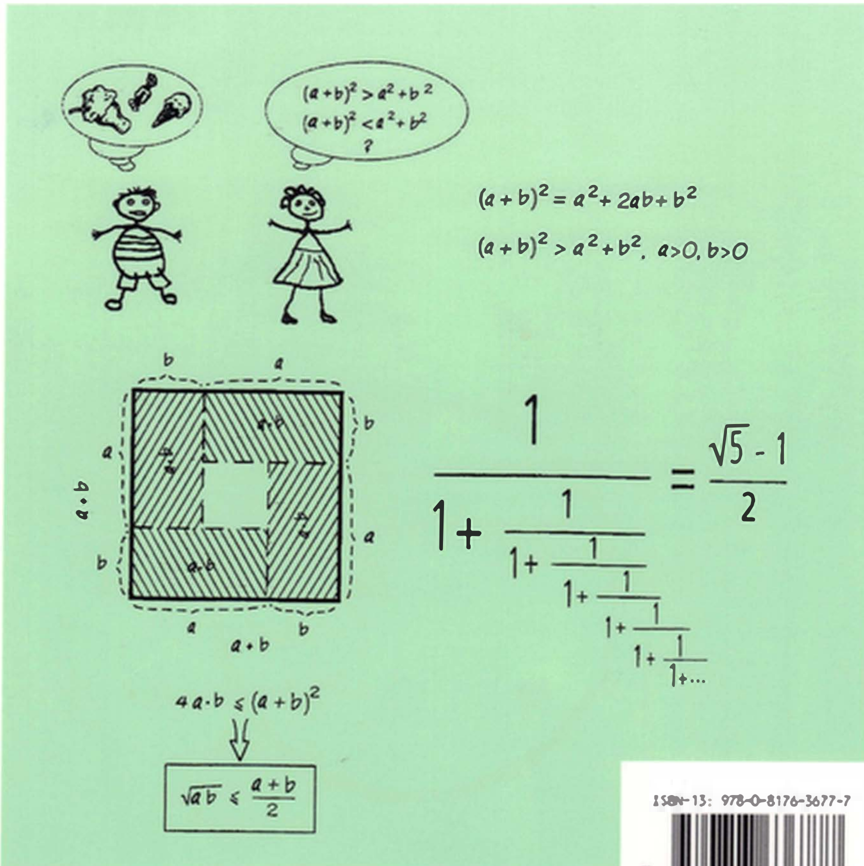
Gelfand continues:
"I would like to make one comment here. Some of my American colleagues have been explained to me that American students are not really accustomed to thinking and working hard, and for this reason we must make the material as attractive as possible. Permit me to not completely agree with this opinion. From my long experience with young students all over the world, I know that they are curious and inquisitive and I believe that if they have some clear material presented in a simple form, they will prefer this to all artificial means of attracting their attention—much as one buys books for their content and not for their dazzling jacket designs that engage only for the moment. The most important thing a student can get from the study of mathematics is the attainment of a higher intellectual level."

"The idea behind teaching is to expect students to learn why things are true, rather than have them memorize ways of solving a few problems, as most of our books have done. [This] same philosophy lies behind the current text by Gelfand and Shen. There are specific 'practical' problems but there is much more development of the ideas... [The authors] have shown how to write a serious yet lively book on algebra."

R. Askey, *The American Mathematical Monthly*

"Were Algebra to be used solely for supplementary reading, it could be wholeheartedly recommended to any high school student or any teacher... In fact, given the long tradition of mistreating algebra as a disjointed collection of techniques in the schools, there should be some urgency in making this book compulsory reading for anyone interested in learning mathematics."

H. Wu, *The Mathematical Intelligencer*



ISBN-13: 978-0-8176-3677-7

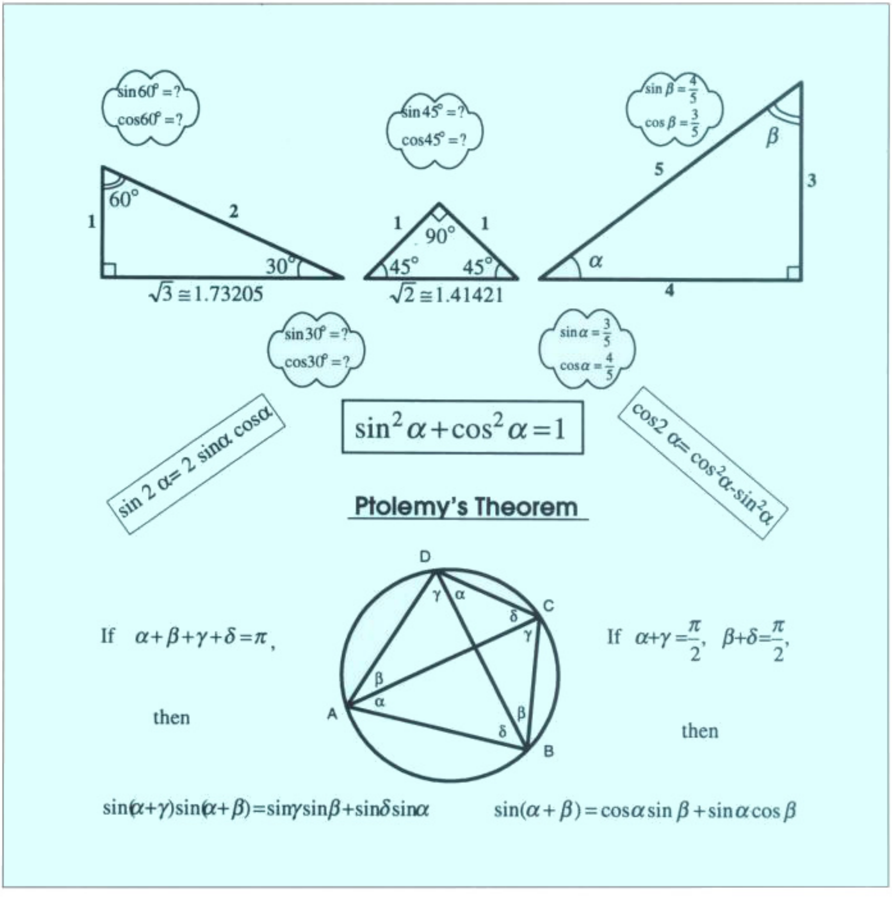
EAN

9 780817 636777

Birkhäuser
Boston • Basel • Berlin

I.M. Gelfand M. Saul

Trigonometry



Cover illustration by Tanya Alekseyevskaya

For

Madge Goldman

who has done so much
to bring so many people together

I.M. Gelfand
Mark Saul

Trigonometry

Birkhäuser
Boston • Basel • Berlin

I.M. Gelfand
Department of Mathematics
Rutgers University
Piscataway, NJ 08854-8019
U.S.A.

Mark Saul
Bronxville Schools
Bronxville, NY 10708
U.S.A.

Library of Congress Cataloging-in-Publication Data

Gel'fand, I.M. (Izrail' Moiseevich)
Trigonometry / I.M. Gel'fand, Mark Saul.
p. cm.
ISBN 0-8176-3914-4 (alk. paper)
1. Trigonometry, Plane. I. Saul, Mark E. II. Title.
QA533.G45 1999
516.24'2–dc21 99-32245
CIP

Printed on acid-free paper
© 2001 I.M. Gelfand
Birkhäuser

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Birkhäuser Boston, c/o Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

ISBN 0-8176-3914-4 SPIN 10536370
ISBN 3-7643-3914-4

Typeset in LaTeX by Martin Stock, Cambridge, MA
Printed and bound by Hamilton Printing, Rensselaer, NY
Printed in the United States of America

9 8 7 6 5 4 3 2 1

Contents

Preface

In a sense, trigonometry sits at the center of high school mathematics. It originates in the study of geometry when we investigate the ratios of sides in similar right triangles, or when we look at the relationship between a chord of a circle and its arc. It leads to a much deeper study of periodic functions, and of the so-called *transcendental* functions, which cannot be described using finite algebraic processes. It also has many applications to physics, astronomy, and other branches of science.

It is a very old subject. Many of the geometric results that we now state in trigonometric terms were given a purely geometric exposition by Euclid. Ptolemy, an early astronomer, began to go beyond Euclid, using the geometry of the time to construct what we now call tables of values of trigonometric functions.

Trigonometry is an important introduction to calculus, where one studies what mathematicians call *analytic* properties of functions. One of the goals of this book is to prepare you for a course in calculus by directing your attention away from particular values of a function to a study of the function as an object in itself. This way of thinking is useful not just in calculus, but in many mathematical situations. So trigonometry is a part of pre-calculus, and is related to other pre-calculus topics, such as exponential and logarithmic functions, and complex numbers. The interaction of these topics with trigonometry opens a whole new landscape of mathematical results. But each of these results is also important in its own right, without being "pre-" anything.

We have tried to explain the beautiful results of trigonometry as simply and systematically as possible. In many cases we have found that simple problems have connections with profound and advanced ideas. Sometimes we have indicated these connections. In other cases we have left them for you to discover as you learn more about mathematics.

About the exercises: We have tried to include a few problems of each "routine" type. If you need to work more such problems, they are easy to find. Most of our problems, however, are more challenging, or exhibit a new aspect of the technique or object under discussion. We have tried to make each exercise tell a little story about the mathematics, and have the stories build to a deep understanding.

We will be happy if you enjoy this book and learn something from it. We enjoyed writing it, and learned a lot too.

Acknowledgments

The authors would like to thank Martin Stock, who took a very ragged manuscript and turned it into the book you are now holding. We also thank Santiago Samanca and the late Anneli Lax for their reading of the manuscript and correction of several bad blunders. We thank Ann Kostant for her encouragement, support, and gift for organization. We thank the students of Bronxville High School for their valuable classroom feedback. Finally, we thank Richard Askey for his multiple readings of the manuscripts, for correcting some embarrassing gaffes, and for making important suggestions that contributed significantly to the content of the book.

Israel M. Gelfand
Mark Saul
March 20, 2001

Chapter 0

Trigonometry

f. Gr. $\tau\rho\acute{\iota}\gamma\omega\nu o$ - ν triangle + - $\mu\epsilon\tau\rho\acute{\iota}\alpha$ measurement.

— *Oxford English Dictionary*

In this chapter we will look at some results in geometry that set the stage for a study of trigonometry.

1 What is new about trigonometry?

Two of the most basic figures studied in geometry are the triangle and the circle. Trigonometry will tell us more than we learned in geometry about each of these figures.

For example, in geometry we learn that if we know the lengths of the three sides of a triangle, then the measures of its angles are completely determined[1] (and, in fact, almost everything else about the triangle is determined). But, except for a few very special triangles, geometry does not tell us how to compute the measures of the angles, given the measures of the sides.

Example 1 The measures of the sides of a triangle are 6, 6, and 6 centimeters. What are the measures of its angles?

[1] It is sometimes said that the lengths of three sides determine a triangle, but one must be careful in thinking this way. Given three arbitrary lengths, one may or may not be able to form a triangle (they form a triangle if and only if the sum of any two of them is greater than the third). But if one can form a triangle, then the angles of that triangle are indeed determined.

Solution. The triangle has three equal sides, so its three angles are also equal. Since the sum of the angles is 180°, the degree-measure of each angle is $180/3 = 60°$. Geometry allows us to know this without actually measuring the angles, or even drawing the triangle. □

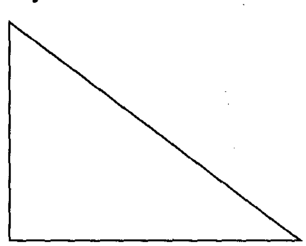
Example 2 The measures of the sides of a triangle are 5, 6, and 7 centimeters. What are the measures of its angles?

Solution. We cannot find these angle measures using geometry. The best we can do is to draw the triangle, and measure the angles with a protractor. But how will we know how accurately we have measured? We will answer this question in Chapter 3. □

Example 3 Two sides of a triangle have length 3 and 4 centimeters, and the angle between them is 90°. What are the measures of the third side, and of the other two angles?

Solution. Geometry tells us that if we know two sides and an included angle of a triangle, then we ought to be able to find the rest of its measurements. In this case, we can use the Pythagorean Theorem (see page 7) to tell us that the third side of the triangle has measure 5. But geometry will not tell us the measures of the angles. We will learn how to find them in Chapter 2. □

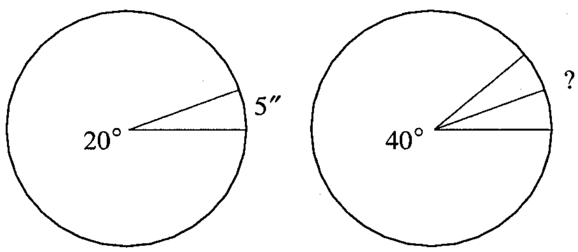
Exercise Using a protractor, measure the angles of the triangle below as accurately as you can. Do your measurements add up to 180°?



Let us now turn our attention to circles.

Example 4 In a certain circle, a central angle of 20° cuts off an arc that is 5 inches long. In the same circle, how long is the arc cut off by a central angle of 40°?

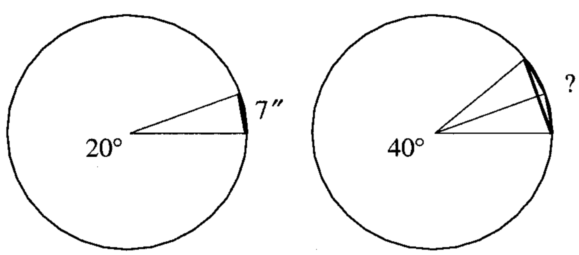
Solution. We can divide the 40°angle into two angles of 20°. Each of these angles cuts off an arc of length 5″, so the arc cut off by the 40°angle is $5 + 5 = 10$ inches long.



That is, if we double the central angle, we also double the length of the arc it intercepts. □

Example 5 In a certain circle, a central angle of 20° determines a chord that is 7 inches long. In the same circle, how long is the chord determined by a central angle of 40°?

Solution. As with Example 4, we can try to divide the 40° angle into two 20° angles:



However, it is not so easy to relate the length of the chord determined by the 40° angle to the lengths of the chords of the 20° angles. Having doubled the angle, we certainly have not doubled the chord. □

Exercises

1. In a circle, suppose we draw any central angle at all, then draw a second central angle which is larger than the first. Will the arc of the second central angle always be longer than the arc of the first? Will the chord of the second central angle also be larger than the chord of the first?

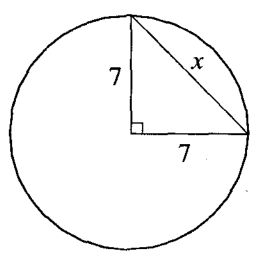
2. What theorem in geometry guarantees us that the chord of a 40° angle is less than double the chord of a 20° angle?

3. Suppose we draw any central angle, then double it. Will the chord of the double angle always be less than twice the chord of the original central angle?

Trigonometry and geometry tell us that any two equal arcs in the same circle have equal chords; that is, if we know the measurement of the arc, then the length of the chord is determined. But, except in special circumstances, geometry does not give us enough tools to calculate the length of the chord knowing the measure of the arc.

Example 6 In a circle of radius 7, how long is the chord of an arc of $90°$?

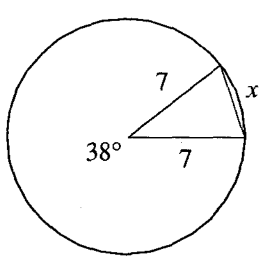
Solution. If we draw radii to the endpoints of the chord we need, we will have an isosceles right triangle:



Then we can use the Pythagorean Theorem to find the length of the chord. If this length is x, then $7^2 + 7^2 = x^2$, so that $x = \sqrt{98} = 7\sqrt{2}$. □

Example 7 In a circle of radius 7, how long is the chord of an arc of $38°$?

Solution. Geometry does not give us the tools to solve this problem. We can draw a triangle, as we did in Example 6:



But we cannot find the third side of this triangle using only geometry. However, this example does illustrate the close connection between measurements in a triangle and measurements in a circle. □

Exercises

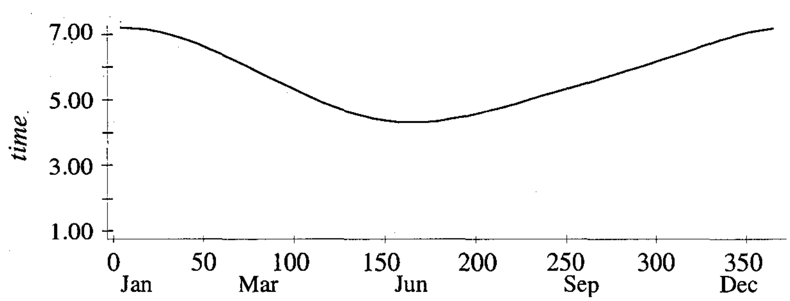
1. What theorem from geometry guarantees us that the triangle in the diagram for Example 7 is completely determined?

2. Note that the triangle in Example 7 is isosceles. Calculate the measure of the two missing angles.

Trigonometry will help us solve all these kinds of problems. However, trigonometry is more than just an extension of geometry. Applications of trigonometry abound in many branches of science.

Example 8 Look at any pendulum as it swings. If you look closely, you will see that the weight travels very slowly at either end of its path, and picks up speed as it gets towards the middle. It travels fastest during the middle of its journey. □

Example 9 The graph below shows the time of sunrise (corrected for daylight savings) at a certain latitude for Wednesdays in the year 1995. The data points have been joined by a smooth curve to make a continuous graph over the entire year.



We expect this curve to be essentially the same year after year. However, neither geometry nor algebra can give us a formula for this curve. In Chapter 8 we will show how trigonometry allows us to describe it mathematically. Trigonometry allows us to investigate any periodic phenomenon – any physical motion or change that repeats itself. □

2 Right triangles

We will start our study of trigonometry with triangles, and for a while we will consider only right triangles. Once we have understood right triangles, we will know a lot about other triangles as well.

Suppose you wanted to use e-mail to describe a triangle to your friend in another city. You know from geometry that this usually requires three pieces of information (three sides; two sides and the included angle; and so on). For a right triangle, we need only two pieces of information, since we already know that one angle measures 90°.

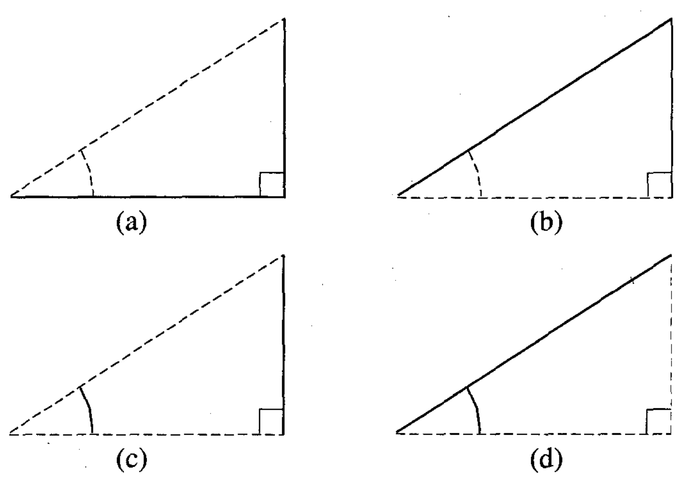
In choosing our two pieces of information, we must include at least one side, so there are four cases to discuss:

a) the lengths of the two legs;

b) the lengths of one leg and the hypotenuse;

c) the length of one leg and the measure of one acute angle;

d) the length of the hypotenuse and the measure of one acute angle.



Suppose we want to know the lengths of all the sides of the triangle. For cases (a) and (b) we need only algebra and geometry. For cases (c) and (d), however, algebraic expressions do not (usually) suffice. These cases will introduce us to trigonometry, in Chapter 1.

3 The Pythagorean theorem

We look first at the chief geometric tool which allows us to solve cases (a) and (b) above. This tool is the famous *Pythagorean Theorem.* We can separate the Pythagorean theorem into two statements:

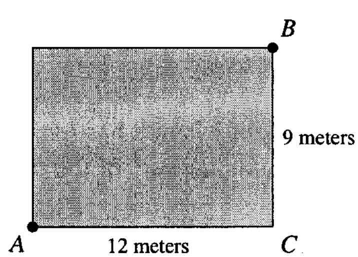
Statement I: If a and b are the lengths of the legs of a right triangle, and c is the length of its hypotenuse, then $a^2 + b^2 = c^2$.

Statement II: If the positive numbers a, b, and c satisfy $a^2 + b^2 = c^2$, then a triangle with these side lengths has a right angle opposite the side with length c.[2]

These two statements are *converses* of each other. They look similar, but a careful reading will show that they say completely different things about triangles. In the first statement, we know something about an angle of a triangle (that it is a right angle) and can conclude that a certain relationship holds among the sides. In the second statement, we know something about the sides of the triangle, and conclude something about the angles (that one of them is a right angle).

The Pythagorean theorem will allow us to reconstruct a triangle, given two legs or a leg and the hypotenuse. This is because we can find, using this information, the lengths of all three sides of the triangle. As we know from geometry, this completely determines the triangle.

Example 10 In the English university town of Oxford, there are sometimes lawns occupying rectangular lots near the intersection of two roads (see diagram).



[2]In fact, we can make a stronger statement than statement II:
Statement II$'$: If the positive numbers a, b, and c satisfy $a^2 + b^2 = c^2$, then there exists a triangle with sides a, b, and c, and this triangle has a right angle opposite the side with length c.

This statement includes, for example, the fact that if $a^2 + b^2 = c^2$, then $a + b > c$.

In such cases, professors (as well as small animals) are allowed to cut across the lawn, while students must walk around it. If the dimensions of the lawn are as shown in the diagram, how much further must the students walk than the professors in going from point A to point B?

Solution. Triangle ABC is a right triangle, so statement I of the Pythagorean theorem applies:

$$\begin{aligned} AB^2 &= AC^2 + BC^2 \\ &= 12^2 + 9^2 \\ &= 144 + 81 = 225\,. \end{aligned}$$

So $AB = 15$ meters, which is how far the professor walks.

On the other hand, the students must walk the distance $AC + CB = 12 + 9 = 21$. This is 6 meters longer than the professor's walk, or 40% longer. □

Example 11 Show that a triangle with sides 3, 4, and 5 is a right triangle

Solution. We can apply statement II to see if it is a right triangle. In fact, $5^2 = 25 = 3^2 + 4^2$, so the angle opposite the side of length 5 is a right angle. Notice that we cannot use statement I of the Pythagorean theorem to solve this problem. □

Exercises The following exercises concern the Pythagorean theorem. In solving each problem, be sure you understand which of the two statements of this theorem you are using.

1. Two legs of a right triangle measure 10 and 24 units. Find the length of the hypotenuse in the same units.
2. The hypotenuse of a right triangle has length 41 units, and one leg measures 9 units. Find the measure of the other leg.
3. Show that a triangle with sides 5, 12, and 13 is a right triangle.
4. One leg of a right triangle has length 1 unit, and the hypotenuse has length 3 units. What is the length of the other leg of the triangle?
5. The hypotenuse of an isosceles right triangle has length 1. Find the length of one of the legs of this triangle.

6. In a right triangle with a 30°angle, the hypotenuse has length 1. Find the lengths of the other two legs.

   Hint: Look at the diagram in the footnote on page 11.

7. Two points, A and B, are given in the plane. Describe the set of points X such that $AX^2 + BX^2 = AB^2$.

   (Answer: A circle with its center at the midpoint of AB.)

8. Two points, A and B, are given in the plane. Describe the set of points for which $AX^2 - BX^2$ is constant.

4 Our best friends (among right triangles)

There are a few right triangles which have a very pleasant property: their sides are all integers. We have already met the nicest of all (because its sides are small integers): the triangle with sides 3 units, 4 units and 5 units. But there are others.

Exercises

1. Show that a triangle with sides 6, 8, and 10 units is a right triangle.

2. Look at the exercises to section 3. These exercises use three more right triangles, all of whose sides are integers. Make a list of them. (Later, in Chapter 7, we will discover a way to find many more such right triangles.)

3. The legs of a right triangle are 8 and 15 units. Find the length of the hypotenuse.

4. We have used right triangles with the following sides:

Leg	Leg	Hypotenuse
3	4	5
6	8	10
9	12	15

   By continuing this pattern, find three more right triangles with integer sides.

5. We have seen that a triangle with sides 5, 12, and 13 is a right triangle. Can you find a right triangle, with integer sides, whose shortest side has length 10? length 15?

6. Exercises 4 and 5 suggest that we can construct one integer-side right triangle from another by multiplying each side by the same number (since the new triangle is *similar* to the old, it is still a right triangle). We can also reverse the process, dividing each side by the same number. Although we won't always get integers, we will always get *rational* numbers. Show that a triangle with sides 3/5, 4/5, and 1 is a right triangle.

7. Using the technique from Exercise 6, start with a 3-4-5 triangle and find a right triangle with rational sides whose shorter leg is 1. Then find a right triangle whose longer leg is 1.

8. Start with a 5-12-13 right triangle, and find a right triangle with rational sides whose hypotenuse is 1. Then find one whose shorter leg is 1. Finally, find a right triangle whose longer leg is 1.

9. Note that the right triangles with sides equal to 5, 12, 13 and 9, 12, 15 both have a leg equal to 12. Using this fact, find the area of a triangle with sides 13, 14, and 15.

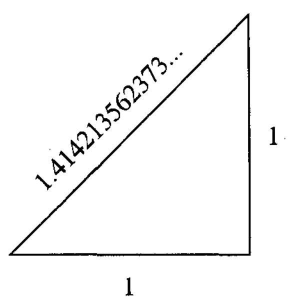
10. (a) Find the area of a triangle with sides 25, 39, 56
    (b) Find the area of a triangle with sides 25, 39, 16.

5 Our next best friends (among right triangles)

In the previous section, we explored right triangles with nice sides. We will now look at some triangles which have nice angles. For example, the two acute angles of the right triangle might be equal. Then the triangle is isosceles, and its acute angles are each 45°.

Or, we could take one acute angle to be double the other. Then the triangle has acute angles of 30 and 60°.

But nobody is perfect. It turns out that the triangles with nice angles never have nice sides. For example, in the case of the 45° right triangle, we have two equal legs, and a hypotenuse that is longer:



If we suppose the legs are each 1 unit long, then the hypotenuse, measured in the same units, is about 1.414213562373 units long, not a very nice number.

For a 30°right triangle, if the shorter leg is 1, the hypotenuse is a nice length[3]: it is 2. But the longer leg is not a nice length. It is approximately 1.732 (you can remember this number because its digits form the year in which George Washington was born – and the composer Joseph Haydn).

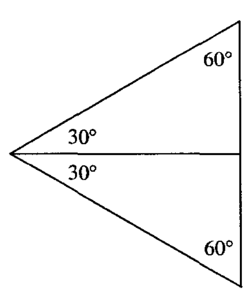
It also turns out that triangles with nice sides never have nice angles.

If we want an example of some theorem or definition, we will look at how the statement applies to our friendly triangles.

Exercises

1. Find the length of each leg of an isosceles right triangle whose hypotenuse has length 1. Challenge: Find the length, correct to nine decimal places without using your calculator (but using information contained in the text above!).
2. Using the Pythagorean theorem, find the hypotenuse of an isosceles right triangle whose legs are each three units long.

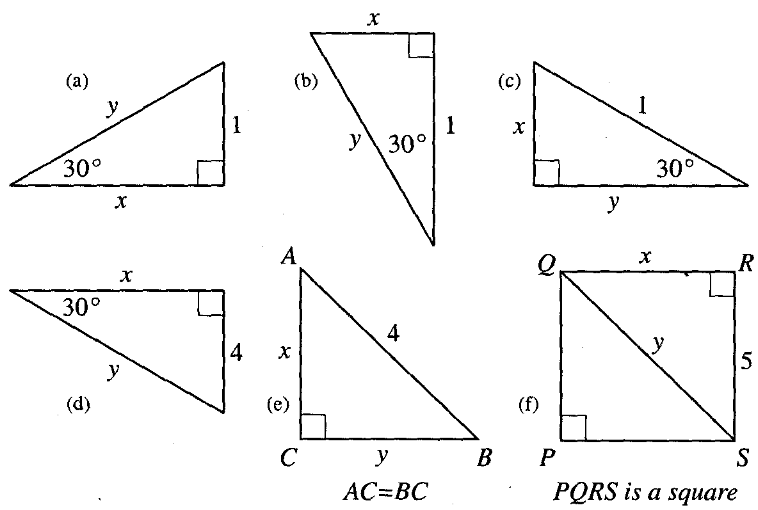
[3]If you don't remember the proof, just take two copies of such a triangle, and place them back-to-back:



You will find that they form an equilateral triangle. The side opposite the 30°angle is half of one side of this equilateral triangle, and therefore half of the hypotenuse.

3. The shorter leg of a 30-60-90°triangle is 5 units. Using the Pythagorean theorem (and the facts about a 30°-60°-90° triangle referred to above), find the lengths of the other two sides of the triangle.

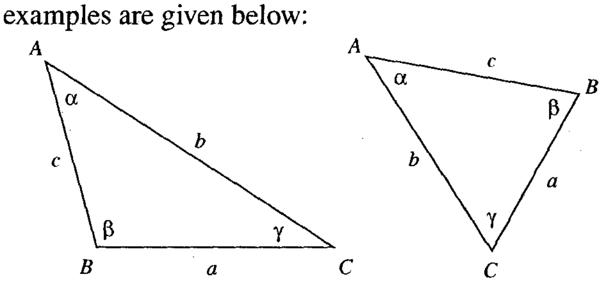
4. In each of the diagrams below, find the value of x and y:



6 Some standard notation

A triangle has *six elements* ("parts"): three sides and three angles. We will agree to use capital letters, or small Greek letters, to denote the measures of the angles of the triangle (the same letters with which we denote the vertices of the angles). To denote the lengths of the sides of the triangle, we will use the small letter corresponding to the name of the angle opposite this side.

Some examples are given below:



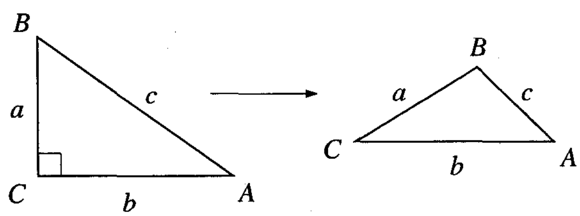
Appendix

I. Classifying triangles

Because the angles of any triangle add up to 180°, a triangle can be classified as acute (having three acute angles), right (having one right angle), or obtuse (having one obtuse angle). We know from geometry that the lengths of the sides of a triangle determine its angles. How can we tell from these side lengths whether the triangle is acute, right, or obtuse?

Statement II of the Pythagorean theorem gives us a partial answer: If the side lengths a, b, c satisfy the relationship $a^2 + b^2 = c^2$, then the triangle is a right triangle. But what if this relationship is not satisfied?

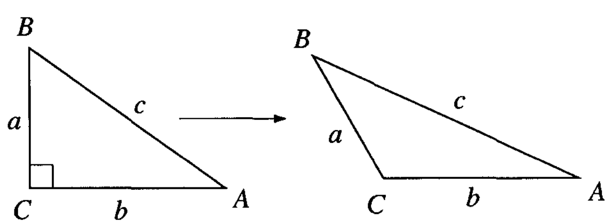
We can tell a bit more if we think of a right triangle that is "hinged" at its right angle, and whose hypotenuse can stretch (as if made of rubber). The diagrams below show such a triangle. Sides a and b are of fixed length, and the angle between them is "hinged."



As you can see, if we start with a right triangle, and "close down" the hinge, then the right angle becomes acute. When this happens, the third side (labeled c) gets smaller. In the right triangle, $c^2 = a^2 + b^2$, so we can see that:

Statement III: If angle C of $\triangle ABC$ is acute, then $c^2 < a^2 + b^2$.

In the same way, if we open the hinge up, angle C becomes obtuse, and the third side gets longer:



So we see that

Statement IV: If angle C of $\triangle ABC$ is obtuse, then $c^2 > a^2 + b^2$.

Exercise Write the converse of statements III and IV above.

While the converses of most statements require a separate proof, for these particular cases, the converses follow from the original statements. For example, if, in $\triangle ABC$, $c^2 < a^2+b^2$, then angle C cannot be right (this would contradict statement II of the Pythagorean Theorem) and cannot be obtuse (this would contradict statement IV above). So angle C must be acute, which is what the converse of statement III says.

Statements III and IV, together with their converses, allow us to decide whether a triangle is acute, right, or obtuse, just by knowing the lengths of its sides.

Some examples follow:

1. Is a triangle with side lengths 2, 3, and 4 acute, right or obtuse?

   Solution. Since $4^2 = 16 > 2^2 + 3^2 = 4 + 9 = 13$, the triangle is obtuse, with the obtuse angle opposite the side of length 4.

   Question: Why didn't we need to compare 3^2 with $2^2 + 4^2$, or 2^2 with $3^2 + 4^2$?

2. Is a triangle with sides 4, 5, 6 acute, right, or obtuse?

   Solution. We need only check the relationship between 6^2 and $4^2 + 5^2$. Since $6^2 = 36 < 4^2 + 5^2 = 41$, the triangle is acute.

3. Is the triangle with side lengths 1, 2, and 3 acute, right, or obtuse?

   Solution. We see that $3^2 = 9 > 2^2 + 1^2 = 5$, so it looks like the triangle is obtuse.

   Question: This conclusion is *incorrect.* Why?

Exercise

If a triangle is constructed with the side lengths given below, tell whether it will be acute, right, or obtuse.

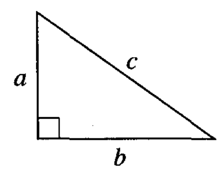
a) $\{6, 7, 8\}$ b) $\{6, 8, 10\}$ c) $\{6, 8, 9\}$ d) $\{6, 8, 11\}$
e) $\{5, 12, 12\}$ f) $\{5, 12, 14\}$ g) $\{5, 12, 17\}$

II. Proof of the Pythagorean theorem

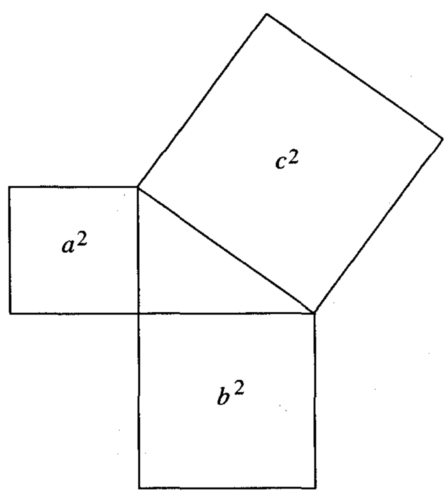
There are many proofs of this classic theorem. Our proof follows the Greek tradition, in which the squares of lengths are interpreted as areas. We first recall statement I from the text:

If a and b are the lengths of the legs of a right triangle, and c is the length of its hypotenuse, then $a^2 + b^2 = c^2$.

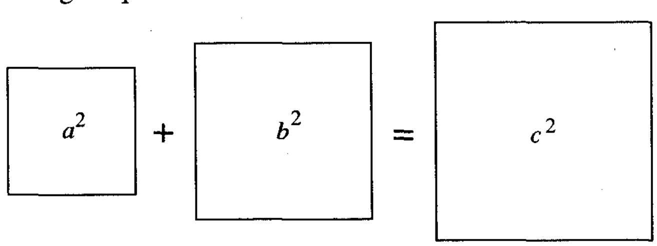
Let us start with any right triangle. The lengths of its legs are a and b, and the length of its hypotenuse is c:



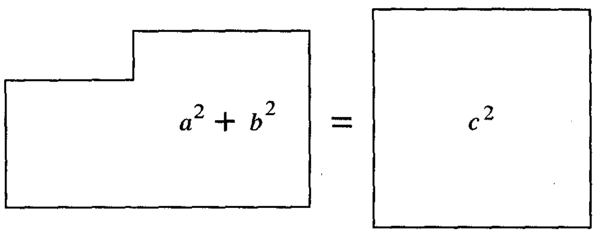
We draw a square (outside the triangle), on each side of the triangle:



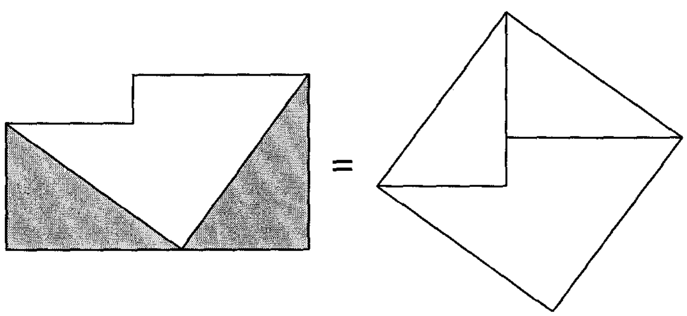
We must show that the sum of the areas of the smaller squares equals the area of the larger square:



Or:

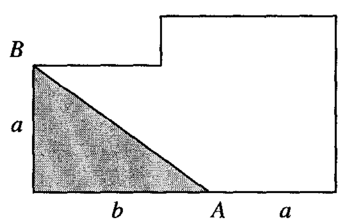


The diagram below gives the essence of the proof. If we cut off two copies of the original triangle from the first figure, and paste them in the correct niches, we get a square with side c:



We fill in some details of the proof below.

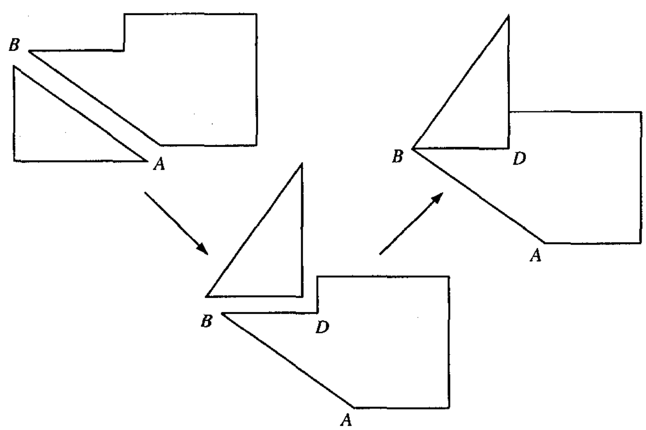
We started with an oddly shaped hexagon, created by placing two squares together. To get the shaded triangle, we lay off a line segment equal to b, starting on the lower left-hand corner. Then we draw a diagonal line. This will leave us with a copy of the original triangle in the corner of the hexagon:



(Notice that the piece remaining along the bottom side of the hexagon has length a, since the whole bottom side had length $a + b$.)

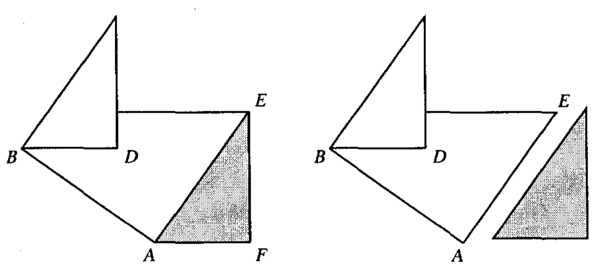
Triangle ABC is congruent to the one we started with, because it has the same two legs, and the same right angle. Therefore hypotenuse AB will have length c.

Next we cut out the copy of our original triangle, and fit it into the niche created in our diagram:

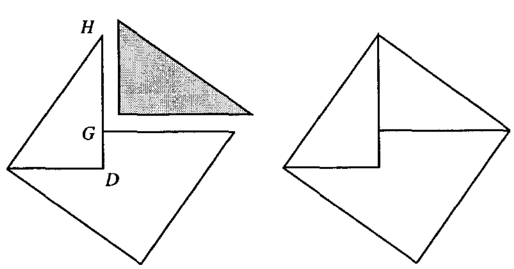


The right angle inside the triangle fits onto the right angle outside the hexagon (at D), and the leg of length a fits onto segment BD, which also has length a.

Connecting A to E, we form another triangle congruent to the original (we have already seen that $AF = a$, and $EF = b$ because each was a side of one of the original squares).

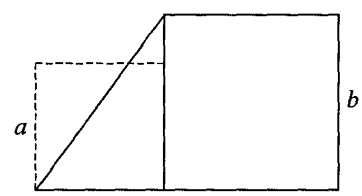


This new copy of the triangle will fit nicely in the niche created at the top of the diagram:

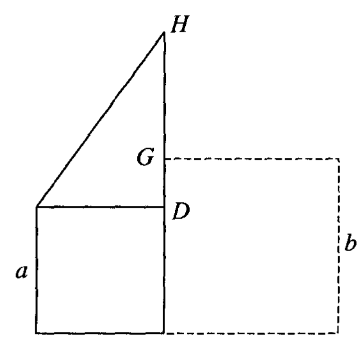


Why will it fit? The longer leg, of length b, is certainly equal to the upper side of the original hexagon. And the right angles at G must fit together. But why does GH fit with the other leg of the triangle, which is of length a?

Let us look again at the first copy of our original triangle. If we had placed it alongside the square of side b, it would have looked like this:

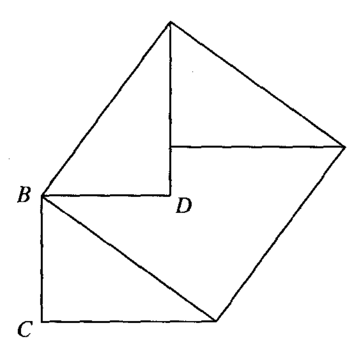


But in fact we draw it sitting on top of the smaller square, so it was pushed up vertically by an amount equal to the side of this square, which is a:



So the amount that it protrudes above point G must be equal to a. This is the length of GH, which must then fit with the smaller leg of the second copy of our triangle.

One final piece remains: why is the final figure a square? Certainly, it has four sides, all equal to length c. But why are its angles all right angles?



Let us look, for example, at vertex B. Angle CBD was originally a right angle (it was an angle of the smaller square). We took a piece of it away when we cut off our triangle, and put the same piece back when we pasted the triangle back in a different position. So the new angle, which is one in our new figure, is still a right angle. Similar arguments hold for other vertices in our figure, so it must be a square.

In fact all the pieces of our puzzle fit together, and we have transformed the figure consisting of squares with sides a and b into a square with side c. Since we have not changed the area of the figure, it must be true that a $a^2 + b^2 = c^2$.

Finally, we prove statement II of the text:

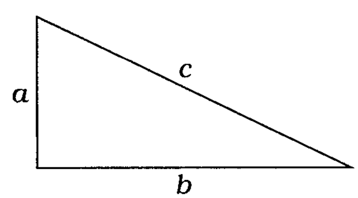
> If the positive numbers a, b, and c satisfy $a^2 + b^2 = c^2$, then a triangle with these side lengths has a right angle opposite the side with length c.

We prove this statement in two parts. First we show that the numbers a, b, and c are sides of some triangle, then we show that the triangle we've created is a right triangle.

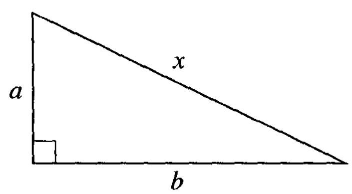
Geometry tells us that three numbers can be the sides of a triangle if and only if the sum of the smallest two of them is greater than the largest. But can we tell which of our numbers is the largest? We can, if we remember that for *positive* numbers, $p^2 > q^2$ implies that $p > q$. Since $c^2 = a^2 + b^2$, and $b^2 > 0$, we see that $c^2 > a^2$, so $c > a$. In the same way, we see that $c > b$.

Now we must show that $a + b > c$. Again, we examine the squares of our numbers. We find that $(a+b)^2 = a^2 + 2ab + b^2 > a^2 + b^2 = c^2$ (since $2ab$ is a positive number). So $a + b > c$ and segments of lengths a, b, and c form a triangle.

What kind of triangle is it? Let us draw a picture:



Does this triangle contain a right angle? We can test to see if it does by copying parts of it into a new triangle. Let us draw a new triangle with sides a and b, and a right angle between them:



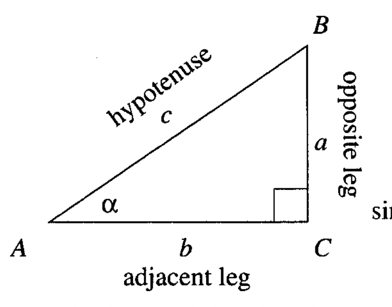
How long is the hypotenuse of this new triangle? If its length is x, then statement I of the Pythagorean theorem (which we have already proved) tells us that $x^2 = a^2 + b^2$. But this means that $x^2 = c^2$, or $x = c$. It remains to note that this new triangle, which has the same three sides as the original one, is congruent to it. Therefore the sides of length a and b in our original triangle must contain a right angle, which is what we wanted to prove.

Chapter 1

Trigonometric Ratios in a Triangle

1 Definition of $\sin \alpha$

Definition: For any acute angle α, we draw a right triangle that includes α. The *sine* of α, abbreviated $\sin \alpha$, is the ratio of the length of the leg opposite this angle to the length of the hypotenuse of the triangle.



$$\sin \alpha = \frac{a}{c} = \frac{\text{opposite leg}}{\text{hypotenuse}}$$

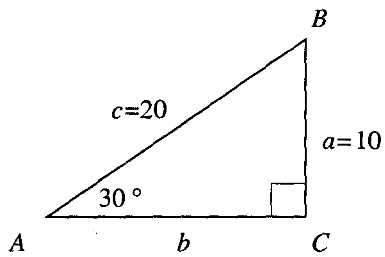
For example, in the right triangle ABC (diagram above), $\sin \alpha = a/c$. We can see immediately that this definition has a weak point: it does not tell us exactly which right triangle to draw. There are many right triangles, large ones and small ones that include a given angle α.

Let us try to answer the following questions.

Example 12 Find $\sin 30°$.

"Solution" 1. Formally, we are not obliged to solve the problem, since we are given only the measure of the angle, without a right triangle that includes it. □

Solution 2. Draw some right triangle with a 30° angle:

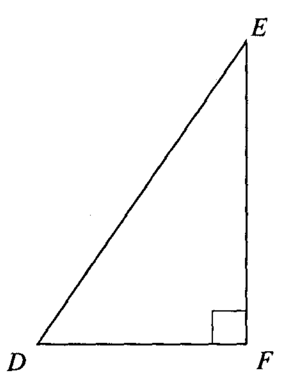


For example, we might let the length of the hypotenuse be 20. Then the length of the side opposite the 30° angle measures 10 units. So

$$\sin 30^\circ = \frac{10}{20} = \frac{1}{2} = 0.5\,.$$

We know, from geometry, that whatever the value of the hypotenuse, the side opposite the 30° angle will be half this value,[1] so $\sin 30^\circ$ will always be $1/2$. This value depends only on the measure of the angle, and not on the lengths of the sides of the particular triangle we used. □

Example 13 An American student is writing by e-mail to her friend in France, and they are doing homework together. The American student writes to the French student: "Look at page 22 of the Gelfand–Saul Trigonometry book. Let's get the sine of angle D."



The French student measured EF with his ruler, then measured ED, then took the ratio EF/ED and sent the answer to his American friend. A

[1] A theorem in geometry tells us that in a right triangle with a 30° angle the side opposite this angle is half the hypotenuse (see Chapter 0, page 11).

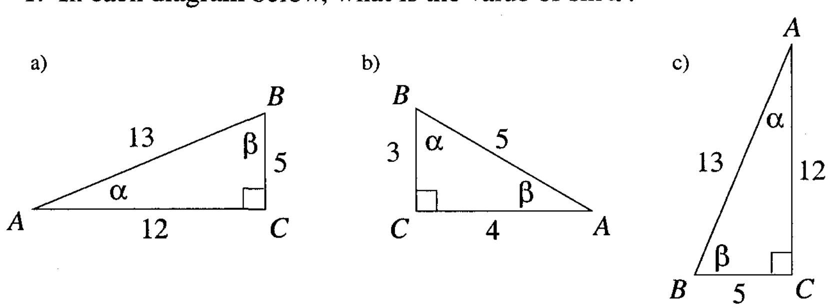
few days later, he woke up in the middle of the night and realized, "*Sacré bleu!* I forgot that Americans use inches to measure lengths, while we use centimeters. I will have to tell my friend that I gave her the wrong answer!" What must the French student do to correct his answer?

Solution. He does not have to do anything – the answer is correct. The sine of an angle is a ratio of two lengths, which does not depend on any unit of measurement. For example, if one segment is double another when measured in centimeters, it is also double the other when measured in inches. □

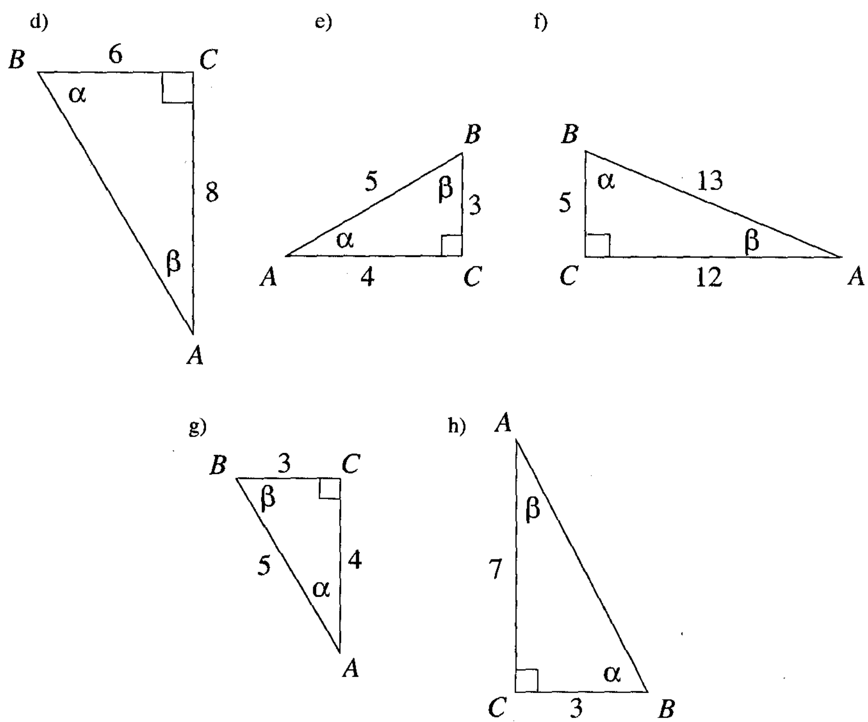
In general, for any angle of $\alpha°$ (for $0 < \alpha < 90$), the value of $\sin\alpha$ depends only on α, and not on the right triangle containing the angle.[2] This is true because any two triangles containing acute angle α are similar, so the ratios of corresponding sides are equal. $\operatorname{Sin}\alpha$ is merely a name for one of these ratios.

Exercises

1. In each diagram below, what is the value of $\sin\alpha$?



[2]Example 12 shows that the value of $\sin\alpha$ does not depend on the particular triangle which contains α. Example 13 shows that the value of $\sin\alpha$ does not depend on the unit of measurement for the sides of the triangle. In fact, we can examine Example 12 more closely. To determine the value of $\sin 30°$, we need three pieces of information: (a) the angle; (b) the right triangle containing the angle; (c) the unit of measurement for the sides of the triangle. We have just shown that the value of $\sin\alpha$ does not in fact depend on the last two pieces of information.



2. In each of the diagrams above, find sin β.

3. In the following list, cross off each number which is less than the sine of 60°. Then check your work with a calculator.

   0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

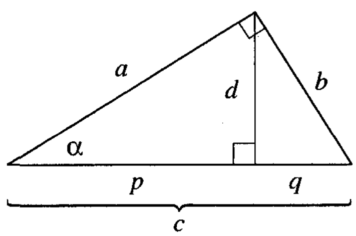
   Hint: Remember the relationships among the sides of a 30-60-90 triangle.

2 Find the hidden sine

Sometimes the sine of an angle lurks in a diagram where it is not easy to spot. The following exercises provide practice in finding ratios equal to the sine of an angle, and lead to some interesting formulas.

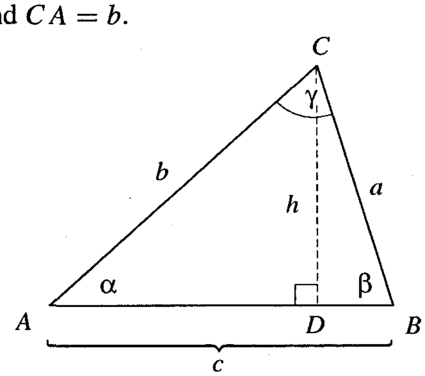
Exercises

1. The diagram below shows a right triangle with an altitude drawn to the hypotenuse. The small letters stand for the lengths of certain line segments.



   a) Find a ratio of the lengths of two segments equal to $\sin\alpha$.
   b) Find another ratio of the lengths of two segments equal to $\sin\alpha$.
   c) Find a third ratio of the lengths of two segments equal to $\sin\alpha$.

2. The three angles of triangle ABC below are acute (in particular, none of them is a right angle), and CD is the altitude to side AB. We let $CD = h$, and $CA = b$.



   a) Find a ratio equal to $\sin\alpha$.
   b) Express h in terms of $\sin\alpha$ and b.
   c) We know that the area of triangle ABC is $hc/2$. Express this area in terms of b, c, and $\sin\alpha$.
   d) Express the area of triangle ABC in terms of a, c, and $\sin\beta$.
   e) Express the length of the altitude from A to BC in terms of c and $\sin\beta$. (You may want to draw a new diagram, showing the altitude to side BC.)

3. a) Using the diagram above, write two expressions for h: one using side b and $\sin\alpha$ and one using side a and $\sin\beta$.

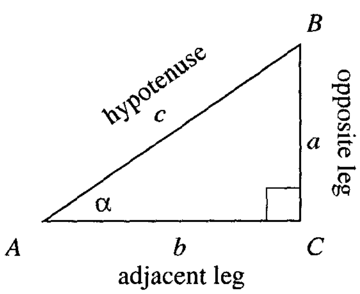
   b) Using the result to part (a), show that $a\sin\beta = b\sin\alpha$.

   c) Using the result of part (e) in problem 2 above, show that $c\sin\beta = b\sin\gamma$.

   d) Prove that $\dfrac{a}{\sin\alpha} = \dfrac{b}{\sin\beta} = \dfrac{c}{\sin\gamma}$. This relation is true for any acute triangle (and, as we will see, even for any obtuse triangle). It is called the *Law of Sines.*

3 The cosine ratio

Definition: In a right triangle with acute angle α, the ratio of the leg adjacent to angle α to the hypotenuse is called the *cosine* of angle α, abbreviated $\cos\alpha$.



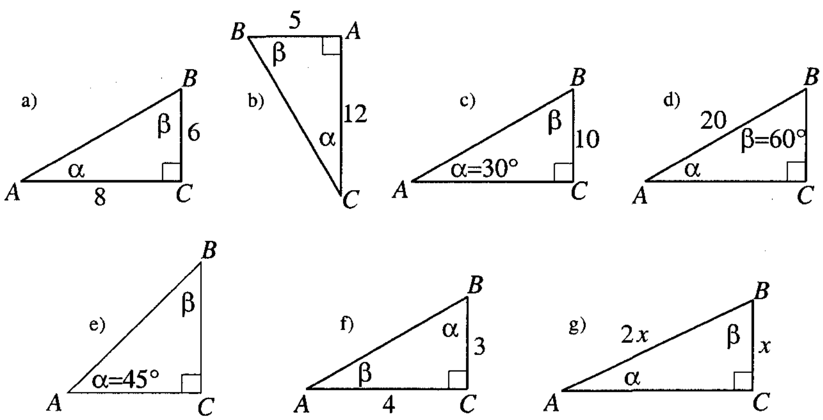
$$\cos\alpha = \frac{b}{c} = \frac{\text{adjacent leg}}{\text{hypotenuse}}$$

Notice that the value of $\cos\alpha$, like that of $\sin\alpha$, depends only on α and not on the right triangle that includes α. Any two such triangles will be similar, and the ratio $\cos\alpha$ will thus be the same in each.

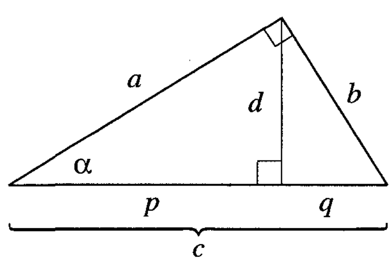
Exercises

1. Find the cosines of angles α and β in each of the triangle figures in Exercise 1 beginning on page 23.

2. Find the cosines of angles α and β in each triangle below.



3. The diagram below shows a right triangle with an altitude drawn to the hypotenuse. The small letters stand for the lengths of certain line segments.



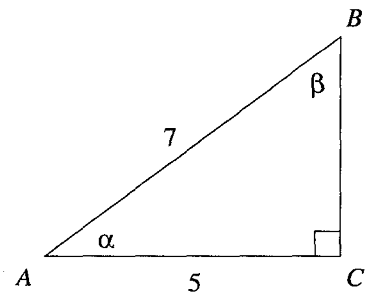
a) Find a ratio of the lengths of two segments equal to $\cos\alpha$.

b) Find another ratio of the lengths of two segments equal to $\cos\alpha$.

c) Find a third ratio of the lengths of two segments equal to $\cos\alpha$.

4 A relation between the sine and the cosine

Example 14 In the following diagram, $\cos\alpha = 5/7$. What is the numerical value of $\sin\beta$?



Solution. By the definition of the sine ratio

$$\sin\beta = \frac{AC}{AB}.$$

The value of this ratio is $5/7$, which is the same as $\cos\alpha$. □

Is this a coincidence? Certainly not. If α and β are acute angles of the same right triangle, $\sin\alpha = \cos\beta$, no matter what lengths the sides of the triangle may have. We state this as a

Theorem If $\alpha + \beta = 90°$, then $\sin\alpha = \cos\beta$ and $\cos\alpha = \sin\beta$.

Exercises

1. Show that $\sin 29° = \cos 61°$.
2. If $\sin 35° = \cos x$, what could the numerical value of x be?
3. Show that we can rewrite the theorem of the above section as: $\sin\alpha = \cos(90 - \alpha)$.

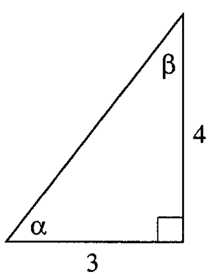
5 A bit of notation

If we are not careful, ambiguity arises in certain notation. What does $\sin x^2$ mean? Do we square the angle, then take its sine? Or do we take $\sin x$ first, then square this number? The first case is very rare: why should we want to square an angle? What units could we use to measure such a quantity?

The second case happens very often. Let us agree to write $\sin^2 x$ for $(\sin x)^2$, which is the case where we take the sine of an angle, then square the result. For example, $\sin^2 30° = 1/4$.

Exercise

In the diagram below, find the numerical value of the following expressions:



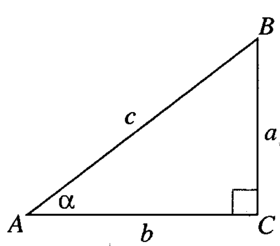
1. $\sin^2 \alpha$
2. $\sin^2 \beta$
3. $\cos^2 \alpha$
4. $\cos^2 \beta$
5. $\sin^2 \alpha + \cos^2 \alpha$
6. $\sin^2 \alpha + \cos^2 \beta$
7. $\cos^2 \alpha + \sin^2 \beta$

6 Another relation between the sine and the cosine

If you look carefully among the exercises of the previous section, you will see examples of the following result:

Theorem For any acute angle α, $\sin^2 \alpha + \cos^2 \alpha = 1$.

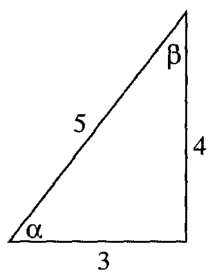
Proof As usual, we draw a right triangle that includes the angle α:



Suppose the legs have lengths a and b, and the hypotenuse has length c. Then $\sin^2 \alpha + \cos^2 \alpha = (a/c)^2 + (b/c)^2 = (a^2 + b^2)/c^2$. But the Pythagorean theorem tells us that $a^2 + b^2 = c^2$, so the last fraction is equal to 1; that is, $\sin^2 \alpha + \cos^2 \alpha = 1$. □

Exercises

1. Verify that $\sin^2\alpha + \cos^2\alpha = 1$, where α is the angle in the following diagram:



2. Did you notice that no right angle is indicated in the diagram above? Is that an error?

3. Verify that $\sin^2\beta + \cos^2\beta = 1$, where β is the other angle in the diagram.

4. Find the value of $\cos\alpha$ if α is an acute angle and $\sin\alpha = 5/13$.

5. Find the value of $\cos\alpha$ if α is an acute angle and $\sin\alpha = 5/7$.

6. If α and β are acute angles in the same right triangle, show that $\sin^2\alpha + \sin^2\beta = 1$.

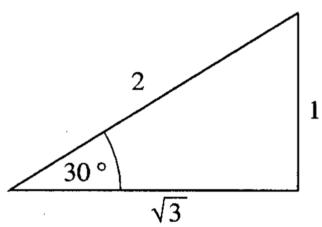
7. If α and β are acute angles in the same right triangle, show that $\cos^2\alpha + \cos^2\beta = 1$.

7 Our next best friends (and the sine ratio)

It is usually not very easy to find the sine of an angle, given its measure. But for some special angles, it is not so difficult. We have already seen that $\sin 30° = 1/2$.

Example 15 Find $\cos 30°$.

Solution. To use our definition of the cosine of an angle, we must draw a right triangle with a $30°$ angle, a triangle with which we are already friendly.



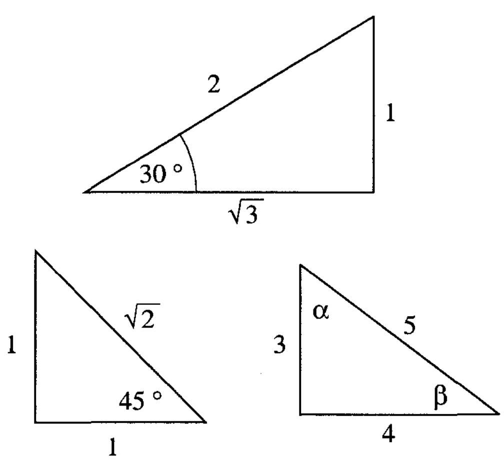
We know that the value of $\cos 30°$ depends only on the shape of this triangle, and not on its sides. So we can assume that the smaller leg has length 1. Then the lengths of the other sides are as shown in the diagram, and we see that $\cos 30° = \sqrt{3}/2$. □

Example 16 Show that $\cos 60° = \sin 30°$.

Solution. In the 30-60-90 triangle we've drawn above, one acute angle is 30°, and the other is 60°. Standing on the vertex of the 30° angle, we see that the opposite leg has length 1, and the hypotenuse has length 2. Thus $\sin 30° = 1/2$. But if we walk over to the vertex of the 60° angle, the opposite leg becomes the adjacent leg, and we see that the ratio that was $\sin 30°$ earlier is also $\cos 60°$. □

Exercises

1. Fill in the following table. You may want to use the model triangles given in the diagram below.



(The angles α and β are angles in a 3-4-5 right triangle.)

angle x	$\sin x$	$\cos x$
30°		
45°		
60°		
α		
β		

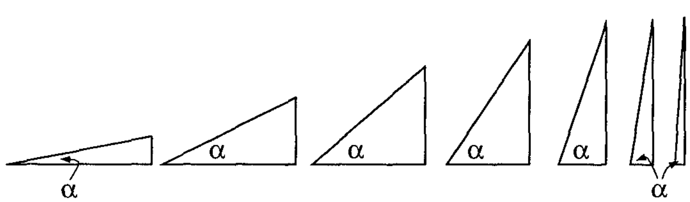
2. Verify that $\sin 60^\circ = \cos 30^\circ$.

3. Verify that $\sin^2 30^\circ + \cos^2 30^\circ = 1$.

4. Let the measure of the smaller acute angle in a 3-4-5 triangle be α. Looking at the values for $\sin\alpha$ and $\cos\alpha$, how large would you guess α is? Is it larger or smaller than 30°? Than 45°? Than 60°?

8 What is the value of $\sin 90^\circ$?

So far we have no answer to this question: We defined $\sin\alpha$ only for an acute angle. But there is a reasonable way to define $\sin 90^\circ$. The picture below shows a series of triangles with the same hypotenuse, but with different acute angles α:



As the angle α gets larger, the ratio of the opposite side to the hypotenuse approaches 1. So we make the following definition.

Definition $\sin 90^\circ = 1$.

The diagram above also suggests something else about $\sin\alpha$. Remember that the hypotenuse of a right triangle is longer than either leg. Since $\sin\alpha$ is the ratio of a leg of a right triangle to its hypotenuse, $\sin\alpha$ can never be larger than 1. So if someone tells you that, for a certain angle α, $\sin\alpha = 1.2$ or even 1.01, you can immediately tell him or her that a mistake has been made.

The same series of triangles lets us make a definition for $\cos 90°$. As the angle α gets closer and closer to $90°$, the hypotenuse remains the same length, but the adjacent leg gets shorter and shorter. This same diagram leads us to the following definition.

Definition $\cos 90° = 0$.

Exercises

1. How does the diagram above lead us to make the definition that $\sin 0° = 0$?
2. What definition does the diagram in this section suggest for $\cos 0°$?

   Answer: $\cos 0° = 1$.
3. Check, using our new definitions, that $\sin^2 0° + \cos^2 0° = 1$.
4. Check, using our new definitions, that $\sin^2 90° + \cos^2 90° = 1$.
5. Your friend tells you that he has calculated the cosine of a certain angle, and his answer is 1.02. What should you tell your friend?

9 An exploration: How large can the sum be?

We have seen that the value of the expression $\sin^2 \alpha + \cos^2 \alpha$ is always 1. Let us now look at the expression $\sin \alpha + \cos \alpha$. What values can this expression take on? This question is not a simple one, but we can start thinking about it now.

Exercises

1. We can ask our best friends for some information. Fill in the blank spaces in the following table, using a calculator when necessary .

$\sin 0^\circ + \cos 0^\circ$	$0+1$	1
$\sin 30^\circ + \cos 30^\circ$		
$\sin 45^\circ + \cos 45^\circ$		
$\sin 60^\circ + \cos 60^\circ$	$\frac{\sqrt{3}}{2} + \frac{1}{2}$	1.366 (approximately)
$\sin 90^\circ + \cos 90^\circ$		
$\sin\alpha + \cos\alpha$, where α is the smaller acute angle in a 3-4-5 right triangle	$\frac{3}{5} + \frac{4}{5}$	1.4
$\sin\alpha + \cos\alpha$, where α is the larger acute angle in a 3-4-5 right triangle		

2. Prove that $\sin\alpha + \cos\alpha$ is always less than 2.

   Hint: Geometry tells us that $\sin\alpha \leq 1$ and $\cos\alpha \leq 1$. Can they both be equal to 1 for the same angle?

3. Show that $\sin\alpha + \cos\alpha \geq 1$ for any acute angle α.

   Hint: Notice that $(\sin\alpha + \cos\alpha)^2 = 1 + 2\sin\alpha\cos\alpha$, and think about how this shows what we wanted.

4. For what value of α is $\sin\alpha + \cos\alpha = \sqrt{2}$?

5. We can see, from the table above, that $\sin\alpha + \cos\alpha$ can take on the value 1.4. Can it take the value 1.5? We will return to this problem a bit later. For now, use your calculator to see how large a value you can get for the expression $\sin\alpha + \cos\alpha$.

10 More exploration: How large can the product be?

Now let us consider the product $(\sin\alpha)(\cos\alpha)$. How large can this be?

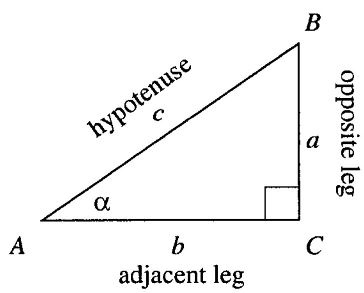
Fill in the table below:

$(\sin 0°)(\cos 0°)$	$0 \cdot 1$	0
$(\sin 30°)(\cos 30°)$		
$(\sin 45°)(\cos 45°)$		
$(\sin 60°)(\cos 60°)$		0.43301 (approximately)
$(\sin 90°)(\cos 90°)$		
$(\sin\alpha)(\cos\alpha)$, where α is the smaller acute angle in a 3-4-5 right triangle	$\frac{3}{5} \cdot \frac{4}{5}$	0.48
$(\sin\alpha)(\cos\alpha)$, where α is the larger acute angle in a 3-4-5 right triangle		

How large do you think the product $(\sin\alpha)(\cos\alpha)$ can get? We will return to this problem later on.

11 More names for ratios

In a right triangle,



we have a total of six different ratios of sides. Each of these ratios has been given a special name.

We have already given a name to two of these ratios:

$$\sin\alpha = \frac{a}{c}, \qquad \cos\alpha = \frac{b}{c}.$$

Below we give names for the remaining four ratios. The first two are very important.

The ratio

$$\frac{\text{leg opposite angle } \alpha}{\text{leg adjacent to angle } \alpha} = \frac{a}{b} \qquad \boxed{\tan\alpha = \frac{\text{opposite leg}}{\text{adjacent leg}}}$$

is called the *tangent* of α, abbreviated $\tan\alpha$.

The ratio

$$\frac{\text{leg adjacent to angle } \alpha}{\text{leg opposite angle } \alpha} = \frac{b}{a} \qquad \boxed{\cot\alpha = \frac{\text{adjacent leg}}{\text{opposite leg}}}$$

is called the *cotangent* of α, abbreviated $\cot\alpha$.

Two more ratios are used in some textbooks, but are not so important mathematically. We list them here for completeness, but will not be working with them:

The ratio

$$\frac{\text{hypotenuse}}{\text{leg adjacent to angle } \alpha} = \frac{c}{b} \qquad \boxed{\sec\alpha = \frac{\text{hypotenuse}}{\text{adjacent leg}}}$$

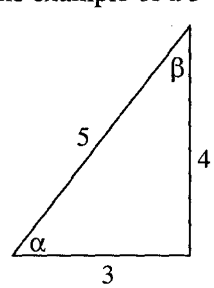
is called the *secant* of α, abbreviated $\sec\alpha$.

The ratio

$$\frac{\text{hypotenuse}}{\text{leg opposite angle } \alpha} = \frac{c}{a} \qquad \boxed{\csc\alpha = \frac{\text{hypotenuse}}{\text{opposite leg}}}$$

is called the *cosecant* of α, abbreviated $\csc\alpha$.

For practice, let's take the example of a 3-4-5 right triangle:

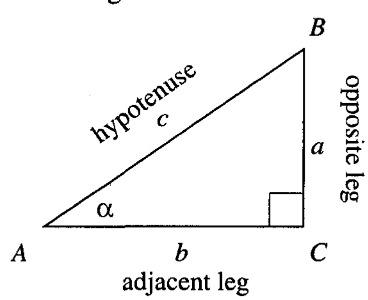


We have six ratios and six names:

$$\sin\alpha = \frac{4}{5} \qquad \cos\alpha = \frac{3}{5} \qquad \tan\alpha = \frac{4}{3}$$

$$\cot\alpha = \frac{3}{4} \qquad \sec\alpha = \frac{5}{3} \qquad \csc\alpha = \frac{5}{4}$$

To sum up, given the triangle



we have:

$\sin\alpha$	opposite leg/hypotenuse	a/c
$\cos\alpha$	adjacent leg/hypotenuse	b/c
$\tan\alpha$	opposite leg/adjacent leg	a/b
$\cot\alpha$	adjacent leg/opposite leg	b/a
$\sec\alpha$	hypotenuse/adjacent leg	c/b
$\csc\alpha$	hypotenuse/opposite leg	c/a

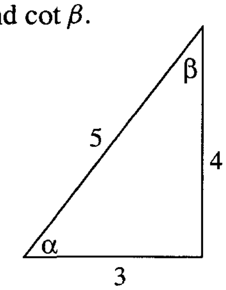
As before, these ratios depend only on the size of the angle α, and not on the lengths of the sides of the particular triangle we are using, or on how we measure the sides. The following theorem generalizes our statement of this fact for $\sin\alpha$.

Theorem The values of the trigonometric ratios of an acute angle depend only on the size of the angle itself, and not on the particular right triangle containing the angle.

Proof Any two triangles containing a given acute angle are similar, so ratios of corresponding sides are equal. The trigonometric ratios are just names for these ratios. □

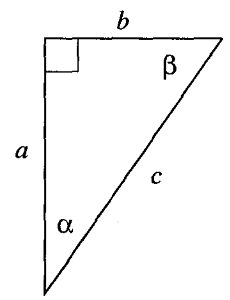
Exercises

1. For the angles in the figure below, find $\cos\alpha$, $\cos\beta$, $\sin\alpha$, $\sin\beta$, $\tan\alpha$, $\tan\beta$, $\cot\alpha$ and $\cot\beta$.

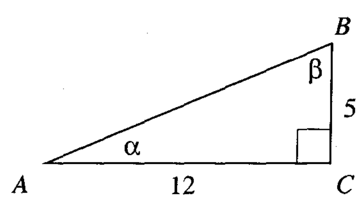


2. Did you assume that the triangle in the figure above is a right triangle? Why is this assumption correct?

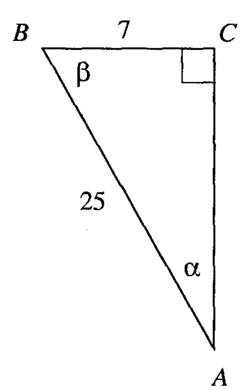
3. Given the figure below, express the quantities $\cos\alpha$, $\cos\beta$, $\sin\alpha$, $\sin\beta$, $\tan\alpha$, $\tan\beta$, $\cot\alpha$ and $\cot\beta$ in terms of a, b, and c:



4. In the diagram below, find the numerical value of $\cos\alpha$, $\cos\beta$, $\cot\alpha$ and $\cot\beta$.

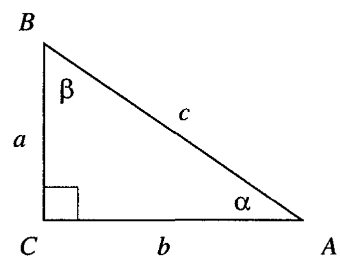


5. In the diagram below, find the numerical value of $\cos\alpha$, $\cos\beta$, $\cot\alpha$ and $\cot\beta$.



6. Find two names for each ratio given. The first example is already filled in.

$$\frac{a}{c} = \underline{\sin\alpha} = \underline{\cos\beta}$$
$$\frac{b}{c} = \underline{\qquad} = \underline{\qquad}$$
$$\frac{a}{b} = \underline{\qquad} = \underline{\qquad}$$
$$\frac{b}{a} = \underline{\qquad} = \underline{\qquad}$$



7. The sine of an angle is 3/5. What is its cosine? What is its cotangent?

8. If $\tan\alpha = 1$, what is $\cos\alpha$? What is $\cot\alpha$?

9. What is the numerical value of $\tan 45°$?

10. What is the numerical value of $\tan 30°$? Express this number using radicals. Then use a calculator to get an approximate numerical value.

11. What is the numerical value of $\tan 45° + \sin 30°$? Why don't you need a calculator to compute this?

Chapter 2

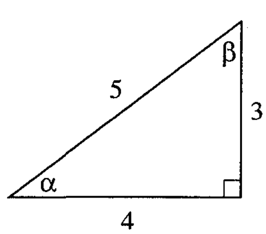
Relations among Trigonometric Ratios

1 The sine and its relatives

We have studied four different trigonometric ratios: sine, cosine, tangent, and cotangent. These four are closely related, and it will be helpful to explore their relationships. We have already seen that $\sin^2\alpha + \cos^2\alpha = 1$, for any acute angle α. The following examples introduce us to a number of other relationships.

Example 17 If $\sin\alpha = 3/5$, find the numerical value of $\cos\alpha$, $\tan\alpha$, and $\cot\alpha$.

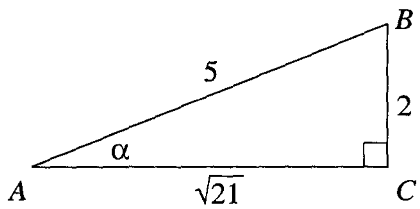
Solution. The fraction 3/5 reminds us of our best friend, the 3-4-5 triangle:



In fact, α is the measure of one of the angles in such a triangle: the one opposite the side of length 3 (see the diagram above). Having drawn this triangle, we easily see that $\cos\alpha = 4/5$, $\tan\alpha = 3/4$, and $\cot\alpha = 4/3$. □

Example 18 If $\sin\alpha = 2/5$, find the numerical value of $\cos\alpha$, $\tan\alpha$, and $\cot\alpha$.

Solution. We can again draw a right triangle with angle α:



We know only two sides of the triangle. To find $\cos\alpha$, we need the third side. The Pythagorean theorem will give this to us. Because $a^2 + b^2 = c^2$, we have $b^2 + 2^2 = 5^2$, so $b^2 = 21$ and $b = \sqrt{21}$. Now we know all the sides of this triangle, and it is clear that

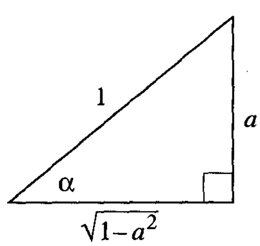
$$\cos\alpha = \frac{\sqrt{21}}{5}, \qquad \tan\alpha = \frac{2}{\sqrt{21}}, \qquad \cot\alpha = \frac{\sqrt{21}}{2}.$$

□

Example 19 We can find an answer to the question in Example 18 in a different way. For the same angle α, we have the right to draw a different right triangle, with hypotenuse 1, and leg 2/5. Do the calculation in this case for yourself. It will produce the same result. □

Example 20 If $\sin\alpha = a$, where $0 < a < 1$, express in terms of a the value of $\cos\alpha$, $\tan\alpha$, and $\cot\alpha$.

Solution. As before, we choose a right triangle with an acute angle equal to α:



The simplest is one in which the hypotenuse has length 1 and the leg opposite α is a. Let the other leg be x. Then $x^2 + a^2 = 1$, so $x = \sqrt{1 - a^2}$. Now we know all the sides of this triangle, and we can write everything down easily:

$$\cos\alpha = \sqrt{1-a^2}, \qquad \tan\alpha = \frac{a}{\sqrt{1-a^2}}, \qquad \cot\alpha = \frac{\sqrt{1-a^2}}{a}. \quad \square$$

Exercises

1. Suppose $\sin\alpha = 8/17$. Find the numerical value of $\cos\alpha$, $\tan\alpha$, and $\cot\alpha$.

2. Suppose $\cos\alpha = 3/7$. Find the numerical value of $\sin\alpha$, $\tan\alpha$, and $\cot\alpha$.

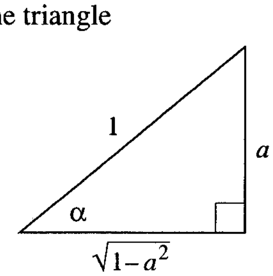
3. Suppose $\cos\alpha = b$. In terms of b, express $\sin\alpha$, $\tan\alpha$, and $\cot\alpha$.

4. Suppose $\tan\alpha = d$. In terms of d, express $\sin\alpha$, $\cos\alpha$, and $\cot\alpha$.

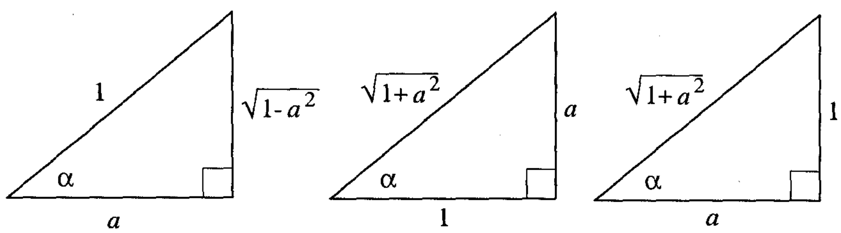
5. Fill in the following table. In each row, the value of one trigonometric function is assigned a variable. Express each of the other trigonometric functions in terms of that variable. The work for one of the rows is already done.

	$\sin\alpha$	$\cos\alpha$	$\tan\alpha$	$\cot\alpha$
$\sin\alpha$	a	$\sqrt{1-a^2}$	$\frac{a}{\sqrt{1-a^2}}$	$\frac{\sqrt{1-a^2}}{a}$
$\cos\alpha$		a		
$\tan\alpha$			a	
$\cot\alpha$				a

Please do not try to memorize this table. Its first row can be filled with the help of the triangle



whose sides can be found from the Pythagorean theorem. For the other rows, you can use use the triangles



This is all you'll ever need.

Remark. We have implicitly assumed that for every number a between 0 and 1, there exists a right triangle that contains an angle whose sine is a. But this is clear from geometry: we can construct such a triangle by taking the hypotenuse to be 1, and one leg to be a.[1]

2 Algebra or geometry?

Example 21 Suppose $\sin\alpha = 1/2$. Find the numerical value of $\cos\alpha$, $\tan\alpha$, and $\cot\alpha$.

Solution. We can do this geometrically, by drawing a triangle (as in Exercise 5 above). Or we can do this algebraically, using the results of Example 20. For instance,

$$\cos\alpha = \sqrt{1-a^2} = \sqrt{1-\left(\frac{1}{2}\right)^2} = \sqrt{1-\frac{1}{4}} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}.$$

Or did you notice right away that α is an angle in one of our friendly triangles? □

[1] You may have some objection to taking the hypotenuse of our triangle to have length 1. If you insist, we can take some length c for this hypotenuse. We will then get the same results, but the calculations will be longer. For example, suppose $\sin\alpha = x$. Choose a right triangle that contains a, and that has hypotenuse of length c. Suppose the leg opposite α has length a. Then $a/c = x$, since $\sin\alpha = x$. So $a = cx$. If we are asked for $\cos\alpha$, we can suppose the length of the other leg is b. Then $b^2 = c^2 - a^2 = c^2 - c^2x^2$, and

$$\cos a = \frac{b}{c} = \frac{\sqrt{c^2-c^2x^2}}{c} = \frac{c\sqrt{1-x^2}}{c} = \sqrt{1-x^2}.$$

Exercises

1. Find $\sin^2 30°$.

   Solution. Since $\sin 30° = 1/2$, we see that
   $$\sin^2 30° = \left(\frac{1}{2}\right)^2 = \frac{1}{4}. \qquad \square$$

2. Find $\sin^2 45°$. Check the result with a calculator.

3. Rewrite the table from Exercise 5 on page 43, but using the names of the trigonometric ratios. The first row below has been filled in as an example.

	$\sin\alpha$	$\cos\alpha$	$\tan\alpha$	$\cot\alpha$
$\sin\alpha$	$\sin\alpha$	$\sqrt{1-\sin^2\alpha}$	$\dfrac{\sin\alpha}{\sqrt{1-\sin^2\alpha}}$	$\dfrac{\sqrt{1-\sin^2\alpha}}{\sin\alpha}$
$\cos\alpha$				
$\tan\alpha$				
$\cot\alpha$				

3 A remark about names

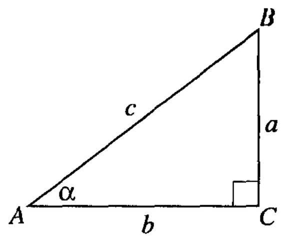
We have already seen that if α and β are complementary, then $\sin\alpha = \cos\beta$. Historically, the prefix "co-" stands for "complement," since two acute angles in the same right triangle are complementary.

The following exercises extend this note.

Exercises

Using the diagram on the right, show that if α and β are complementary, then:

1. $\tan\alpha = \cot\beta$.
2. $\cot\alpha = \tan\beta$.
3. $\sec\alpha = \csc\beta$.
4. $\csc\alpha = \sec\beta$.



4 An identity crisis?

From the table on page 45 we see, for example, that $\cos\alpha = \sqrt{1-\sin^2\alpha}$. We have also seen that, for any angle α, $\sin^2\alpha + \cos^2\alpha = 1$. Such equations, which are true for every value of the variable, are called *identities*.

From the identities we have, we can derive many more. But there is no need for anxiety. We will not have an identity crisis. If you forget all these identities, they are easily available from the three fundamental identities below:

$$\boxed{\begin{aligned} \sin^2\alpha + \cos^2\alpha &= 1 \\ \tan\alpha &= \frac{\sin\alpha}{\cos\alpha} \\ \cot\alpha &= \frac{1}{\tan\alpha} \end{aligned}}$$

From these simple identities we can derive many others involving the sine, cosine, tangent, and cotangent of a single angle. One way to derive a new identity is to draw a right triangle with an acute angle equal to α, and substitute a/c for $\sin\alpha$, b/c for $\cos\alpha$, and so on.

Example 22 Prove that $\tan\alpha = \dfrac{\sin\alpha}{\cos\alpha}$.

Solution. We can draw a right triangle with legs a, b, hypotenuse c, and acute angle α opposite the leg of length a. Then we have

$$\sin\alpha/\cos\alpha = \frac{(a/c)}{(b/c)} = (a/c)(c/b) = a/b = \tan\alpha\,.$$

□

Another way to prove a new identity is to show that it follows from other identities that we know already.

Example 23 Prove the identity $\tan\alpha \cot\alpha = 1$.

Solution. From our table, we see that $\tan\alpha = \sin\alpha/\cos\alpha$. We also see that $\cot\alpha = \cos\alpha/\sin\alpha$. Therefore,

$$\tan\alpha \cot\alpha = \left(\frac{\sin\alpha}{\cos\alpha}\right)\left(\frac{\cos\alpha}{\sin\alpha}\right) = 1. \qquad \square$$

Example 24 Show that $\tan^2\alpha + 1 = 1/\cos^2\alpha$.

Solution. We know that $\sin^2\alpha + \cos^2\alpha = 1$, so

$$\frac{\sin^2\alpha}{\cos^2\alpha} + \frac{\cos^2\alpha}{\cos^2\alpha} = \frac{1}{\cos^2\alpha},$$

or

$$\tan^2\alpha + 1 = \frac{1}{\cos^2\alpha}. \qquad \square$$

You will have a chance to practice both these techniques in the exercises below.

Exercises

1. Verify that $\sin^2\alpha + \cos^2\alpha = 1$ if α equals $30°$, $45°$, and $60°$.
2. The sine of an angle is $\sqrt{5}/4$. Express in radical form the cosine of this angle.
3. The cosine of an angle is $2/3$. Express in radical form the sine of the angle.
4. The tangent of an angle is $1/\sqrt{3}$. Find the numerical value of the sine and cosine of this angle.
5. Prove the following identities for an acute angle α:

   a) $\cot x \sin x = \cos x$.

b) $\dfrac{\tan x}{\sin x} = \dfrac{1}{\cos x}$.

c) $\cos^2 \alpha - \sin^2 \alpha = 2\cos^2 \alpha - 1$.

d) $\dfrac{\sin \alpha}{1 + \cos \alpha} = \dfrac{1 - \cos \alpha}{\sin \alpha}$.

e) $(\sin^2 \alpha + 2\cos^2 \alpha - 1)/\cot^2 \alpha = \sin^2 \alpha$.

f) $\cos^2 \alpha = 1/(1 + \tan^2 \alpha)$.

g) $\sin^2 \alpha = 1/(\cot^2 \alpha + 1)$.

h) $\dfrac{1 - \cos \alpha}{1 + \cos \alpha} = \left(\dfrac{\sin \alpha}{1 + \cos \alpha}\right)^2$.

i) $\dfrac{\sin^3 \alpha - \cos^3 \alpha}{\sin \alpha - \cos \alpha} = 1 + \sin \alpha \cos \alpha$.

6. a) For which angles α is $\sin^4 \alpha - \cos^4 \alpha > \sin^2 \alpha - \cos^2 \alpha$?

   b) For which angles α is $\sin^4 \alpha - \cos^4 \alpha \geq \sin^2 \alpha - \cos^2 \alpha$?

7. If $\tan \alpha = 2/5$, find the numerical value of $2 \sin \alpha \cos \alpha$.

8. a) If $\tan \alpha = 2/5$, find the numerical value of $\cos^2 \alpha - \sin^2 \alpha$.

   b) If $\tan \alpha = r$, write an expression in terms of r that represents the value of $\cos^2 \alpha - \sin^2 \alpha$.

9. If $\tan \alpha = 2/5$, find the numerical value of $\dfrac{\sin \alpha - 2\cos \alpha}{\cos \alpha - 3\sin \alpha}$.

10. If $\tan \alpha = 2/5$, and a, b, c, d are arbitrary rational numbers, with $5c + 2d \neq 0$, show that $\dfrac{a \sin \alpha + b \cos \alpha}{c \cos \alpha + d \sin \alpha}$ is a rational number.

11. For what value of α is the value of the expression $(\sin \alpha + \cos \alpha)^2 + (\sin \alpha - \cos \alpha)^2$ as large as possible?

5 Identities with secant and cosecant

While we do not often have to use the secant and cosecant, it is often convenient to express the fundamental identities above in terms of these two ratios. We can always restate the results as desired, using the fact that $\sec \alpha = 1/\cos \alpha$ and $\csc \alpha = 1/\sin \alpha$.

Example 25 Show that $\sec^2\alpha = 1 + \tan^2\alpha$ for any acute angle α.

Solution. We know that

$$\sec\alpha = \frac{1}{\cos\alpha},$$

so the given identity is equivalent to the statement that

$$\frac{1}{\cos^2\alpha} = 1 + \tan^2\alpha\,.$$

This last identity was proven in Example 24, page 47. □

Exercises

1. Rewrite each given identity using only sine, cosine, tangent or cotangent.

   a) $\tan\alpha\csc\alpha = \sec\alpha$.

   b) $\cot\alpha\sec\alpha = \csc\alpha$.

   c) $\dfrac{1}{\sec\alpha}\csc\alpha = \cot\alpha$.

   d) $\tan^2\alpha = (\sec\alpha + 1)(\sec\alpha - 1)$.

   e) $\csc^2\alpha = 1 + \cot^2\alpha$.

2. Rewrite in terms of secant and cosecant, tangent or cotangent. Simplify your answers so that they do not involve fractions.

   a) $\dfrac{\tan\alpha}{\sin\alpha} = \dfrac{1}{\cos\alpha}$.

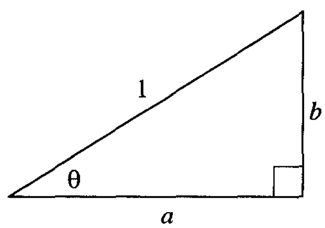
   b) $\dfrac{1}{\sin\alpha}\cos\alpha = \cot\alpha$.

   c) $\tan^2\alpha + 1 = \dfrac{1}{\cos^2\alpha}$.

   d) $\dfrac{1}{\sin^2\alpha} = 1 + \cot^2\alpha$.

6 A lemma

We have already seen that if $a = \cos\alpha$ and $b = \sin\alpha$ for some acute angle α, then $a^2 + b^2 = 1$. We can also prove the converse of this statement: If a and b are some pair of positive numbers such that $a^2 + b^2 = 1$, then there exists an angle θ such that $a = \cos\theta$ and $b = \sin\theta$. Indeed, if we draw a triangle with sides a, b, and 1, the Pythagorean theorem (statement II) guarantees us that this is a right triangle. Then the angle θ that we are looking for appears in the triangle "automatically."



Exercises

1. Suppose α is some angle less than 45°. If $a = \cos^2\alpha - \sin^2\alpha$ and $b = 2\sin\alpha\cos\alpha$, show that there is an angle θ such that $a = \cos\theta$ and $b = \sin\theta$.

2. Suppose that α is some angle. If $a = \sqrt{(1+\cos\alpha)/2}$ and $b = \sqrt{(1-\cos\alpha)/2}$, show that there is an angle θ such that $a = \cos\theta$ and $b = \sin\theta$.

3. Suppose that α is some angle. If $a = 4\cos^3\alpha - 3\cos\alpha$ and $b = 3\sin\alpha - 4\sin^3\alpha$, show that there is an angle θ such that $a = \cos\theta$ and $b = \sin\theta$.

4. Suppose that t is a number between 0 and 1. If
$$a = \frac{1-t^2}{1+t^2} \quad \text{and} \quad b = \frac{2t}{1+t^2},$$
show that there is an angle θ such that
$$a = \cos\theta \quad \text{and} \quad b = \sin\theta.$$

5. (A non-trigonometric identity) If $p^2 + q^2 = 1$, show that $(p^2 - q^2)^2 + (2pq)^2 = 1$ also. Which trigonometric identity, of those in the exercises above, is this similar to?

7 Some inequalities

You may remember from geometry that the hypotenuse is the largest side in a right triangle (since it is opposite the largest angle). So the ratio of any leg to the hypotenuse of a right triangle is less than 1. It follows that $\sin\alpha < 1$ and $\cos\alpha < 1$ for any acute angle α.

That is all the background you need to do the following exercises.

Exercises

1. For any acute angle α, show that $1 - \sin\alpha \geq 0$. For what value(s) of α do we have equality?

2. For any acute angle α, show that $1 - \cos\alpha \geq 0$. For what value(s) of α do we have equality?

3. Which of the following statements are true for all values of α?

   a) $\sin^2\alpha + \cos^2\alpha = 1$.
   b) $\sin^2\alpha + \cos^2\alpha \geq 1$.
   c) $\sin^2\alpha + \cos^2\alpha \leq 1$.
   Answer. They are all correct. Can you see why?

4. There are 4 supermarkets having a sale. Which of these are offering the same terms for their merchandise?

   - In supermarket A, everything costs no more than \$1.
   - In supermarket B, everything costs less than \$1.
   - In supermarket C, everything costs \$1 or less.
   - In supermarket D, everything costs more than \$1.

5. Which inequality is correct?

   a) For any angle α, $\sin\alpha + \cos\alpha < 2$.
   b) For any angle α, $\sin\alpha + \cos\alpha \leq 2$.

6. What is the largest possible value of $\sin\alpha$? Of $\cos\alpha$?

8 Calculators and tables

It is, in general, very difficult to get the numerical value of the sine of an angle given its degree measure. For example, how can we calculate $\sin 19°$?

One way would be to draw a right triangle with a 19° angle, and measure its sides very accurately. Then the ratio of the side opposite the angle to the hypotenuse will be the sine of 19°.

But this is not a method that mathematicians like. For one thing, it depends on the accuracy of our diagram, and of our rulers. We would like to find a way to calculate $\sin 19°$ using only arithmetic operations. Over the centuries, mathematicians have devised some very clever ways to calculate sines, cosines, and tangents of any angle without drawing triangles.

We can benefit from their labors by using a calculator. Your scientific calculator probably has a button labeled "sin," another labeled "cos," and a third labeled "tan." These give approximate values of the sine, cosine, and tangent (respectively) of various angles.

Warning: Most "nice" angles do not have nice values for sine, cosine, or tangent. The values of $\tan 61°$ or $\sin 47°$ will not be rational, and will not even be a square root or cube root of a rational number. There are a very few angles with integer degree measures and "nice" values for sine, cosine, or tangent.

Exercises

1. Find a handheld scientific calculator, and get from it the values of $\sin 30°$, $\sin 45°$, and $\sin 60°$. Compare these values with those we found in Chapter 1.

2. Betty thinks that the tangent of 60° is $\sqrt{3}$. How would you check this using a scientific calculator?

3. How would you use a calculator to get the cotangent of an angle of 30°? of 20°?

   Hint: Many calculators have a button labeled "1/x." If you press this, the display shows the reciprocal of the number previously displayed.

4. Fill in the following tables:

in radical or rational form				
α	$\sin\alpha$	$\cos\alpha$	$\tan\alpha$	$\cot\alpha$
30°				
45°				
60°				

in decimal form, from calculator				
α	$\sin\alpha$	$\cos\alpha$	$\tan\alpha$	$\cot\alpha$
30°				
45°				
60°				

9 Getting the degree measure of an angle from its sine

Example 26 What is the degree measure of the smaller acute angle of a right triangle with sides 3, 4, and 5?

Solution. We could draw a very accurate diagram, and use a very accurate protractor to answer this question. But again, mathematicians have developed methods that do not depend on the accuracy of our instruments. Your calculator uses these methods, but you must know how the buttons work.

The sine of the angle we want is $3/5 = .6$. Enter the number $.6$, then look for the button marked "arcsin" or "$\sin^{-1}$" (for some calculators, you must press this button first, then enter .6). You will find that pushing this button gives a number close to 36°. This is the angle whose sine is $.6$. □

On a calculator, you can read the symbol "arcsin" or "$\sin^{-1}$" as "the angle whose sine is ... " Similarly, "arccos" means "the angle whose cosine is ... " and "$\tan^{-1}$" means "the angle whose tangent is ... "

Exercises

1. In the text, we found an estimate for the degree-measure of the smaller acute angle in a 3-4-5 triangle. Using your calculator, find, to the nearest degree, the measure of the larger angle. Using your estimate, does the sum of the angles of such a triangle equal 180 degrees?

2. Using your calculator, find

   a) arcsin 1.

   b) arccos 0.7071067811865.

3. Using your calculator, find the angle whose cosine is .8 .

4. Using your calculator, find the angle whose sine is .6 .

5. We know that $\sin 30^\circ = .5$. Write down your estimate for $\sin 15^\circ$, then check your estimate with the value from a table or calculator.

6. Suppose $\sin x = .3$. Use your calculator to get the degree-measure of x. Now check your answer by taking the sine of the angle you found.

7. Suppose $\arcsin x = 53^\circ$. Use your calculator to get an estimate for the value of x. Now check your answer by taking the arcsin of the number you found.

8. If $\arcsin x = 60^\circ$, find x, without using a calculator.

9. Using your calculator, find arcsin ($\sin 17^\circ$).

10. Using your calculator, find sin (arcsin 0.4).

11. Find arcsin ($\sin 30^\circ$) without using your calculator. Then find sin (arcsin $1/2$), without using the calculator. Explain your results.

12. With a calculator, check that $\cos^2 A + \sin^2 A = 1$ if A equals 20° and if A equals 80°.

13. With a calculator, check that $\tan A = \sin A / \cos A$ if A equals 20° and if A equals 80°.

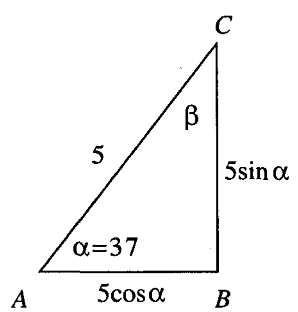
14. Using a calculator to get numerical values, draw a graph of the value of $\sin x$ as x varies from 0° to 90°.

10 Solving right triangles

Many situations in life call for the solution of problems like the following.

Example 27 The hypotenuse of a right triangle is 5, and one of its acute angles is 37 degrees. Find the other two sides.

Solution. From a calculator, we obtain $\sin 37° \approx 0.6018$ and $\cos 37° \approx 0.7986$.



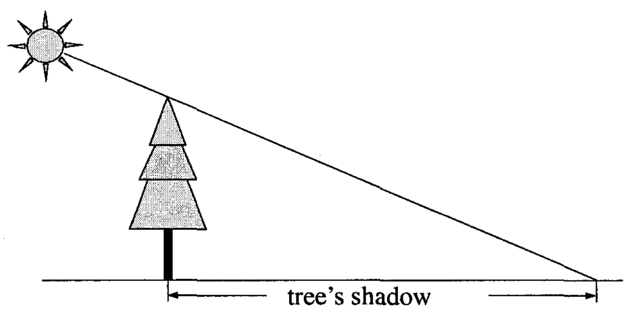
Since $\sin\alpha = BC/AC = BC/5$, we have that $BC = 5\sin 37° = 5 \times 0.6018 = 3.009$. Similarly, $AC = 5\cos 37° = 5 \times 0.7986 = 3.993$. Both values are correct to the nearest thousandth. □

Exercises

1. Find the legs of a right triangle with hypotenuse 9 and an acute angle of 72 degrees.
2. The two legs of a right triangle are 7 and 10. Find the hypotenuse and the two acute angles.
3. A right triangle has a leg of length 12. If the acute angle opposite this leg measures 27 degrees, find the other leg, the other acute angle, and the hypotenuse.
4. A right triangle has a leg of length 20. If the acute angle adjacent to this leg measures 73 degrees, find the other leg, the other acute angle, and the hypotenuse.

11 Shadows

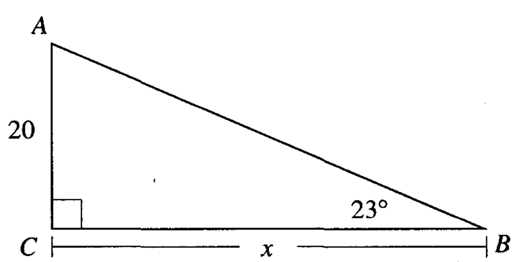
Because the sun is so far away from the earth, the rays of light that reach it from the earth are almost parallel. If we think of a small area of the earth as flat (and we usually do!), then the sun's rays strike this small region at the same angle:



So we can, for example, tell how long the shadow of an object will be, given its length.

Example 28 If the rays of the sun make a 23° angle with the ground, how long will the shadow be of a tree which is 20 feet high?

Solution. In the diagram below, AC is the tree, and BC is the shadow:



We have that

$$\tan 23^\circ = \frac{AC}{BC} = \frac{20}{x}, \text{ or } x = \frac{20}{\tan 23^\circ}.$$

From a calculator, $\tan 23^\circ \approx 0.4244$, so $x \approx 20/0.4244 \approx 47.13$ feet. □

Exercises

1. When the sun's rays make an angle with the ground of 46 degrees, how long is the shadow cast by a building 50 feet high?

2. At a certain moment, the sun's rays strike the earth at an angle of 32 degrees. At that moment, a flagpole casts a shadow which is 35 feet long. How tall is the flagpole?

3. Why are shadows longest in the morning and evening? When would you expect the length of a shadow to be the shortest?

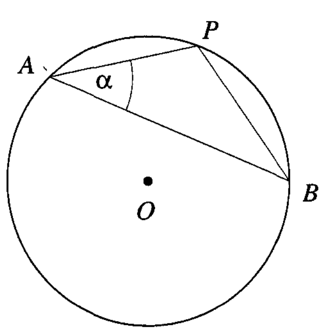
4. Can it happen that an object will not cast any shadow at all? When and where? You may need to know something about astronomy to investigate this question.

12 Another approach to the sine ratio

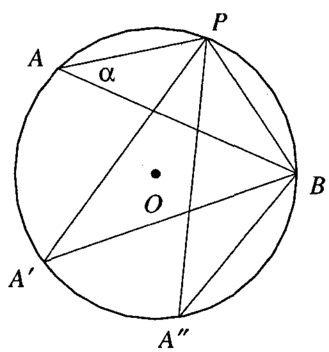
There is a simple connection between the sine of an angle and chords in a circle.

Theorem If α is the angle subtended by a chord PB at a point on a circle of radius r (such as point A in the diagram below), then

$$\sin\alpha = \frac{PB}{2r}.$$

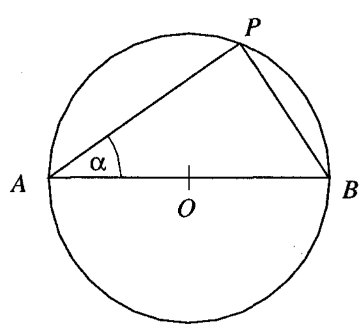


Before we prove this theorem, let us resolve a problem in the way it is stated. We can pick different points on the circle (such as A' and A'' in the figure below), and consider the different angles subtended by the same chord PB:



Does it matter which point we pick? No, it does not. An important theorem of geometry asserts that all the angles subtended by a given chord in a circle are equal[2], so that it makes no difference which point on (major) arc PB we choose.

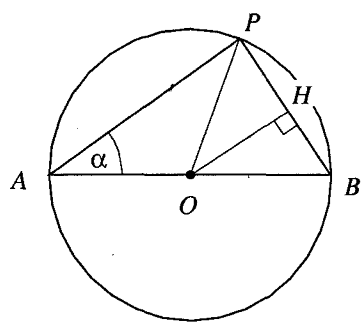
For this reason, we can prove our theorem by making a very special choice: for point A, we choose the point diametrically opposite to point B:



The same geometric theorem about inscribed angles assures us that $\angle APB$ is a right angle, so $\sin\alpha = PB/AB = PB/2r$. This completes the proof.

We can give another proof of this theorem. Consider the diagram

[2]See the theorem on inscribed angles in the appendix to this chapter.



We have shown here a perpendicular from the center of the circle O to the chord PB. Geometry tell us that the foot of this perpendicular H is the midpoint of PB.

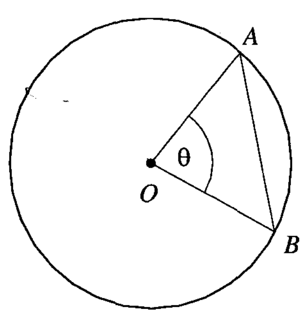
Then, as before, $\angle APB = 90°$, which implies that AP and HO are parallel. Thus, $\angle HOB = \angle PAB = \alpha$. Now in right triangle HOB, we have

$$\sin\alpha = \frac{HB}{OB} = \frac{1}{2} \cdot \frac{PB}{OB} = \frac{PB}{2r},$$

as we know from the first proof.

Exercises

1. The diagram on the left below shows a chord AB and its central angle $\angle AOB = \theta$:



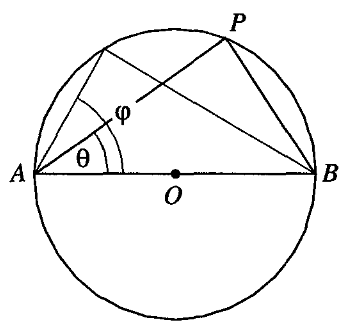
Suppose the diameter of the circle is 1. How is the length of AB related to θ?

Answer. $AB = \sin(\theta/2)$.

2. Now, using the same diagram, suppose the *radius* of the circle is 1. How is the length of AB related to θ now?

Answer. $AB = 2\sin(\theta/2)$.

3. The diagram below shows a circle of diameter 1, and two acute angles θ and φ:



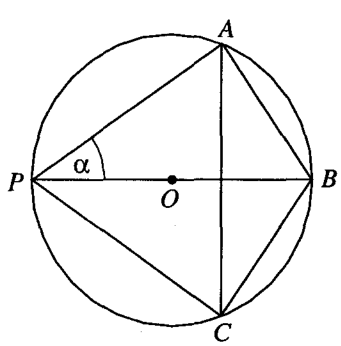
How does the diagram suggest that if $\varphi > \theta$, then $\sin\varphi > \sin\theta$?

4. We know from geometry that a circle may be drawn through the three vertices of any triangle. Find the radius of such a circle if the sides of the triangle are 6, 8, and 10.

5. Starting with an acute triangle, we can draw its circumscribed circle (the circle that passes through its three vertices). If α is any one of the angles of the triangle, show that the ratio $a : \sin\alpha$ is equal to the diameter of the circle.

6. Use Exercise 5 to show that if α, β, γ are three angles of an acute triangle, and a, b, and c are the sides opposite them respectively, then $\dfrac{a}{\sin\alpha} = \dfrac{b}{\sin\beta} = \dfrac{c}{\sin\gamma}$.

7. The diagram below shows a circle with center O, and chords AB and AC:



Arc AC is double arc AB. Diameter BP, chord AP and chord CP are drawn in, and $BP = 1$ (the diameter of the circle has unit length). If angle APB measures α degrees, use this diagram to show that $\sin 2\alpha < 2 \sin \alpha$.

You may need the theorem known as the *triangle inequality*: The sum of the lengths of any two sides of a triangle is greater than the length of the third side.

8. In a circle of diameter 10 units, how long is a chord intercepted by an inscribed angle of 60 degrees?

9. In a circle of diameter 10 units, how long is a chord intercepted by a central angle of 60 degrees?

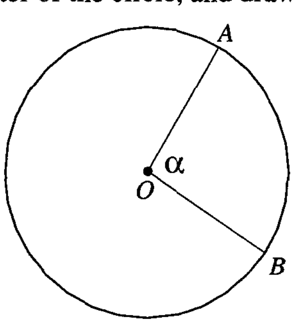
10. Find the length of a side of a square inscribed in a circle of diameter 10 units.

11. If you knew the exact numerical value of $\sin 36°$, how could you calculate the side of a regular pentagon inscribed in a circle of diameter 10?

Appendix – Review of Geometry

I. Measuring arcs

One natural way to measure an arc of a circle is to ask what portion of its circle the arc covers. We can look at the arc from the point of view of the center of the circle, and draw the central angle that cuts off the arc:



If central angle AOB measures α degrees, then we say that arc $\overset{\frown}{AB}$ measures α degrees as well.

Exercises

1. What is the degree measure of a semicircle? A quarter of a circle?
2. What is the degree measure of the arc cut off by one side of a regular pentagon inscribed in a circle? A regular hexagon? A regular octagon?

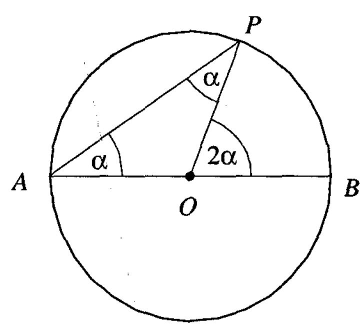
II. Inscribed angles and their arcs

An important theorem of geometry relates the degree-measure of an arc not to its central angle, but to any *inscribed* angle which intercepts that arc:

Theorem The degree measure of an inscribed angle is half the degree measure of its intercept arc.

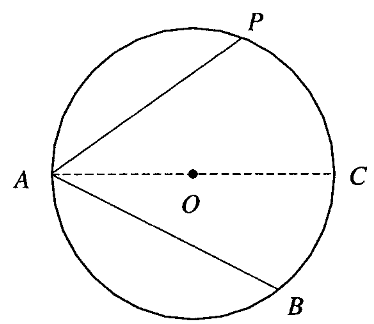
Proof We divide the proof into three cases.

1: First we prove the statement for the case in which one side of the inscribed angle is a diameter.



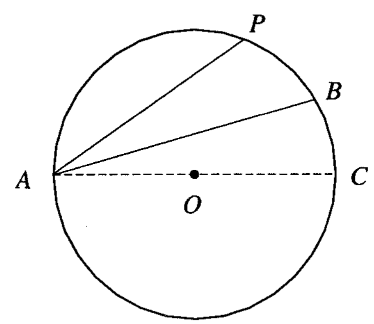
Take inscribed angle PAB, and draw PO (where O is the center of the circle). Since OP and OA are radii of the circle, they are equal, and triangle PAO is isosceles. Hence $\angle APO = \angle PAO = \alpha$. But $\angle POB$ is an exterior angle of this triangle, and so is equal to the sum of the remote interior angles, which is $\alpha + \alpha = 2\alpha$. So the degree-measure of arc $\widehat{PB}$ is also 2α, which proves the theorem for this case.

2: Suppose the center of the circle is not on one side of the inscribed angle, but inside it.



If we draw diameter AC, then angles PAC, BAC are inscribed angles covered by Case 1, so $\angle PAC = \frac{1}{2}\overset{\frown}{PC}$ and $\angle CAB = \frac{1}{2}\overset{\frown}{CB}$. Now $\angle PAB = \angle PAC + \angle CAB = \frac{1}{2}\overset{\frown}{PC} + \frac{1}{2}\overset{\frown}{CB} = \frac{1}{2}\overset{\frown}{PB}$, which is what we wanted to prove.

3: Suppose the center of the circle is outside the inscribed angle.



If we draw diameter AC, then angles PAC, BAC are inscribed angles covered by Case 1, so $\angle PAC = \frac{1}{2}\overset{\frown}{PC}$ and $\angle CAB = \frac{1}{2}\overset{\frown}{CB}$. Now $\angle PAB = \angle PAC - \angle CAB = \frac{1}{2}\overset{\frown}{PC} - \frac{1}{2}\overset{\frown}{CB} = \frac{1}{2}\overset{\frown}{PB}$, which is what we had to prove. □

As a corollary to the theorem above, we state Thales's theorem, one of the oldest mathematical results on record:

Theorem An angle inscribed in a semicircle is a right angle.

The proof is a simple application of the previous result, and is left for the reader as an exercise.

Exercises

1. If two inscribed angles intercept the same arc, show that they must be equal.

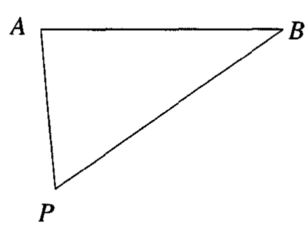
2. Find the degree-measure of an angle of a regular pentagon.

   Hint: Any regular pentagon can be inscribed in a circle.

3. If a quadrilateral is inscribed in a circle, show that its opposite angles must be supplementary.

III. ... and conversely

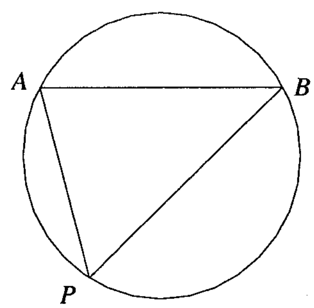
If we have a particular object, which we will represent as a line segment, we are sometimes not so much interested in how big it *is*, but how big it *looks*. We can measure this by seeing how much of our field of vision the object takes up. If we think of standing in one place and looking all around, our field of vision is 2π. The object (AB in the diagram below) is *seen* at the angle APB if you are standing at point P. We often say that AB *subtends* angle APB at point P.



For example, viewed from the earth, the angle subtended by a star is very, very small, although we know that the star is actually very large. And the angle subtended by the sun is much greater, although we know that the sun, itself a star, is not the largest one.

Suppose the angle subtended by object AB at P measures $60°$. Can we find other points at which AB subtends the same angle? From what positions does it subtend a larger angle? From what positions a smaller angle?

The answer is interesting and important. If we draw a circle through points A, B and P, then AB will subtend a $60°$ angle at any point on the circle, to one side of line AB:



Also, AB will subtend an angle greater than 60° at any point inside the circle (to one side of line AB), and will subtend an angle less than 60° at any point outside the circle.

All this follows from the converse theorem to the one in the previous section:

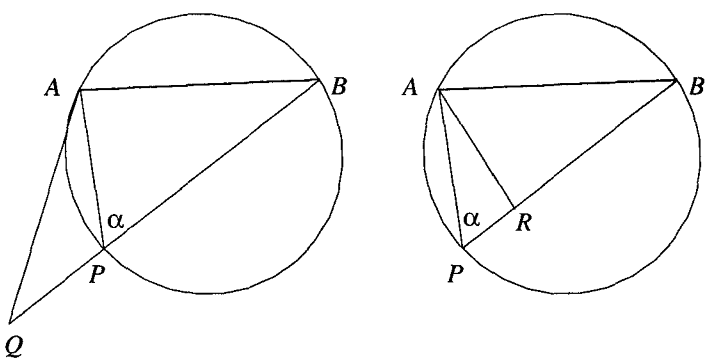
Theorem Let AB subtend a given angle at some point P. Choose another point Q on the same side of the line AB as point P. Then

- If AB subtends the same angle at point Q as at point P, then Q is on the circle through A, B and P.
- If AB subtends a greater angle at Q, then Q is inside the circle through A, P and B.
- If AB subtends a smaller angle at Q, then Q is outside the circle through A, P and B.

(Remember that Q and P must be on the same side of line AB.) The proof of this converse will emerge from the exercises below.

Exercises

1. From what points will the object AB subtend an angle of 120°?
2. From what points will the object AB subtend an angle of 90°?
3. The diagrams below show an object AB, which subtends angle α at point P. Using these diagrams below, show that if point Q is outside the circle, then AB subtends an angle less than α at point Q, and if point R is inside the circle, AB subtends an angle greater than α at point R.



4. How does Exercise 3 prove that if an angle is half of a given arc, then it is inscribed in the circle of that arc?

5. The set of points P at which object AB subtends an angle equal to α is not the whole circle, but only the arc APB. What angle does AB subtend from the other points on the circle?

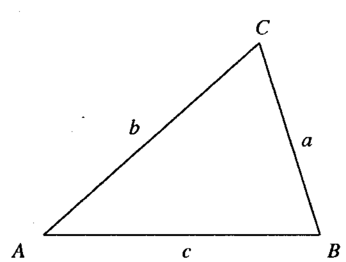
Chapter 3

Relationships in a Triangle

1 Geometry of the triangle

We would like to develop some applications of the trigonometry we've learned to geometric situations involving a triangle.

Let us work with the three sides and three angles of the triangle.



How many of these measurements do we need in order to reconstruct the triangle?[1]

This question is the subject of various "congruence theorems" in geometry. For example, if we know a, b and c (the three sides), the "SSS theorem" tells us that the three angles are determined. Any two triangles with the same three side-lengths are congruent.

But can we use any three side-lengths we like to make up our triangle? The "triangle inequality" of geometry tells us no. We must be sure that the

[1]Remember that if points A, B, and C are the vertices of a triangle, then we will also call the measures of the angles of the triangle A, B, and C. Then the length of the sides opposite angles A, B, and C are called a, b, and c, respectively. There are also other "parts" of a triangle: its area, angle bisectors, altitudes, medians, and still more interesting lines and measurements.

sum of any two of our three lengths is greater than the third; otherwise, the sides don't make a triangle. With this restriction, we can say that the three side-lengths of a triangle determine the triangle.

What other sets of measurements can determine a triangle? A little reflection will show that we will always need at least three parts (sides or angles), and various theorems from geometry will help us in answering this question.

Exercise

1. The table below gives several sets of data about a triangle. For example, "ABa" means that we are discussing two angles and the side opposite one of these angles. Some of the cases listed below are actually duplicates of others.

	Data	Determine a triangle?	Restrictions?
1	ABa		
2	ABb		
3	ABc		
4	AbC		
5	ABC		
6	Abc		
7	Bbc		
8	Cbc		

For each case, decide whether the given data determines a triangle. What restrictions must we place on the data so that a triangle can be formed? Some of these restrictions are a bit tricky. The case "abc" was discussed above.

Please do not memorize this table! We just want you to recall what geometry tells us – and what it does not tell us – about a triangle.

2 The congruence theorems and trigonometry

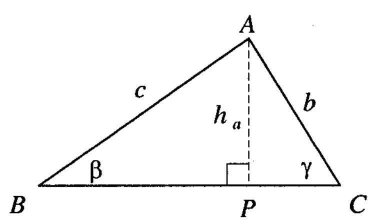
Some of the sets of data described above determine a triangle. For example, "SAS" data (the lengths of two sides of a triangle and the measure of the angle between them) always determines a triangle, and there is always a triangle which three such measurements specify.

But suppose we are given the lengths of two sides and the included angle in a triangle. How can we compute the length of the third side, or the degree-measures of the other two angles? The SAS statement of geometry doesn't tell us this. The next series of results will allow us to find missing parts of a triangle in this situation, and also in many others.

3 Sines and altitudes

A triangle has six basic elements: the three sides and the three angles. We would like to explore the relationship between these six basic elements and other elements of a triangle.

We begin with altitudes



The diagram shows triangle ABC and the altitude to side BC. We use the symbol h_a to denote the length of this altitude, since it is the height to side a in the triangle. Similarly, we use h_b and h_c to denote the altitudes to sides b and c, respectively.

We can use the sine ratio to express h_a in terms of our six basic elements. In fact, we can do this in two different ways: from right triangle ABP, we have $\sin\beta = h_a/c$, so that

$$h_a = c\sin\beta\,.$$

From right triangle ACP, we have $\sin\gamma = h_a/b$, so

$$h_a = b\sin\gamma\,.$$

We can get formulas for each of the other altitudes by replacing each side with another side and the corresponding angles. This replacement is made easier if we think of it as a "cyclic" substitution. That is, we replace:

i) a with b, b with c, and c with a, and

ii) α with β, β with γ, and γ with α.

We obtain the following two new sets of relations:

$$\begin{aligned} h_b &= c\sin\alpha &= a\sin\gamma\,, \\ h_c &= b\sin\alpha &= a\sin\beta\,. \end{aligned}$$

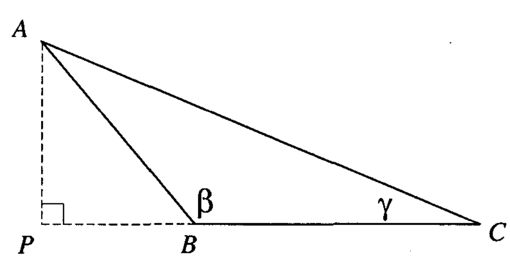
Exercises

1. By drawing diagrams showing h_b and h_c, check that these last two sets of relations are correct.
2. In triangle ABC, $\alpha = 70°$ and $b = 12$. Find h_c.
3. Check to see that the expressions for the altitudes of a triangle are correct when the triangle is right-angled. (Remember how we defined the sine of a right angle on page 32.)
4. In triangle PQR, $p = 10$, $q = 12$ and $\angle PRQ = 30°$. Find its area.

4 Obtuse triangles

If a triangle contains an obtuse angle, two of its altitudes will fall outside the triangle. In the section above, we have not taken this possibility into account. Let us now correct this oversight.

In triangle ABC below, angle β is obtuse. Let us again try to express its altitude h_a in terms of its basic elements (sides and angles). (Note that h_a lies outside the triangle.)



As before, triangle APC is a right triangle, so we have

$$AP = h_a = b\sin\gamma\,.$$

By the result in the previous section, we would also expect that

$$AP = h_a = c \sin\beta\,.$$

But we have no definition for $\sin\beta$, since β is an obtuse angle.

We can remedy the situation by looking at right triangle APB, in which $AB = c$. We find that $AP = AB \times \sin(\angle ABP)$, or

$$h_a = c \sin(180^\circ - \beta)\,.$$

This formula is a little bit cumbersome, so we take a rather daring step. We *define* $\sin\beta$ to be the same as $\sin(180^\circ - \beta)$.

In fact, we make the following general agreement:

Definition The sine of an obtuse angle is equal to the sine of its supplement.

Then we can write $h_a = c\sin\beta$ even when β is an obtuse angle. The remaining relations in such a case will follow from our rule for cyclic substitution, which still holds.

As we will see, this definition is convenient, not just to obtain this formula, but for other applications of trigonometry as well.

Exercise

1. Check to see that our new definition allows us to write

$$\begin{aligned} h_b &= c\sin\alpha &= a\sin\gamma\,, \\ h_c &= b\sin\alpha &= a\sin\beta\,, \end{aligned}$$

as a cyclic substitution would produce.

5 The Law of Sines

In a triangle, we have two expressions for h_c:

$$h_c = a\sin\beta = b\sin\alpha\,.$$

We obtain an interesting relationship if we divide the last two equal quantities by the product $\sin\alpha\sin\beta$:

$$\frac{a\sin\beta}{\sin\alpha\sin\beta} = \frac{b\sin\alpha}{\sin\alpha\sin\beta}\,,$$

or

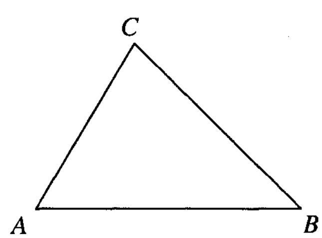
$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta}.$$

And we can get corresponding proportions from the other pairs of sides by making the same cyclic substitution as before. We obtain

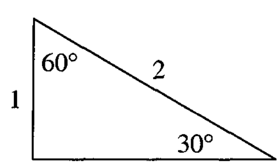
$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma}.$$

This is a very important relationship among the sides and angles of a triangle. It is known as the *Law of Sines*.

This formula has many interesting connections. For example, you may have learned in geometry that if two sides of a triangle are unequal, then the greater side lies opposite the greater angle: If $\beta < \alpha$ then $AC < BC$.



But it is *not* true that if angle β is double the angle α, then side BC is double the side AC. This is shown clearly in the figure below, with a 30-60-90 triangle. As we know, there is a side double the smallest, but it's not the one opposite the 60 degree angle. It's the hypotenuse.



The Law of Sines generalizes correctly the fact that the greater side lies opposite the greater angle, because it tells us that the ratio of two sides of a triangle is the ratio of the sines of the opposite angles. And, as we have seen, the sines of two angles are not in the same ratio as their degree-measures.

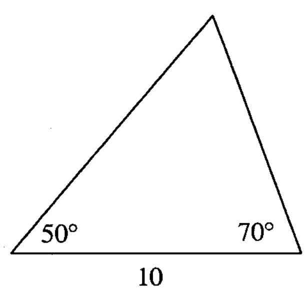
The Law of Sines can help us in another way too, which we mentioned at the start of this chapter. We know from geometry that two triangles are congruent if two pairs of corresponding angles are equal, and a pair of corresponding sides are equal (in many textbooks, this is called ASA or SAA, depending on whether the side is between or outside of the two angles). Another way to say this is to assert that the measure of two angles and one side determines the triangle. Geometry shows us one way to get the third angle (using the fact that the three angles of a triangle sum to 180°). But geometric methods do not let us compute the lengths of the other two sides. The law of sines allows us to do this.

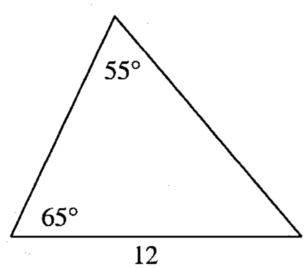
Exercises

1. Verify that the cyclic substitutions give the equalities shown above.

2. Check that the Law of Sines holds in a 30-60-90 triangle.

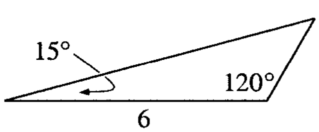
3. Use the Law of Sines in the triangles below to determine the lengths of the missing sides. (Use your calculator for the computations.)

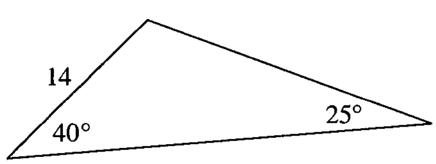




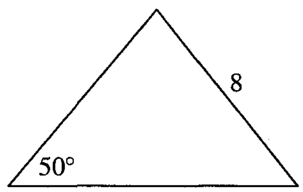
4. We have defined the sine of an obtuse angle as equal to the sine of its supplement. With this definition, show that the law of sines is true for an obtuse triangle.

5. Use the Law of Sines in the triangles below to determine the lengths of the missing sides.





6. Use the Law of Sines to find the two missing angles in the triangle below:



7. Recall from geometry that SSA does not guarantee congruence. That is, if two triangles match in two sides and an angle not included between these two sides, then the triangles may not be congruent. Look back at Problem 6. Is the triangle determined uniquely? How many possible values are there for the degree measurements of the remaining angles?

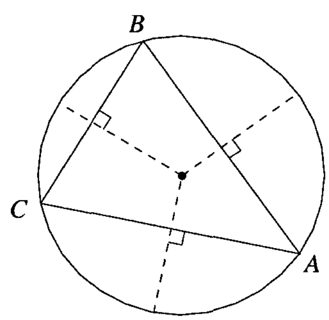
8. Suppose triangle ABC is inscribed in a circle of radius R. Prove the *extended Law of Sines*:

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} = 2R\,.$$

6 The circumradius

We can learn more about the Law of Sines another way if we give a geometric interpretation of the ratio $a/\sin\alpha$ in any triangle ABC.

We construct the circle circumscribing the triangle:[2]



Suppose the radius of this circle (the *circumradius* of the triangle) is R. We know from the result on page 57 that

$$BC = a = 2R\sin\alpha\,.$$

[2]Recall that the perpendicular bisectors of the three sides of a triangle coincide at a point equidistant from all three vertices. This point is the center of the triangle's circumscribed circle.

So the ratio $a/\sin\alpha$ is simply equal to $2R$.

Exercises

1. Find the circumradius of a triangle in which a 30° angle lies opposite a side of length 10 units. Note that this information does not determine the triangle.

2. Find the circumradius of a 30-60-90 triangle with hypotenuse 8. Do you really need the result of this section to find this circumradius?

7 Area of a triangle

Our altitude formulas have given us one interesting result: the Law of Sines. We now show how they lead to a new formula for the area of a triangle. But in fact, the formula we present is not really new. It is just the usual formula from geometry, written in trigonometric form.

If S denotes the area of a triangle, we know that

$$S = \frac{1}{2}ah_a\,.$$

But $h_a = b\sin\gamma$, so we can write

$$S = \frac{1}{2}ab\sin\gamma\,.$$

This is our "new" formula. As with our other formulas, we can use "cyclic substitutions" (see page 69) to get two more formulas:

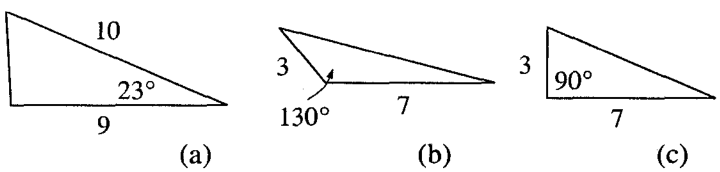
$$S = \frac{1}{2}bc\sin\alpha$$

$$S = \frac{1}{2}ca\sin\beta\,.$$

Exercises

1. Find the area of a triangle in which two sides of length 8 and 11 include an angle of 40° between them.

2. Find the areas of the triangles shown:



Can you use our new formula for part (c)? Is it necessary to use this formula?

3. The area of triangle ABC is 40. If side AB is 6 and angle A is 40 degrees, find the length of side AC.

4. In triangle PQR, side $PQ = 5$, and side $PR = 6$. If the area of the triangle is 9, find the degree-measure of angle P.

   Hint: There are two possible answers. Can you find them both?

5. Two sides of a triangle are a and b. What is the largest area the triangle can have? What is the shape of the triangle with largest area?

   Answer: The largest area is $ab/2$, achieved when the angle between the two sides is a right angle.

   Challenge: There is another right triangle with sides a and b. Find this triangle and its area.

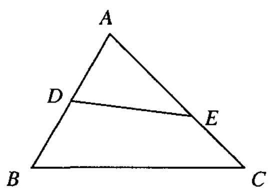
6. The length of a leg of an isosceles triangle is x. Express in terms of x the largest possible area the triangle can have.

7. Show that the area of a parallelogram is $ab \sin C$, where a and b are two adjacent sides and C is one of the angles. Does it matter which angle we use?

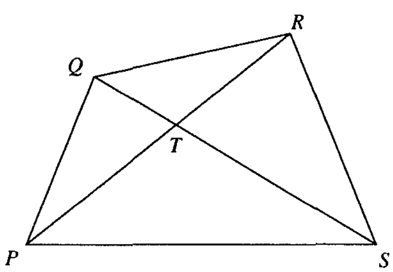
8. We start with any quadrilateral whose diagonals are contained inside the figure. Show that the area of the quadrilateral is equal to half the product of the diagonals times the sine of the angle between the diagonals. Should we take the acute angle formed by the diagonals, or the obtuse angle?

9. Show that we can use the same formula to get the area of a quadrilateral whose diagonals (when extended) intersect outside the figure.

10. In the figure, $AD = 4$, $AE = 6$, $AB = 8$, $AC = 10$. Find the ratio of the area of triangle ADE to that of triangle ABC.



11. In quadrilateral $PQRS$, diagonals PR and QS intersect at point T. The sum of the areas of triangles PQT and RST is equal to the sum of the areas of triangles PST and QRT. Show that T is the midpoint of (at least) one of the quadrilateral's diagonals.



Solution. The sines of angles PTQ, QTR, RTS, STP are all equal. If this sine is s, and using absolute value for area, we have $|PQT| + |RST| = (1/2)PT \times QT \times s + (1/2)ST \times RT \times s = |QRT| + |PTS| = (1/2)QT \times RT \times s + (1/2)PT \times ST \times s$, so $PT \times QT + RT \times ST = QT \times RT + PT \times ST$, or $PT \times QT + RT \times ST - QT \times RT - PT \times ST = 0$, or $(PT - RT)(QT - ST) = 0$. But this means that one of the factors must be zero, so that T is the midpoint of at least one of the diagonals. □

12. In quadrilateral $ABCD$, diagonals AC and BD meet at point P. Again using absolute value for area, show that $|APB| \times |CPD| = |BPC| \times |DPA|$. Is this true if the intersection point of the diagonals is outside the quadrilateral?

13. In acute triangle ABC, show that $c = a\cos B + b\cos A$.

Hint: Draw the altitude to side c. How must we change this result if angle A or angle B is obtuse?

8 Two remarks

Remark 1: Note that these formulas express the area of a triangle in terms of two sides and an included angle. We knew from geometry that these three pieces of information determine the triangle (and therefore its area), but we need trigonometry to actually compute the area. We will see in the next section how trigonometry allows us to compute other elements of a triangle determined by two sides and their include angle.

Remark 2: We now have three ways to think of $\sin\alpha$ geometrically:

1. In a right triangle with an acute angle α, $\sin\alpha$ is the ratio of the leg opposite α to the hypotenuse:
$$\sin\alpha = \frac{\text{opposite leg}}{\text{hypotenuse}}.$$
2. In any triangle, $\sin\alpha$ is the ratio of the side opposite α to the diameter of the circumscribed circle:
$$\sin\alpha = \frac{a}{2R}.$$
3. In any triangle, $\sin\alpha$ is the ratio of twice the area to the product of the two sides which include α:
$$\sin\alpha = \frac{2S}{bc}.$$

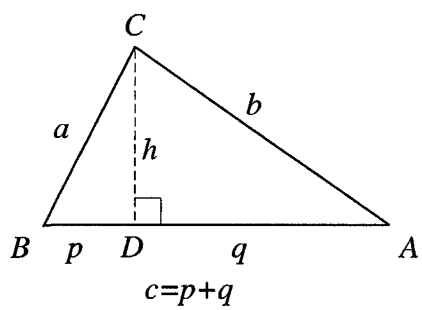
We can use whichever fits the situation we are working in. Indeed, it turns out that any of these could actually function as the definition of $\sin\alpha$.

9 Law of cosines

The law of cosines is a very old theorem. It appears in Euclid's *Elements*, the very first textbook of geometry, although Euclid does not use the term cosine. It is a generalization of the Pythagorean theorem.

In triangle ABC, if angle B is an acute angle, then
$$b^2 = a^2 + c^2 - 2ac\cos B.$$

In fact, this is not difficult to prove:



From right triangle BDC, we have $p = a\cos B$. Using the Pythagorean theorem twice, in triangles ACD and BCD, we have $b^2 = h^2 + q^2 = a^2 - p^2 + (c-p)^2 = a^2 + c^2 - 2pc = a^2 + c^2 - 2ca\cos B$, which is what we wanted to prove.

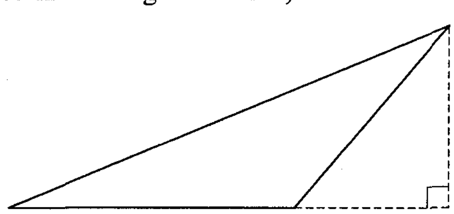
The Pythagorean theorem says that the square of a side of a triangle opposite a right angle is equal to the sum of the squares of the other two sides.

One of the ways in which the law of cosines generalizes the Pythagorean theorem is by showing that the square of a side of a triangle opposite an acute angle is less than the sum of the squares of the other two sides.

What if we take the side of a triangle opposite an obtuse angle?

Exercise Show that if b is a side opposite an obtuse angle of a triangle, then $b^2 = a^2 + c^2 + 2ac\cos B'$, where B' is the measure of the supplement of obtuse angle B.

(A hint is contained in the diagram below.)



On the basis of this result, we make a second daring definition (to follow our daring definition of the sine of an obtuse angle):

Definition The cosine of an obtuse angle is the cosine of its supplement, multiplied by -1.

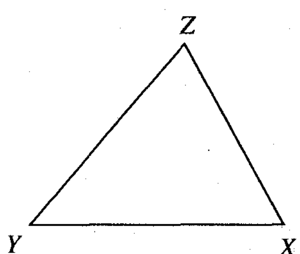
So we have three results: the Pythagorean theorem for a right angle, and the two new results for an acute and an obtuse angle. Just as with the sine function, we can make all these results into a single formula, the so-called *Law of Cosines*: In triangle ABC,

$$b^2 = a^2 + c^2 - 2ac\cos B\,.$$

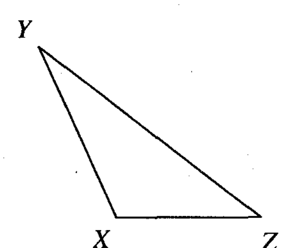
Exercises

1. Check to see that this is correct, whether angle B is acute, right, or obtuse.

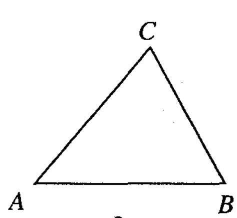
2. In each of the triangles below, use the Law of Cosines to express the square of the indicated side in terms of the other two sides and their included angle:



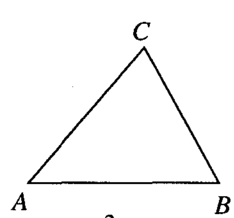
$x^2 =$



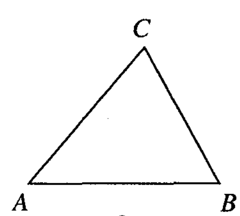
$x^2 =$



$a^2 =$



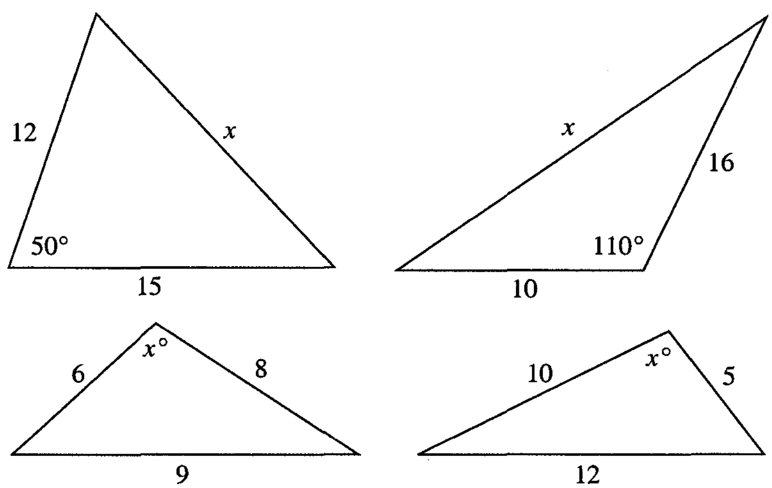
$b^2 =$



$c^2 =$

3. We know, from geometry, that a triangle is determined by SAS (the lengths of two sides and the angle between them). Explain how the Law of Cosines allows us to calculate the missing parts of a triangle, if we are given SAS.

4. Find the side or angle marked x in each diagram below:



5. In triangle ABC, $AB = 10$, $AC = 7$, and $BC = 6$. Find the measures of each angle of the triangle.

6. Peter's teacher gave the following problem:

   > A parallelogram has sides 3 and 12. Find the sum of the squares of its diagonals.

   But Peter had trouble even drawing the diagram. He knew that opposite sides of a parallelogram are equal, so he knew where to put the numbers 3 and 12. But then he didn't know what kind of parallelogram to draw. He drew a rectangle (which, he knew, is a kind of parallelogram). Then he drew a parallelogram with a 30° angle, and another parallelogram with a 60° angle. But he didn't know which one to use to do the computation.

   Can you help Peter out?

7. Show that the sum of the squares of the sides of any parallelogram is equal to the sum of the squares of the diagonals.

8. If M is the midpoint of side BC in triangle ABC, then AM is called a median of triangle ABC. Show that for median AM, $4AM^2 = 2AB^2 + 2AC^2 - BC^2$.

   Hint: The diagram for this problem is "half" of the diagram for Exercise 7 above.

9. Show that the sum of the squares of the three medians of a triangle is 3/4 the sum of the squares of its sides.

10. The diagonals of quadrilateral $ABCD$ intersect inside the figure. Show that the sum of the squares of the sides of the quadrilateral is equal to the sum of the squares of its diagonals, plus four times the length of the line segment connecting the midpoints of the diagonals (notice that this generalizes problem 6).

11. In triangle ABC, angle C measures 60 degrees, $a = 1$ and $b = 4$. Find the length of side c.

12. In triangle ABC, angle C measures 60 degrees. Show that $c^2 = a^2 + b^2 - ab$. What is the corresponding result for triangles in which angle C measures 120 degrees?

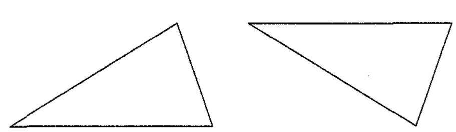
13. Three riders on horseback start from a point X and travel along three different roads. The roads form three 120° angles at point X. The first rider travels at a speed of 60 MPH, the second at a speed of 40 MPH, and the third at a speed of 20 MPH. How far apart is each pair of riders after 1 hour? After 2 hours?

Appendix – Three big ideas and how we can use them

I. Invariants: Motions in the plane

We often talk about the congruence of triangles. Two triangles are congruent if one can be moved so that it fits exactly on the other. So we can say that two congruent triangles are exactly the same, except for their position.

The two triangles below cannot be considered congruent if we confine our motions to the plane. To move one of them onto the other, we must flip it around (reflect it in a line) before we can make it fit. These triangles are *mirror images* of each other.



Exercises

1. Most triangles cannot be placed on their mirror images without reflecting them in a line. However, there are certain special triangles that can be placed onto their mirror images without using reflections at all. Draw one such triangle.

2. Describe the set of all triangles that can be placed onto their mirror images without reflection in a line.

3. Draw some quadrilaterals that can be placed onto their mirror images without reflection in a line.

I.1 Triangle invariants

A triangle invariant is a quantity associated with a triangle that is unaffected by its position. Thus the value of a triangle invariant for any two congruent triangles will be equal. Some examples of triangle invariants are the lengths of the sides, the measures of the angles, and the area. While these seven invariants are basic, there are many others (such as the lengths of the altitudes, or the radius of the circumscribed circle).

When we work with the relationships among triangle invariants, we are connecting the geometry of the triangle with algebra and trigonometry. A mathematician would say that algebra and trigonometry are the *analytical tools* of geometry.

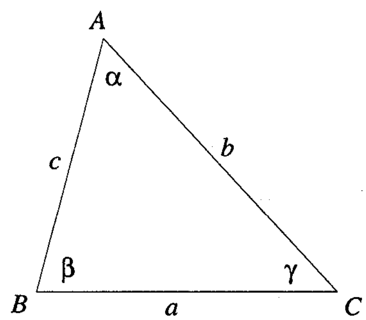
I.2 The sine and triangle invariants

We can give an alternative definition of the sine of an angle in terms of triangle invariants. Indeed, we have already seen how.

We have seen (Chapter 3, section 7) that for any triangle,

$$S = \frac{1}{2}ab\sin\gamma$$

where γ is the angle included between two sides of lengths a and b. So if we start with any angle γ, draw a triangle (not necessarily a right triangle!) including it, and denote by a and b the measures of the sides surrounding γ, then we can define $\sin\gamma$ as equal to $2S/ab$, where S is the area of the triangle.



Whether you think of this statement as a consequence of our original development, or as a definition of the sine of an angle, is your choice. In either case, we have the following:

$$\sin\alpha = \frac{2S}{bc}, \quad \sin\beta = \frac{2S}{ac}, \quad \sin\gamma = \frac{2S}{ab}.$$

Exercises

1. Note that if α is a right angle in the diagram above, then the area of $\triangle ABC$ is $bc/2$. Show that in this case, the formulas for $\sin\beta$ and $\sin\gamma$ given above are just what were given in Chapter 1.

2. Using the law of cosines, show that in any triangle ABC of area S,

$$c^2 = a^2 + b^2 - 4S\cot\gamma.$$

II. Symmetry

Let us look once more at the formula which expresses the area of a triangle s in terms of two sides and the sine of the included angle. We already know that there are three of these formulas:

$$S = \frac{1}{2}ab\sin\gamma\,, \quad S = \frac{1}{2}bc\sin\alpha\,, \quad S = \frac{1}{2}ca\sin\beta\,.$$

Each of them is obtained from the others by *cyclic substitutions* of a, b and c, and α, β and γ, respectively.

In general, when we have a formula for a triangle, we can expect this sort of symmetry. No one of the sides and angles plays a special role with respect to the others, so if in the formula we perform a cyclic substitution of them, we should get a valid formula as well.

We can also look at this process in reverse. We can consider the three equal quantities:

$$ab \sin \gamma = bc \sin \alpha = ca \sin \beta .$$

When we have three symmetric expressions that are equal, sometimes we can find a geometric reason why they are all equal. In this case, they are all equal to twice the area of the triangle.

In applying cyclic substitutions, we must be sure that each variable in our formula can represent *any* side or angle in a triangle. For instance, the Pythagorean theorem says that if a and b are the legs of a right triangle, and c the hypotenuse, then $a^2 + b^2 = c^2$. We cannot subsitute a for b, b for c, and c for a, because c cannot be any side of the triangle: it must be the hypotenuse. (However we can substitute a for b and b for a.)

Exercises

1. The law of cosines says that

$$c^2 = a^2 + b^2 - 2ab \cos \gamma .$$

   Using cyclic substitutions, write down two more formulas like this one.

2. The law of sines says that

$$\frac{a}{\sin \alpha} = \frac{b}{\sin \beta} = \frac{c}{\sin \gamma} .$$

   Can you find a geometric reason why these three quantities are equal? **Hint:** See page 74.

III. The sine and its dimension

Physicists often deal with dimensions as well as numbers, since the numbers they use are often the result of some measurement. For example, sides of triangles are measured in units of length (such as centimeters), while their areas are measured in units of length squared (such as square centimeters). We can borrow this idea from the physicist, by noting that in an algebraic or trigonometric identity, both sides should have the same dimension. For example, in the Pythagorean theorem, the dimension of both sides is length squared, the same dimension as areas. And in fact, our proof of this theorem interpreted it as a statement about areas (as does the proof in Euclid's *Elements*).

What is the dimension of $\sin\alpha$? As we originally defined it, $\sin\alpha$ is the ratio of two lengths, so in fact its dimension is 0. (This is another way of saying that the unit of length used to measure the sides of a triangle does not affect the value of the sines of its angles.)

Let us check that the dimensions are correct in our new formulas. We have written $\sin\alpha = 2S/bc$. Now S has dimension length squared, and the product bc has the same dimension (length times length), so the dimensions cancel out, and $\sin\alpha$ has dimension 0. This agrees with our previous result.

Exercise

1. Check that the dimensions of each side are the same in the following formulas:

   a) $\dfrac{a}{\sin\alpha} = \dfrac{b}{\sin\beta}$.

   b) $\dfrac{a}{b} = \dfrac{\sin\alpha}{\sin\beta}$.

   c) $S = \frac{1}{2}ab\sin\gamma$.

   d) $c^2 = a^2 + b^2 - 2ab\cos\gamma$.

IV. Hero's formula

We know that any two triangles with the same three side lengths are congruent. This means that they will give the same value for any triangle invariant, such as the area. That is, the lengths of the sides of a triangle *determine* its area.

There is a wonderful formula, credited to Hero (or Heron) of Alexandria, which expresses the area of a triangle in terms of the lengths of its sides. If these lengths are a, b and c, and if $s = (a+b+c)/2$, we have that

$$S = \sqrt{s(s-a)(s-b)(s-c)}\,.$$

Let us prove this formula.

We know that $\sin\gamma = 2S/ab$, so that

$$\sin^2\gamma = \frac{4S^2}{a^2b^2}\,.$$

From the law of cosines, we have

$$c^2 = a^2 + b^2 - 2ab\cos\gamma\,,$$

or

$$\cos\gamma = \frac{a^2 + b^2 - c^2}{2ab}\,.$$

Hence,

$$\cos^2\gamma = \frac{(a^2 + b^2 - c^2)^2}{4a^2b^2}\,.$$

Finally, we remember that $\sin^2\gamma + \cos^2\gamma = 1$, and now we have all that we need. If we substitute the results above for $\sin^2\gamma + \cos^2\gamma$, we will have a relationship that includes only S, a, b, and c, just what we want.

Indeed,

$$\sin^2\gamma + \cos^2\gamma = \frac{4S^2}{a^2b^2} + \frac{(a^2 + b^2 - c^2)^2}{4a^2b^2} = 1\,,$$

or

$$16S^2 + (a^2 + b^2 - c^2)^2 = 4a^2b^2\,,$$

or

$$16S^2 = 4a^2b^2 - (a^2 + b^2 - c^2)^2\,.$$

This is the relationship we need, but it doesn't look very "nice." In particular, it doesn't look symmetric in a, b and c.

But in fact it is. We can show this by factoring the right-hand side as the difference of two squares:

$$\begin{aligned} 16S^2 &= 4a^2b^2 - (a^2 + b^2 - c^2)^2 \\ &= \big(2ab + (a^2 + b^2 - c^2)\big)\big(2ab - (a^2 + b^2 - c^2)\big)\,, \end{aligned}$$

and so

$$\begin{aligned} 16S^2 &= \big(2ab + (a^2 + b^2 - c^2)\big)\big(2ab - (a^2 + b^2 - c^2)\big) \\ &= (a^2 + 2ab + b^2 - c^2)(-a^2 + 2ab - b^2 + c^2)\,. \end{aligned}$$

Each of the factors above on the right is again the difference of two squares:

$$16S^2 = \big((a+b)^2 - c^2\big)\big(c^2 - (a-b)^2\big)\,,$$

so we can factor once more:

$$\begin{aligned} 16S^2 &= ((a+b)+c)((a+b)-c)(c+(a-b))(c-(a-b)) \\ &= (a+b+c)(a+b-c)(a-b+c)(-a+b+c), \end{aligned}$$

and we now see the beautiful symmetry of the expression. We can write this relationship as

$$S^2 = \frac{(a+b+c)(a+b-c)(c-a+b)(c+a-b)}{16},$$

or

$$S = \frac{\sqrt{(a+b+c)(a+b-c)(c-a+b)(c+a-b)}}{4}.$$

This formula is nice, but it can be made even nicer if we set $s = (a+b+c)/2$. Then we have:

$$\begin{aligned} a+b-c &= 2s-2c, \\ c-a+b &= 2s-2a, \\ c+a-b &= 2s-2b. \end{aligned}$$

Substituting these results into the formula above, we then obtain

$$S = \frac{\sqrt{(2s)2(s-a)2(s-b)2(s-c)}}{4} = \sqrt{s(s-a)(s-b)(s-c)}.$$

Exercises

1. Show that Hero's formula gives the correct value for the area of a triangle with sides 3, 4 and 5.

2. Show that Hero's formula gives the correct value for the area of a triangle with sides 5, 12 and 13.

3. Using Hero's formula, show that the area of an equilateral triangle with side of length l is given by $l^2\sqrt{3}/4$.

4. Show that the formula $S = \frac{1}{2}ab\sin\gamma$ also leads to the formula in Problem 3 above.

5. Use Hero's formula to solve Problems 9 and 10 on page 10.

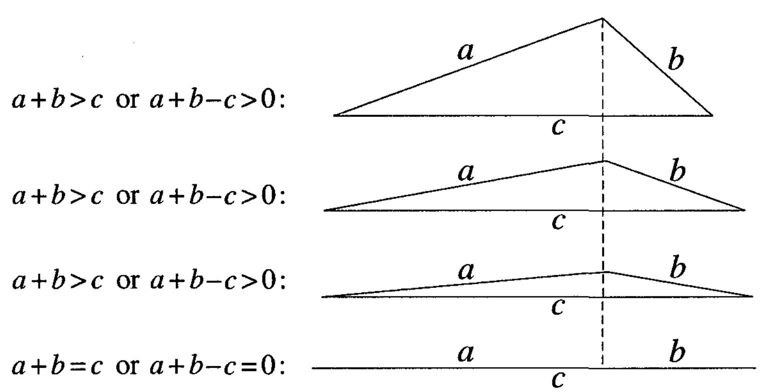
V. A physicist's interpretation

Hero's formula may seem strange. In most area formulas, you multiply just two quantities together, but here we multiply four quantities together. We make up for it by taking a square root, but this is also unusual for an area formula.

To help us make a bit more sense of this formula, we can imagine Richard Feynman, a Nobel Laureate in Physics, who was very skilled at explaining subtle ideas simply. He might have explained Hero's formula in the following way:

In high school I had a very good course in geometry, and I remember studying Hero's formula, which relates the lengths of the side of a triangle to its area. But I've forgotten its exact form. Let's see what I can recall. I know it had a square root in it.[3] *Now the dimension of the area is length squared, so under the square root we must have a polynomial of degree four.*[4] *We can get such a polynomial by multiplying together four factors, each of degree* 1.

What could these factors be? Well, if $a + b = c$, *then our* triangle *is actually a line segment (as the triangle inequality tell us), which has area* 0. *So when* $a + b - c = 0$, *the whole polynomial is zero. This means that* $a+b-c$ *is a factor of the polynomial. Similarly,* $a-b+c$ *must be a factor, and so must be* $-a + b + c$.



[3]In fact, Feynman's mathematician friends could explain why there must be a square root in the formula. The explanation involves attaching a sign to the triangle's area, depending on the orientation of the triangle.

[4]For a polynomial of several variables, the degree of each term is the sum of the exponents of all the variables that appear in it, and the degree of the polynomial is the highest of the degrees of its terms.

So we have three of the four factors of the polynomial under the square root. What can the fourth factor be? It must be linear in a, b *and* c, *and it must be symmetric in these three variables.*[5] *This means that the fourth factor must be of the form* $k(a+b+c)$, *for some constant* k.

So we must have

$$S = C\sqrt{(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}\,,$$

for some constant C. *We can determine this constant by examining one particular triangle, and I remember that a triangle with sides* 3, 4 *and* 5 *is a right triangle. The area of this triangle is* 6, *and the expression*

$$C\sqrt{(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}$$

has the value $C\sqrt{(12)(2)(4)(6)} = C \times 24$ *for this particular triangle. Therefore,* $C = 6/24 = 1/4$. *I also remember that we can clean this up algebraically by introducing* $s = (a+b+c)/2$, *but I will leave this to my friends the mathematicians. And now that I've had fun figuring out what the formula must be, I also leave to them the actual proof. They are good at that.*

[5] What Feynman would mean here is that if we interchange any two of the variables a, b and c, the value of the polynomial would be unaffected.

Chapter 4

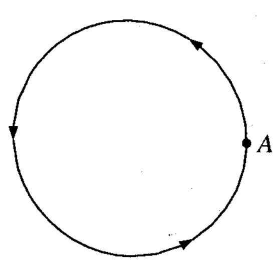
Angles and Rotations

1 Measuring rotations

In previous chapters we explored the meaning of expressions such as $\sin 30°$, $\cos 45°$ and $\tan 60°$. In this chapter and the next we show how we can use expressions such as $\sin 180°$, $\tan 300°$ or even $\sin 1000°$.

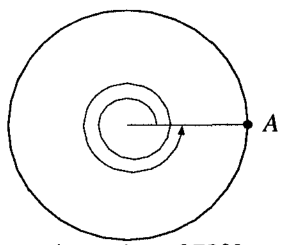
But what might $1000°$ measure? Certainly it is not the measure of the angle of a triangle. These can only be between $0°$ and $180°$ (acute, right or obtuse). Nor can it be the measure of an angle (or an arc) in a circle. These can only be between $0°$ and $360°$.

If you have ever owned toy electric trains, you may have set up the tracks in a circle, and run the trains around the circle. The diagram below shows a circular track. If a train starts at point A, travels around the circle, and arrives back at point A, we say that it has made one full rotation around the circle.



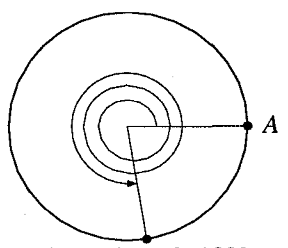
Since we divide a circle into 360 degrees, it is natural to say that the train has rotated around the circle by 360 degrees.

Now suppose the train continues past point A, and travels around the circle again. Then we can say that it has rotated through more than 360 degrees. If it travels around the circle twice, returning to point A, we say that it has rotated $360 + 360 = 720$ degrees.



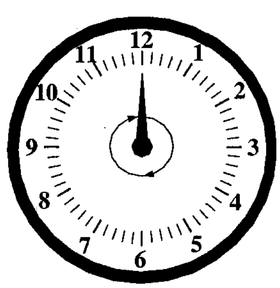
A rotation of 720°

And if it travels a bit further around the circle, along an arc measuring 280°, we say that it has rotated $720° + 280° = 1000°$:



A rotation of 1000°

Here is another example. Look at the hour hand of a clock. In 12 hours it has made a full rotation, or rotated by 360°.



But this time the rotation is clockwise (by definition!), while our train was rotating counterclockwise. In a plane, there are two different directions of rotation, and it turns out to be important to distinguish between them. Mathematicians call a counterclockwise rotation *positive* and a clockwise rotation *negative*. So we say that in 12 hours, the hour hand of a clock performs a rotation of $-360°$.

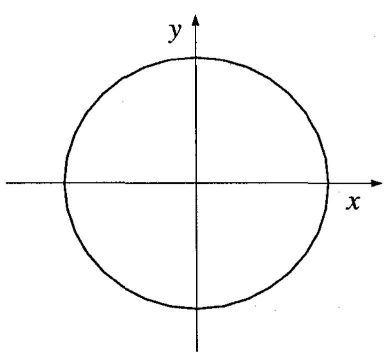
Exercise

1. Draw diagrams showing the following rotations:

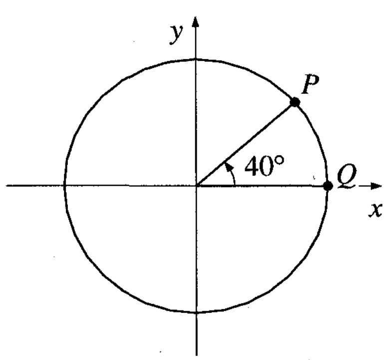
a) $160°$	b) $190°$	c) $400°$
d) $600°$	e) $1200°$	f) $-70°$
g) $-400°$	h) $360°$	i) $-270°$

2 Rotation and angles

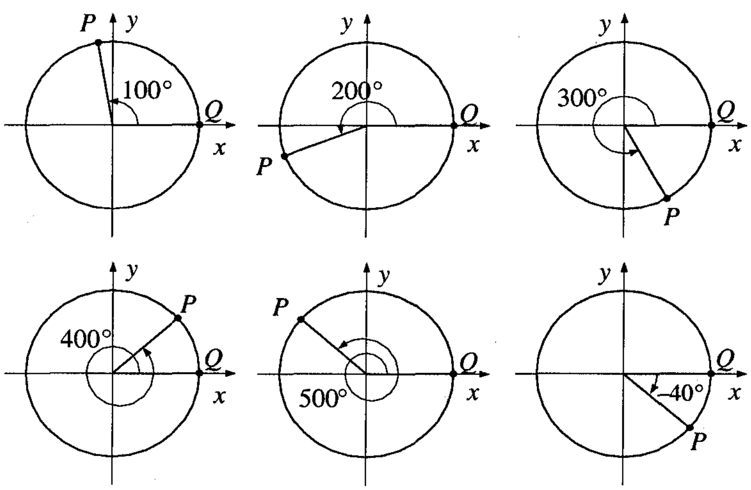
Picture a circle of radius 1, with its center at the origin of a system of coordinates:



We take an acute angle with one leg along the x-axis. The other leg will end up someplace in the first quadrant. If the measure of this acute angle is, say, $40°$, then we can get from point Q to point P by rotating through $40°$.



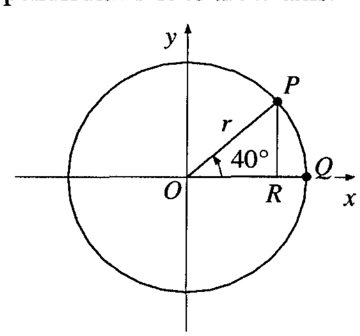
So we can associate angles with rotations. Even if the rotation exceeds 180°, we sometimes talk about the "angle" instead of the "rotation." The figure below gives some examples.



3 Trigonometric functions for all angles

Let us look again at what we mean by $\sin 40°$. We will do so in such a way that it will help us understand what is meant by $\sin 300°$, $\cos 1100°$, or $\tan(-240°)$.

We draw a circle of radius r centered at the origin of coordinates. To find $\sin 40°$, we mark the point P in the first quadrant such that $\angle POR = 40°$, and drop the perpendicular PR to the x-axis:



From right triangle POR, we see that

$$\sin 40° = \frac{PR}{OP} = \frac{PR}{r}.$$

Similarly, we can write

$$\cos 40^\circ = \frac{OR}{r}.$$

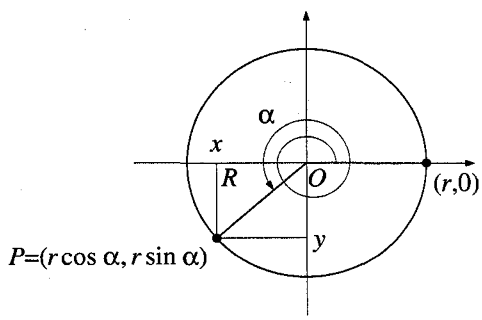
But if the coordinates of P are (x, y), we see that $y = PR$ and $x = OR$. So we can write

$$\sin 40^\circ = \frac{y}{r}, \quad \cos 40^\circ = \frac{x}{r}.$$

So far we have said nothing new.

Or have we?

We can use this observation to extend our definitions of sine and cosine to our new angles, which measure rotations. Suppose a point P starts at position $(r, 0)$, and rotates through an angle α.



If P has coordinates (x, y), we define $\cos\alpha$ and $\sin\alpha$ by writing

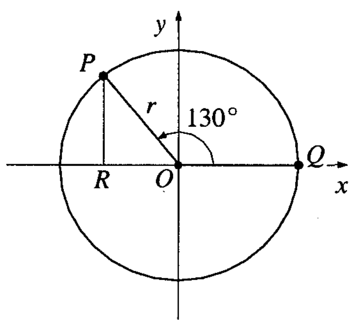
$$\begin{aligned}\cos\alpha &= \frac{x}{r}, \\ \sin\alpha &= \frac{y}{r}.\end{aligned}$$

Note that these new definitions give the same values as the old definitions when α is an acute angle.

Example 29 Find the numerical values of $\sin 130^\circ$ and $\cos 130^\circ$.

Note that by Chapter 3, page 71, we already know that $\sin 130^\circ = \sin(180^\circ - 130^\circ) = \sin 50^\circ$, so in fact, this quantity had been defined already. But let us see if our new definition gives the same result.

Solution. The diagram below shows $\angle QOP = 130^\circ$:



The circle in the diagram has radius r, and the point P has rotated through an angle of 130° from point Q. If the coordinates of P are (x, y), our new definition tells us that

$$\begin{aligned}\cos 130^\circ &= \frac{x}{r},\\ \sin 130^\circ &= \frac{y}{r}.\end{aligned}$$

Now we look at right triangle OPR, in which $\angle POR = 50^\circ$, and note that

$$\frac{y}{r} = \frac{PR}{OP} = \sin 50^\circ .$$

So this is the value of $\sin 130^\circ$.

Triangle OPR will also give us the numerical value of $\cos 130^\circ$, but we must be careful. Since the x-coordinate of point P is negative, we must write

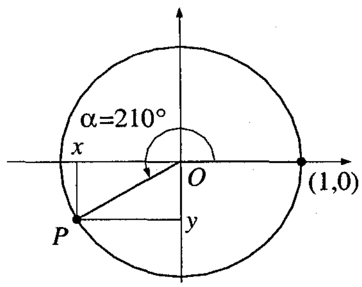
$$\cos 130^\circ = \frac{x}{r} = -\frac{OR}{OP} = -\cos 50^\circ .$$

Thus, $\cos 130^\circ = -\cos(180^\circ - 130^\circ) = -\cos 50^\circ$. This value agrees with the one given by the definition on page 79. □

It is important to note that the result above does not depend on the length of OP. We can choose a circle of any radius and draw the corresponding diagram for a 130° angle. Triangle OPR will always have the same angles, and the computation will be the same.

Example 30 What are the values of $\cos 210^\circ$ and $\sin 210^\circ$?

Solution. Since the radius of the circle will not matter, we are free to choose, for example, a circle of radius 1. Then our new definitions lead to the diagram below:



$P=(x, y)=(\cos 210°, \sin 210°)$

The geometry of a 30-60-90 triangle shows that the coordinates of point P are $(-\frac{\sqrt{3}}{2}, -\frac{1}{2})$. Then

$$\cos 210° = -\frac{\sqrt{3}}{2} \quad \text{and} \quad \sin 210° = -\frac{1}{2}.$$

Notice that both the sine and cosine of 210° are negative numbers. □

Example 31 Find the values of $\cos 360°$ and $\sin 360°$.

Solution. We choose a circle of radius 1. For a rotation of 360°, the coordinates of point P are $(1, 0)$. Therefore, $\cos 360° = 1$ and $\sin 360° = 0$. □

Now that we have definitions for sine and cosine of any angle, we can make definitions for the other trigonometric functions of these angles.

For any angle α,

$$\begin{aligned}
\tan\alpha &= \frac{\sin\alpha}{\cos\alpha}, \\
\cot\alpha &= \frac{\cos\alpha}{\sin\alpha}, \\
\sec\alpha &= \frac{1}{\cos\alpha}, \\
\csc\alpha &= \frac{1}{\sin\alpha}.
\end{aligned}$$

Example 32 Find the numerical value of $\tan 210°$.

Solution. From the results of Example 30, we have that

$$\tan 210° = \frac{\sin 210°}{\cos 210°} = \frac{-1/2}{-\sqrt{3}/2} = \frac{1}{\sqrt{3}}.$$

Note that this value is positive. □

Our new definitions of sine and cosine give values for any angle α. But this is not quite true for our new definitions of tangent, cotangent, secant and cosecant, because they involve division. We must be sure that we are not dividing by 0.

Indeed, we will not define $\tan\alpha$ if $\cos\alpha = 0$. Expressions such as $\tan 90°$, $\tan 270°$, and $\tan(-90°)$ must remain undefined.

For similar reasons, we cannot define $\cot 0°$, or $\csc 180°$.

Exercises

1. Find the numerical value of the following expressions. Do this without using your calculator, then check your answers with your calculator.

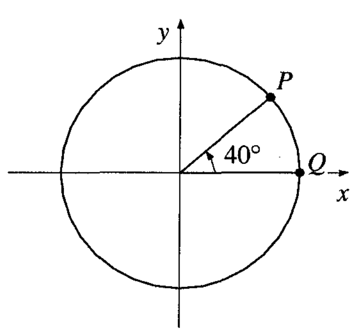
a) $\sin 390°$	b) $\cos 3720°$	c) $\tan 1845°$
d) $\sin 315°$	e) $\cot 420°$	f) $\tan(-30°)$

2. Find the numerical value of each expression below, or indicate if the given expression is undefined.

a) $\tan 360°$	b) $\sin 180°$	c) $\cos 180°$
d) $\cot 90°$	e) $\cot 360°$	f) $\tan(-270°)$

4 Calculations with angles of rotations

Let us look back at our original picture of an angle in a circle:



Originally, we thought of this as an angle of 40°. But a diagram of a 400° angle would look exactly the same, as would a diagram for 760° or −320°.

The diagram will look the same for any two angles which differ by a full rotation. Therefore, $\sin\alpha = \sin(\alpha + 360°)$.

Similarly, $\cos\alpha = \cos(\alpha + 360°)$ for any angle α.

These observations allow us to find the sine, cosine, tangent, or cotangent of very large angles easily.

Example 33 What is $\cos 1140°$?

Solution. If we divide 1140 by 360, the quotient is 3 and the remainder is 60, that is, $1140 = 3 \cdot 360 + 60$. So $\cos 1140° = \cos(3 \cdot 360 + 60°) = \cos 60° = 1/2$. □

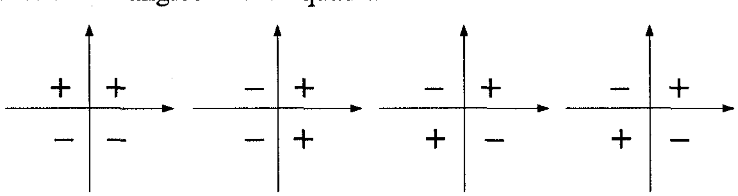
Example 34 Is the sine of 100,000° positive or negative?

Solution. If we divide 100,000 by 360, we get 277, with a remainder of 280. So the sine of 100,000° is the same as $\sin 280°$. Since the position of point P is in the fourth quadrant, its y-coordinate is a negative number. The sine of 100,000° is therefore negative. □

You can check the logic of these solutions using your calculator, which already "knows" if the sine of an angle is positive or negative. That is, the people who designed it did exercises like yours before they built the calculator.

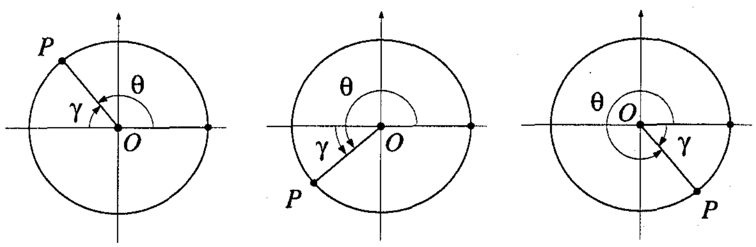
But it is also important to be able to "predict" certain values of the trigonometric functions, or at least tell whether their values will be positive or negative. It's not difficult to see that if point P ends up in quadrant I, all functions of the angles are positive. If point P lies in quadrant II, the sine and cosecant are positive, and all other functions are negative, and so on.

Exercise Check to see that the diagrams below give the correct signs for the functions of angles in each quadrant.



sine cosine tangent cotangent

We can even, sometimes, predict a bit more about the values of the trigonometric functions. If you look at each of the diagrams below, you may see that $\sin\theta$ is equal, in absolute value, to the sine of the acute angle made by one side of the angle and the x-axis (the angle marked γ in each diagram below):



Example 35 Find $\sin 300°$.

Solution. Point P, having rotated through $300°$, will end up in quadrant IV. So $\sin 300°$ is negative. Furthermore, the angle made by one side and the x-axis is $60°$. Hence $\sin 300° = -\sin 60° = -\sqrt{3}/2$. □

Exercises

1. In what quadrant will the point P lie after a rotation of $400°$? $3600°$? $1845°$? $-30°$? $-359°$?

2. Fill in the table below (you won't need a calculator). What is the relationship between $\sin\alpha$ and $\sin(-\alpha)$?

$\sin 30°$		$\sin(-30°)$	
$\sin 135°$		$\sin(-135°)$	
$\sin 210°$		$\sin(-210°)$	
$\sin 300°$		$\sin(-300°)$	
$\sin 390°$		$\sin(-390°)$	
$\sin 480°$		$\sin(-480°)$	

3. Solve the following equations for α, where $0 < \alpha < 360°$:

a) $\sin\alpha = 0$ b) $\cos\alpha = 0$ c) $\sin\alpha = 1$

d) $\cos\alpha = 1$ e) $\sin\alpha = -1$ f) $\cos\alpha = \frac{1}{2}$

g) $\sin\alpha = -\frac{1}{2}$ h) $\sin^2\alpha = \frac{1}{2}$ i) $\cos^2\alpha = -\frac{3}{4}$

4. a) If $\sin\alpha = 5/13$, in what quadrant can α lie? What are the possible values of $\cos\alpha$?

b) If $\sin\alpha = -5/13$, in what quadrant can α lie? What are the possible values of $\cos\alpha$?

5. We have seen (Chapter 2, page 50) that if a and b are non-negative numbers such that $a^2 + b^2 = 1$, then there exists an angle θ such that $\sin\theta = a$ and $\cos\theta = b$. Show that this is true, even if a or b is negative.

5 Odd and even functions

Consider the result of Exercise 2 on page 100. If you have filled in the table correctly, you will note that, for the angles given there, $\sin\alpha$ and $\sin(-\alpha)$ have opposite signs. This relationship holds in general:

$$\sin(-\alpha) = -\sin\alpha \quad \text{for any angle } \alpha\,.$$

Similarly, we find the following:

$$\begin{aligned}\tan(-\alpha) &= -\tan\alpha\\ \cot(-\alpha) &= -\cot\alpha\,.\end{aligned}$$

However, the cosine function is different. We have

$$\cos(-\alpha) = \cos\alpha\,.$$

In general, we can distinguish two type of functions.

A function is called **even** *if, for every x, $f(-x) = f(x)$.*
A function is called **odd** *if, for every x, $f(-x) = -f(x)$.*

So, for example, the functions

$$f(x) = \cos x\,, \quad f(x) = x^2 + 3\,, \quad \text{and } f(x) = \frac{1}{x^6}$$

are all even, while the functions

$$f(x) = \tan x\,, \quad f(x) = x^3 + 4x\,, \quad \text{and } f(x) = \frac{1}{x^7}$$

are all odd. The following functions are neither even nor odd:

$$f(x) = x^3 + x^2\,, \quad f(x) = \sin x + \cos x\,.$$

In summary:

$\cos x$ *is an even function, while* $\sin x$, $\tan x$, *and* $\cot x$ *are odd functions.*

This may be the reason that some mathematicians prefer to work with the cosine function, rather than the sine.

Exercises

1. Which of the following functions are even? Which are odd? Which are neither?

   a) $f(x) = x^6 - x^2 + 7$
   b) $f(x) = x^3 - \sin x$
   c) $f(x) = \frac{1}{x+1}$
   d) $f(x) = \sec x$
   e) $f(x) = \csc x$
   f) $f(x) = 2\sin x \cos x$
   g) $f(x) = \sin^2 x$
   h) $f(x) = \cos^2 x$
   i) $f(x) = \sin^2 x + \cos^2 x$

2. If $f(x)$ is any function, show that

$$g(x) = \tfrac{1}{2}(f(x) + f(-x))$$

   is an even function, and that

$$h(x) = \tfrac{1}{2}(f(x) - f(-x))$$

   is an odd function. Use these results to show that every function can be written as the sum of an even and an odd function.

3. Express the following functions as the sum of an even and odd function:

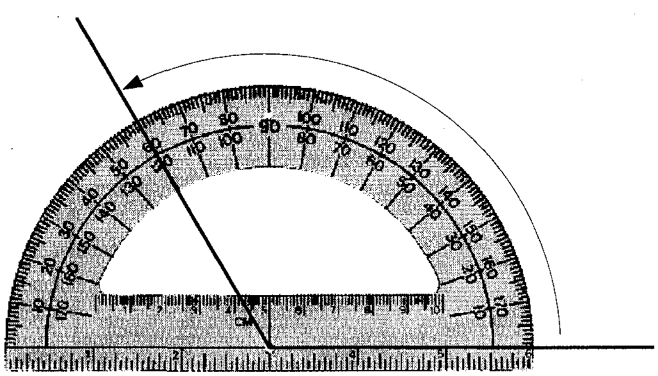
   a) $f(x) = \sin x + \cos x$
   b) $f(x) = x^3 + x^2 + x + 1$
   c) $f(x) = 2^x$
   d) $f(x) = \dfrac{1 - \sin x}{1 + \sin x}$
   e) $f(x) = \dfrac{1}{x+2}$

Chapter 5

Radian Measure

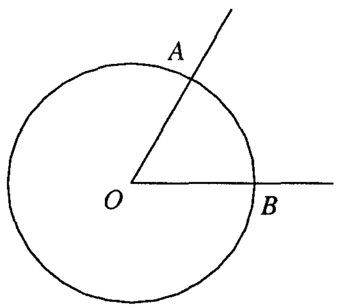
1 Radian measure for angles and rotations

So far, our unit of measurement for angles and rotations is the degree. We measure an angle in degrees using a *protractor*:

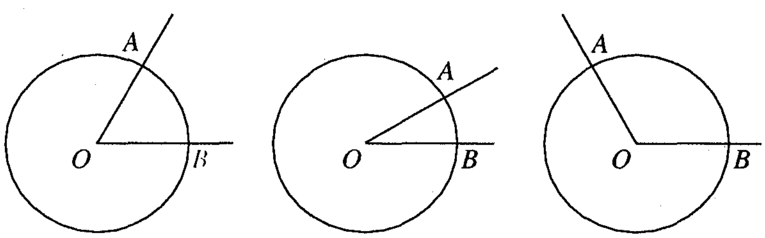


Why are there 360 degrees in a full rotation? The answer to this lies in history, not in mathematics. It turns out that there is a more convenient way to measure angles and rotations called *radian* measure. Mathematicians and scientists find it natural to use radian measure to express relationships. Unlike degree measure, radian measure does not depend on an arbitrary unit.

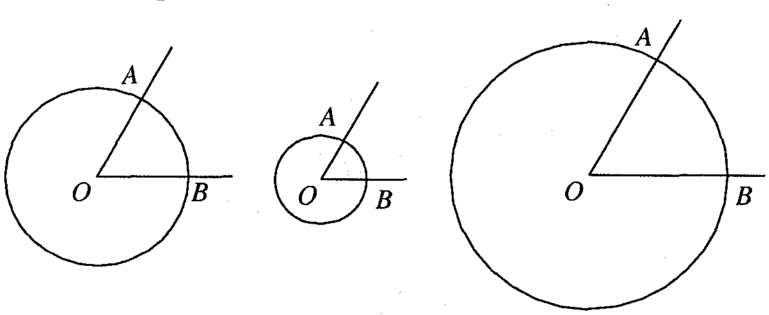
To measure an angle in radians, we place its vertex at the center of any circle, and think about the length of arc AB, as measured in inches, centimeters, or some other unit of length:



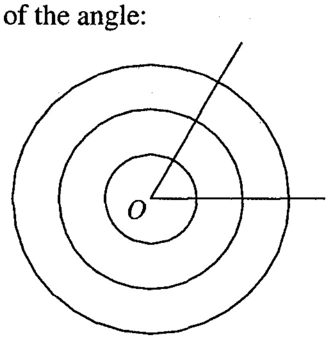
The length of this arc depends on the size of angle AOB:



But is also depends on the size of the radius of the circle:



So we cannot simply take the length of this arc as the measure of the angle. But the *ratio* of the length of the arc to the radius of the circle depends only on the size of the angle:



Definition The radian measure of an angle is the ratio of the arc it cuts off to the radius of any circle whose center is the vertex of the angle.[1]

This definition reminds us of the definition of the sine of an angle, which is also a ratio, and which does not depend on the particular right triangle that the angle belongs to.

Example 36 What is the radian measure of an angle of $60°$?

Answer. We place the vertex of our $60°$ angle at the center of a circle of radius r, and examine the arc it cuts off. Since $60/360 = 1/6$, this arc is 1/6 of the circumference of the circle. So its length is $(1/6)(2\pi r) = \pi r/3$ units.[2] By our definition, the radian measurement of our $60°$ is the ratio

$$\frac{\pi r/3}{r} = \frac{\pi}{3}.$$

Numerically, this is approximately 1.0471976, or a little more than 1 radian. □

Example 37 What is the radian measure of an angle of $360°$?

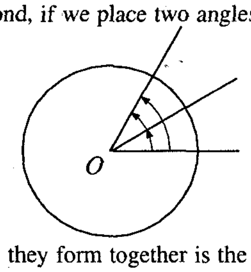
Answer. $2\pi r/r = 2\pi$ radians. □

Example 38 What is the degree measure of an angle of 1 radian?

Answer. An angle of 2π radians is 360 degrees (see Example 37). So an angle of 1 radian is $360/2\pi = 180/\pi$ degrees. □

In Example 38, we have used the very important fact that radian measure is *proportional* to degree measure. In fact, it is not hard to see that

[1] There are two simple tests that this measurement passes. First, the bigger the angle, the bigger its radian measure. Second, if we place two angles next to each other (see figure),



the measure of the larger angle they form together is the sum of the measures of the two original angles.

[2] Recall that if r is the radius of a circle, the length of its circumference is given by the formula $2\pi r$.

the ratio of the radian measure of an angle to its degree measure is always $\pi/180$:

$$\boxed{\frac{\text{radians}}{\text{degrees}} = \frac{\pi}{180}}$$

In general, if we are measuring an angle in radians, we do not use any special symbol like the "degree" sign.

Example 39 Find $\sin(\pi/6)$.

Solution. We first express the angle in degrees. If D is the required degree measure, we have

$$\frac{\text{radians}}{\text{degrees}} = \frac{\pi/6}{D} = \frac{\pi}{180},$$

which leads to $D = 30°$, and we know that $\sin 30° = 1/2$. □

Example 40 In a circle of radius 1, what is the length of the arc cut off by a central angle of 2 radians?

Solution 1 (the long and hard way). We saw above that the degree-measure of this angle is about 114°. So the arc cut off by this angle is approximately

$$\frac{114}{360} \times 2\pi \approx 1.989675347274$$

units long. □

Solution 2 (the neat and easy way). In a unit circle (whose radius is 1 unit), the radian measure of a central angle is just the length of the arc it cuts off. This tells us that the required arc is exactly two units long (and gives us an idea of the error we made in using the approximate degree measure in Solution 1). □

Example 41 A central angle in a circle of radius 2 units cuts off an arc 5 units long. What is the radian measure of this angle?

Solution. By definition, this radian measure is $5/2$. □

Exercises

1. What is the radian measure of an angle of 180°? 90°?
2. What is the degree measure of an angle of 2 radians?
3. What is the radian measure of $1/4$ of a full rotation?
4. What is the radian measure of a rotation through an angle of 45°?
5. Fill in the following table:

Degree measure	Radian Measure
90	
180	
270	
360	
	$\pi/2$
	π
	$3\pi/2$
	2π

6. Fill in the following tables:

Degree measure	Radian measure
0	
30	
72	
120	
135	
	$\pi/6$
	$\pi/5$
	$\pi/4$
	$\pi/3$
	$2\pi/3$
	$7\pi/10$

Degree measure	Radian measure
198	
210	
216	
225	
240	
	$11\pi/10$
	$10\pi/9$
	$7\pi/6$
	$6\pi/5$
	$5\pi/4$
	$4\pi/3$

7. What is the radian measure of an angle of 1 degree?

8. Using your calculator, find the sine of an angle of (a) 1 radian; (b) 1 degree.

9. Without using your calculator, fill in the following table:

α (in radian)	$\sin\alpha$	$\cos\alpha$
$\pi/6$		
$\pi/3$		
$\pi/2$		
$2\pi/3$		
$7\pi/6$		
$5\pi/4$		
$3\pi/2$		
$11\pi/6$		

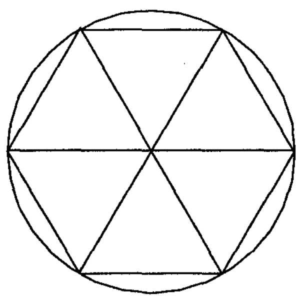
10. In a circle of radius 1, what is the length of an arc cut off by a central angle of 2 radians? Of 3 radians? Of π radians?

11. In a circle of radius 3, what is the length of an arc cut off by a central angle of 2 radians? Of 3 radians? Of π radians?

12. If $\sin\pi/9 = \cos\alpha$ and α is acute, what is the radian measure of α?

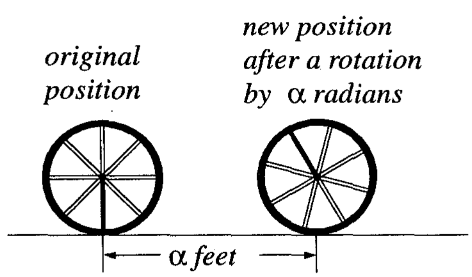
13. If α is an angle between 0 and $\pi/2$ (in radian measure), which is bigger: $\sin\alpha$ or $\cos(\pi/2 - \alpha)$?

14. Let us take an angle whose radian measure is 1. Using the picture below, prove that its degree-measure is less than 60°. (In fact, an angle of radian measure 1 is approximately 57 degrees.)



2 Radian measure and distance

Imagine a wheel whose radius is 1 foot. Let this wheel roll, without slipping, along a straight road:[3]



Since the wheel does not slip as it rolls, the distance it rolls, in feet, is just the length of the arc that the angle α cuts off.

Example 42 How far will a wheel of radius 1 foot travel after 1 rotation?

Solution. Because it rotates without slipping, the wheel will travel exactly the length of the circumference of the circle. But if the radius of the circle is r, then the circumference is $2\pi r$. Since $r = 1$, the answer is just 2π, or approximately 6.28 feet. □

[3]Sometimes a car wheel slips as it rolls. This is called a *skid*, and it happens when there is not enough friction between the wheel and the road (for example when the road is icy or wet). A car wheel can also turn without rolling: sometimes, a car stuck in deep snow will *spin its wheels*. We assume that neither of these things is happening to our wheel.

Example 43 A wheel of radius 1 foot rotates through 1/2 a rotation. How far will it travel?

Answer. It will travel a distance of π, or about 3.14 feet. □

Example 44 How far does a wheel of radius 1 foot travel along a line, if it rotates through an angle of 2 radians?

Answer. Two feet.

Example 45 Through how many radians does a circle of radius 1 foot rotate, if it travels 5 feet down a road?

Answer. Five radians.

Example 46 How much does a wheel with radius 1 foot rotate if it travels 1000 feet along a road? Give the answer in radians and also in degrees.

Solution. In radians, this is easy: it has rotated through 1000 radians.

In degrees, the answer is more difficult to find. Each full rotation covers 2π feet. So in traveling 1000 feet, our wheel has rotated through $1000/2\pi \approx 159.155$ rotations. Since each rotation is $360°$, the degree measure of a rotation of 1000 radians is

$$159.155 \times 360 \approx 57296° .$$ □

Example 47 What is the radian measure of an angle of $1000°$?

Solution. A rotation of $1000°$ is $1000/360 \approx 2.77$ rotations, and each rotation is 2π radians. So $1000°$ is

$$2.77 \times 2\pi \approx 17.405$$

in radian measure. □

Example 48 Is $\sin 500$ (in radian measure) a positive or a negative number?

Solution. Since $500/2\pi \approx 79.577$, 500 radians is 79 full rotations, plus approximately 0.57 of one more rotation. Since $0.5 < 0.57 < 0.75$, this is between 1/2 and 3/4 of one rotation. So a rotation of 500 radians will end up in the third quadrant, and its sine is a negative number. □

Exercises

1. Through how many radians does a circle of radius 1 foot rotate, if it travels 5 feet down a road?

2. Through how many degrees does a circle of radius 1 foot rotate, if it travels 5 feet down a road?

3. How far does a circle of radius 1 foot travel, if it turns through an angle of 4 radians?

4. How far does a circle of radius 1 foot travel, if it turns through an angle of $120°$?

5. In a circle of radius 1, what is the length of an arc cut off by an angle with radian measure 1/2? $\pi/2$? α?

6. What is the radian measure of an angle of 720 degrees? 1440 degrees? 3600 degrees? 15120 degrees? What is the degree measure of an angle whose radian measure is 12π? 15π? 100π?

7. In a circle of radius 3, what is the length of an arc cut off by an angle with radian measure 1/2? $\pi/2$? α?

8. In a circle of radius 3, how long is the arc cut off by an angle with radian measure 1.5?

9. In a circle of radius 5, how long is the arc cut off by an angle of 80 degrees?

10. In a circle of radius 2, what is the radian measure of a central angle whose arc has length 3 units?

11. In a circle of radius 6, what is the degree-measure of a central angle whose arc has length 2 units?

12. A circle of radius 7 units rolls along a straight line. If it covers a distance of 20 units, what is the radian measure of the rotation it has made?

13. A circle of radius 8 rolls along a straight line, through an angle of 150 degrees. How far does it roll?

14. Through what angle does the hour hand of a watch rotate in one hour? Give your answer in radians.

15. Through what angle does the minute hand of a watch rotate in one hour? The second hand?

16. In answering problem 14, Joe Blugg gave the following solution: One hour on the face of a watch is 1/12 of the circle, so it is $2\pi/12 = \pi/6$ radians. In degrees, the answer is $360/12 = 30$ degrees. But Joe is *not* correct, either in degrees or in radians. Find and correct his mistake. Did you make the same error here and in similar problems about a watch?

    Hint: Do the hands of a clock rotate? In which direction?

17. Let us look at a pocket watch whose hour hand is exactly one inch long. Suppose the tip of this hour hand travels a distance of 1000 inches as it goes around. How long does this trip take?

18. Suppose the length of the hour hand of Big Ben is exactly one yard long. How long it will it take Big Ben's hour hand to turn through 1000 radians?

19. A wheel with radius 1 meter is rolling along a straight line. One of its spokes is painted red. At the starting position this spoke is vertical, with its endpoint towards the ground. How many radians does the wheel turn before the spoke is again in this position? How many radians does the wheel turn before the spoke is vertical, with its endpoint towards the sky?

20. A wheel whose radius is 1 meter rolls along a straight path. The path is marked out in 3-meter lengths, with red dots three meters apart. The wheel has a wet spot of blue paint on one point. When it starts rolling, that point is touching the ground. As the wheel rolls, it leaves a blue mark every time the initial point touches the ground again.

    a) How far apart are the blue marks?
    b) Through what angle has the wheel rolled between the time it makes a blue mark and the time it makes the next blue mark?
    c) Will a blue mark ever coincide with a red mark?

d) When the blue marks do not coincide with the red marks, how close do they come to the red marks? (If you know how to program a computer or calculator, you may need to write a program to answer this question.)

e) Now suppose each interval between the red dots is divided into four equal sub-intervals, say by pink dots in between. A blue mark is created as the wheel completes its 100th rotation. Between what two dots does this blue mark occur?

3 Interlude: How to explain radian measure to your younger brother or sister

When you drive with Mom or Dad in the car, did you ever notice the odometer? That's the little row of numbers in front of the steering wheel. It measures the distance covered by the car, in miles.

But how does it know this? The odometer cannot read the road signs, telling us how far we've come. It must get the information from the car's wheels. But the car's wheels can only tell the odometer how much they have turned. The more the car's wheels turn, the more distance we cover. The odometer knows how to convert rotations to miles. In geometry, we learn that a circle of radius r has a circumference of $2\pi r$. This, and the radius of the wheel in feet, is all the odometer really needs to know.

Suppose the wheel tells the odometer that it has rotated 50 times. Then the odometer knows that the wheel has traveled $50 \times 2\pi r$ feet, where r is the radius of the car's wheels in feet (it must convert this number to miles). And if the wheel has rolled only 1/4 of the way around, the odometer reports a distance of $(1/4) \times 2\pi r$ feet, again with a conversion to miles.

But suppose you want to know how much wear the tires have had. Then we must read the odometer, and figure out how many rotations the tires have made from the distance they traveled. So if the odometer says that the car has traveled 200 feet (we have to convert from miles, again), then 200 is $2\pi r$ times the amount of rotation the wheels have made. So the wheels have made $200/2\pi r$ rotations.

And this is what we call the radian measure of this rotation.

4 Radian measure and calculators

Most calculators, and all scientific calculators, know about radian measure. You can switch your calculator between "degree mode" and "radian mode" (and sometimes there are still other ways to measure angles). But each calculator does this in a different way. It is important that you know how to tell which mode your calculator is working in, and also how to switch from one mode to another.

Exercises

1. A student asked his calculator for the sin of 1. The answer was 0.8414709848079. Was the calculator in radian mode or in degree mode?

2. For small angles, $\sin x$ is approximately equal to x, when x is given in radian measure. Use your calculator to find out how big the difference is between x and $\sin x$ for angles of radian measure 0.2, 0.15, 0.05.

   In each case, which is bigger, x or $\sin x$?

3. A better approximation to $\sin x$ (when measured in radians) is given by $x - x^3/6$. Find the difference between this value and the actual value of $\sin x$ for the three angles above.

4. In the old schools of artillery, the officers would use a version of the approximation $\sin x \approx x$ However, they had to measure x in degrees, so they used $\sin x = x/60$. What is the error in this approximation, if $x = 10°$?

5. a) Without your calculator, make a guess for the value of $\sin 0.1$ (in radian measure). Then use a calculator to check your guess.

   b) Now perform the same experiment for $\sin 0.1$ (in degree measure).

6. a) Find the sine of an angle whose degree measure is 1000.

   b) Find the sine of an angle whose radian measure is 1000.

7. a) Find sin (sin 1000), where radian measure is used for the angle.

   b) Find $\sin 3.14$, where radian measure is used for the angle.

8. Without looking up this number on your calculator, prove that cos 1.5707 is less than 0.0001.

   Hint: Do you recognize the number 1.5707?

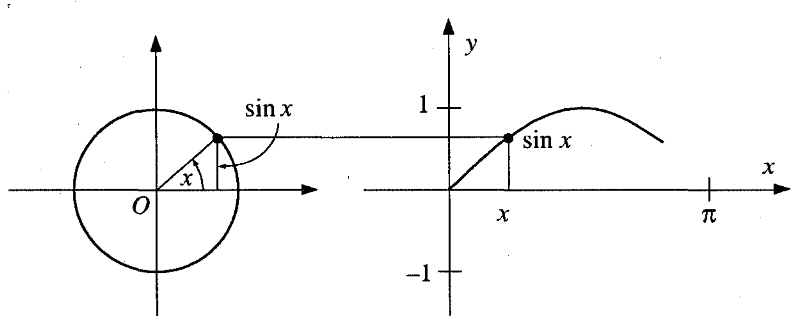
5 An important graph

Let us summarize our knowledge of the sine function by drawing its graph.

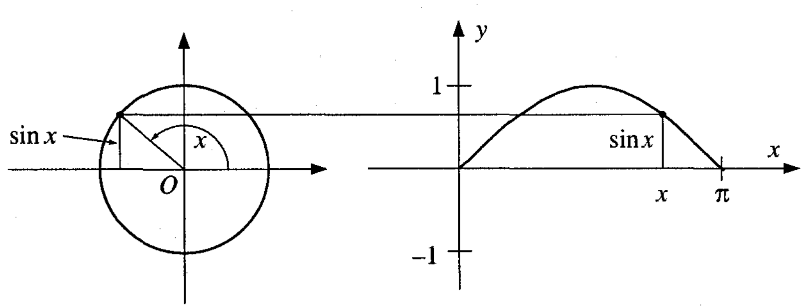
The integer multiples of π will give us a convenient scale for the x-axis, since the values of $\sin x$ at these points are easy to calculate. For the y-axis, we need only values from -1 to 1, since $\sin x$ can only take on these values.

We can draw the graph by looking at a unit circle (drawn on the right below), and recording the height of a point which makes an angle α with the x-axis. Here is what it looks like for a typical acute angle α.

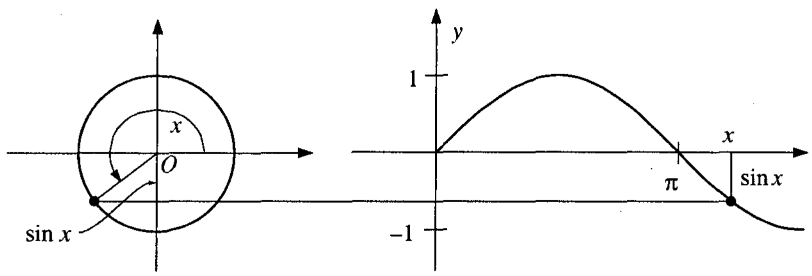
As α varies from 0 to $\pi/2$, the graph of $y = \sin x$ increases.

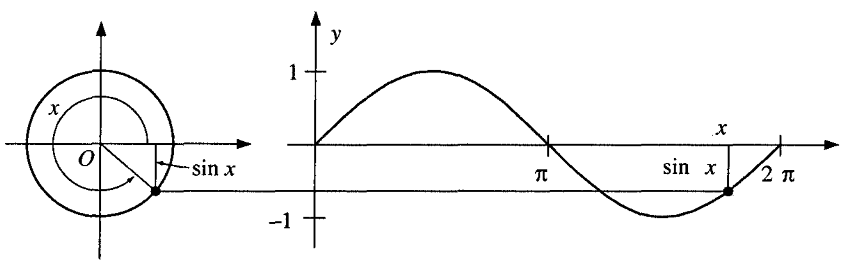


Here is a typical scene from the second quadrant:

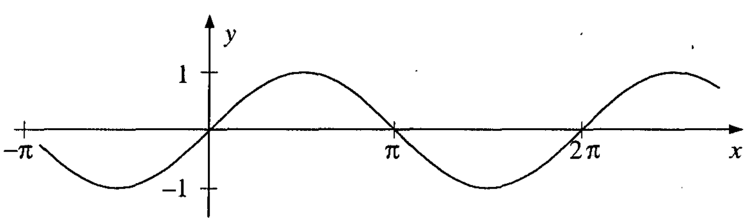


And in the third and fourth quadrants, the situation is like this:





After α has rotated through 2π radians, the whole cycle repeats itself. For negative values of α, the situation is the same. Here is the complete curve:



Exercise

1. Use the graph above to answer the following questions. You can check some of the answers using your calculator.

   a) Is $\sin 7\pi/5$ positive or negative? Estimate its value.

   b) Is $\sin(-3\pi/7)$ positive or negative? Estimate its value.

   c) We know that $\sin\pi/6 = 1/2$. Check this on the graph. Where else does the sine function achieve a value of $1/2$?

   d) For what values of x does $\sin x = \sin\pi/12$? Mark, on the x-axis, as many of these values as you can find.

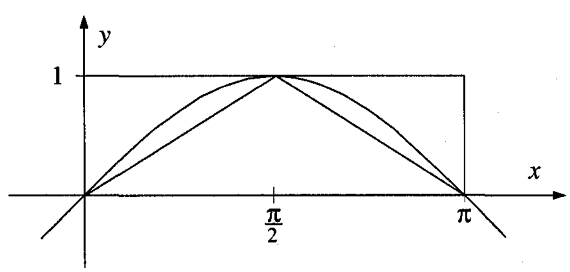
   e) For what values of x does $\sin x = 0.8$? Estimate a value of x for which this is true. Then locate, on the x-axis, as many other values as you can find.

6 Two small miracles

We pause here to describe two remarkable relationships, so remarkable that they seem like miracles. An explanation (that is, a mathematical proof) of these miracles is postponed for later.

Miracle 1: The area under the sine curve

Look at the first arch of the curve $y = \sin x$. What can we tell about the area under this arch? The area is certainly less than π, since it fits into a rectangle whose dimensions are 1 and π:

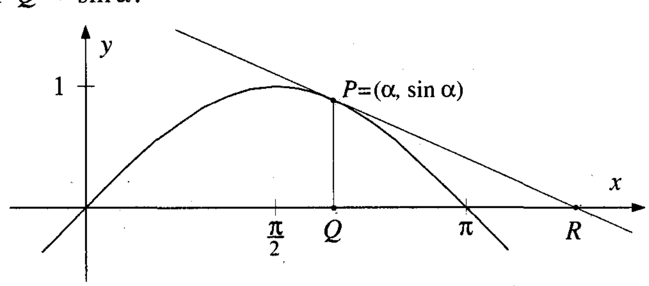


And the area is greater than that of the isosceles triangle shown in the diagram, whose area is $\pi/2$. So if we wanted to approximate the area under this arch, we would say that it is between $\pi/2$ and π. We could go further with the approximations, taking more and more triangles which would "fill" the area below the curve. Something like this is in fact done, in calculus.

The result is a small miracle: The area under one arch of the sine curve is exactly 2.

Miracle 2: The tangent to the sine curve

Let us take a point $P = (\alpha, \sin\alpha)$ on the curve $y = \sin x$. Let the perpendicular from P meet the x-axis at the point Q. Let us draw the tangent to the curve at point P, and extend it to meet the x-axis at R. It is easy to see that $PQ = \sin\alpha$.



But, by a small miracle, we can also find the length of QR. It is just $|\tan\alpha|$, the absolute value of $\tan\alpha$.

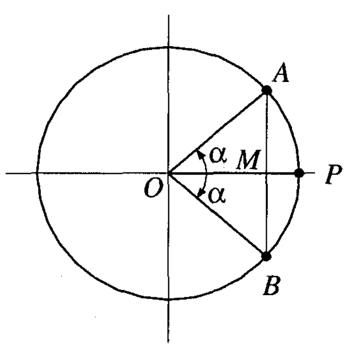
Appendix – Some advantages of radian measure

Notice that the radian measure of angles, like their degree measure, is additive. That is, if two angles are placed so as to "add up" to a larger angle, the sum of the angles corresponds to the sum of the arcs.[4]

Another good thing about radian measure is that it is *dimensionless*. That is, it is independent of any unit of measurement. Length, for instance, can be measured in centimeters, inches, or miles, and we get different numbers. The same is true of area, volume, and many other quantities. But radian measure, like the sine of an angle, is a ratio, and so does not depend on the units used to measure the arc of the circle or its radius. This is another reason why physicists, and other scientists too, like to use radians.

Since the radian measure and the sine of an angle are both dimensionless, we can compare them. For an acute angle α, which is larger, $\sin\alpha$ or the radian measure of α?

Geometry can help us answer this, if the angle is small. In the diagram below, we took a circle of unit radius, and drew a tiny angle AOP. Then we made another copy of this angle (back-to-back with the first copy) and labeled it POB. Then arc AP = arc PB.



[4]Whenever we decide how to measure something, we would like the measure to be additive. Length is additive, as is area and volume. However, a trip to the grocery will quickly confirm that the price of Coca-Cola is not additive. The price of two 6-ounce bottles is likely to be more than the price of one 12-ounce bottle, because you are paying for packaging, labeling, shipping, and so on.

If the length of this arc is α, then the radian measure of $\angle AOP = \angle POB = 2\alpha/1$ (since the circle has unit radius, and $\angle AOP = \angle POB = \alpha$). From right triangle AOM, we see that $AM = \sin\alpha$, so $AB = 2\sin\alpha$. Since arc AB is longer than line segment AB, we see that $2\sin\alpha < 2\alpha$, or $\sin\alpha < \alpha$.

From this picture we also see that if the angle α is small enough, then 2α and $2\sin\alpha$ are very close to each other.[5]

Once again, we can see the advantage of radian measure. If the angle α were measured in degrees, the best statement we could make would be that $\sin\alpha < \alpha\pi/180$.

But with radian measure we can even prove a bit more. Later on we will see that for a small angle α measured in radians, the ratio $\sin\alpha/\alpha$ is very close to 1. For example, for $\alpha = 0.1$, $\sin\alpha$ is more than 99% of α itself.

Radian measure also goes well with the trigonometric ratios. We have already seen that $\sin x$ is approximately close to x for small angles. It is even closer to $x - x^3/6$, an excellent and simple approximation. We can even show that the error is less than $x^5/120$, which, for small angles, is a very tiny number.

But this is true only if we use radian measure for x. In degrees, as we have seen, this formula would be terrible.

It is true that the nicest angles have radian measures which involve the number π. And we admit that sometimes it is difficult to deal with π because it is an irrational number, and our decimal notational system for numbers doesn't provide us with a good symbol for it[6](this is why we use a Greek letter). But it's even less convenient for English-speaking people to convert miles to kilometers, or pounds to kilograms. So please don't let this slight inconvenience stop you from using radian measurement.

---

[5]What does it mean for two numbers to be "close"? For example, 1 and 0.99 are certainly close: their difference is 0.01, a tiny number. But 1000 and 998 are also close. Their difference is 2, which is a much larger number than 0.01. However, the ratio $998 : 1000 = 0.998$ is very close to 1. So sometimes we should measure "closeness" by seeing how close the *ratio* of two numbers is to 1. Thinking this way, we would not say that, 0.1 and 0.0001 are close. Although both these numbers are small, and their difference is small, their ratio is 1000, which is not small. In the diagram it is true that if α is small, not only is $\sin\alpha$ also small, but the two numbers are close, since their ratio is close to 1.

[6]The number π is one of two irrational constants that come up quite naturally. The other is e, which is approximately 2.71828, and is also irrational. The number e comes up in calculus as naturally as the number π does in geometry. About 250 years ago, it was discovered that these two numbers are related by the remarkable equation $e^{i\pi} = -1$.

Exercises

1. Use your calculator to fill in the following table (of course, the second and third columns will be numerical approximations):

α (radians)	α (degrees)	$\sin\alpha$
1	57.29578	
0.5		
0.2		
0.1		
0.01		
0.02		
0.001		
0.002		
0.005		

2. Without using your calculator, give an estimate for the value of

$$\sin 0.00123456\,.$$

Is this estimate too large or too small? Check this using your calculator, after you've answered the question.

3. a) Use your calculator to fill in the following table:

α	$\alpha - \dfrac{\alpha^3}{6}$	$\sin\alpha$
1		
0.5		
0.2		
0.1		
0.05		
0.01		
0.001		

b) The table above shows that $\sin\alpha$ is approximately equal to $\alpha - \alpha^3/6$, if α is a small angle measured in radians. Write the corresponding approximation for $\sin D$, where D is a small angle measured in degrees. Your approximation should be an expression in the variable D. Then check your expression for $D = 1°$.

4. The error in the above estimate is always less than $\alpha^5/120$. What is the largest possible error if the angle is measured in degrees, instead of radians?

5. Use your calculator to determine the radian measures of the angles x for which $x^5/120 < 0.001$.

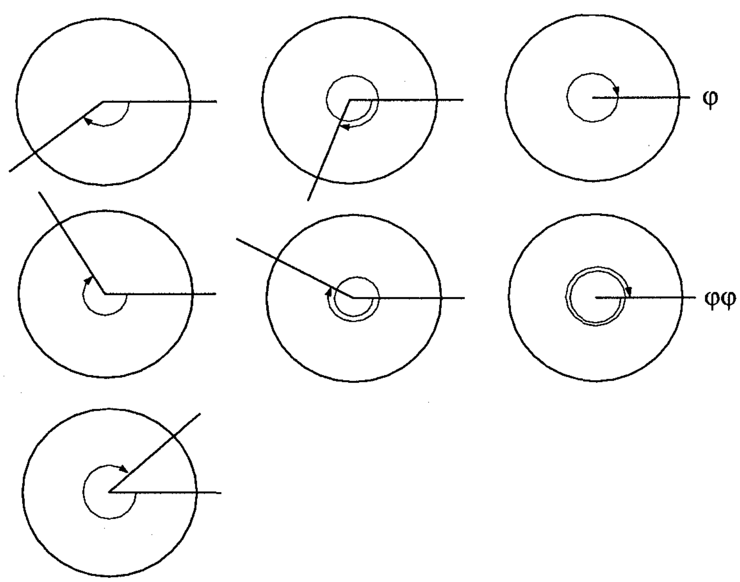
6. We have discussed the formula

$$\sin x \approx x - \frac{x^3}{6},$$

which is proven in calculus. Can you guess the next term of this approximation?

If you can do this, you will have a formula which gives $\sin x$ for small values of x to more decimal places than most calculators can display!

7. In the year 2096, a space capsule landed on earth, with artifacts from a distant alien civilization. Here are some diagrams found in the space capsule:



Experts believe that this chart shows how they measure angles. Tell as much as you can about the system of angle measure in this civilization. What do you think the symbol φ stands for?

8. In which quadrant do each of the following angles lie?

$$1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 1000 \text{ (all these in radians)}, \qquad 1000^\circ .$$

9. Suppose you answered the question above for angles of radian measure $1, 2, 3, 4, \ldots, 100$. What fraction of these angles do you suppose would lie in quadrant 1? quadrant 2? quadrant 3? quadrant 4?

   Solution. You can guess that approximately 1/4 of the angles lie in each quadrant – there is no reason for the angles to "favor" one quadrant in particular. In fact, this guess is correct. It is a special case of the important Ergodic Theorem of higher mathematics. If you took angles of radian measure $1, 2, 3, \ldots$ up to 1000, your approximation would be even closer to 1/4 for each quadrant. □

Chapter 6

The Addition Formulas

1 More identities

We now come to an important and fundamental property of the sine and cosine functions. If we know the values of $\sin\alpha$ and $\cos\alpha$, and also the values of $\sin\beta$ and $\cos\beta$, then we can calculate the values of $\sin(\alpha+\beta)$, $\cos(\alpha+\beta)$, $\sin(\alpha-\beta)$, and $\cos(\alpha-\beta)$.

But perhaps this is easy. Perhaps $\sin(\alpha+\beta)$ is simply equal to $\sin\alpha + \sin\beta$. Let us test this guess by setting $\alpha=\beta=\pi/2$. Then $\sin(\alpha+\beta) = \sin(\pi/2+\pi/2) = \sin\pi = 0$, while $\sin\pi/2+\sin\pi/2 = 1+1 = 2$. Since these two values are not equal, our guess is wrong.

Exercises

1. Complete the following table:

α	β	$\sin\alpha$	$\sin\beta$	$\sin\alpha+\sin\beta$	$\sin(\alpha+\beta)$
$60°$	$30°$				
$\pi/4$	$\pi/4$				
$\pi/6$	$\pi/3$				

2. Note that $\sin(\alpha+\beta)$ is not equal to $\sin\alpha+\sin\beta$ for these values of α and β. Which expression has the larger value in each case?

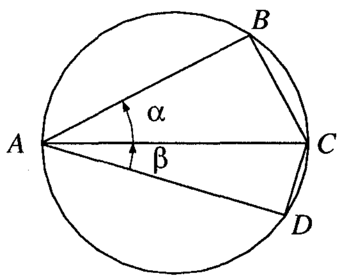
3. Check which of the following identities are correct, and which are not, using the angles $\alpha = 60°$, $\beta = 30°$:

   a) $\sin\alpha+\sin\beta = \sin(\alpha+\beta)$.

b) $\sin(\alpha - \beta) = \sin\alpha - \sin\beta$.

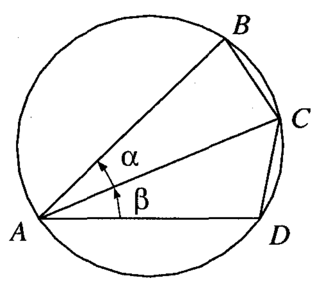
c) $\sin^2\alpha - \sin^2\beta = \sin(\alpha+\beta)\sin(\alpha-\beta)$.

4. a) The diagram below shows a circle with diameter $AC = 1$.



Find a line segment in the diagram equal in length to $\sin\alpha$ and one equal to $\sin\beta$.

b) The diagram below shows the same circle as above. Its diameter is still 1, but AC is not a diameter. Angles α and β are the same acute angles as before.



Find line segments in the diagram equal in length to $\sin\alpha$ and to $\sin\beta$.

c) In the figures for parts a) and b), draw in a line segment equal in length to $\sin(\alpha+\beta)$.

5. Recall that $\sin 45^\circ = \frac{\sqrt{2}}{2} \approx 0.707$ and $\sin 60^\circ = \frac{\sqrt{3}}{2} \approx 0.866$. Note that both these values are greater than 1/2. How can you tell immediately, without much calculation, that $\sin 45^\circ + \sin 60^\circ$ cannot equal $\sin 105^\circ$, although $45 + 60 = 105$?

2 The addition formulas

So far, most of what we have done is to give new names to familiar objects. But now we will explore the following *addition formulas* for sines and cosines:

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta\,,$$
$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta\,.$$

In a sense, they are the key reason why the sine and cosine functions find so many uses in physics, and in mathematics as well.[1]

There are also two related formulas for differences:

$$\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta\,,$$
$$\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta\,.$$

Exercises

1. Check the formulas given above by letting $\alpha = 60°$, $\beta = 30°$.

2. Check that these formulas say something true (if not enlightening) when $\alpha = 0$ and β is any angle. What happens if $\beta = 0$?

   Note: If you ever forget which formula is which, you can quickly look at what happens if $\beta = 0$. The formula for $\sin(\alpha + 0)$, for example, should give you the value $\sin\alpha$.

3. Check the formulas for $\sin(\alpha+\beta)$ and $\cos(\alpha+\beta)$ when $\alpha + \beta = \pi/2$.

   Hint: Assume α and β are acute angles in the same triangle, and compare $\sin\alpha$ and $\cos\beta$.

4. Check that the addition formulas are true if $\alpha = \beta = \pi/4$.

5. Check that $\sin^2(\alpha+\beta) + \cos^2(\alpha+\beta) = 1$ using the formulas above. That is, show that $(\sin\alpha\cos\beta + \cos\alpha\sin\beta)^2 + (\cos\alpha\cos\beta - \sin\alpha\sin\beta)^2 = 1$.

---

[1] But these uses for sine and cosine were not the earliest. The astronomer Ptolemy, in the second century CE, used these addition formulas, although he didn't have the names sine and cosine that we use now. As an astronomer, he needed equivalent concepts to locate the stars and planets, and to describe their periodic motions.

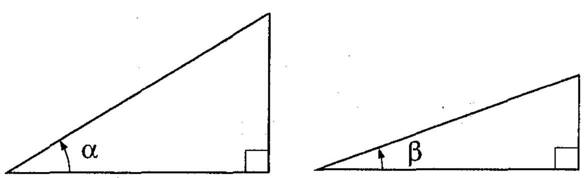
6. Using the formulas given in the text above, prove that $\sin(\alpha+\beta)\sin(\alpha-\beta) = \sin^2\alpha - \sin^2\beta$. That is, show that

$$(\sin\alpha\cos\beta+\cos\alpha\sin\beta)(\sin\alpha\cos\beta-\cos\alpha\sin\beta) = \sin^2\alpha-\sin^2\beta\,.$$

3 Proofs of the addition formulas

The exercises above have shown that the addition and subtraction formulas we propose are reasonable, but if we are to do mathematics, we must have a proof.

We will first prove the addition formula for $\sin(\alpha+\beta)$ in the case where α, β, and $\alpha+\beta$ are all acute angles. We will need two right triangles: one containing an acute angle equal to α, and another containing an acute angle equal to β.



We must put these triangles together in some way, so that the resulting diagram includes an angle equal to $\alpha+\beta$ (we assumed that this angle is also acute). There are only three ways to do this, so that they have a common side:

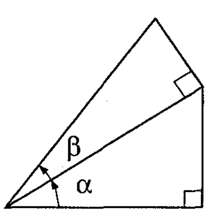


Fig. 1

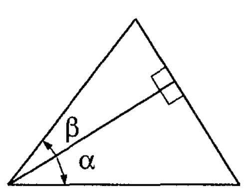


Fig. 2

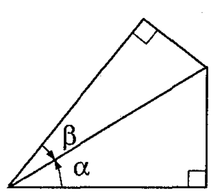
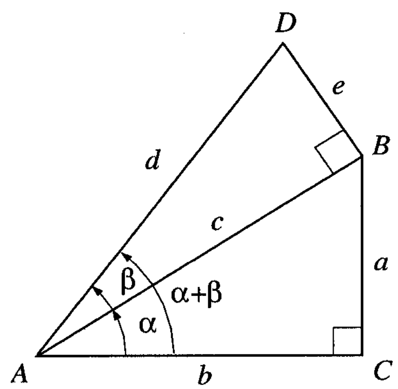


Fig. 3

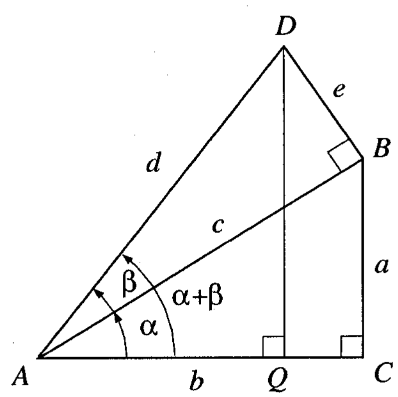
Each of these pictures gives us a different beautiful proof of the formula for $\sin(\alpha+\beta)$. We explore here the first two. We postpone the third, which is perhaps the most interesting, for another occasion (see the appendix of this chapter).

4 A first beautiful proof

We start with Fig. 1. Let us label the sides of the triangle as shown. Then $\sin\alpha = a/c$, $\sin\beta = e/d$. We need to represent $\sin(\alpha+\beta)$ in the diagram.



Let us draw line DQ perpendicular to the segment marked b:



Now we can write

$$\sin(\alpha+\beta) = \frac{DQ}{d}.$$

But DQ, which is related to $\sin(\alpha+\beta)$, is not related to the ratios representing $\sin\alpha$ and $\sin\beta$. To establish this relationship, we divide DQ into two parts, with a perpendicular from point B:

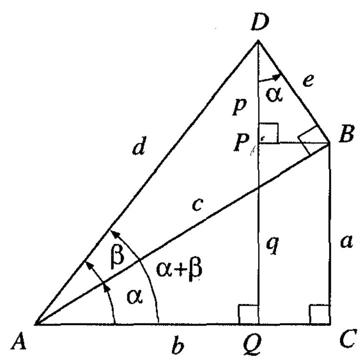


Fig. 4

Then

$$\sin(\alpha+\beta) = \frac{p+q}{d} = \frac{p}{d} + \frac{q}{d} = \frac{p}{d} + \frac{a}{d}.$$

Now we must relate p/d and a/d to $\sin\alpha$ and $\sin\beta$. We start with the second fraction. The segment a is in triangle ACB, and the segment d is in triangle ABD. We relate the fraction a/d to both triangles by introducing c as an intermediary (since c is in both triangles):

$$\frac{a}{d} = \frac{a}{c} \cdot \frac{c}{d} = \sin\alpha\cos\beta.$$

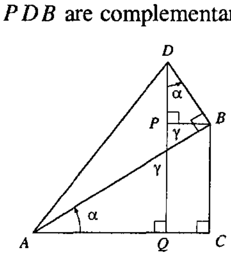
It is a bit more difficult to work with the fraction p/d. The segment d is in triangle ABD (which includes angle β), and the segment p is in triangle DPB. Happily, this last triangle contains an angle equal to α; namely[2] $\angle PDB$. Now we use segment e as an intermediary, and write

$$\frac{p}{d} = \frac{p}{e} \cdot \frac{e}{d} = \cos\alpha\sin\beta.$$

Putting this all together, we find that

$$\sin(\alpha+\beta) = \frac{a}{c}\frac{c}{d} + \frac{p}{e}\frac{e}{d} = \sin\alpha\cos\beta + \cos\alpha\sin\beta.$$

[2]If you don't see that $\angle PDB = \angle BAC = \alpha$ right away, look at the diagram below. You will see that both $\angle BAC$ and $\angle PDB$ are complementary to the angles marked γ, which are equal.



Exercises

1. We can also use the same diagram (Fig. 4) to derive a formula for $\cos(\alpha+\beta)$, where α, β, and $\alpha+\beta$ are acute angles. Let $AQ = q$, $CQ = r$. Fill in the gaps in the following proof:

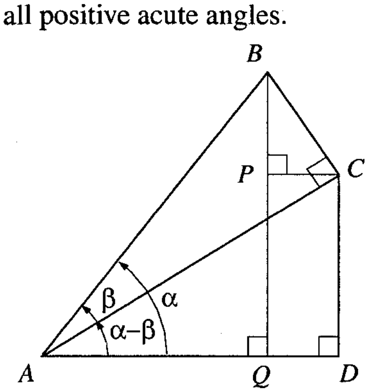
$$\begin{aligned}\cos(\alpha+\beta) &= \frac{AQ}{AD}\\ &= \frac{AC - QC}{AD}\\ &= \frac{AC}{AD} - \frac{BP}{AD}\\ &= \frac{AC}{AB}\cdot\frac{AB}{AD} - \frac{BP}{BD}\cdot\frac{BD}{AD}\\ &= \cos\alpha\cos\beta - \sin\alpha\sin\beta\,.\end{aligned}$$

Notes:

a) Here we used two different "intermediaries": AB for the first ratio and BD for the second. Again, each intermediary plays two different roles, in two different triangles.

b) The segment QD appears with a minus sign. This is how the final formula ends up having a term subtracted rather than added.

2. Derive formulas for $\sin(\alpha-\beta)$ and $\cos(\alpha-\beta)$, using the diagram below, in terms of $\sin\alpha$, $\sin\beta$, $\cos\alpha$, and $\cos\beta$, assuming that α, β, and $\alpha-\beta$ are all positive acute angles.



Here are the formulas to derive:

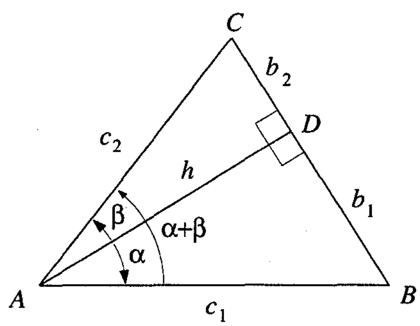
$$\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta\,,$$
$$\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta\,.$$

5 A second beautiful proof

For our second proof, we use the following theorem from Chapter 3 (see page 75).

Theorem The area of a triangle is equal to half the product of two sides and the sine of the angle between them.

In our diagram (see page 126) we have two right triangles, one including an acute angle α and the other including an acute angle β. If we place them so that they have a common side, then we get a new triangle, with one angle equal to $\alpha + \beta$:



In this new triangle, the common leg of the two right triangles is an altitude, labeled h in the diagram. Each original hypotenuse is a side of the new triangle, labeled c_1 and c_2 in the diagram.

The theorem above tells us that the area of the new triangle is

$$(1/2)c_1c_2 \sin(\alpha + \beta) .$$

Let us also calculate the area of this triangle using methods of elementary geometry. The comparison of these two results will give us our formula.

Let $BD = b_1$ and $DC = b_2$. In right triangle ABD, we have

$$\frac{b_1}{c_1} = \sin\alpha ,$$

so $b_1 = c_1 \sin\alpha$. Similarly, in triangle ACD, $b_2 = c_2 \sin\beta$. Also, (from right triangle ABD), $h = c_1 \cos\alpha$, and (from right triangle ADC) $h = c_2 \cos\beta$.

Using these relationships, we can express the area of triangle ABC as:[3]

$$\begin{aligned}\tfrac{1}{2}AD \cdot BC = \tfrac{1}{2}h(b_1 + b_2) &= \tfrac{1}{2}hb_1 + \tfrac{1}{2}hb_2 \\ &= \tfrac{1}{2}c_2 \cos\beta c_1 \sin\alpha + \tfrac{1}{2}c_2 \cos\alpha c_1 \sin\beta \,.\end{aligned}$$

Equating our two expressions for the area of triangle ABC, we have

$$\tfrac{1}{2}c_1c_2 \sin(\alpha + \beta) = \tfrac{1}{2}c_1c_2 \cos\beta \sin\alpha + \tfrac{1}{2}c_1c_2 \cos\alpha \sin\beta \,.$$

Finally, dividing through by $(1/2)c_1c_2$ gives us the desired result. Note that $\alpha + \beta$ need not be acute for the proof to be correct (although α and β must be acute).

Exercises

1. If $\alpha = 30°$ and $\beta = 30°$, what values do our formulas give us for $\sin(\alpha + \beta)$ and $\cos(\alpha + \beta)$? Do these values agree with the values that you already know?

2. If $\sin\alpha = 3/5$ and $\sin\beta = 5/13$, what values do our formulas give us for $\sin(\alpha + \beta)$ and $\cos(\alpha + \beta)$?

3. Show that $\sin 75° = \dfrac{\sqrt{6} + \sqrt{2}}{4}$ and $\cos 75° = \dfrac{\sqrt{6} - \sqrt{2}}{4}$.

4. Find expressions in radicals (similar to those in Problem 3) for $\sin 15°$ and $\cos 15°$. Explain the coincidences.

5. a) Suppose α and β are acute angles. Can $\cos(\alpha + \beta)$ be zero?
   b) Suppose α and β are acute angles. Can $\sin(\alpha + \beta)$ be zero? Remember that neither 0 nor $\pi/2$ are considered acute.
   c) We know that if α and β are acute angles, then $\sin\alpha$, $\sin\beta$, $\cos\alpha$ and $\cos\beta$ are all positive. For acute angles α and β, must $\sin(\alpha + \beta)$ be positive? Must $\cos(\alpha + \beta)$ be positive?

6. Phoebe set out to prove the identity

$$\sin^2\alpha - \sin^2\beta = \sin(\alpha + \beta)\sin(\alpha - \beta)\,.$$

[3] Remember that the area of a triangle is half the product of any altitude and the side to which it is drawn.

She reasoned as follows:

$$\begin{aligned}\sin^2\alpha - \sin^2\beta &= (\sin\alpha + \sin\beta)(\sin\alpha - \sin\beta)\\ &= \sin(\alpha+\beta)\sin(\alpha-\beta)\,.\end{aligned}$$

What criticism do you have of her reasoning?

7. Check the identity in Problem 6 using $\alpha = 30°$, $\beta = 60°$, on your calculator. You will find that, despite Phoebe's specious reasoning, the identity is true for these values. Is this a coincidence?

8. Prove that $\sin^2\alpha - \sin^2\beta = \sin(\alpha+\beta)\sin(\alpha-\beta)$.

9. Prove that $\cos^2\beta - \cos^2\alpha = \sin(\alpha+\beta)\sin(\alpha-\beta)$.

10. Without using your calculator, find the numerical value of $\sin 18° \cos 12° + \cos 18° \sin 12°$.

11. a) Without using your calculator, try to find the numerical value of $\sin 113° \cos 307° + \cos 113° \sin 307°$.
    b) Now use your calculator to check the result.
    c) Did you use the addition formulas in part (a)? Remember that we have proved the addition formulas only for positive acute angles. Doesn't it look like they work for larger angles as well?

12. Simplify the expression $\sin 2\alpha \cos\alpha - \cos 2\alpha \sin\alpha$.

13. Simplify the expression $\sin(\alpha+\beta)\sin\beta + \cos(\alpha+\beta)\cos\beta$.

14. Simplify the expression

$$\frac{\sin(\alpha+\beta) - \cos\alpha\sin\beta}{\cos(\alpha+\beta) + \sin\alpha\sin\beta}\,.$$

15. For any angle $\alpha < \pi/4$, show that

$$\sin\left(\alpha + \frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}(\sin\alpha + \cos\alpha)\,.$$

16. For any acute angles α and β for which $\cos\alpha\cos\beta \neq 0$, show that

$$\frac{\cos(\alpha+\beta)}{\cos\alpha\cos\beta} = 1 - \tan\alpha\tan\beta\,.$$

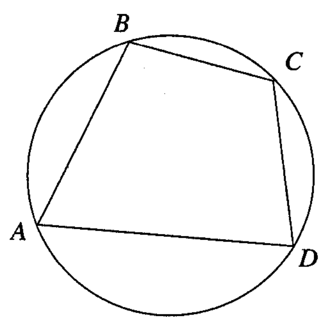
17. Use the law of cosines and the figure drawn for the second beautiful proof to give a direct derivation of the formula for $\cos(\alpha+\beta)$.

Appendix – Ptolemy's theorem and its connection with the addition formulas

In this appendix we explore the connection between the formula for $\sin(\alpha + \beta)$ and a remarkable geometric theorem of Ptolemy.

1. The angles of a quadrilateral inscribed in a circle

Ptolemy's theorem concerns quadrilaterals that are inscribed in circles. Suppose we have a quadrilateral $ABCD$, and we want to inscribe it in a circle. This is not always possible. In fact, if there is such a circle, then $\angle A + \angle B = \angle C + \angle D = \pi$.



Indeed, $\angle A = \frac{1}{2}\widehat{BCD}$, and $\angle C = \frac{1}{2}\widehat{BAD}$, so $\angle A + \angle C = \frac{1}{2}(\widehat{BCD} + \widehat{BAD}) = \frac{1}{2}(2\pi) = \pi$, and similarly, $\angle B + \angle D = \pi$.

We can also show that this condition is sufficient: If the opposite angles of a quadrilateral are supplementary, then the quadrilateral can be inscribed in a circle.

To prove this, let us take some quadrilateral $ABCD$ in which $\angle B + \angle D = \pi$, and draw a circle through A, B, and C (we know that any three non-collinear points lie on a circle).

Then we can show that point D also lies on this circle. Indeed, $\angle B = \frac{1}{2}\widehat{AC}$ (the arc not containing point B), so $\widehat{ABC} = 2\pi - \widehat{AC} = 2\pi - 2\angle B$. Point D is on the circle if $\angle D = \frac{1}{2}\widehat{ABC}$ (see page 65). But in fact this is true, since $\frac{1}{2}\widehat{ABC} = \pi - \angle B = \angle D$.

So we have the following results:

Theorem A quadrilateral can be inscribed in a circle if and only if its opposite angles are supplementary.

Example 49 Suppose we want to inscribe a parallelogram in a circle. The result above tell us that its opposite angles must be supplementary, so this parallelogram must be a rectangle. Then the intersection of its diagonals will be the center of the circle, and half the diagonal will be its radius. □

2. The sides of a quadrilateral inscribed in a circle

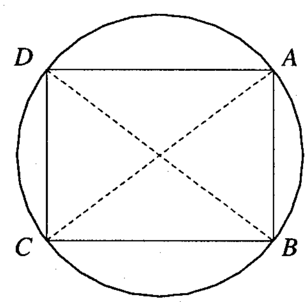
The theorem of the last section characterizes inscribed quadrilaterals in terms of their angles. Ptolemy's theorem characterizes them in terms of the length of their sides.

A quadrilateral has four vertices, and so pairs of vertices determine six lengths. Four of these lengths are sides of the quadrilateral, and two of these lengths are diagonals. Ptolemy's theorem will use these six lengths to tell us whether or not the quadrilateral can be inscribed in a circle.

Ptolemy's Theorem A quadrilateral can be inscribed in a circle if and only if the product of its diagonals equals the sum of the products of its opposite sides.

That is, quadrilateral $ABCD$ can be inscribed in a circle if and only if $AB \times CD + AD \times BC = AC \times BD$.

Example 50 What does Ptolemy's theorem tell us for a rectangle? We know that a rectangle can be inscribed in a circle.



If the rectangle is $ABCD$, then we have

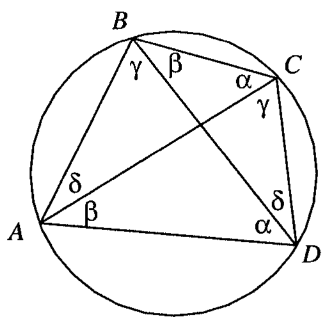
$$AB \times CD + AD \times BC = AC \times BD\,, \quad \text{or}$$

$$AB^2 + BC^2 = AC^2\,.$$

That is, Ptolemy's theorem here reduces to the theorem of Pythagoras.

We will not give a geometric proof of Ptolemy's theorem here. Rather, we will show that it is equivalent to the addition formula for $\sin(\alpha + \beta)$.

Ptolemy's theorem concerns the sides of a quadrilateral. Trigonometry, of course, works with angles. So our first job is to reformulate Ptolemy's theorem in terms of angles. Let us take a quadrilateral inscribed in a circle of diameter 1.



We know (Chapter 0, page 62) that in such a circle, the length of a chord is equal to the sine of its inscribed angle. If we look at the inscribed angles in the diagram, we find pairs of equal angles. These are labeled with the same Greek letter.

If we have four points A, B, C, and D, then we can divide them into pairs in three different ways:

$$\begin{array}{cc} AB & CD \\ AC & BD \\ AD & BD \end{array}$$

Each pair of points determines a length. If we take the product of these lengths, then Ptolemy's theorem says that a circle exists passing through the four points if and only if the sum of two of these products minus the third equals 0. Similarly, if we have n points, a necessary and sufficient condition that they lie on a circle is that the condition of Ptolemy's theorem is fulfilled for every choice of four of the given points.

Now we can "translate" the lengths of the quadrilateral's sides into trigonometric expressions. We have

$$AB = \sin\alpha \qquad BC = \sin\delta$$

$$CD = \sin\beta \qquad DA = \sin\gamma\ .$$

What about the diagonals? Diagonal BD is subtended by $\angle BAD$, and AC by $\angle ABC$, so we have

$$BD = \sin(\delta+\beta) = \sin(\gamma+\alpha)$$

$$AC = \sin(\alpha+\delta) = \sin(\beta+\gamma)\ .$$

Now we can write Ptolemy's theorem in trigonometric form:

$$AB \times CD + AD \times BC = AC \times BD$$

$$\sin\alpha \sin\beta + \sin\gamma \sin\delta = \sin(\beta+\gamma)\sin(\alpha+\gamma)\ .$$

Let us put this another way. If we have four angles α, β, γ, δ such that $\alpha+\beta+\gamma+\delta=\pi$, then we can divide a circle of diameter 1 into arcs of length $2\alpha, 2\beta, 2\gamma, 2\delta$ and use this circle to recreate the above figure. Since the resulting quadrilateral is inscribed in a circle, we have:

Ptolemy's Identity If $\alpha+\beta+\gamma+\delta=\pi$, then $\sin\alpha\sin\beta+\sin\gamma\sin\delta = \sin(\alpha+\gamma)\sin(\beta+\gamma)$.

This statement is equivalent to the part of Ptolemy's theorem that says that if a quadrilateral is inscribed in a circle, then the product of the diagonals equals the sum of the products of the opposite sides.

Ptolemy's theorem is a bit more general than the usual addition formula for $\sin(\alpha+\beta)$, and looks a bit nicer, since it uses only sines, and not cosines.

3. Ptolemy's identity implies the addition formula for sines

What happens to our old formula for $\sin(\alpha+\beta)$? It is a particular case of Ptolemy's identity. Indeed, suppose, in quadrilateral $ABCD$, $\alpha+\delta = \beta+\gamma = \pi/2$. Then $\sin(\beta+\gamma)=1$, and because $\alpha+\delta=\pi/2$, we have $\sin\delta=\cos\alpha$. And since $\beta+\gamma=\pi/2$, we have $\sin\beta=\cos\gamma$. For this special case, Ptolemy's identity reduces to

$$\sin\alpha\cos\gamma+\cos\alpha\sin\gamma = 1\cdot\sin(\alpha+\gamma)\ ,$$

which is the usual addition formula. Thus Ptolemy's theorem implies Ptolemy's identity, which implies the addition formula for sines. □

4. The addition formulas imply Ptolemy's theorem

Suppose we know the addition formulas for $\sin(\alpha+\beta)$ and $\cos(\alpha+\beta)$. Let us show that we can use them to prove Ptolemy's identity.

In Ptolemy's identity, every term is the product of two sines. In order to derive this identity from the addition formulas, we need to convert these products into sums. The reader is invited to verify, using the formulas for $\cos(x \pm y)$, that

$$\sin\alpha\sin\beta = \tfrac{1}{2}\big[\cos(\alpha-\beta)-\cos(\alpha+\beta)\big]\ , \qquad (1)$$

and to recall that

$$\cos(\pi + \alpha) = -\cos\alpha\,. \tag{2}$$

We use these results to prove Ptolemy's identity.

We want to show that if $\alpha + \beta + \gamma + \delta = \pi$, then

$$\sin\alpha \sin\beta + \sin\gamma \sin\delta = \sin(\beta + \gamma)\sin(\alpha + \gamma)\,.$$

The right side is a product of two sines. We use (1) to convert this to a sum of cosines:

$$\begin{aligned}\sin(\beta + \gamma)\sin(\alpha + \gamma) &= \tfrac{1}{2}\big[\cos\big((\beta + \gamma) - (\alpha + \gamma)\big)\\ &\qquad - \cos\big((\beta + \gamma) + (\alpha + \gamma)\big)\big]\\ &= \tfrac{1}{2}\big[\cos(\beta - \alpha) - \cos(\alpha + \beta + 2\gamma)\big]\,.\end{aligned}$$

We must do something about the expression $\alpha + \beta + 2\gamma$. We have $\alpha + \beta + 2\gamma = \alpha + \beta + \gamma + \delta + \gamma - \delta = \pi + \gamma - \delta$, and by (2), $\cos(\alpha + \beta + 2\gamma) = \cos(\pi + \gamma - \delta) = -\cos(\gamma - \delta)$. So if $\alpha + \beta + \gamma + \delta = \pi$, we have

$$\sin(\alpha + \beta)\sin(\alpha + \delta) = \tfrac{1}{2}\big[\cos(\beta - \alpha) + \cos(\gamma - \delta)\big]\,.$$

Since the cosine is an even function, we can write this as

$$\sin(\alpha + \beta)\sin(\alpha + \delta) = \tfrac{1}{2}\big[\cos(\alpha - \beta) + \cos(\gamma - \delta)\big]\,.$$

Now let us look at the left side of Ptolemy's identity. We have

$$\begin{aligned}\sin\alpha \sin\beta + \sin\gamma \sin\delta &= \tfrac{1}{2}\big[\cos(\alpha - \beta) - \cos(\alpha + \beta)\big]\\ &\qquad + \tfrac{1}{2}\big[\cos(\gamma - \delta) - \cos(\gamma + \delta)\big]\,.\end{aligned}$$

Now if $x + y = \pi$, we know that $\cos y = -\cos x$. Here, $(\alpha + \beta) + (\gamma + \delta) = \pi$, so $\cos(\gamma + \delta) = -\cos(\alpha + \beta)$, and we can write

$$\begin{aligned}\sin\alpha \sin\beta + \sin\gamma \sin\delta &= \tfrac{1}{2}\big[\cos(\alpha - \beta) - \cos(\alpha + \beta)\big]\\ &\qquad + \tfrac{1}{2}\big[\cos(\gamma - \delta) + \cos(\alpha + \beta)\big]\\ &= \tfrac{1}{2}\big[\cos(\alpha - \beta) + \cos(\gamma - \delta)\big]\,.\end{aligned}$$

But this is the same expression that we found equal to the left side. So Ptolemy's identity follows from the formulas for sine and cosine. □

Chapter 7

Trigonometric Identities

1 Extending the identities

Let us look back at some of our trigonometric identities. We first noted that $\sin^2\alpha+\cos^2\alpha = 1$ for any acute angle α. When we extended the definition of $\sin\alpha$ and $\cos\alpha$ to angles greater than $90°$ and less than $0°$, we noted that the identity still held true.

We have shown (in Chapter 6) that

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$$

for α and β positive acute angles. Is this identity still true for any angle at all?

Using the definition from Chapter 4, we can see that this formula works, for example, when $\alpha = 150°$ and $\beta = 300°$. And in fact, it will always work for angles of any size. Why is this true?

2 The Principle of Analytic Continuation: Higher mathematics to the rescue

Checking the formula for $\sin(\alpha+\beta)$ for general angles becomes very tedious. You can try it for other angles, reducing each sine or cosine to a function of a positive acute angle. But pack a lunch, because such a procedure takes a long time.

For this situation, a theorem from higher mathematics comes to our rescue. Called the Principle of Analytic Continuation, it says, roughly, that most of our identities will be preserved under the new definitions of the trigonometric functions.

More precisely, the Principle of Analytic Continuation says that *any identity involving rational trigonometric functions that is true for positive acute angles is true for any angle at all.*

Since a proof of this statement will involve results from a course in calculus and another in complex analysis, we will only state this principle here. But to understand the statement above, we must explore some terminology. A *rational trigonometric function* is a function you can get by taking the sine and cosine of various angles, together with all the constant functions, and adding, subtracting, multiplying, or dividing them.[1] Some examples of rational trigonometric functions are:

$$\frac{2\sin\alpha+3\cos\alpha}{3\sin\alpha-2\cos\alpha}, \qquad \sin(\alpha+\beta), \qquad \frac{2\sin\alpha+3\cos\beta}{3\sin\alpha-2\cos\beta},$$

$$\sin\alpha\cos\beta+\cos\alpha\sin\beta, \qquad \frac{\cos x+\sqrt{3}\sin x}{2}, \qquad \tan\alpha,$$

$$\frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}.$$

Here are some examples which *are not* rational trigonometric functions:

$$\sqrt{\sin x}, \qquad \sqrt{1-\sin^2 x}, \qquad \sqrt{\sin^2 x-3},$$

$$\log(\sin x), \qquad \cos(\sin x), \qquad \frac{\sqrt{\sin x}}{1-\cos x}.$$

Some of our examples should seem familiar to you. In fact, you can check that most of our identities so far have involved rational trigonometric functions.

The *Principle of Analytic Continuation* tells us that *if two such trigonometric rational functions are equal for numbers in any one interval (all the numbers between two real numbers) then they are equal for any numbers.*

For example, in our list above of rational trigonometric functions, we have the examples $\sin(\alpha+\beta)$ and $\sin\alpha\cos\beta+\cos\alpha\sin\beta$. Using geometry, we have already proved (three times!) that these two functions are equal for $0° < \alpha, \beta < 45°$ (so that α, β, and $\alpha+\beta$ are all acute angles). The Principle of Analytic Continuation says that these two functions must

[1] In the same way, if you start with integers, you can get all the *rational numbers* by adding, subtracting, multiplying, and dividing.

then be equal for any values of α and β, and not just for the ones in the interval between 0° and 45°.

Exercises

1. For each of the functions below, state whether or not it is a rational function of $\sin\alpha$:

   a) $\sqrt{2}\sin\alpha$ b) $\sqrt{2\sin\alpha}$ c) $\dfrac{1-\sin^2\alpha}{\sin\dfrac{\pi}{2}}$

   d) $\dfrac{1}{1+\dfrac{1}{\sin\alpha}}$ e) $\sqrt{1-\sin^2\alpha}$ f) $\sqrt{\dfrac{1-\cos\alpha}{2}}$.

2. Write each of the following expressions as rational functions of sines and cosines:

   a) $\tan\alpha$ b) $(1+\tan\alpha)(1-\tan\alpha)$

   c) $\dfrac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}$ d) $\tan^2\alpha+\cot^2\alpha$

   e) $\tan\alpha\cot\alpha$ f) $1+\tan^2\alpha$.

3. For any angle α between 0 and $\pi/2$, we know $\cos\alpha=\sqrt{1-\sin^2\alpha}$. Does the Principle of Analytic Continuation guarantee that this statement is true for any angle? For example, is this identity correct if $\alpha=2\pi/3$?

4. For any angle α between 0 and $\pi/2$, we know $\sin^2\alpha+\cos^2\alpha=1$. Does the Principle of Analytic Continuation guarantee that this statement is true for any angle? For example, is this identity correct if $\alpha=2\pi/3$?

3 Back to our identities

You may imagine that a general statement such as the Principle of Analytic Continuation (and the full statement of this principle is even more general!) must have its roots in rather deep properties of functions. And in fact it does. This is why one needs to follow two advanced courses of mathematics before understanding it fully.

So we can continue to work with our identities, with the assurance of the mathematicians, who have proved the Principle of Analytic Continuation, that our work is valid for angles of any measure, and not just for positive acute angles.

Here, once again, are our formulas. We repeat them to emphasize their added meaning. Because of the Principle of Analytic Continuation, they are true for angles of any measure, and not just acute angles:

$$\boxed{\begin{aligned}
\sin(\alpha+\beta) &= \sin\alpha\cos\beta+\cos\alpha\sin\beta\\
\sin(\alpha-\beta) &= \sin\alpha\cos\beta-\cos\alpha\sin\beta\\
\cos(\alpha+\beta) &= \cos\alpha\cos\beta-\sin\alpha\sin\beta\\
\cos(\alpha-\beta) &= \cos\alpha\cos\beta+\sin\alpha\sin\beta
\end{aligned}}$$

Exercises

1. If α and β are acute angles such that $\sin\alpha = 3/5$ and $\sin\beta = 5/13$, find the numerical value of $\sin(\alpha+\beta)$ amd $\cos(\alpha+\beta)$. In what quadrant does the angle $\alpha+\beta$ lie?

2. If α and β are acute angles such that $\sin\alpha = 4/5$ and $\sin\beta = 12/13$, find the numerical value of $\sin(\alpha+\beta)$ and $\cos(\alpha+\beta)$. In what quadrant does the angle $\alpha+\beta$ lie?

3. If α and β are angles such that $\sin\alpha = 3/5$ and $\sin\beta = 5/13$, find $\sin(\alpha+\beta)$. (Note that we don't specify here that α and β are acute angles.) How many possible answers are there?

4. Verify that $\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta$ for:

   a) $\alpha = \dfrac{2\pi}{3}, \beta = \dfrac{\pi}{3}$.

   b) $\alpha = \dfrac{\pi}{4}, \beta = \dfrac{3\pi}{4}$.

   c) $\alpha = -\dfrac{\pi}{6}, \beta = \dfrac{3\pi}{2}$.

5. Show that $\cos^2\alpha + \cos^2(2\pi/3+\alpha) + \cos^2(2\pi/3-\alpha) = 3/2$.

6. Show that $\sin(x+y) + \sin(x-y) = 2\sin x\cos y$.

7. Simplify $\cos(x+y) + \cos(x-y)$.

8. Show that $\cos(x+y)\cos(x-y) = \cos^2 x \cos^2 y - \sin^2 x \sin^2 y$.
9. Show that $\sin(x+y)\sin(x-y) = \sin^2 x \cos^2 y - \cos^2 x \sin^2 y$.
10. Using the previous two exercises, show that

$$\cos(x+y)\cos(x-y) - \sin(x+y)\sin(x-y) = \cos^2 x - \sin^2 x\,.$$

Note that the left side depends on both x and y, but the right side depends only on x.

Remark We can simplify the expression

$$\cos(x+y)\cos(x-y) - \sin(x+y)\sin(x-y)$$

in another way. Let us put $A = x+y$, $B = x-y$. Then we have $\cos(x+y)\cos(x-y) + \sin(x+y)\sin(x-y) = \cos A \cos B - \sin A \sin B$. But this is just $\cos(A+B)$. However, $A+B = (x+y)+(x-y) = 2x$. Hence,

$$\cos(x+y)\cos(x-y) + \sin(x+y)\sin(x-y) = \cos 2x\,.$$

We see once more that the value of the expression

$$\cos(x+y)\cos(x-y) + \sin(x+y)\sin(x-y)$$

is independent of y.

11. We now have a slight misunderstanding. From Exercise 10 we see that the expression we are interested in equals $\cos^2 x - \sin^2 x$. And in the remark to that same exercise we see that it is equal to $\cos 2x$. Is this an error? Try to prove that it is not.
12. Show that $\cos(\alpha+\beta)\cos\beta + \sin(\alpha+\beta)\sin\beta$ does not depend on β.

4 A formula for $\tan(\alpha + \beta)$

Let us now show that $\tan(\alpha+\beta) = \dfrac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta}$. We use the addition formulas to write:

$$\tan(\alpha+\beta) = \frac{\sin(\alpha+\beta)}{\cos(\alpha+\beta)} = \frac{\sin\alpha\cos\beta + \cos\alpha\sin\beta}{\cos\alpha\cos\beta - \sin\alpha\sin\beta}.$$

We now can divide the numerator and denominator by $\cos\alpha\cos\beta$:

$$\begin{aligned}\frac{\sin\alpha\cos\beta + \cos\alpha\sin\beta}{\cos\alpha\cos\beta - \sin\alpha\sin\beta} &= \frac{\dfrac{\sin\alpha\cos\beta}{\cos\alpha\cos\beta} + \dfrac{\cos\alpha\sin\beta}{\cos\alpha\cos\beta}}{\dfrac{\cos\alpha\cos\beta}{\cos\alpha\cos\beta} - \dfrac{\sin\alpha\sin\beta}{\cos\alpha\cos\beta}} \\ &= \frac{\dfrac{\sin\alpha}{\cos\alpha} + \dfrac{\sin\beta}{\cos\beta}}{1 - \dfrac{\sin\alpha\sin\beta}{\cos\alpha\cos\beta}}.\end{aligned}$$

This leads to:

$$\boxed{\tan(\alpha+\beta) = \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta}}$$

In a way, this is nicer than the formula for $\sin(\alpha+\beta)$ and $\cos(\alpha+\beta)$, since it uses only the tangents of α and β. The formula for $\sin(\alpha+\beta)$, on the other hand, uses $\cos\alpha$ and $\cos\beta$ as well as $\sin\alpha$ and $\sin\beta$.

Exercises

1. Check that our formula for $\tan(\alpha+\beta)$ is correct for $\alpha = 7\pi/6$, $\beta = 5\pi/3$.

2. Find a formula for $\tan(\alpha-\beta)$ in terms of $\tan\alpha$ and $\tan\beta$.

3. Show that
$$\tan\left(\frac{\pi}{4}+\alpha\right) = \frac{1+\tan\alpha}{1-\tan\alpha}.$$

4. Show that
$$\tan\left(\frac{\pi}{4}-\alpha\right) = \frac{1-\tan\alpha}{1+\tan\alpha}.$$

5. If $\alpha+\beta = \pi/4$, prove that $(1+\tan\alpha)(1+\tan\beta) = 2$.

6. Find an expression for $\tan(\alpha+\beta+\gamma)$ which involves only $\tan\alpha$, $\tan\beta$, and $\tan\gamma$.

7. Using the result of Problem 6, or otherwise, show that if $\alpha+\beta+\gamma=\pi$ (for example, if they are the three angles of a triangle), then $\tan\alpha+\tan\beta+\tan\gamma=\tan\alpha\tan\beta\tan\gamma$.

8. Show that $\tan\alpha\tan 2\alpha\tan 3\alpha = \tan 3\alpha - \tan 2\alpha - \tan\alpha$ whenever all these expressions are defined. For what values of α are some of these expressions not defined?

5 Double the angle

If we know $\sin\alpha$ and $\cos\alpha$, we can find the value of $\sin 2\alpha$ and $\cos 2\alpha$.

We know that

$$\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta$$

and

$$\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta\,.$$

Let $\alpha=\beta$. Then we have:

$$\begin{aligned}\sin 2\alpha &= \sin(\alpha+\alpha) &&= \sin\alpha\cos\alpha+\cos\alpha\sin\alpha\\ & &&= 2\sin\alpha\cos\alpha\,,\\ \cos 2\alpha &= \cos(\alpha+\alpha) &&= \cos\alpha\cos\alpha-\sin\alpha\sin\alpha\\ & &&= \cos^2\alpha-\sin^2\alpha\,.\end{aligned}$$

The formula for $\cos 2\alpha$ is particularly interesting. Since $\cos^2\alpha = 1 - \sin^2\alpha$, we can write $\cos 2\alpha = 1-2\sin^2\alpha$.

Similarly, since $\sin^2\alpha = 1-\cos^2\alpha$, we can write $\cos 2\alpha = 2\cos^2\alpha - 1$. The reader is invited to check these computations.

So we have four beautiful and useful formulas:

$$\boxed{\begin{aligned}\sin 2\alpha &= 2\sin\alpha\cos\alpha\\ \cos 2\alpha &= \cos^2\alpha-\sin^2\alpha\\ \cos 2\alpha &= 2\cos^2\alpha-1\\ \cos 2\alpha &= 1-2\sin^2\alpha\end{aligned}}$$

Problem 1. If $\cos\alpha = \sqrt{3}/2$, find $\cos 2\alpha$.

Solution. We have

$$\cos 2\alpha = 2\cos^2\alpha - 1 = 2\left(\frac{\sqrt{3}}{2}\right)^2 - 1 = \frac{1}{2}.$$

(The reader should check that the other two formulas for $\cos 2\alpha$ lead to the same answer.) □

Problem 2. If $\cos\alpha = \sqrt{3}/2$, find $\sin 2\alpha$.

Solution. We have $\sin 2\alpha = 2\sin\alpha\cos\alpha$, and before we go any further we must compute $\sin\alpha$. But $\sin\alpha$ is not uniquely determined. (After all, we are not given the value of α, but only of $\cos\alpha$. More than one angle has a cosine equal to $\sqrt{3}/2$.)

To compute $\sin\alpha$, we recall that

$$\sin^2\alpha = 1 - \cos^2\alpha = 1 - \left(\frac{\sqrt{3}}{2}\right)^2 = \frac{1}{4},$$

so

$$\sin\alpha = \pm\frac{1}{2}.$$

Then

$$\sin 2\alpha = 2\sin\alpha\cos\alpha = 2\left(\pm\frac{1}{2}\right)\left(\frac{\sqrt{3}}{2}\right) = \pm\frac{\sqrt{3}}{2}.$$

The reader should check that in fact there are values of α for which each of our two answers is correct. If we are given the value of $\cos\alpha$, then the value of $\cos 2\alpha$ is determined, but the value of $\sin 2\alpha$ is not. (And certainly the value of α itself is not determined.) □

The "double angle" formulas are often used in the following form. If we write $\alpha = 2\beta$, then $\beta = \alpha/2$, and we have:

$$\sin\beta = 2\sin\frac{\beta}{2}\cos\frac{\beta}{2},$$

$$\cos\beta = \cos^2\frac{\beta}{2} - \sin^2\frac{\beta}{2} = 2\cos^2\frac{\beta}{2} - 1 = 1 - 2\sin^2\frac{\beta}{2}.$$

Exercises

1. a) If $\sin\alpha = 7/25$ and $\cos\alpha$ is positive, find $\sin 2\alpha$ and $\cos 2\alpha$.

   b) If $\sin\alpha = 7/25$ and $\cos\alpha$ is negative, find $\sin 2\alpha$ and $\cos 2\alpha$.

2. If $\sin\alpha$ and $\cos\alpha$ are both rational numbers, can $\sin 2\alpha$ be irrational? Can $\cos 2\alpha$? Check your answer with the examples given in the text, and with Exercise 1 above.

3. In doing a certain problem, a student accidentally wrote $\cos^2\alpha$ instead of $\cos 2\alpha$. But for the particular angle he was using, the answer turned out to be correct. What could these values of α have been? That is, for what values of α is $\cos^2\alpha = \cos 2\alpha$?

4. If $\sin\alpha + \cos\alpha = 0.2$, find the numerical value of $\sin 2\alpha$.

5. If $\sin\alpha - \cos\alpha = -0.3$, find the numerical value of $\sin 2\alpha$.

6. Show that $\cos 2\alpha \cos\alpha + \sin 2\alpha \sin\alpha = \cos\alpha$.

7. Show that $\sin 2\alpha \cos\alpha + \cos 2\alpha \sin\alpha = \sin 4\alpha \cos\alpha - \cos 4\alpha \sin\alpha$.

8. Prove that $\cos^2\alpha \le \cos 2\alpha$.

9. Express $\left(\sin(\alpha/2) - \cos(\alpha/2)\right)^2$ in terms of $\sin\alpha$ only.

10. Find the numerical value of $\sin 10^\circ \sin 50^\circ \sin 70^\circ$.

    Hint: If the value of the given expression is M, find $M\cos 10^\circ$.

11. Find the numerical value of $\cos 20^\circ \cos 40^\circ \cos 80^\circ$.

12. Show that $\sin\dfrac{\pi}{10}\cos\dfrac{\pi}{5} = \dfrac{1}{4}$.

6 Triple the angle

Let us now find formulas for $\sin 3\alpha$ and $\cos 3\alpha$.

We can write

$$\begin{aligned}\sin 3\alpha &= \sin(2\alpha+\alpha)\\ &= \sin 2\alpha\cos\alpha+\cos 2\alpha\sin\alpha\\ &= 2\sin\alpha\cos^2\alpha+(1-2\sin^2\alpha)\sin\alpha\\ &= 2\sin\alpha(1-\sin^2\alpha)+(1-2\sin^2\alpha)\sin\alpha\\ &= 3\sin\alpha-4\sin^3\alpha\,.\end{aligned}$$

(The reader should check the details.)

In the same way, we can show that

$$\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha\,.$$

Exercises

1. Complete the derivation of the formula for $\cos 3\alpha$ given above.
2. If $\sin\alpha = 3/5$, what are the possible values of $\sin 3\alpha$? Of $\cos 3\alpha$?
3. If $\cos\alpha = 4/5$, what are the possible values of $\sin 3\alpha$? Of $\cos 3\alpha$?
4. Derive formulas for $\cos 4\alpha$ in terms of (a) $\cos\alpha$ only; (b) $\sin\alpha$ only.
5. Show that $\sin 3\alpha\cos\alpha - \cos 3\alpha\sin\alpha = \sin 2\alpha$.
6. Show that
$$\frac{\sin 3\alpha}{\sin\alpha} - \frac{\cos 3\alpha}{\cos\alpha} = 2$$
for any angle α.
7. a) Show that $\sin 3\alpha = 4\sin\alpha\sin(60^\circ+\alpha)\sin(60^\circ-\alpha)$.
   b) Show that $\cos 3\alpha = 4\cos\alpha\cos(60^\circ+\alpha)\cos(60^\circ-\alpha)$.
8. Derive a formula for the ratio $\sin 4\alpha/\sin\alpha$ in terms of $\cos\alpha$.
9. Show that $\sin 3\alpha\sin^3\alpha + \cos 3\alpha\cos^3\alpha = \cos^3 2\alpha$.

7 Derivation of the formulas for $\sin\alpha/2$ and $\cos\alpha/2$

Let us now derive formulas for $\sin\alpha/2$ and $\cos\alpha/2$ in terms of trigonometric functions of α.

We being with the formula for $\cos\alpha$ in terms of $\cos\alpha/2$ (see page 146):

$$\cos\alpha = 2\cos^2\left(\frac{\alpha}{2}\right) - 1\,.$$

This can be written as $2\cos^2\left(\frac{\alpha}{2}\right) = 1 + \cos\alpha$, which leads to

$$\boxed{\cos\left(\tfrac{\alpha}{2}\right) = \pm\sqrt{\frac{1+\cos\alpha}{2}}}$$

To get a formula for $\sin\alpha/2$, we proceed similarly:

$$\cos\alpha = 1 - 2\sin^2\left(\frac{\alpha}{2}\right)\,,$$

or $2\sin^2\left(\frac{\alpha}{2}\right) = 1 - \cos\alpha$, which leads to

$$\boxed{\sin\left(\tfrac{\alpha}{2}\right) = \pm\sqrt{\frac{1-\cos\alpha}{2}}}$$

To show that $\tan\left(\frac{\alpha}{2}\right) = \pm\sqrt{\frac{1-\cos\alpha}{1+\cos\alpha}}$, we write

$$\begin{aligned}\tan\left(\frac{\alpha}{2}\right) = \frac{\sin\left(\frac{\alpha}{2}\right)}{\cos\left(\frac{\alpha}{2}\right)} &= \frac{\pm\sqrt{\frac{1-\cos\alpha}{2}}}{\pm\sqrt{\frac{1+\cos\alpha}{2}}} \\ &= \pm\sqrt{\frac{\frac{1-\cos\alpha}{2}}{\frac{1+\cos\alpha}{2}}} \\ &= \pm\sqrt{\frac{1-\cos\alpha}{2}\cdot\frac{2}{1+\cos\alpha}} \\ &= \pm\sqrt{\frac{1-\cos\alpha}{1+\cos\alpha}}\,.\end{aligned}$$

In the next section, we will see two formulas for $\tan(\alpha/2)$ that are more convenient.

Exercises

1. If $\cos\alpha = 1$, find all possible values of $\cos(\alpha/2)$. You will find that there are two possible values. Give an example of a value for α which leads to each of these values.

2. Try out the formula given above for $\cos(\alpha/2)$ if

$$\text{a) } \alpha = 60^\circ, \qquad \text{b) } \alpha = 120^\circ, \qquad \text{c) } \alpha = 240^\circ.$$

For which of these angles must we take the positive square root, and for which angles must we take the negative?

3. Fill in the following table. Note that for each of the given values of α, $\cos\alpha = \frac{1}{2}$.

α	Quadrant α?	$\alpha/2$	Quadrant $\alpha/2$?	$\cos \alpha/2$
780°				
1020°				
1140°				
1380°				
-60°				
-300°				
-420°				
-660°				
-780°				

4. Find radical expressions for $\sin 15^\circ$ and $\cos 15^\circ$.

5. Each of the "half-angle formulas" we have developed includes the square root of a trigonometric expression. Why don't we have to worry about the possibility that we are taking the square root of a negative number?

6. For positive acute angles, we can write $\cos\frac{\alpha}{2} = \sqrt{\dfrac{1+\cos\alpha}{2}}$, without the ambiguity of sign. If we could apply the Principle of Analytic Continuation to this identity, we would conclude (erroneously) that

this statement, without ambiguity of sign, was true for any angle. What is it about this identity that prevents us from applying the Principle of Analytic Continuation?

7. Suppose that the angles α, β, γ are such that $\alpha + \beta + \gamma = \pi$. (For example, α, β, γ could be the three angles of a triangle.) Show that:

   a) $\tan \frac{\alpha}{2} \tan \frac{\beta}{2} + \tan \frac{\beta}{2} \tan \frac{\gamma}{2} + \tan \frac{\gamma}{2} \tan \frac{\alpha}{2} = 1$.

   Hint: Note that $\tan \frac{\alpha+\beta}{2} = \cot \frac{\gamma}{2}$, so that
   $$\tan \frac{\alpha + \beta}{2} \tan \frac{\gamma}{2} = 1 .$$

   b) $\sin \alpha + \sin \beta + \sin \gamma = 4 \cos \frac{\alpha}{2} \cos \frac{\beta}{2} \cos \frac{\gamma}{2}$.

8 Another formula for $\tan \alpha/2$

We showed that
$$\tan \frac{\alpha}{2} = \pm\sqrt{\frac{1 - \cos \alpha}{1 + \cos \alpha}} .$$

We can write this as
$$\begin{aligned}
\tan \frac{\alpha}{2} = \pm\sqrt{\frac{1 - \cos \alpha}{1 + \cos \alpha}} &= \pm\sqrt{\left(\frac{1 - \cos \alpha}{1 + \cos \alpha}\right)\left(\frac{1 + \cos \alpha}{1 + \cos \alpha}\right)} \\
&= \pm\sqrt{\frac{1 - \cos^2 \alpha}{(1 + \cos \alpha)^2}} \\
&= \pm\sqrt{\frac{\sin^2 \alpha}{(1 + \cos \alpha)^2}} \\
&= \pm\frac{\sin \alpha}{1 + \cos \alpha} .
\end{aligned}$$

So we have another formula for $\tan(\alpha/2)$, without radicals, but with an ambiguous sign. But in fact there is a small miracle here: we don't need the ambiguous sign! This miracle can easily be understood by looking at analytic continuation.

If the angle is positive and acute, that is, between 0° and 90°, we must select the positive sign. In other words, in this case we have

$$\tan\left(\frac{\alpha}{2}\right) = \frac{\sin\alpha}{1+\cos\alpha},$$

(without the ambiguous sign). Unlike the formula we started with, this new formula is a rational trigonometric expression, so the Principle of Analytic Continuation guarantees that in fact the equation is true for any angle.

Exercises

1. In this exercise, we check the result of the section above directly. We have shown that $\sin\alpha$ is twice the product of two particular numbers (they are $\sin\alpha/2$ and $\cos\alpha/2$), and we know that $\tan\alpha/2$ is the quotient of the same two numbers. But the product and the quotient of any two numbers always have the same sign. So $\sin\alpha$ and $\tan\alpha/2$ have the same sign. How does it now follow that

$$\tan\left(\frac{\alpha}{2}\right) = \frac{\sin\alpha}{1+\cos\alpha},$$

without ambiguity of the sign?

2. Show that

$$\tan\left(\frac{\alpha}{2}\right) = \frac{1-\cos\alpha}{\sin\alpha}.$$

9 Products to sums

We can get some further useful results by working with the formulas for $\sin(\alpha+\beta)$ and $\cos(\alpha+\beta)$. For example, we can write

$$\cos(\alpha+\beta) + \cos(\alpha-\beta) = 2\cos\alpha\cos\beta.$$

This simple yet remarkable formula says that the sum of the cosines of two angles can be written as the product of the cosines of two other angles. Perhaps this is clearer if we write it as follows:

$$\boxed{\cos\alpha\cos\beta = \tfrac{1}{2}\cos(\alpha+\beta) + \tfrac{1}{2}\cos(\alpha-\beta)}$$

So the cosine function, in a rather complicated way, "converts" products to sums. You may know that the logarithm function also "converts" products to sums, although in a much simpler fashion. In fact, people used to use cosine tables, like logarithm tables, to perform tedious multiplications by turning them into addition. If you study complex analysis you will learn of the rather deep relationship between the trigonometric functions and the exponential or logarithmic functions.

In the same way, we can write

$$\boxed{\begin{aligned} \sin\alpha\sin\beta &= \tfrac{1}{2}\cos(\alpha-\beta) - \tfrac{1}{2}\cos(\alpha+\beta) \\ \sin\alpha\cos\beta &= \tfrac{1}{2}\sin(\alpha+\beta) + \tfrac{1}{2}\sin(\alpha-\beta) \end{aligned}}$$

Exercises

1. Prove the last two identities referred to in the text.
2. Show that $\sin 75^\circ \sin 15^\circ = \frac{1}{4}$.
3. Show that $\sin 75^\circ \cos 15^\circ = \frac{2+\sqrt{3}}{4}$.
4. Find the numerical value of

   a) $\cos 75^\circ \cos 15^\circ$, b) $\cos 75^\circ \sin 15^\circ$.

5. Show that
   $$2\cos\left(\frac{\pi}{4}+\alpha\right)\cos\left(\frac{\pi}{4}-\alpha\right) = \cos 2\alpha\,,$$
   for any angle α.
6. For any three angles α, β, γ, show that
   $$\sin(\alpha+\beta)\sin(\alpha-\beta) + \sin(\beta+\gamma)\sin(\beta-\gamma) + \sin(\gamma+\alpha)\sin(\gamma-\alpha) = 0\,.$$
7. For any three angles α, β, γ, show that
   $$\sin\alpha\sin(\beta-\gamma) + \sin\beta\sin(\gamma-\alpha) + \sin\gamma\sin(\alpha-\beta) = 0\,.$$

10 Sums to products

It is sometimes useful to convert sums of sines and cosines to products. The following series of examples shows how this can be done.

Example 51 Factor $\sin(\gamma+\delta)+\sin(\gamma-\delta)$.

Solution. We begin by using the addition formulas

$$\begin{aligned}\sin(\gamma+\delta) &= \sin\gamma\cos\delta+\sin\delta\cos\gamma\,,\\ \sin(\gamma-\delta) &= \sin\gamma\cos\delta-\sin\delta\cos\gamma\,.\end{aligned}$$

Adding, we find that $\sin(\gamma+\delta)+\sin(\gamma-\delta)=2\sin\gamma\cos\delta$, which represents a factored form of the given expression. □

Example 52 A bottle and a cork together cost \$1.10. The bottle costs \$1 more than the cork. How much does the cork cost?

Solution. It is tempting to say immediately that the bottle costs \$1 and the cork costs 10 cents, but this is incorrect. With those prices, the bottle would cost only 90 cents more than the cork.

Algebra will quickly supply the correct answer. If the price of the bottle is b, and the price of the cork is c, then we have

$$\begin{aligned}b+c &= 1.1\\ b-c &= 1\,.\end{aligned}$$

We may solve for b and c by adding these two equations. We find that $2b=2.1$, so $b=1.05$. Using this result, we know how to calculate the value of c from either equation. For example, using the first equation, we obtain $c=1.1-b=1.1-1.05=0.05$.

Thus, the bottle costs \$1.05 and the cork costs 5 cents. □

Example 53 If $x+y=a$ and $x-y=b$, express x and y separately in terms of a and b.

Solution. Proceeding as in the problem with the bottle and the cork, we add the two equations to obtain $2x=a+b$, so $x=\frac{1}{2}(a+b)$. Then, instead of adding, we can subtract the equations, to obtain $2y=a-b$, so $y=\frac{1}{2}(a-b)$. In general, we have

$$\boxed{\text{If } x+y=a \text{ and } x-y=b, \text{ then } x=\tfrac{1}{2}(a+b) \text{ and } y=\tfrac{1}{2}(a-b)}$$

Please remember this result. It will be useful in many applications of algebra and trigonometry, and not just in problems about bottles (of undetermined contents). □

Example 54 Write the expression $\sin\alpha + \sin\beta$ as a product of sines and cosines.

Solution. With the experience of the previous examples, this is not difficult to do. We may use Example 51 if we can find angles γ and δ such that $\gamma + \delta = \alpha$ and $\gamma - \delta = \beta$. Example 53 shows us how to do this. We just need to choose

$$\gamma = \frac{\alpha+\beta}{2}, \qquad \delta = \frac{\alpha-\beta}{2}.$$

Substituting into the result of Example 51, we obtain the useful formula

$$\boxed{\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}}$$ □

We may also express the difference $\sin\alpha - \sin\beta$ as a product of sines and cosines. We use the angles γ and δ found before, such that $\gamma + \delta = \alpha$ and $\gamma - \delta = \beta$, and write

$$\sin\alpha - \sin\beta = \sin(\gamma+\delta) - \sin(\gamma-\delta) = 2\cos\gamma\sin\delta\,.$$

We now express this result in terms of the original variables α and β, and find that

$$\boxed{\sin\alpha - \sin\beta = 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}}$$

In the same way, we can prove the formulas

$$\boxed{\begin{aligned}\cos\alpha + \cos\beta &= 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}\\ \cos\alpha - \cos\beta &= -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}\end{aligned}}$$

Exercises

1. Give a detailed derivation of each of the last two formulas mentioned above.

2. Show that $\cos 70° + \sin 40° = \cos 10°$.

3. Find an acute angle α such that $\cos 55° + \cos 65° = \cos\alpha$.

4. Show that $\cos 20° + \cos 100° + \cos 140° = 0$.

5. Show that $\sin 78° + \cos 132° = \sin 18°$.

6. Show that
$$\frac{\cos 15° + \sin 15°}{\cos 15° - \sin 15°} = \sqrt{3}\,.$$

7. If $\alpha + \beta + \gamma = \pi$, show that
   a) $\sin(\alpha + \beta) = \sin\gamma$.
   b) $\cos(\alpha + \beta) = -\cos\gamma$.
   c) $\sin 2\alpha + \sin 2\beta + \sin 2\gamma = 4\sin\alpha\sin\beta\sin\gamma$.

8. For any angle α, show that
$$\sin\alpha + \sin(\alpha + 2\pi/3) + \sin(\alpha + 4\pi/3) = 0\,.$$

9. For any angle α, show that
$$\sin\alpha + 2\sin 3\alpha + \sin 5\alpha = 4\cos^2\alpha\sin 3\alpha\,.$$

10. For any three angles α, β, γ, show that
$$\frac{\sin(\beta - \gamma)}{\sin\beta\sin\gamma} + \frac{\sin(\gamma - \alpha)}{\sin\gamma\sin\alpha} + \frac{\sin(\alpha - \beta)}{\sin\alpha\sin\beta} = 0\,.$$

11. For any three angles α, β, γ, show that
$$\begin{aligned}&\sin(\alpha - \beta) + \sin(\alpha - \gamma) + \sin(\beta - \gamma)\\ &\quad = 4\cos\frac{(\alpha - \beta)}{2}\sin\frac{(\alpha - \gamma)}{2}\cos\frac{(\beta - \gamma)}{2}\,.\end{aligned}$$

12. For any three angles α, β, γ, show that
$$\begin{aligned}&\sin(\alpha + \beta + \gamma) + \sin(\alpha - \beta - \gamma) + \sin(\alpha + \beta - \gamma)\\ &\quad + \sin(\alpha - \beta + \gamma) = 4\sin\alpha\cos\beta\cos\gamma\,.\end{aligned}$$

Appendix

I. 1. Expressions for $\sin\beta$**,** $\cos\beta$**, and** $\tan\beta$ **in terms of** $\tan\beta/2$**.**

We can use our results in trigonometry to obtain some results in number theory. Let us begin by reviewing some results obtained earlier.

Example 55 Show that $\tan^2\beta + 1 = 1/\cos^2\beta$.

Solution. $\tan^2\beta + 1 = \dfrac{\sin^2\beta}{\cos^2\beta} + 1 = \dfrac{\sin^2\beta + \cos^2\beta}{\cos^2\beta} = \dfrac{1}{\cos^2\beta}$. □

Example 56 Show that $\cos^2\beta = \dfrac{1}{1+\tan^2\beta}$.

Solution. This result follows from the previous one. □

Example 57 If $\tan\beta = a$, express in terms of a the value of $\sin 2\beta$.

Solution. We have

$$\begin{aligned}
\sin 2\beta &= 2\sin\beta\cos\beta \\
&= 2\sin\beta\cos\beta\frac{\cos\beta}{\cos\beta} \\
&= \frac{2\sin\beta\cos^2\beta}{\cos\beta} \\
&= 2\tan\beta\ \cos^2\beta \\
&= \frac{2\tan\beta}{1+\tan^2\beta},
\end{aligned}$$

this last because of the result of Example 56. Then, since $\tan\beta = a$, we have that

$$\sin 2\beta = \frac{2a}{1+a^2}.$$ □

In working Example 57, we have found a way to express $\sin 2\beta$ as a rational function of $\tan\beta$:

$$\sin 2\beta = \frac{2\tan\beta}{1+\tan^2\beta}.$$

Similarly, we can express $\cos 2\beta$ in terms of $\tan\beta$:

$$\begin{aligned}\cos 2\beta = \cos^2\beta - \sin^2\beta &= \left(\frac{\cos^2\beta}{\cos^2\beta} - \frac{\sin^2\beta}{\cos^2\beta}\right)(\cos^2\beta)\\ &= (1-\tan^2\beta)(\cos^2\beta)\\ &= \frac{1-\tan^2\beta}{1+\tan^2\beta}.\end{aligned}$$

We can also express $\tan 2\beta$ in terms of $\tan\beta$. The simplest way to do this is to use the formula we have derived for $\tan(\alpha+\beta)$:

$$\tan(\alpha+\beta) = \frac{\tan\alpha + \tan\beta}{1-\tan\alpha\tan\beta}.$$

Letting $\alpha = \beta$, we find that

$$\tan 2\beta = \frac{2\tan\beta}{1-\tan^2\beta}.$$

If we let $\tan\beta = a$, we can write

$$\boxed{\sin 2\beta = \frac{2a}{1+a^2}, \quad \cos 2\beta = \frac{1-a^2}{1+a^2}, \quad \tan 2\beta = \frac{2a}{1-a^2}}$$

which are all rational expressions in a.

Exercises

Using the above rational expressions, verify that:

1. $\sin^2\beta + \cos^2\beta = 1$
2. $\tan 2\beta = \sin 2\beta / \cos 2\beta$.

I.2. Uniformization of $\sin\alpha$, $\cos\alpha$, and $\tan\alpha$

We can rewrite our new identities by letting $\alpha = 2\beta$:

$$\sin\alpha = \frac{2\tan\frac{\alpha}{2}}{1+\tan^2\frac{\alpha}{2}},$$

$$\cos\alpha = \frac{1-\tan^2\frac{\alpha}{2}}{1+\tan^2\frac{\alpha}{2}},$$

$$\tan\alpha = \frac{2\tan\frac{\alpha}{2}}{1-\tan^2\frac{\alpha}{2}}.$$

These formulas provide a *uniformization* of the trigonometric functions. That is, they allow us to represent all these functions using rational expressions of a single function, $\tan\alpha/2$. So, for instance, if we have a trigonometric identity, or an equation involving trigonometric functions, we can rewrite these functions as rational functions of this single variable. Then the trigonometric equation or identity becomes an algebraic equation or identity.

While this may be important theoretically, it rarely makes things easier when we have an actual problem to solve. However, this uniformization yields some very interesting results in a most unexpected area. We can use it to find *Pythagorean triples*: solutions in natural numbers to the equation $a^2 + b^2 = c^2$.

We know that if the numbers a, b, and c form a Pythagorean triple, then there is a right triangle with legs a and b and hypotenuse c. Then each acute angle of this triangle has a rational sine, cosine, and tangent. For example, we are familiar with the fact that the numbers 3, 4, and 5 satisfy the equation $a^2 + b^2 = c^2$. We can build a triangle with legs 3 and 4 and hypotenuse 5. For the smaller acute angle α of this triangle, $\sin\alpha = 3/5$, $\cos\alpha = 4/5$, and $\tan\alpha = 3/4$.

We can use our uniformization to find other triangles with angles whose sine, cosine, and tangent are rational by following this process backwards. If we let $\tan\alpha/2$ be some rational number, then our uniformization tells us that $\sin\alpha$, $\cos\alpha$, and $\tan\alpha$ will also be rational. We can then form a right triangle with rational sides and, by scaling it up, we can form a right triangle with integer sides. The sides of this triangle will be a Pythagorean triple.

For example, let

$$\tan\frac{\alpha}{2} = \frac{2}{3}.$$

Then we have

$$\sin\alpha = \frac{\frac{4}{3}}{1+\frac{4}{9}} = \frac{12}{13},$$

$$\cos\alpha = \frac{1-\frac{4}{9}}{1+\frac{4}{9}} = \frac{5}{13},$$

$$\tan\alpha = \frac{12}{5}.$$

In this case, a right triangle with an acute angle α can have sides $12/13$, $5/13$, and 1. Multiplying each side by 13, we form a similar right triangle with sides 12, 5, and 13. Because they are the sides of a right triangle, these three natural numbers satisfy the equation $a^2 + b^2 = c^2$.

Let us do this in general. Suppose

$$\tan\frac{\alpha}{2} = \frac{p}{q}.$$

Then

$$\sin\alpha = \frac{\dfrac{p}{q}}{1+\dfrac{p^2}{q^2}} = \frac{2pq}{q^2+p^2},$$

$$\cos\alpha = \frac{q^2-p^2}{q^2+p^2}.$$

Then the triangle has rational sides $2pq/(q^2+p^2)$, $(q^2-p^2)/(q^2+p^2)$, and 1, and the triangle with integer sides has sides $2pq$, q^2-p^2, and q^2+p^2.

Exercises

1. If $\tan(\alpha/2) = 3/2$, find the values of $\sin\alpha$, $\cos\alpha$ and $\tan\alpha$. Do these values provide us with a Pythagorean triple? with an integer right triangle?
2. What right triangle with integer sides results from letting $\tan(\alpha/2) = 5/8$ in our formulas above?
3. Verify that the numbers $2pq$, q^2-p^2, and q^2+p^2 satisfy the Pythagorean relationship. Which side is the hypotenuse?

II. Themes and variations

We return to a theme that we introduced in Chapter 1, and develop it more fully.

Theme: The maximum value of $\sin x \cos x$

Variation 1: Find the largest possible value of the expression $\sin x \cos x$.

Certainly $\sin x \cos x < 1$, since both $\sin x$ and $\cos x$ are at most 1 (and cannot be equal to 1 for the same angle). But is this the best estimate?

Exercises

1. With your calculator, find the value of $\sin x \cos x$ for the following values of x:

$$20^\circ, \quad 10^\circ, \quad 5^\circ \quad 1^\circ, \quad 70^\circ, \quad 80^\circ, \quad 85^\circ, \quad 89^\circ.$$

2. Without your calculator, find the value of $\sin x \cos x$ for the following values of x:

$$30^\circ, \quad 45^\circ, \quad 60^\circ.$$

Variation 2: Perhaps you have noticed some patterns in the numerical examples above. Let us see what is going on mathematically.

The product $\sin x \cos x$ reminds us of the formula $\sin 2x = 2 \sin x \cos x$. In fact, $\sin x \cos x = \sin 2x \,/\, 2$. But $\sin 2x$, like the sine of any angle, is less than 1. Hence,

$$\sin x \cos x = \frac{\sin 2x}{2} \le \frac{1}{2}.$$

As we have seen, the value $1/2$ occurs, for example, if $x = 45^\circ$, so this is the maximum value of our expression.

Exercises

1. Find all x for which

   a) $\sin x \cos x = \frac{1}{2}$.

   b) $\sin x \cos x = \frac{\sqrt{3}}{2}$.

   c) $\sin x \cos x = \frac{\sqrt{3}}{4}$.

2. Which of the following equations has no solutions at all?

$$\begin{array}{ll} \text{a)} & \sin x \cos x = 0.4\,, \\ \text{b)} & \sin x \cos x = 0.5\,, \\ \text{c)} & \sin x \cos x = 0.6\,. \end{array}$$

3. For what values of N does the equation

$$\sin x \cos x = N$$

   have a solution? How would you solve it?

Theme: The maximum value of $\sin x + \cos x$

Variation 1: For any x, $\sin x + \cos x < 2$, of course, since each addend on the left is at most 1 (and the addends cannot equal 1 simultaneously). Can the value be as much as $\frac{1}{2}$? Certainly: if $x = 30°$, then $\sin x = \frac{1}{2}$ and $\cos x > 0$, so $\sin x + \cos x$ is certainly greater than $\frac{1}{2}$.

Exercises

1. Check that if $x = 30°$, $\sin x + \cos x$ is greater than 1.
2. Find at least one value of x for which $\sin x + \cos x = 1$.
3. Find at least one value of x for which $\sin x + \cos x = \sqrt{2}$.

Now let us do things mathematically. Notice that $(\sin x + \cos x)^2 = \sin^2 x + \cos^2 x + 2\sin x \cos x = 1 + \sin 2x$. Since the maximum value of $\sin 2x$ is 1, the maximum value of $(\sin x + \cos x)^2 = 2$, and $\sin x + \cos x \leq \sqrt{2}$.

Exercises

1. Can $\sin x + \cos x = 1.414$?
2. Can $\sin x + \cos x = 1.415$?
3. For what values of x does $\sin x + \cos x = \sqrt{2}$?
4. What is the smallest possible value of the expression $\sin x + \cos x$? For what value of x is this minimum achieved?

Variation 2: Let us find the maximum value of $\sin x + \cos x$ in a different way, by comparing this with the formula $\sin(x + a) = \sin x \cos a + \cos x \sin a$. We can do this by using a trick. We will write

$$\sin x + \cos x = \sqrt{2}\left(\frac{1}{\sqrt{2}}\sin x + \frac{1}{\sqrt{2}}\cos x\right).$$

Why do we do this strange thing? The answer is that $\frac{1}{\sqrt{2}}$ is $\sin\frac{\pi}{4}$ and also $\cos\frac{\pi}{4}$. So we can write

$$\sin x + \cos x = \sqrt{2}\left(\sin x \cos\frac{\pi}{4} + \cos x \sin\frac{\pi}{4}\right) = \sqrt{2}\sin\left(x + \frac{\pi}{4}\right).$$

Now the largest possible value for the sine of any angle is 1, so the largest possible value for $\sin x + \cos x$ is $\sqrt{2}$.

Exercises

1. For what values of x is the maximum of $\sin x + \cos x$ achieved?

2. What is the minimum possible value of $\sin x + \cos x$? When is this minimum achieved?

Variation 3: Now let us look at the expression $3\sin x + 4\cos x$. What is its maximum value? This time, it won't help to square the quantity (try it!), so we can't use our first method.

We can compare the expression $3\sin x + 4\cos x$ to $\cos a \sin x + \sin a \cos x$. But the numbers 3 and 4 are not the sine and cosine of the same angle. However, the numbers 3 and 4 remind us of our "best friend", the 3-4-5 right triangle. In fact, the larger acute angle of this triangle has a cosine of 3/5 and a sine of 4/5. So, if we call this angle α, we can write

$$\begin{aligned} 3\sin x + 4\cos x &= 5\left(\tfrac{3}{5}\sin x + \tfrac{4}{5}\cos x\right) \\ &= 5(\cos\alpha \sin x + \sin\alpha \cos x) = 5\sin(\alpha + x)\,. \end{aligned}$$

The maximum value of this expression is 5.

Exercises

1. In the above argument, must α be positive and acute?

2. What is the minimum value of $3\sin x + 4\cos x$? For what values of x does this occur?

3. What are the maximum and minimum values of $2\sin x + 7\cos x$?

   Hint: Take $\sqrt{53} = \sqrt{2^2 + 7^2}$, and investigate the corresponding question for $\sqrt{53}\left(\frac{2}{\sqrt{53}}\sin x + \frac{7}{\sqrt{53}}\cos x\right)$.

III. An approximation to π

We can use the half-angle formulas to find a numerical approximation to the number π.

Let us begin by checking our formulas for $\cos x/2$ and $\sin x/2$ when $x = \pi/2$. We have

$$\cos\frac{\pi/2}{2} = \cos\frac{\pi}{4} = \sqrt{\frac{1 + \cos\frac{\pi}{2}}{2}} = \sqrt{\frac{1+0}{2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}\,,$$

which, as we already know, is correct (note that we choose the positive sign for the radical).

Similarly, we have

$$\sin\frac{\pi/2}{2} = \sin\frac{\pi}{4} = \sqrt{\frac{1-0}{2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2},$$

which we also expected.

Now let us get radical expressions for $\cos\pi/8$ and $\sin\pi/8$:

$$\cos\frac{\pi}{8} = \cos\frac{\pi/4}{2} = \sqrt{\frac{1+\cos\frac{\pi}{4}}{2}} = \sqrt{\frac{1+\frac{\sqrt{2}}{2}}{2}} = \sqrt{\frac{2+\sqrt{2}}{4}}$$
$$= \frac{1}{2}\sqrt{2+\sqrt{2}}.$$

$$\sin\frac{\pi}{8} = \sin\frac{\pi/4}{2} = \sqrt{\frac{1-\cos\frac{\pi}{4}}{2}} = \sqrt{\frac{1-\frac{\sqrt{2}}{2}}{2}} = \sqrt{\frac{2-\sqrt{2}}{4}}$$
$$= \frac{1}{2}\sqrt{2-\sqrt{2}}.$$

Note that the expressions we get contain "nested radicals."

Exercises

1. Finish the derivations below of radical expressions for $\cos\frac{\pi}{16}$ and $\sin\frac{\pi}{16}$:

$$\cos\frac{\pi}{16} = \cos\frac{\pi/8}{2} = \sqrt{\frac{1+\cos\frac{\pi}{8}}{2}} = \cdots = \frac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2}}}.$$

$$\sin\frac{\pi}{16} = \sin\frac{\pi/8}{2} = \sqrt{\frac{1-\cos\frac{\pi}{8}}{2}} = \cdots = \frac{1}{2}\sqrt{2-\sqrt{2+\sqrt{2}}}.$$

2. Fill in the table with nested radical expressions for the values of the indicated trigonometric functions. Two of the values have been filled for you.

α	$\cos\alpha$	$\sin\alpha$
$\dfrac{\pi}{16}$	$\frac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2}}}$	
$\dfrac{\pi}{32}$		$\frac{1}{2}\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2}}}}$
$\dfrac{\pi}{64}$		
$\dfrac{\pi}{128}$		

Now we know that $\cos 0 = 1$, and it is also true that the cosine of a very small angle (one whose measure is close to 0) is close to 1. The sequence of angles

$$\frac{\pi}{2},\quad \frac{\pi}{4},\quad \frac{\pi}{8},\quad \frac{\pi}{16},\quad \ldots,\quad \frac{\pi}{2^n},\quad \ldots$$

get closer and closer to 0 (approaches 0). So it is reasonable to expect that the sequence

$$\cos\frac{\pi}{2},\quad \cos\frac{\pi}{4},\quad \cos\frac{\pi}{8},\quad \cos\frac{\pi}{16},\quad \ldots,\quad \cos\frac{\pi}{2^n},\quad \ldots$$

approaches 1. In fact, this is the case. That is, the sequence:

$$0,\quad \frac{1}{2}\sqrt{2},\quad \frac{1}{2}\sqrt{2+\sqrt{2}},\quad \frac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2}}},\quad \ldots$$

approaches 1. Mathematicians express this by writing

$$\lim_{n\to\infty}\frac{1}{2}\underbrace{\sqrt{2+\sqrt{2+\sqrt{2+\cdots+\sqrt{2}}}}}_{n \text{ radicals}} = 1\,.$$

Now let us look at another sequence:

$$\frac{\sin\frac{\pi}{2}}{\frac{\pi}{2}},\quad \frac{\sin\frac{\pi}{4}}{\frac{\pi}{4}},\quad \frac{\sin\frac{\pi}{8}}{\frac{\pi}{8}},\quad \ldots,\quad \frac{\sin\frac{\pi}{2^n}}{\frac{\pi}{2^n}},\quad \ldots$$

We have seen (Chapter 5) that for very small angles α, the ratio $\sin\alpha/\alpha$ is very close to 1. So the sequence above should approach 1. One way of saying this is to assert that the value $\sin\frac{\pi}{2^n}$ approaches the value $\frac{\pi}{2^n}$ (for large values of n), or that the value of

$$2^n \sin\frac{\pi}{2^n}$$

approaches π, and mathematicians have in fact proved this.

That is, they have shown (using our nested radical expressions for $\sin\frac{\pi}{2^n}$) that

$$\lim_{n\to\infty} 2^n \underbrace{\sqrt{2-\sqrt{2+\sqrt{2+\cdots+\sqrt{2}}}}}_{n \text{ radicals}} = \pi\,.$$

Exercises

1. Using your calculator or a computer, check that the expressions

$$\begin{aligned}&\tfrac{1}{2}\sqrt{2},\\ &\tfrac{1}{2}\sqrt{2+\sqrt{2}},\\ &\tfrac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2}}},\\ &\tfrac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2}}}},\ \ldots\end{aligned}$$

   approach 1.

2. Using your calculator or a computer, check that the expressions

$$\begin{aligned}&2^2\sqrt{2-\sqrt{2}},\\ &2^3\sqrt{2-\sqrt{2+\sqrt{2}}},\\ &2^4\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2}}}},\\ &2^5\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2}}}}},\ \ldots\end{aligned}$$

   approach π. You will have to think a bit about how to organize the computation. (The value of π is approximately $3.141592653589793\ldots$)

3. We know that

$$\cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}.$$

Show that:

a) $\cos\frac{\pi}{12} = \frac{1}{2}\sqrt{2+\sqrt{3}}.$

b) $\cos\frac{\pi}{24} = \frac{1}{2}\sqrt{2+\sqrt{2+\sqrt{3}}}.$

c) $\cos\frac{\pi}{48} = \frac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{3}}}}.$

d) $\cos\frac{\pi}{96} = \frac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{3}}}}}.$

By evaluating these expressions (with a calculator or computer), observe that they are approaching 1. Can you explain why?

4. We can find another approximation to π by finding nested radical expressions for $\sin\pi/12$, $\sin\pi/24$, $\sin\pi/48$, $\sin\pi/96$, etc. Using a calculator or computer, find the values of the expressions:

a) $12\sin\frac{\pi}{12} = 6\sqrt{2-\sqrt{3}}.$

b) $24\sin\frac{\pi}{24} = 12\sqrt{2-\sqrt{2+\sqrt{3}}}.$

c) $48\sin\frac{\pi}{48} = 24\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{3}}}}.$

d) $96\sin\frac{\pi}{96} = 48\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{3}}}}}.$

IV. Trigonometric series

In this section we use the identities we have learned to the find the sum of series whose terms involve trigonometric expressions. This topic turns out to be of great importance in later work.

We introduce some of the techniques used by first looking at some purely algebraic problems.

Example 58 Find the sum $x + x^2 + x^3 + x^4 + \cdots + x^{100}$.

Solution. Let $S = x + x^2 + x^3 + x^4 + \cdots + x^{100}$ and multiply S by x:

$$xS = x^2 + x^3 + x^4 + \cdots + x^{101}.$$

Things get very simple if we subtract

$$\begin{aligned} S - xS = S(1-x) &= x - x^2 + x^2 - \cdots + x^{100} - x^{101} \\ &= x - x^{101}. \end{aligned}$$

Most of the terms drop out, and we find that

$$S = \frac{x - x^{101}}{1-x}. \qquad \square$$

Of course, if you already know the general formula for the sum of a geometric progression, this result is not unexpected. But if you don't already know the general formula for the sum of a geometric progression, you have essentially learned it above: the general case will work in just the same way.

The key to this trick is forming a "telescoping" sum: a sum of terms in which many pairs add up to zero.

Exercises

1. Find the sum

$$\frac{1}{\sqrt{1}+\sqrt{2}} + \frac{1}{\sqrt{2}+\sqrt{3}} + \cdots + \frac{1}{\sqrt{99}+\sqrt{100}}.$$

   Hint: Rationalize the denominators to get a telescoping sum.

2. Express in terms of n the sum $1 + 3 + 5 + \cdots + (2n + 1)$.

   Hint: Write each odd integer as the difference of consecutive squares.

3. Find the product $(1 + x)(1 + x^2)(1 + x^4)(1 + x^8)(1 + x^{16})$.

   Hint: Call this product P, and multiply P by $(1 - x)$.

4. Without using your calculator, find the numerical value of the product $\cos 20^\circ \cos 40^\circ \cos 80^\circ$.

   Hint: Call this product P, and multiply P by the sine of a certain well-chosen angle.

V. Summing a trigonometric series

We would like to find the sum of the series

$$S = \sin x + \sin 2x + \sin 3x + \cdots + \sin nx\,.$$

We can form a telescoping sum, as in Example 58 above. The trick is to multiply by $2\sin(x/2)$:

$$2\sin\frac{x}{2}S = 2\sin\frac{x}{2}\sin x + 2\sin\frac{x}{2}\sin 2x + \cdots + 2\sin\frac{x}{2}\sin nx\,.$$

Now we turn the products into sums. The reader can recall, or check, that

$$\begin{aligned} 2\sin A\sin B &= \cos(A-B) - \cos(A+B) \\ &= \cos(B-A) - \cos(B+A)\,. \end{aligned}$$

So we can write

$$\begin{aligned} 2\sin\frac{x}{2}S &= 2\sin\frac{x}{2}\sin x + 2\sin\frac{x}{2}\sin 2x + \cdots + 2\sin\frac{x}{2}\sin nx \\ &= \left(\cos\tfrac{1}{2}x - \cos\tfrac{3}{2}x\right) + \cdots + \left(\cos(n-\tfrac{1}{2})x - \cos(n+\tfrac{1}{2})x\right) \\ &= \cos\tfrac{1}{2}x - \cos(n+\tfrac{1}{2})x\,, \end{aligned}$$

and so,

$$S = \frac{\cos\frac{1}{2}x - \cos(n+\frac{1}{2})x}{2\sin\frac{x}{2}}\,.$$

Sometimes this formula is more useful if we convert the sum in the numerator to a product. We find that

$$S = \frac{\sin\frac{n+1}{2}x\,\sin(\frac{n}{2}x)}{\sin\frac{x}{2}}\,.$$

This technique is quite general, and can be used to sum the sines or cosines of angles which are in *arithmetic progression*. We can find a general formula for the sum

$$S = \sin x + \sin(x+\alpha) + \sin(x+2\alpha) + \cdots + \sin(x+n\alpha)$$

by multiplying this sum by $2\sin\alpha/2$ and "telescoping" the result. We find that

$$S = \frac{\sin\frac{n+1}{2}\alpha\,\sin(x+\frac{n}{2}\alpha)}{\sin\frac{\alpha}{2}}\,.$$

Similarly, we can find a general formula for the sum

$$S = \cos x + \cos(x + \alpha) + \cos(x + 2\alpha) + \cdots + \cos(x + n\alpha),$$

again by multiplying by $2\sin\alpha/2$, and using the identity

$$2\cos A \sin B = \sin(A + B) - \sin(A - B).$$

We find that

$$C = \frac{\sin\frac{n+1}{2}\alpha \cos(x + \frac{n}{2}\alpha)}{\sin\frac{\alpha}{2}}.$$

In the following exercises, we recommend using the hints provided, then checking the results by applying the formulas directly.

Exercises

1. Find the sum

$$\sin x + \sin 3x + \sin 5x + \cdots + \sin 99x.$$

Hint: Multiply this sum by $2\sin x$.

2. Find the sum

$$\sin x + \sin(x + \frac{\pi}{4}) + \sin(x + \frac{2\pi}{4}) + \cdots + \sin(x + \frac{99\pi}{4}).$$

Hint: Multiply this sum by $2\sin(\pi/8)$.

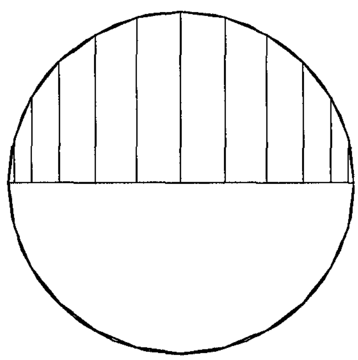
3. Find the sum

$$\cos 2x + \cos 4x + \cos 6x + \cdots + \cos 2nx.$$

4. Find the sum

$$\cos\frac{\pi}{k} + \cos\frac{2\pi}{k} + \cos\frac{3\pi}{k} + \cdots + \cos\frac{n\pi}{k}.$$

5. The diagram below shows a regular 24-sided polygon inscribed in a circle. A diameter of the circle is drawn, and perpendiculars are dropped from all the vertices of the polygon that lie on one side of this diameter. Find the sum of the lengths of these perpendiculars.



Chapter 8

Graphs of Trigonometric Functions

One of the most important uses of trigonometry is in describing periodic processes. We find many such processes in nature: the swing of a pendulum, the tidal movement of the ocean, the variation in the length of the day throughout the year, and many others.

All of these periodic motions can be described by one important family of functions, which all physicists use. These are the functions of the form

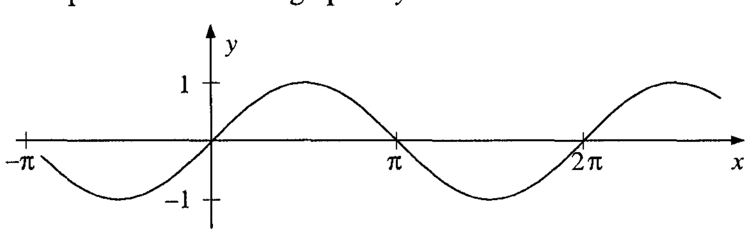
$$y = a \sin k(x - \beta)\ ,$$

where the constants a and k are positive, and β is arbitrary. In this chapter, we will describe their graphs, which we will call *sinusoidal* curves. Since they are so important, we will discuss them step-by-step, analyzing in turn each of the parameters a, k, and β.

1 Graphing the basic sine curve

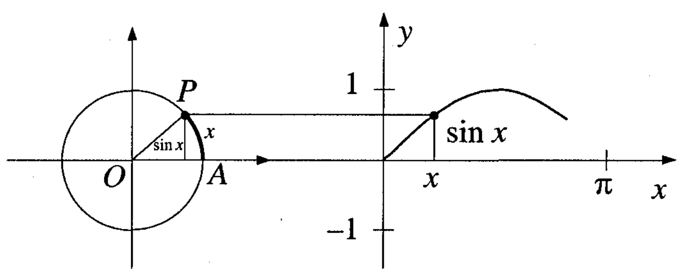
$$y = a \sin k(x - \beta) \quad \text{for } a = 1, k = 1, \beta = 0$$

In Chapter 5 we drew the graph of $y = \sin x$:



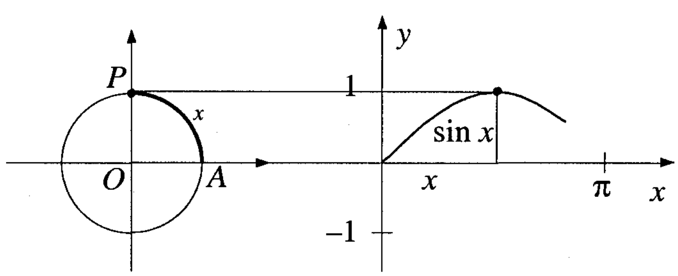
That is, we start with the case $a = 1$, $k = 1$, $\beta = 0$. Recall that we can take the sine of any real number (the *domain* of the function $y = \sin x$ is all real numbers), but that the values we get are all between -1 and 1 (the *range* of the function is the interval $-1 \leq y \leq 1$).

Let us review how we obtained this graph. On the left below is a circle with unit radius. Point P is rotating around it in a counterclockwise direction, starting at the point labeled A. If x is the length of the arc $\overparen{AP}$, then

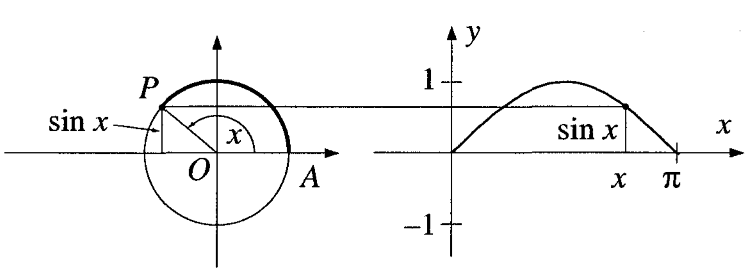


$\sin x$ is the vertical displacement of P. On the right, we have marked off the length x of arc $\overparen{AP}$. The height of the curve above the x-axis is $\sin x$.

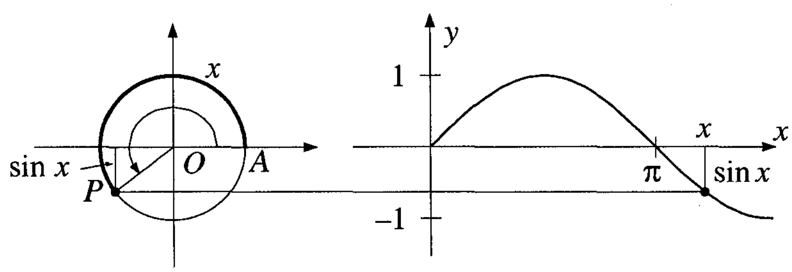
As the angle x goes from 0 to $\pi/2$, $\sin x$ grows from 0 to 1 (the picture for $x = \pi/2$ is shown below).

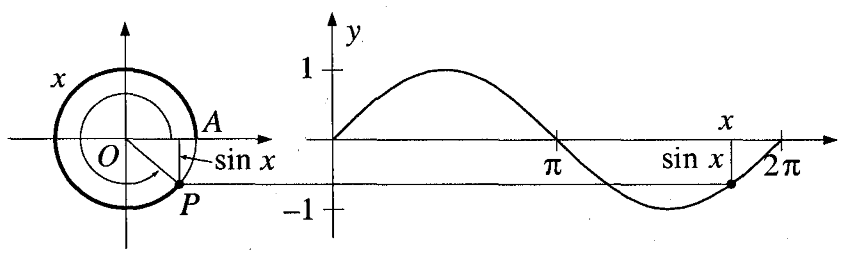


In fact, this is all we need to graph $y = \sin x$. As x goes from $\pi/2$ to π, the values of $\sin x$ repeat themselves "backwards":



And as x goes from π to 2π, the values are the negatives of the values in the first two quadrants:





2 The period of the function $y = \sin x$

As x grows larger than 2π, the values of $\sin x$ repeat on intervals of length 2π. For this reason, we say that the function $y = \sin x$ is *periodic*, with period 2π. Geometrically, this means that if we shift the whole graph 2π units to the right or to the left, we will still have the same graph. Algebraically, this means that

$$\sin(x + 2\pi) = \sin x$$

for any number x.

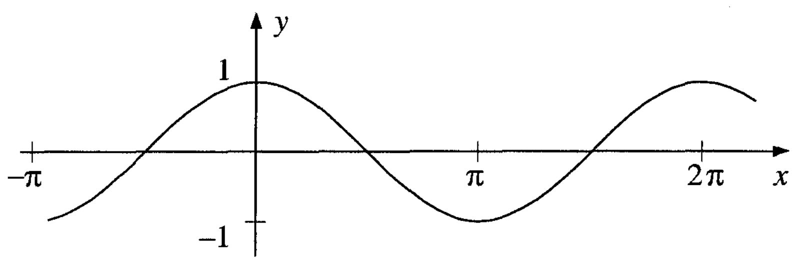
Definition: A function f has a period p if $f(x) = f(x + p)$ for all values of x for which $f(x)$ and $f(x + p)$ are defined.

The function $y = \sin x$ has a period of 2π. You can check that it also has periods of $4\pi, 6\pi, -2\pi$, and in general, $2\pi n$ for any integer n. This is no accident: if $f(x)$ is a periodic function with period p, then $f(x)$ is periodic with period np for any integer n. This is why we make the following definition:

Definition: The period of a periodic function $f(x)$ is the *smallest positive* real number p such that $f(x + p) = f(x)$ for all values of x for which $f(x)$ and $f(x + p)$ are defined.

Using this definition, we say that the period of $y = \sin x$ is 2π.

Let us also draw the graph of the function $y = \cos x$. Following the same methods, we find that the graph is as shown below:



The period of the function $y = \cos x$ is also 2π. We will see later that this curve can be described by an equation of the form $y = a \sin k(x - \beta)$.

3 Periods of other sinusoidal curves

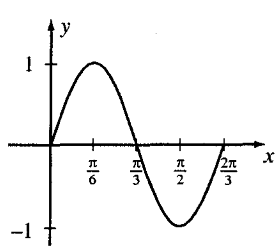
$$y = a \sin k(x - \beta) \quad \text{for } a = 1,\ \beta = 0,\ k > 0$$

Example 59 Find the period of the function $y = \sin 3x$.

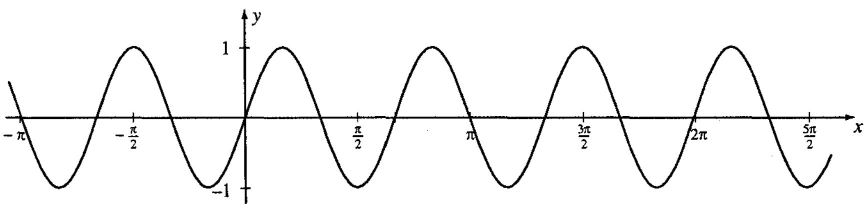
Solution: One period of this function is $2\pi/3$, since $\sin 3(x + 2\pi/3) = \sin(3x + 2\pi) = \sin 3x$. It is not difficult to see that this is the smallest positive period (for example, by looking at the values of x for which $\sin 3x = 0$).

Example 60 Draw the graph of the function $y = \sin 3x$.

Solution: The function $y = \sin x$ takes on certain values as x goes from 0 to 2π. The function $y = \sin 3x$ takes on these same values, but as x goes from 0 to $2\pi/3$. Hence one period of the graph looks like this:



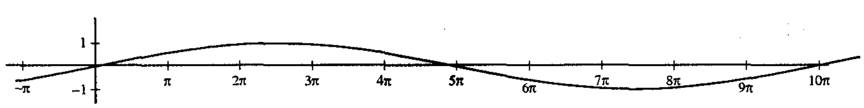
Having drawn one period, of course, it is easy to draw as much of the whole graph as we like (or have room for):



The graph is the same as that of y = sin x, but compressed by a factor of 3 in the x- direction. In general, we have the following result:

> For $k > 1$, the graph of $y = \sin kx$ is obtained from the graph $y = \sin x$ by compressing it in the x-direction by a factor of k.

What if $0 < k < 1$? Let us draw the graph of $y = \sin x/5$. Since the period of $y = \sin x/5$ is 10π, our function takes on the same values as the function $y = \sin x$, but stretched out over a longer period.



Again, we have a general result:

> For $0 < k < 1$, the graph of $y = \sin kx$ is obtained from the graph $y = \sin x$ by stretching it in the x-direction by a factor of k.

Analogous results hold for graphs of the functions $y = \cos kx$, $k > 0$.

Our basic family of functions is $y = a \sin k(x - \pi)$. What is the significance of the constant k here? We have seen that $2\pi/k$ is the period of the function. So in an interval of 2π, the function repeats its period k times. For this reason, the constant k is called the *frequency* of the function.

Exercises

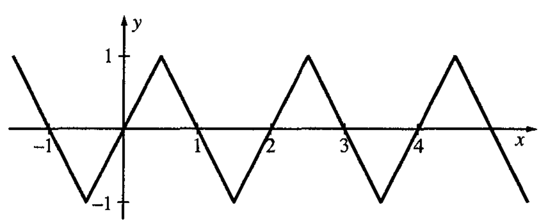
Find the period and frequency of the following functions:

1. $y = \sin 5x$ 2. $y = \sin x/4$ 3. $y = \cos 4x/5$ 4. $y = \cos 5x/4$.

Graph each of the following curves. Indicate the period of each. Check your work with a graphing calculator, if you wish.

5. $y = \sin 3x$ 6. $y = \sin x/3$ 7. $y = \sin 3x/2$ 8. $y = \sin 2x/3$
9. $y = \cos 2x/3$ 10. $y = \cos 3x/2$

11. The graph shown below has some equation $y = f(x)$.



(a) Draw the graph of the function $y = f(3x)$.

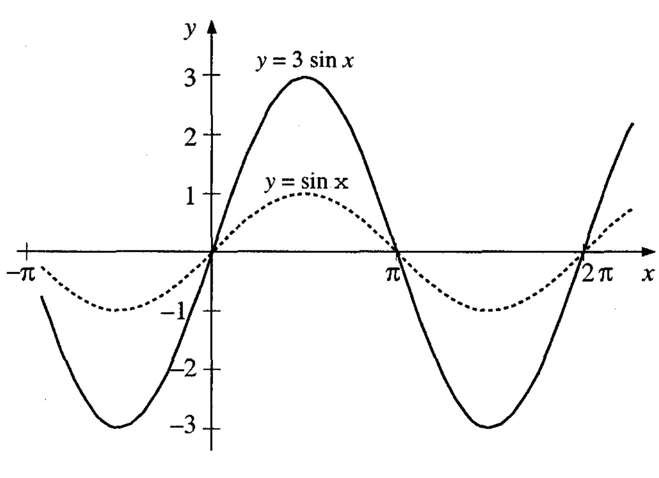
(b) Draw the graph of the function $y = f(x/3)$.

4 The amplitude of a sinusoidal curve

$$y = a \sin k(x - \beta); \quad a > 0, \beta = 0, k > 0$$

Example 61 Draw the graph of the function $y = 3 \sin x$.

Solution: The values of this function are three times the corresponding values of the function $y = \sin x$. Hence the graph will have the same period, but each y-value will be multiplied by 3:



We see that the graph of $y = 3\sin x$ is obtained from the graph of $y = \sin x$ by stretching in the y-direction. Similarly, it is not hard to see that the graph of $y = (1/2)\sin x$ is obtained from the graph of $y = \sin x$ by a compression in the y-direction.

We have the following general result:

> *For $a > 1$, the graph of $y = a\sin x$ is obtained from the graph $y = \sin x$ by stretching in the y-direction. For $0 < a < 1$, the graph of $y = a\sin x$ is obtained from the graph $y = \sin x$ by compressing in the y-direction.*

Analogous results hold for graphs of functions in which the period is not 1, and for equations of the form $y = a\cos x$. The constant a is called the *amplitude* of the function $y = a\sin k(x - \beta)$.

Exercises

Graph the following functions. Give the period and amplitude of each. As usual, you are invited to check your work, after doing it manually, with a graphing calculator.

1. $y = 2\sin x$	2. $y = (1/2)\sin x$	3. $y = 3\sin 2x$
4. $y = (1/2)\sin 3x$	5. $y = 4\cos x$	6. $y = (1/3)\cos 2x$

7. Suppose $y = f(x)$ is the function whose graph is given in Exercise 11 on page 178.

(a) Draw the graph of the function $y = 3f(x)$.

(b) Draw the graph of the function $y = (1/3)f(x)$.

5 Shifting the sine

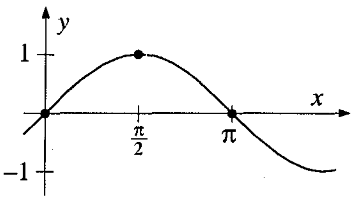
$$y = a\sin k(x - \beta); \quad a = 1, k = 1, \beta \text{ arbitrary}$$

We start with two examples, one in which β is positive and another in which β is negative.

Example 62 Draw the graph of the function $y = \sin(x - \pi/5)$.

Solution: We will graph this function by relating the new graph to the graph of $y = \sin x$. The positions of three particular points[1] on the original graph will help us understand how to do this:

x	$\sin x$
0	0
$\frac{\pi}{2}$	1
π	0

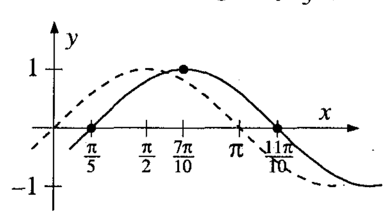


What are the analogous points on the graph of $y = \sin(x - \frac{\pi}{5})$? It is not convenient to use $x = 0$, because then $y = \sin(-\frac{\pi}{5})$, whose value is difficult to work with. Similarly, if we use $x = \frac{\pi}{2}$, we will need the value $y = \sin(\frac{\pi}{2} - \frac{\pi}{5}) = \sin\frac{3\pi}{10}$, which is still less convenient.

But if we let $x = \frac{\pi}{5}, \frac{\pi}{2} + \frac{\pi}{5}, \pi + \frac{\pi}{5}$, things will work out better:

x	$x - \frac{\pi}{5}$	$\sin(x - \frac{\pi}{5})$
$\frac{\pi}{5}$	0	0
$\frac{\pi}{2} + \frac{\pi}{5}$	$\frac{\pi}{2}$	1
$\pi + \frac{\pi}{5}$	π	0

That is, our choice of "analogous" points in our new function are those where the y-values are the same as those of the original function, not where the x-values are the same. The graph of $y = \sin(x - \frac{\pi}{5})$ looks just like the graph of $y = \sin x$, but shifted to the right by $\frac{\pi}{5}$ units:



But we must check this graph for more than three points. Are the other points on the graph shifted the same way? Let us take any point $(x_0, \sin x_0)$ on the graph $y = \sin x$. If we shift it to the right by $\frac{\pi}{5}$, we are merely adding this number to the point's x-coordinate, while leaving its y-coordinate the same. The new point we obtain is $(x_0 + \frac{\pi}{5}, \sin x_0)$, and this is in fact on the graph of the function $y = \sin(x - \frac{\pi}{5})$.

[1] Of course, with a calculator or a table of sines, you can get many more values. Or, if you have a good memory, you can remember the values of the sines of other particular angles. But these three points will serve us well for quite a while.

There is nothing special about the number $\frac{\pi}{5}$, except that it is positive. In general, the following statement is useful:

> *If $\beta > 0$, the graph of $y = \sin(x - \beta)$ is obtained from the graph of $y = \sin x$ by a shift of β units to the right.*

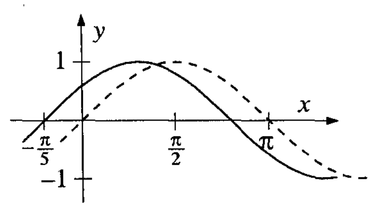
What if β is negative?

Example 63 Draw the graph of the function $y = \sin(x + \frac{\pi}{5})$.

Solution: In this example, $\beta = -\frac{\pi}{5}$. Again, we will relate this graph to the graph of $y = \sin x$. Using the method of the previous example, we seek values of x such that

$$\sin\,(x + \frac{\pi}{5}) = 0\,,\ \sin\,(x + \frac{\pi}{5}) = 1\,,\ \sin\,(x + \frac{\pi}{5}) = 0 \text{ (for a second time).}$$

It is not difficult to see that these values are $x = -\frac{\pi}{5}, \frac{\pi}{2} - \frac{\pi}{5}, \pi - \frac{\pi}{5}$, respectively. Using these values, we find that the graph of $y = \sin\,(x + \frac{\pi}{5})$ is obtained by shifting the graph of $y = \sin x$ by $\frac{\pi}{5}$ units to the *left*:



In general:

> *The graph of the function $y = \sin(x - \beta)$ is obtained from the graph of $y = \sin x$ by a shift of β units. The shift is towards the left if β is negative, and towards the right if β is positive.*

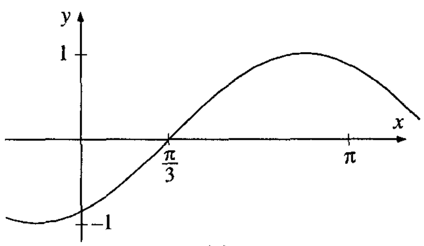
The number β is called the *phase angle* or *phase shift* of the curve. Analogous results hold for the graph of $y = \cos(x - \beta)$.

Exercises

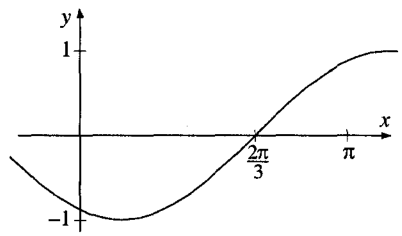
Sketch the graphs of the following functions:

1. $y = \sin(x - \frac{\pi}{6})$ 2. $y = \sin(x + \frac{\pi}{6})$ 3. $y = 2\sin(x - \frac{\pi}{2})$
4. $y = \frac{1}{2}\sin(x + \frac{\pi}{2})$ 5. $y = \cos(x - \frac{\pi}{4})$ 6. $y = 3\cos(x + \frac{\pi}{3})$
7. $y = \sin(x - 2\pi)$

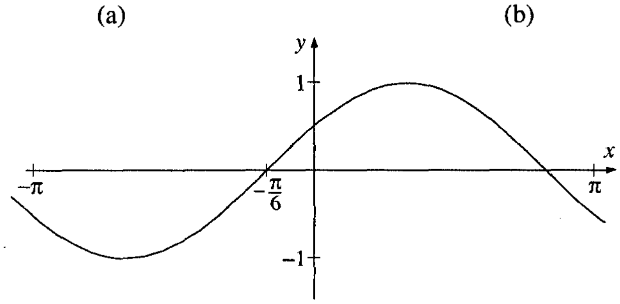
8–11: Write equations of the form $y = \sin(x - \alpha)$ for each of the curves shown below:



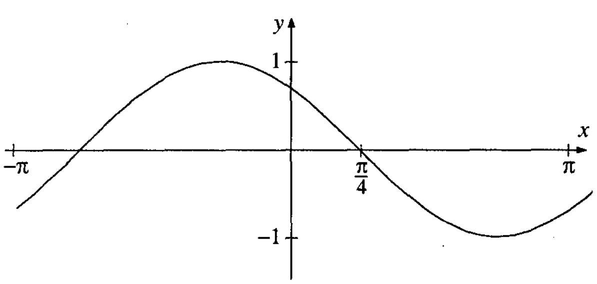
(a)



(b)



(c)



(d)

6 Shifting and stretching

Graphing $y = a \sin k(x - \beta)$

We run into a small difficulty if we combine a shift of the curve with a change in period.

Example 64 Graph the function $y = \sin(2x + \pi/3)$.

Solution: Let us write this equation in our standard form:

$$\sin(2x + \pi/3) = \sin 2(x + \pi/6)$$

We see that the graph is that of $y = \sin 2x$, shifted $\pi/6$ units to the left.

At first glance, one might have thought that the shift is $\pi/3$ units to the left. But this is incorrect. In the original equation, $\pi/3$ is added to $2x$, not to x. The error is avoided if we rewrite the equation in standard form.

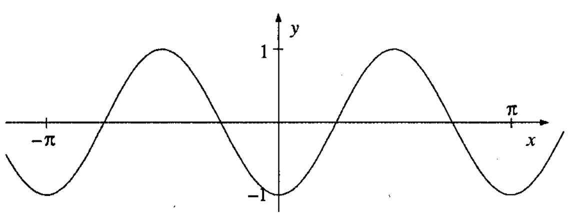
Exercises

Graph the following functions:

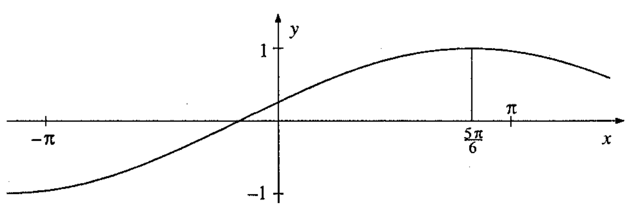
1. $y = \sin \frac{1}{2}(x - \frac{\pi}{6})$ 2. $y = \sin(\frac{1}{2}x - \frac{\pi}{6})$ 3. $y = \cos 2(x + \frac{\pi}{3})$

4–5: Write equations of the form $y = \sin k(x - \beta)$ for the following graphs:

4.



5.



7 Some special shifts: Half-periods

We will see, in this section, that we have not lost generality by restricting a and k to be positive, or by neglecting the cosine function.

It is useful to write our general equation as $y = a \sin k(x + \gamma)$, where $\gamma = -\beta$. Then, for positive values of γ, we are shifting to the left. For the special value $c = 2\pi$, we already know what happens to the graph $y = \sin x$. Since 2π is a period of the function, the graph will coincide with itself after such a shift.

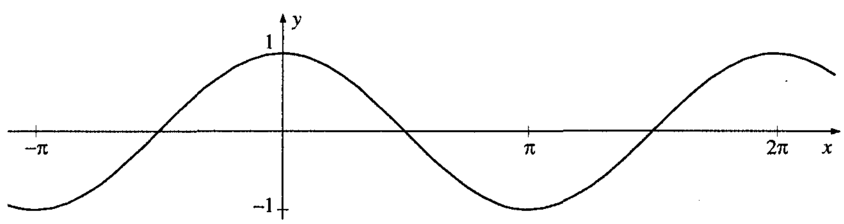
In fact, we can state the following alternative definition of a period of a function:

> *A function $y = f(x)$ has period p if the graph of the function coincides with itself after a shift to the left of p units.*

Our original definition said that a function $f(x)$ is periodic with period p if $f(x) = f(x+p)$ for all values of x for which these expressions are defined. Our new definition is equivalent to the earlier one, since the graph of $y = f(x)$, when shifted to the left by p units, is just the graph $y = f(x+p)$. These graphs are the the same if and only if $f(x) = f(x+p)$.

Let us see what happens when we shift the graph $y = \sin x$ to the left by $n\pi/2$, where n is an integer.

For $n = 1$, we have the graph $y = \sin(x + \pi/2)$, a shift to the left of the graph $y = \sin x$:

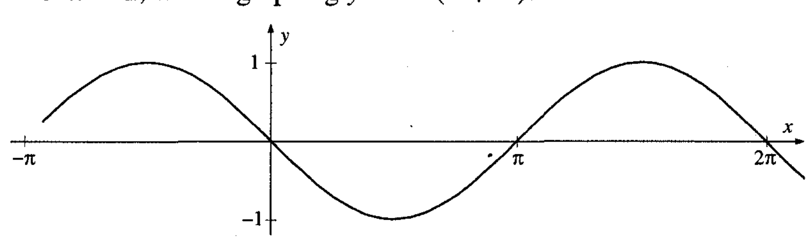


But $\sin(x + \pi/2) = \cos x$. The reader is invited to check this, either by using the addition formulas or by looking at the definitions, quadrant by quadrant. That is:

> *The graph of the function* $y = \cos x$ *can be obtained from the graph* $y = \sin x$ *by a shift to the left of* $\pi/2$.

We don't need to make a separate study of the curves $y = a\cos k(x+\beta)$. Letting $\gamma = x+\beta+\pi/2$, we can write any such curve as $y = a\sin k(x+\gamma)$.

For $n = 2$, we are graphing $y = \sin(x+\pi)$:



But $\sin(x+\pi) = -\sin x$. So we have:

> *The graph of the function* $y = -\sin x$ *can be obtained from the graph* $y = \sin x$ *by a shift to the left of* π.

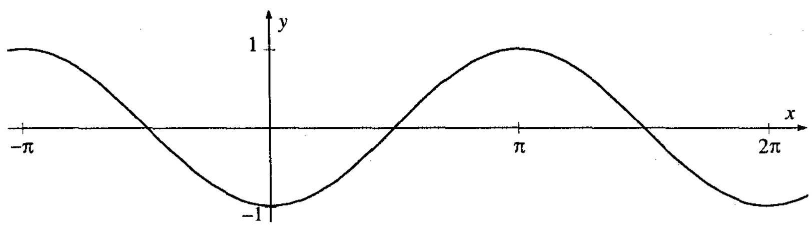
In fact, we do not need to make a separate study of the curves $y = a\sin k(x+\gamma)$ for negative values of a. We need only adjust the value of γ, and we can describe each such curve with an equation in which $a > 0$.

The following general definition is convenient:

The number p is called a half-period of the function f if $f(x+p) = -f(x)$, for all values of x for which $f(x)$ and $f(x+p)$ are defined.

We have shown that π is a half-period of the function $y = \sin x$.

Now let $k = 3$. We obtain the following graph:



It is not difficult to check that

$$\sin(x + 3\pi/2) = -\cos x .$$

If $k = 4$, we will shift by $4(\pi/2) = 2\pi$, which we already know is a full period, and we will have come back to our original sine graph.

What if $k = 5$? Since $5 = 1 + 4$, we have $\sin(x + 5\pi/2) = \sin(x + \pi/2 + 4\pi/2) = \sin(x + \pi/2)$, because $2p$ is a period of the sine function. So $k = 5$ has the same effect as $k = 1$, and the cycle continues.

In general, we can make the following statements:

If $k = 4n$ for some integer n, then $\sin(x + k\pi/2) = \sin x$.

If $k = 4n + 1$ for some integer n, then $\sin(x + k\pi/2) = \cos x$.

If $k = 4n+2$ for some integer n, then $\sin(x+k\pi/2) = -\sin x$.

If $k = 4n+3$ for some integer n, then $\sin(x+k\pi/2) = -\cos x$.

To summarize, we have now examined the whole family of sinusoidal curves $y = a \sin k(x - \beta)$.

The constant a is called the *amplitude* of the curve. It tells us how far from 0 the values of the function can get. Without loss of generality, we may take a to be positive.

The constant k is called the *frequency* of the curve. It tells us how many periods are repeated in an interval of 2π. The period of the curve is $2\pi/k$. Without loss of generality, we can take k to be positive.

The constant β is called the *phase* or *phase shift* of the curve. It tells us how much the curve has been shifted right or left. If we allow β to be arbitrary, we need not consider negative values of a or k, and we need not study separately curves expressed using the cosine function.

Exercises

1–10: These exercises are multiple choice. Choose the answer

(A) if the given expression is equal to $\sin x$,

(B) if the given expression is equal to $\cos x$,

(C) if the given expression is equal to $-\sin x$, or

(D) if the given expression is equal to $-\cos x$.

1. $\sin(x + 2\pi)$	2. $\sin(x + 3\pi)$	3. $\sin(x + 9\pi/2)$
4. $\sin(x - \pi/2)$	5. $\sin(x - 3\pi/2)$	6. $\sin(x + 19\pi/2)$
7. $-\sin(x - 19\pi/2)$	8. $\sin(x + 157\pi/2)$	9. $\sin(x - 157\pi/2)$

11. Prove that π is a half-period of the function $y = \cos x$. Is π a half-period of the function $y = \tan x$? of $y = \cot x$?

12. Prove that if q is a half-period of some function f, then $2q$ is a period of f.

13. Show that for all values of x, $\cos(x + k\pi/2) =$

a) $-\sin x$, if $k = 4n + 1$ for some integer n,

b) $-\cos x$, if $k = 4n + 2$ for some integer n,

c) $\sin x$, if $k = 4n + 3$ for some integer n,

d) $\cos x$, if $x = 4n$ for some integer n.

14. Write each of the following in the form $y = a \sin k(x - \beta)$, where a and k are nonnegative:

a) $y = -2 \sin x$

b) $y = -2 \sin(x - \pi/3)$

c) $y = -2 \sin(x + \pi/4)$

d) $y = 3 \cos x$

e) $y = 3 \cos(x - \pi/6)$

f) $y = -3 \cos(x + \pi/8)$

15. Draw the graph of the function $y = \cos(x - \pi/5)$

16. Suppose we start with the graph of the function $y = \cos x$. By how much must we shift this graph to the right in order to obtain the graph of $y = \sin x$? By how much must we shift to the left to obtain the graph of $y = \sin x$?

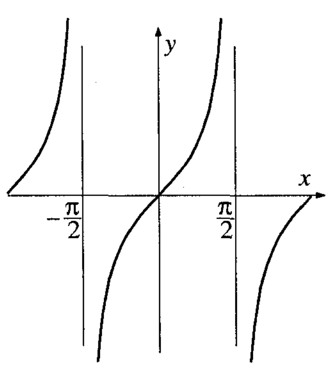
17. Show that if k is odd, $\tan(x + k\pi/2) = -\cot x$. How can we simplify the expression $\tan(x + k\pi/2)$ if k is even?

8 Graphing the tangent and cotangent functions

The function $y = \tan x$ is different from the functions $y = \sin x$ and $y = \cos x$ in two significant ways. First, the domain of definition of the sine and cosine functions is all real numbers. However, $\tan x$ is not defined for $x = n\pi/2$, where n is an odd integer.

Second, the sine and cosine functions are *bounded*: the values they take on are always between -1 and 1 (inclusive). But the function $y = \tan x$ takes on all real numbers as values.

These differences are easily seen in the graph of the function $y = \tan x$:

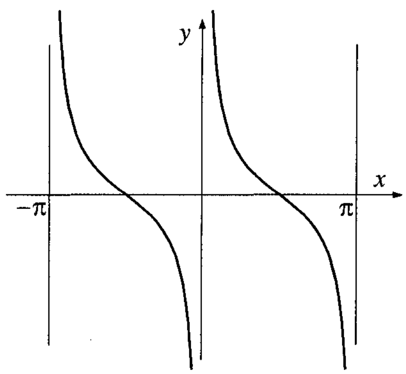


Note that the graph approaches the line $x = \pi/2$, but never reaches it. This line is called a *vertical asymptote* of the curve $y = \tan x$. This graph $y = \tan x$ has a vertical asymptote at every line $y = n\pi/2$, for n an odd integer.

To draw the graph of $y = \cot x$, we note that

$$\cot x = \frac{\cos x}{\sin x} = -\frac{\sin(x - \pi/2)}{\cos(x - \pi/2)} = -\tan(x - \pi/2)\ .$$

Therefore, the graph of $y = \cot x$ also takes on all real numbers as values. It is not defined for $x = n\pi$, where n is any integer, and has vertical asymptotes at $y = n\pi$:



Exercises

1. Draw the graphs of

   a) $y = \tan(x - \pi/6)$ b) $y = 3\tan x$ c) $y = \cot(x + \pi/4)$.

2. Suppose we graphed the equation $y = \tan x$. Is it possible to describe this graph with an equation of the form $y = \cot(x + \varphi)$, for some number φ? Why or why not?

9 An important question about sums of sinusoidal functions

We hope that from this material you have seen the importance, and the beauty, of the family of sinusoidal curves that we have been studying. Physicists call this family the curves of *harmonic oscillation.*

Let us now consider the following question. Suppose we have two sinusoidal curves (harmonic oscillations):

$$y_1 = a_1 \sin k_1(x - \beta_1)$$
$$y_2 = a_2 \sin k_2(x - \beta_2)\,.$$

Will the sum of these two also be a sinusoidal curve (harmonic oscillation)? That is, will

$$y = a_1 \sin k_1(x - \beta_1) + a_2 \sin k_2(x - \beta_2)$$

be a sinusoidal curve? The answer is somewhat surprising. If $k_1 = k_2$, the answer is yes, but if $k_1 \neq k_2$, the answer is no.

That is, the sum of two harmonic oscillations is again a harmonic oscillation if and only if the original frequencies are the same. The results of the next few sections will allow us to explore this situation.

Exercises

Each of these exercises concerns the following three functions:

$$\begin{aligned} y_1 &= 2\sin x \\ y_2 &= \sin(x - \pi/4) \\ y_3 &= 3\sin 2x \end{aligned}$$

1. Use your calculator to draw the graph of (a) $y_1 + y_2$; (b) $y_1 + y_3$; (c) $y_2 + y_3$.
2. Which of the graphs in Exercise 1 appear to be sinusoidal functions?

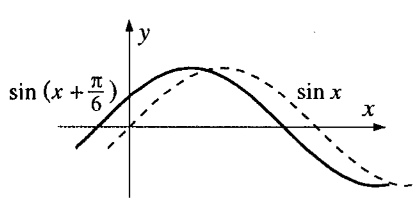
10 Linear combinations of sines and cosines

Definition: If we have two functions $f(x)$ and $g(x)$, and two constants a and b, then the expression $af(x) + bg(x)$ is called a *linear combination* of the functions $f(x)$ and $g(x)$.

Let us look at the graph of a linear combination of sinusoidal curves.

Example 65 Graph the function $y = \frac{\sqrt{3}}{2}\sin x + \frac{1}{2}\cos x$.

Solution. Since $\frac{\sqrt{3}}{2} = \cos\frac{\pi}{6}$ and $\frac{1}{2} = \sin\frac{\pi}{6}$, we use the formula $\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$. Letting $\alpha = x$ and $\beta = \frac{\pi}{6}$, this formula tells us that the given function can be written as $y = \sin(x + \frac{\pi}{6})$. Now we can graph it as we did in Section 5:



This solution may seem artificial, but is in fact a general method. It works because there is an angle φ such that $\cos\varphi = 1/2$ and $\sin\varphi = \sqrt{3}/2$, and this happened because the values $A = 1/2$ and $B = \sqrt{3}/2$ satisfy the equation $A^2 + B^2 = 1$. (The reader is invited to do this computation.)

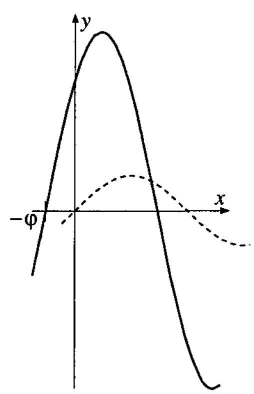
But what if $A^2 + B^2$ is not equal to 1?

Example 66 Draw the graph of the function $f(x) = 3\sin x + 4\cos x$.

Solution. Our "best friends" (of Chapter 1) are hiding in this expression: where we have 3 and 4, we try to look for the number 5. Indeed, $f(x)/5 = \frac{3}{5}\sin x + \frac{4}{5}\cos x$, and $\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2 = 1$, so we can use the method of Example 3. We know that there is an angle φ such that $\cos\varphi = 3/5$ and $\sin\varphi = 4/5$, and so

$$\frac{f(x)}{5} = \cos\varphi\sin x + \sin\varphi\cos x = \sin(x + \varphi)$$

or $f(x) = 5\sin(x + \varphi)$, for a certain angle φ. The graph is a sine curve, shifted to the left φ units, and with amplitude 5:



The same technique will work for linear combinations of $y = \sin kx$ and $y = \cos kx$, as long as the frequency of the two functions is the same. This is important enough to state as a theorem:

Theorem A linear combination of $y = \sin kx$ and $y = \cos kx$ can be expressed as $y = a\sin k(x + \varphi)$, for suitable constants a and φ.

Proof. A linear combination of $y = \sin kx$ and $y = \cos kx$ has the form $y = A\sin kx + B\cos kx$. We can rewrite it as

$$y = \sqrt{A^2 + B^2}\left(\frac{A}{\sqrt{A^2 + B^2}}\sin kx + \frac{B}{\sqrt{A^2 + B^2}}\cos kx\right).$$

Then

$$\left(\frac{A}{\sqrt{A^2 + B^2}}\right)^2 + \left(\frac{B}{\sqrt{A^2 + B^2}}\right)^2 = 1$$

so there exists an angle α such that

$$\cos\alpha = \frac{A}{\sqrt{A^2+B^2}} \text{ and } \sin\alpha = \frac{B}{\sqrt{A^2+B^2}} . \tag{1}$$

Now we can write

$$\begin{aligned} A\sin kx + B\cos kx &= \sqrt{A^2+B^2}(\cos\alpha\sin kx + \sin\alpha\cos kx) \\ &= \sqrt{A^2+B^2}\sin(kx+\alpha) \\ &= \sqrt{A^2+B^2}\sin k(x+\alpha/k) . \end{aligned}$$

Taking $a = \sqrt{A^2+B^2}$ and $\varphi = \alpha/k$, we have the required form. □

We have proved that $A\sin kx + B\cos kx$ can be written in the form $a\sin k(x+\gamma)$, where $a = \sqrt{A^2+B^2}$ and $\varphi = \alpha/k$ (for α defined by equations (1) above).

The converse statement is also correct:

Theorem The function $a\sin k(x+\varphi)$ can be written as a linear combination of the functions $\sin kx$ and $\cos kx$.

Proof. We have $a\sin k(x+\varphi) = a(\sin kx\cos k\varphi + \cos kx\sin k\varphi)$. Taking $A = a\cos k\varphi$ and $B = a\sin k\varphi$, we see that $a\sin k(x+\varphi) = A\sin kx + B\cos kx$. □

We can now write a sinusoidal curve in either of two standard forms: $y = a\sin k(x-\beta)$ or $y = A\sin kx + B\cos kx$.

Example 67 Write the function $y = 2\sin(x+\pi/3)$ as a linear combination of the function $y = \sin x$ and $y = \cos x$.

Solution. We have $2\sin(x+\pi/3) = 2(\sin x\cos\pi/3 + \cos x\sin\pi/3) = 2(1/2)\sin x + 2\sqrt{3}/2\cos x = \sin x + \sqrt{3}\cos x$.

Exercises

1. Write the function $y = 2\sin x + 3\cos x$ in the form $y = a\sin k(x-\beta)$. What is its amplitude?

2. What is the maximum value achieved by the function $y = 2\sin x + 3\cos x$?

3–6: Write each function in the form $y = a \sin k(x - \beta)$. What is the maximum value of each function?

3. $y = \sin x + \cos x$
4. $y = \sin x - \cos x$
5. $y = 4 \sin x + 3 \cos x$
6. $y = \sin 2x + 3 \cos 2x$

7, 8: Write each function in the form $A \sin x + B \cos x$:

7. $y = \sin(x - \pi/4)$
8. $y = 4 \sin(x + \pi/6)$

11 Linear combinations of sinusoidal curves with the same frequency

Now we are ready to address the important question of Section 9.

Theorem The sum of two sinusoidal curves with the same frequency is again a sinusoidal curve with this same frequency.

Proof. Let us take the two sinusoidal curves

$$a_1 \sin k(x - \beta_1) \text{ and}$$
$$a_2 \sin k(x - \beta_2)\ .$$

Using the addition formula, we can write:

$$a_1 \sin k(x - \beta_1) = A_1 \sin kx + B_1 \cos kx$$
$$a_2 \sin k(x - \beta_2) = A_2 \sin kx + B_2 \cos kx$$

for suitable values of A_1, A_2, B_1, and B_2. Then our sum is equal to

$$(A_1 + A_2) \sin kx + (B_1 + B_2) \cos kx\ .$$

But we know, from the theorem of Section 9, that this sum is also a sinusoidal curve. Our theorem is proved. □

We invite the reader to fill in the details, by giving the expressions for A_1, A_2, B_1, and B_2.

Note that the two functions we are adding may have different amplitudes. The result depends only on their having the same period. This result is very important in working with electricity. Alternating electric current is described by a sinusoidal curve, and this theorem says that if we add two currents with the same periods, the resulting current will have this period as well. So if we are drawing electric power from different sources, we need not worry how to mix them (whether their phase shifts are aligned), as long as their periods are the same.

The next result is important in more advanced work:

Theorem If a linear combination of the functions $y = \sin kx$ and $y = \cos kx$ is shifted by an angle β, then the result can be expressed as a linear combination of the same two functions.

Proof. Let us take the linear combination

$$a \sin kx + b \cos kx$$

and shift it by an angle β. The result is

$$a \sin k(x - \beta) + b \cos k(x - \beta) \ .$$

We know that $\cos k(x-\beta)$ can be written as $\sin k(x-\gamma)$, for some angle γ. Thus we can write our shifted linear combination as

$$a \sin k(x - \beta) + b \sin k(x - \gamma) \ .$$

But this is a sum of sinusoidal curves with the same frequency k, so the previous theorem tells us that it can be written as a single sinusoidal curve with frequency k (even though the shifts are different!). And we know, from Section 9, that such a sum can be written as a linear combination of $\sin kx$ and $\cos kx$.

Example 68 Suppose we take the graph of a linear combination of $y = \sin x$ and $y = \cos x$:

$$y = 2 \sin x + 4 \cos x$$

and shift it $\pi/6$ units to the left. We get:

$$\begin{aligned}
y &= 2\sin(x + \pi/6) + 4\cos(x + \pi/6) \\
&= 2(\sin x \cos \pi/6 + \cos x \sin \pi/6) + 4(\cos x \cos \pi/6 - \sin x \sin \pi/6) \\
&= 2(\sqrt{3}/2 \sin x + 2(1/2) \cos x + 4(\sqrt{3}/2) \cos x - 4(1/2) \sin x \\
&= (\sqrt{3} - 2) \sin x + (2\sqrt{3} + 1) \cos x
\end{aligned}$$

which is again a linear combination of $y = \sin x$ and $y = \cos x$.

This technique works whenever we apply a shift to a linear combination of $y = \sin kx$ and $y = \cos kx$. The proof follows the reasoning of the above example.

A final comment: We have not considered linear combinations of sines and cosines with *different* frequencies. This is a more difficult situation, and leads to some very advanced mathematical topics, such as Fourier Series and almost periodic functions. We will return to this question a bit later.

Exercises

1. Express each function in the form $y = A \sin kx + B \cos kx$
   (a) $y = 2\sin(x + \pi/6) + \cos(x + \pi/6)$
   (b) $y = 2\sin 2(x + \pi/4) - \cos 2(x + \pi/4)$
2. Look at the exercises for Section 9 on page 189.
   (a) Write $y_1 + y_2$ as a linear combination of $\sin x$ and $\cos x$.
   (b) What goes wrong when you try to write $y_1 + y_3$ as a linear combination of $\sin x$ and $\cos x$?

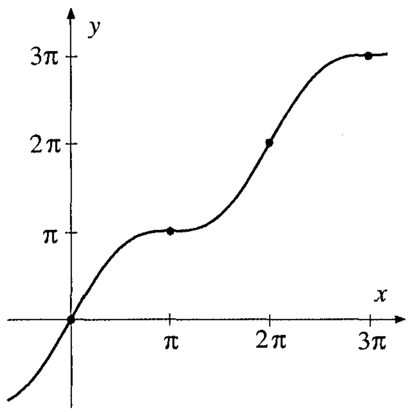
12 Linear combinations of functions with different frequencies

So far, we have some important results about linear combinations of sines and cosines with the same frequency. We would like to investigate the sum of two functions like $y = \sin k_1 x$ and $y = \sin k_2$, where $k_1 \neq k_2$. We start the discussion with some examples which may not at first appear related.

Example 69 Graph the function $y = x + \sin x$.

Solution. Each y-value on this graph is the sum of two other y-values: the value $y = \sin x$ and the value $y = x$. So we can take each point on the curve $y = \sin x$ and "lift it up" by adding the value $y = x$ to the value $y = \sin x$.

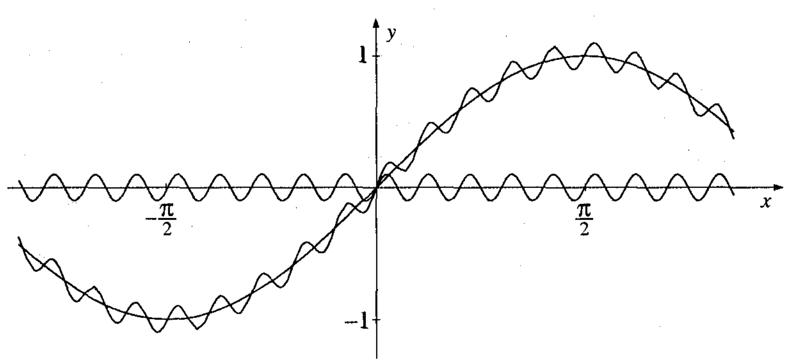
This is particularly easy to see for those points where $\sin x = 0$. For these points, the value of $x + \sin x$ is just x:



In between these points, the line $y = x$ is lifted up slightly, or brought down slightly, by positive or negative values of $\sin x$. We can think of the sine curve as "riding" on the line $y = x$.

Example 70 Graph the function $y = \sin x + 1/10 \sin 20x$.

Solution. This seems much more complicated, but in fact can be solved using the same method as the previous examples. We graph the two curves $y = \sin x$ and $y = 1/10 \sin 20x$ independently, then add their y-values at each point:



Again, we can think of one curve "riding" on the other. This time the curve $y = 1/10 \sin 20x$ "rides" on the curve $y = \sin x$, or *perturbs* it a bit at each point.

Note that our new curve is *not* a sinusoidal curve. We cannot express it either in the form $y = a \sin k(x - \beta)$ or in the form $y = A \sin kx + b \cos kx$.

Exercises

Construct graphs of the following functions.

1. $y = -x + \sin x$
2. $y = x^2 + \sin x$
3. $y = x^2 + \cos x$. Hint: Is the function odd? Is it even?
4. $y = x^3 + \sin x$
5. $y = x^2 + (1/10)\sin x$
6. $y = \cos x + (1/10)\sin 20x$
7. $y = 2\sin x + (1/10)\sin 20x$

13 Finding the period of a sum of sinusoidal curves with different periods

We know that the function $y = \sin 10x + \sin 15x$ is not a sinusoidal curve. Let us show that it is still periodic. Indeed, if we shift the curve by 2π, we have $y = \sin 10(x + 2\pi) + \sin 15(x + 2\pi) = \sin(10x + 20\pi) + \sin(15x + 30\pi) = \sin 10x + \sin 15x$.

But what is its smallest positive period? We can answer this by looking separately at all the periods of the two functions we are adding. Any period of $y = \sin 10x$ must have the form $m(2\pi/10)$, for some integer m. Any period of $y = \sin 15x$ must have the form $n(2\pi/15)$, for some integer n. To be a period of both functions, a number must be of both these forms. That is, we must have integers m and n such that $2m\pi/10 = 2n\pi/15$, or $3m = 2n$. If we take $m = 2$, $n = 3$, our problem is solved. The number $2\pi/5 = 2(2\pi/10) = 3(2\pi/15)$ is a period for both functions. And since we took the smallest positive values of m and n, this is the smallest positive period for the function $y = \sin 10x + \sin 15x$.

The argument above is drawn from number theory, where it is connected with the least common multiple of two numbers. This concept is used in elementary arithmetic, in finding the least common denominator for two fractions. The general statement, proved in number theory, is this:

> *The function* $y = \sin k_1x + \sin k_2x$ *is periodic if and only if the quotient* k_1/k_2 *is rational.*

But a function like $y = \sin x + \sin\sqrt{2}x$ has no period at all.

Exercises

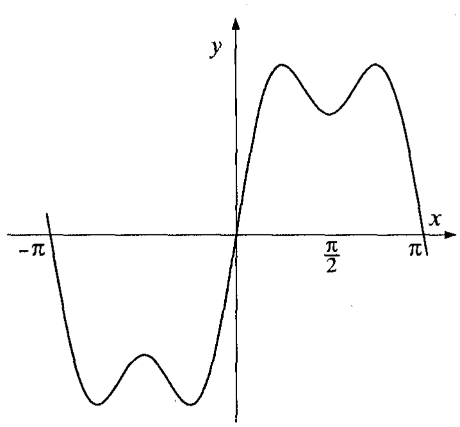
Find the (smallest positive) period for each of the following functions.

1. $y = \sin 2x + \sin 3x$
2. $y = \sin 3x + \sin 6x$
3. $y = \sin 4x + \sin 6x$
4. $y = \sin \sqrt{2}x + \sin 3\sqrt{2}x$

14 A discovery of Monsieur Fourier

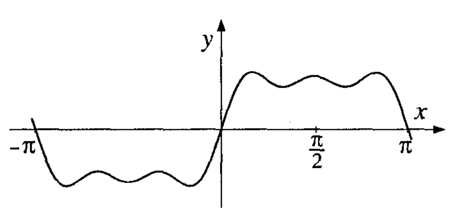
Example 71 Graph the function $y = \sin x + (1/3)\sin 3x$.

Solution. This example is similar to Example 70. The values of $\sin x$ are "perturbed" by those of $\frac{1}{3}\sin 3x$:



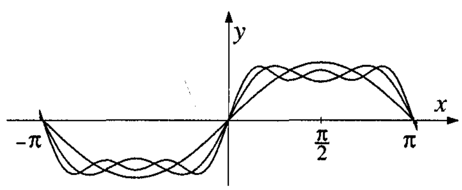
Example 72 Graph the function $y = \sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x$. (Use a graphing calculator or software utility for this complicated function.)

Solution.



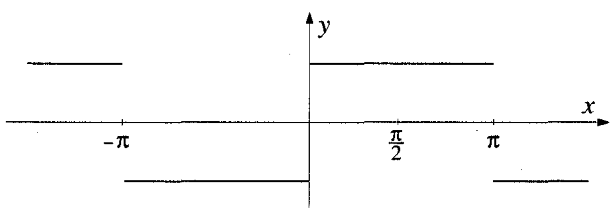
Let us compare the graphs of the three functions:

$$y = \sin x$$
$$y = \sin x + \tfrac{1}{3}\sin 3x$$
$$y = \sin x + \tfrac{1}{3}\sin 3x + \tfrac{1}{5}\sin 5x\ .$$



The formulas for these functions show a pattern. Can you guess what the next formula in the pattern would be? Can you guess what its graph would look like? Check your guess with a graphing calculator or software utility.

It is not difficult to guess that the graphs of these functions will look more and more like the following:



Mathematicians say that this sequence of functions *converges* to a limit, and that this limit is the function whose graph is given above. In fact, this is a special case of the very important mathematical theory of Fourier series. The French physicist Fourier discovered that almost any periodic function, including some with very complicated or bizarre graphs, can be represented as the limit of a sum of sines and cosines (the above example doesn't happen to contain cosines). He also showed how to calculate this sum (using techniques drawn from calculus).

Fourier's discovery allows mathematicians to describe very simply any periodic function, and physicists can use these descriptions to model actions that repeat. For example, sounds are caused by periodic vibrations of particles of air. Heartbeats are periodic motions of a muscle in the body. These phenomena, and more, can be explored using the mathematical tools of Fourier analysis.

Exercises

Please use a graphing calculator or graphing software package for these exercises.

1. Graph the function $y = \sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x$.
2. Graph the function $y = \sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \frac{1}{4}\sin 4x + \frac{1}{5}\sin 5x$.
3. Consider the sequence of functions:

$$y = \sin x$$
$$y = \sin x - \tfrac{1}{2}\sin 2x$$
$$y = \sin x - \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x$$
$$y = \sin x - \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x - \tfrac{1}{4}\sin 4x$$
$$y = \sin x - \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x - \tfrac{1}{4}\sin 4x + \tfrac{1}{5}\sin 5x$$

   Draw the graph of the function that you think is the limit of this sequence of functions.

4. Consider the sequence of functions:

$$y = \cos x$$
$$y = \cos x + \tfrac{1}{9}\cos 3x$$
$$y = \cos x + \tfrac{1}{9}\cos 3x + \tfrac{1}{25}\cos 5x$$
$$y = \cos x + \tfrac{1}{9}\cos 3x + \tfrac{1}{25}\cos 5x + \tfrac{1}{49}\cos 7x$$

   Draw the graph of the function that you think is the limit of this sequence of functions. Do you recognize the pattern in the amplitudes?

Appendix

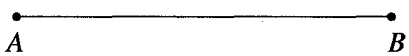
I. Periodic phenomena

Many phenomena in nature exhibit periodic behavior: the motions repeat themselves after a certain amount of time has passed. The sine function, it turns out, is the key to describing such phenomena mathematically.

The following exercises concern certain periodic motions. Their mathematical representations remind us of the sine curve, but are not exactly the same. In more advanced work trigonometric functions can indeed be used to describe these motions.

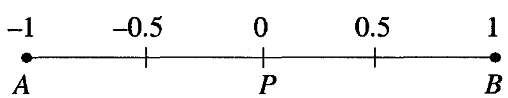
Exercises

1. The diagram represents a line segment 1 foot in length.



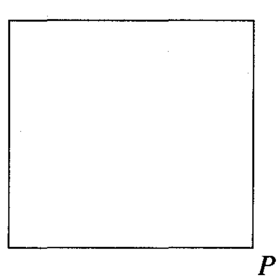
An ant is walking along the line segment, from point A to point B, then back again, at a speed of one foot per minute. Draw a graph showing the distance from point A to the ant's position at a given time t. For example, when $t = 0.5$, the ant is halfway between A and B, and headed towards B.

2. The diagram shows part of a number line, from $A = -1$ to $P = 0$, to $B = +1$.



An ant is walking along the number line, starting from point P. The ant walks to point B, then to A, then back to B, and so on. The ant walks at a speed of one foot per minute. Draw a graph showing the position of the ant on the number line at time t. For example, when $t = 0.5$, the ant is at 0.5, and when $t = 2.5$, the ant is at -0.5.

3. The diagram shows a square wall of a room. Each side of the square is 8 feet long.



An ant is walking along the perimeter of the wall, at a speed of one foot per minute, starting at the point P shown and moving counter-clockwise. Draw a graph showing the height of the ant above the floor (call it h) at any given time t. For example, where $t = 4$ the ant's height is 4 feet, and where $t = 12$ the ant's height is 8 feet.

4. The Bay of Fundy lies between the Canadian provinces of Nova Scotia and New Brunswick. The people who live on its shores experience some of the world's highest tides, which can reach a height of 40 feet. This creates a landscape that shifts twice every day. Beaches become bays, small streams turn into raging rivers, and peninsulas are suddenly islands as the tides rise and fall.

   For anyone who lives near the ocean, it is important to know when high and low tide will occur. But for the Bay of Fundy, it is critical also to know how fast the tide is rising or falling. The inhabitants of this area use the so-called *rule of twelfths* to estimate this. They take the interval between low and high tide to be 6 hours (it is actually a bit more). Then they approximate that:

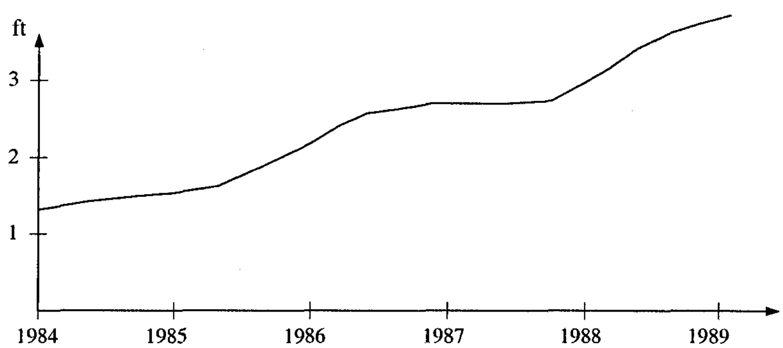
   $\frac{1}{12}$ of the tide will come in during the first hour
   $\frac{2}{12}$ of the tide will come in during the second hour
   $\frac{3}{12}$ of the tide will come in during the third hour
   $\frac{3}{12}$ of the tide will come in during the fourth hour
   $\frac{2}{12}$ of the tide will come in during the fifth hour
   $\frac{1}{12}$ of the tide will come in during the sixth hour

   Assume that the height of a day's tide is 36 feet, and draw a graph of the height of the water at a given point along the Bay of Fundy, using these estimates.

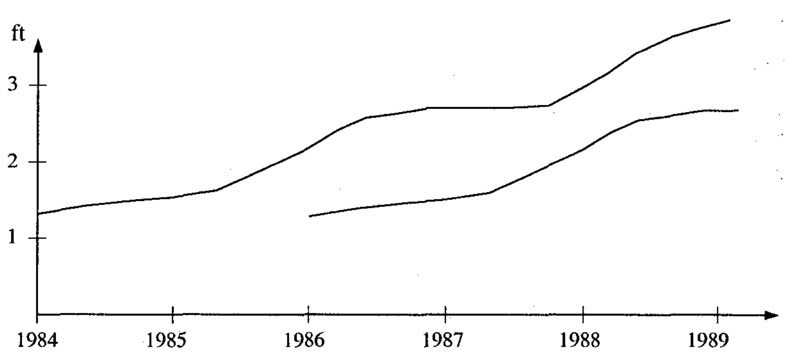
   When is the tide running fastest? Slowest?

II. How to explain the shifting of the graph to your younger brother or sister

When we were little, we used to go every few months to the doctor. The doctor would measure our height, and make a graph showing how tall we were at every visit. Here is the graph for my height :[2]



Two years later, when you were born, our parents asked the doctor if she could predict your growth year-by-year. Well, she couldn't exactly do this, but she said: "If the new baby follows the same growth curve as your older child, then he will be as tall as the older one was three years earlier." So the doctor was predicting a growth curve for you which looks like this:



You will be 3 feet tall exactly two years after I was 3 feet tall, and 4 feet tall also exactly two years after I was, and so on. Your graph is the same as mine, but shifted to the right by two years. If you want to know the prediction for your height, just look at what my height was two years ago. So if my graph is described by the equation height $= f(\text{year})$, then your graph is described by the equation height $= f(\text{year} - 2)$.

Of course, it hasn't quite turned out this way. My growth curve was not exactly the same as yours. So the doctor's prediction was not accurate. But for some families, it is accurate.

[2]Of course, when I was very little, I couldn't stand up, so they measured my "length." When I learned to stand, this became my height.

In just the same way, if you are dealing with the graph $y = \sin(x - \alpha)$, rather than $y = \sin x$, you must "wait" for x to get bigger by α before the height of the new graph is the same as that of the old graph. So the new graph is shifted α units to the right.

III. Sinusoidal curves with rational periods

We have taken, as our basic sine curve, the function $y = \sin x$. The period of this function is 2π, which is an irrational number. The other functions we've investigated also have irrational periods. Can a sine curve have a rational period?

Consider the function $y = \sin 2\pi x$. Using our formula, its period is $2\pi/2\pi = 1$. We can check this directly:

$$\sin\bigl(2\pi(x+1)\bigr) = \sin(2\pi x + 2\pi) = \sin 2\pi x \ .$$

The exercises below require the construction of sinusoidal curves with other rational periods.

Exercises

1. Show that the function $y = \sin \pi x$ has the value 0 when $x = 1$, $x = 2$, $x = 3$, and $x = 4$.

2. Show that the function $y = \sin 4\pi x$ has a period of $\frac{1}{2}$.

3. Write the equation of a sine curve with period 3.

4. Write the equation of a sine curve with period 2.

5. If n is a positive integer, write a function of the form $y = \sin kx$ with period n.

IV. From graphs to equations

A tale is told of the Russian tsar Alexei Mikhaelovitch, the second of the Romanov line (1629–76; reigned 1645–76). His court astronomer came to him one day in December, and told him, "Your majesty, from this day forth the number of hours of daylight will be increasing."

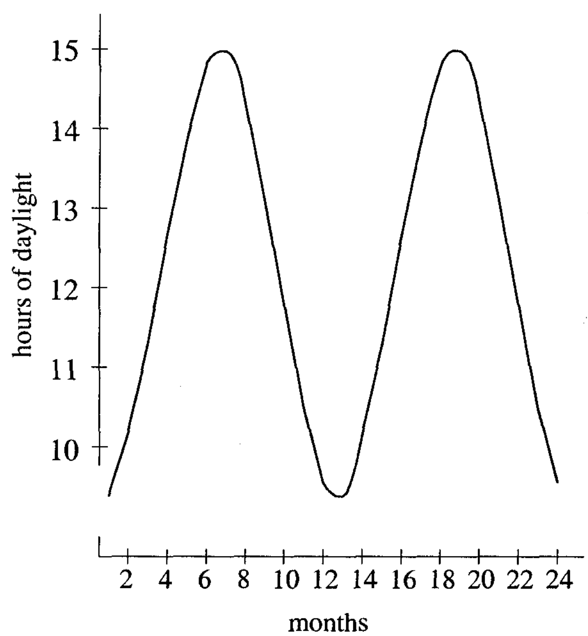
The tsar was pleased. "You have done well, court astronomer. Please accept this gift for your services." And, motioning to a courtier, he presented the astronomer with a valuable gemstone.

The astronomer enjoyed his gift and practiced his arts, until one day in June, when he again reported to the tsar. "Your highness, from this day forth the number of hours of daylight will be be decreasing."

The tsar scowled. "What? More darkness in my realm?" And he ordered the hapless astronomer beaten.

Of course, the variation in the amount of daylight was not the fault of this astronomer, or any other astronomer. It is due to the circumstance that the earth's axis is tilted with respect to the plane in which it orbits the sun. Because of this phenomenon, the days grow longer from December to June, then shorter from June to December.

What is interesting to us is the rate at which the number of hours of daylight changes. It turns out that if we graph the number of hours of daylight in each day, we get a sinusoidal curve:



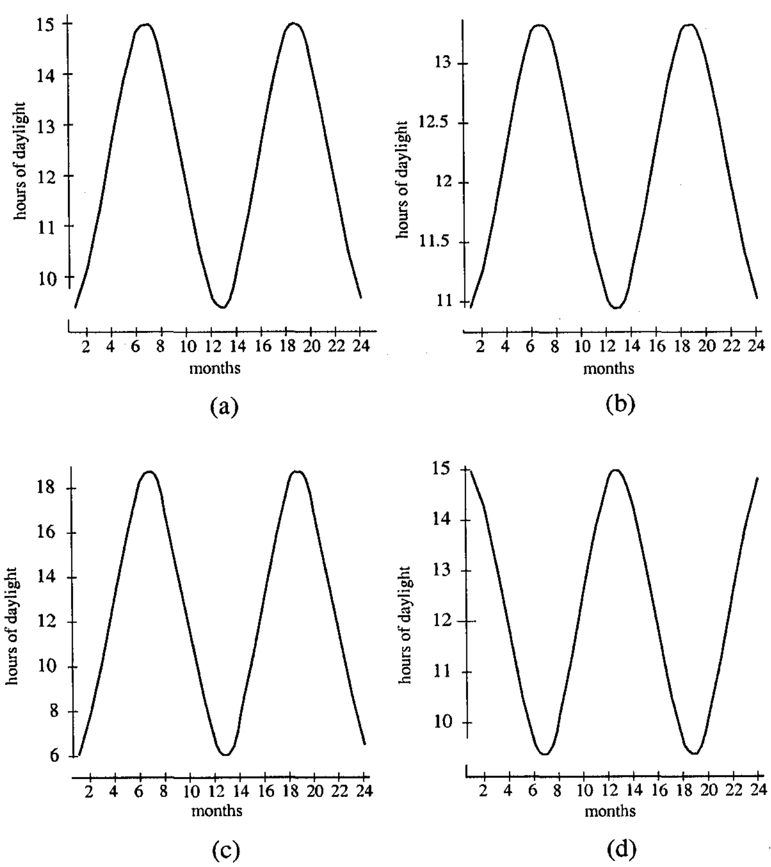
Since this curve is high above the x-axis, we have shown it with the y-axis "broken," so that you can see the interesting part of the graph. If you don't like this, try redrawing the curve without a "broken" y-axis. You will find that most of your diagram is empty.

We will learn more about this curve in the following exercises.

Exercises

1. By estimating the distances on the graph above, find an equation of the form $y = a \sin k(x - \pi)$ which approximates the function whose graph is shown.

2. The curves below give the hours of daylight at certain latitudes.



Notice that the maximum number of hours of daylight occur at the same time of year from graphs (a), (b), and (c), but at different times for graph (d). If graph (a) corresponds to a location in the northern hemisphere, in which hemisphere are the locations of the other graphs?

3. Notice that the "average" number of daylight hours is the same for each graph. This "average" is given by the y-coordinates along the line around which the curve oscillates: On certain days of the year, at each location, the actual number of hours of daylight is the same as the average number. How does the time of year at which this average is actually achieved vary from location to location?

Chapter 9

Inverse Functions and Trigonometric Equations

1 Functions and Inverse Functions

Let us recall the definition of a function. If we have two sets A and B, a function from set A to set B is a correspondence between elements of A and elements of B such that

1. Each element of A corresponds to some element of B, and
2. No element of A corresponds to more than one element of B.

If the element x in set A corresponds to the element y in B, we write $y = f(x)$, where f is the symbol for the function itself.

Example 73 Let us take A as the set of all real numbers, and B as another copy of the set of real numbers. If x is an element of A, then we can make it correspond to an element y in B by taking $y = x^2$. Every element x in A corresponds to some element y in B, (since any number can be squared), and no element x in A corresponds to more than one element y in B (since we get a unique answer when we square a number).

Our definition of a function is not very democratic. For every element of A, we must produce exactly one element of B. But if we have an element of B, we cannot tell if there is an element in A to which it corresponds. An element of B may correspond to no element of A, to one element of A, or to more than one element of A.[1]

[1] In older texts, this undemocratic situation was described by calling x the *independent variable* and y the *dependent variable*.

Example 74 In Example 73, if we were given a number x in A, we are obliged to supply an answer to the question: what number y in B corresponds to A? For example, if $x = 3$, then we can answer that $y = 9$, and if $x = -3$, we can answer $y = 9$ again. This is allowed, under our definition of a function. The only restriction is that our answer must be a number in set B.

But if we choose an element y in set B, we are not obliged to answer the question: what number in A corresponds to it? Certainly, if we chose $y = 9$, we could answer $x = 3$. But we could just as well answer $x = -3$, and so our answer would not be unique. Worse, if we chose $y = -1$, we have no answer at all. There is no real number whose square is -1.

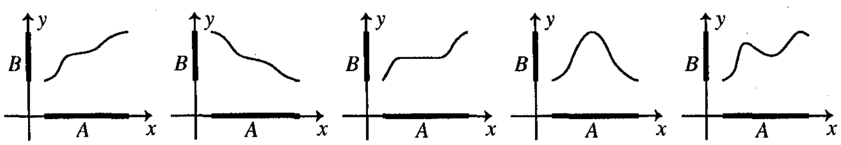
That is, if y is a function of x, it may not be the case that x is also a function of y. However, in some cases, we can improve the situation.

Example 75 Take the set A to be the set of nonnegative real numbers, and for B take another copy of the same set. As before, the correspondence $y = x^2$ is a function: if x is a number in A, then x^2 is a number in B, since the square of a real number cannot be negative. But now, if we take a number y in B, we can always answer the question: What number x in A corresponds to y? For example, if $y = 9$, we can answer that $x = 3$. We are not embarrassed by the possibility of a second answer, since -3 is not in our (new) set A. Nor are we embarrassed by the lack of any answer. Negative numbers, which are not squares of real numbers, do not exist in our new set B.

In general, we can take a function $y = f(x)$, try to start with a value of y, and get the corresponding value of x. If this is possible – if x is a function of y as well – then this new function is called the *inverse function* for $f(x)$.

Thus the function $y = x^2$, where $x \geq 0$ and $y \geq 0$, does have an inverse, given by the formula $x = \sqrt{y}$. This is the reason for insisting, in elementary algebra books, that the symbol $\sqrt{y}$ refers to the *nonnegative* real number whose square is y.

When does a function have an inverse function? This is an important question. We will not give a general answer here. We will, however, observe that if A and B are intervals on the real line, then $y = f(x)$, defined on these intervals, has an inverse if and only if it is *monotone* (steadily increasing or steadily decreasing). The first two graphs below show functions that are monotone, and have inverses. The last three graphs show functions that have no inverse on the sets A and B.

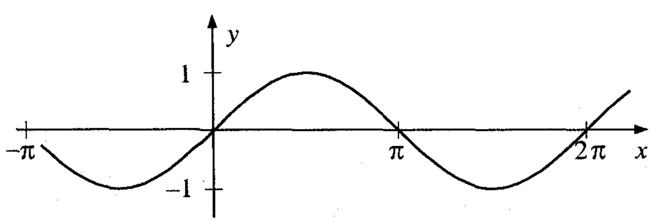


2 Arcsin: The inverse function to sin

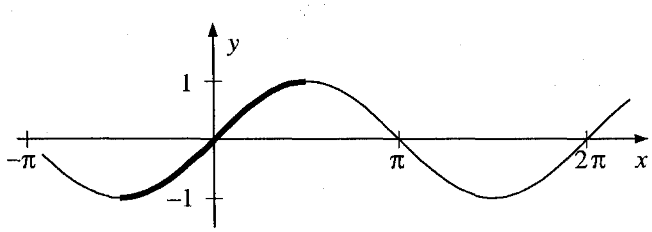
Example 76 The equation $y = \sin x$ defines a function from the set A of real numbers to the set B of real numbers. Does it have an inverse function?

Again, the answer is no, and for the same two reasons as in Example 75. For some values of y in B, such as $y = 5$, there are no values of x such that $\sin x = y$. For other values of y, such as $y = 1/2$, there are many values of x: $\sin \pi/6 = 1/2$, $\sin 5\pi/6 = 1/2$, $\sin 13\pi/6 = 1/2$, and so on.

In Example 75, we were able to overcome these difficulties, by restricting the sets A and B that the function is defined on. Can we do this here? Let us look at the graph of $y = \sin x$.



Let us start by including the number 0 in our set A. We must choose for set A a domain on which the function $y = \sin x$ is monotone, and it's easiest to take the for set A the set $-\pi/2 \le x \le \pi/2$:



Now we can choose for set A our interval $-\pi/2 \le x \le \pi/2$, and for set B the interval $-1 \le y \le 1$, and for every y in set B, there exists exactly one x in A such that $\sin x = y$.

The inverse function to $y = \sin x$, defined in this way, is important enough to merit its own name. It is called the *arcsine* function[2], and if $y = \sin x$ (with x and y in the two sets described above), we write $x = \arcsin y$.

However, sometimes we will discuss the arcsine function in its own right. Then we will write $y = \arcsin x$, where $-1 \leq x \leq 1$, and $-\pi/2 \leq y \leq \pi/2$. We have already had a chance encounter with this function on the calculator. Now we will get to know it much better.

Example 77 Find $\arcsin 1/2$.

Solution: Again, if $y = \arcsin 1/2$, then $\sin y = 1/2$. There are many such angles, but we have agreed to choose the unique y such that $-\pi/2 \leq y \leq \pi/2$. This value is $\pi/6$, so $\arcsin 1/2 = \pi/6$.

Example 78 Find $\arcsin -\sqrt{3}/2$.

Solution: If $y = \arcsin \sqrt{3}/2$, then $\sin y = -\sqrt{3}/2$ and $-\pi/2 \leq y \leq \pi/2$. Hence $y = -\pi/3$.

Example 79 Find $\arcsin(\sin \pi/5)$.

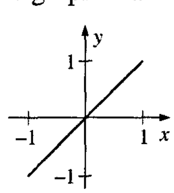
Solution: We let $y = \arcsin(\sin \pi/5)$, and rewrite this statement as $\sin x = \sin(\pi/5)$. We know that there are many solutions to this equation: $x = \pi/5$, $4\pi/5$, and so on. But since we require that $-\pi/2 \leq y \leq \pi/2$, so $\arcsin(\sin \pi/5)$ is just $\pi/5$.

Example 80 Find $\arcsin(\sin 3\pi/5)$.

Solution: As usual, we write $x = \arcsin(\sin 3\pi/5)$, so that $\sin x = \sin 3\pi/5$. But this time we cannot choose $x = 3\pi/5$, since this value is not in the required interval. However, there is a value of x in the interval that satisfies this equation. It is $x = 2\pi/5$, and this is our required value.

Example 81 Draw the graph of the function $y = \sin(\arcsin x)$.

Solution: We first decide what the domain of definition of this function is. Since we are taking $\arcsin x$, we must have $-1 \leq x \leq 1$. And since y is the sine of some angle, $-1 \leq y \leq 1$ as well. On these intervals, $\sin(\arcsin x)$ is simply x, so the graph is as follows:

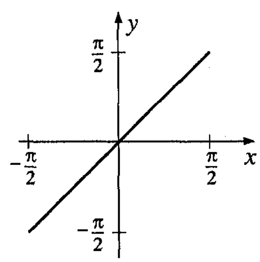


[2]We can explain the odd notation $y = \arcsin x$ by remembering that it stands for the sentence "y is the arc (or angle) whose sine is x".

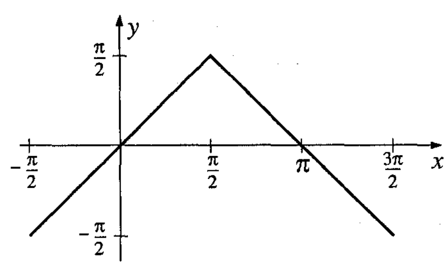
Example 82 Draw the graph of the function $y = \arcsin(\sin x)$.

Solution: We will soon see that this is not the same as the previous example(!). Again, we begin by deciding on the domain of the function. We can take the sine of any real number x. Since the resulting value is in the interval from -1 to 1, we can then take the arcsine of this value. Hence the function $y = \arcsin(\sin x)$ is defined for any real number x. The possible values for y are those of the arcsine function, so $-\pi/2 \leq y \leq \pi/2$.

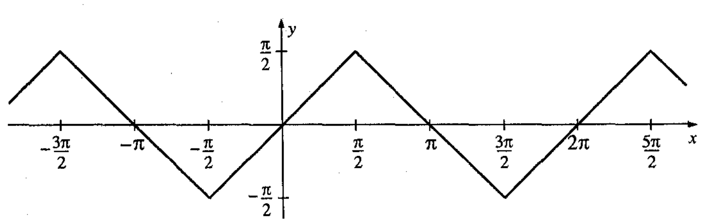
Let us next look at the function for values of x between $-\pi/2$ and $\pi/2$. On this interval, we find that $y = x$, so the graph looks like this:



But x can take on any real value, so we are not finished. Let us look at the function for values of x between $\pi/2$ and $3\pi/2$. In this interval, $\sin x$ decreases from 1 to -1, so the values of $y = \arcsin(\sin x)$ will decrease from $\pi/2$ to $-\pi/2$. The reader is invited to check that the graph is the following:



And now we note that the function $y = \arcsin(\sin x)$ is periodic, with period 2π: $\arcsin(\sin[x+2\pi]) = \arcsin(\sin x)$. The full graph is as follows:



As we mentioned, the arcsine function appears on your calculator, and you will find that the calculator knows how to compute $\arcsin x$ for any number x. The button which does this is marked either $\boxed{\text{arcsin}}$ or $\boxed{\sin^{-1}}$. (We are unhappy with the second notation, and will use only the first.[3])

In just the same way, we can define an inverse of the function $y = \tan x$. We choose an interval near 0 for which the function is monotone. It will be convenient once again to choose the interval $-\pi/2 \leq x \leq \pi/2$. Then the inverse function, which we will call arctan x, will take on all real values.

This is really a very nice function, since it is defined for any real number x. Its y-values, however, are restricted to the interval $-\pi/2 \leq x \leq \pi/2$. Indeed, the function $y = \arctan x$ supplies us with a one-to-one correspondence between all the real numbers and the numbers on that interval.[4]

We can also define an inverse of the function $y = \cos x$. But we cannot choose the same interval we chose for the sine and tangent, since the cosine is not monotone on $-\pi/2 \leq x \leq \pi/2$. Instead we choose the interval $0 \leq x \leq \pi$, on which the function $y = \cos x$ is monotone and decreasing. We write the new function $y = \arccos x$.

Example 83 Find $\sin(\arccos(5/13))$.

Solution: Let $\alpha = \arccos 5/13$. Then $\cos\alpha = 5/13$, $0 \leq \alpha \leq \pi$, and we seek $\sin\alpha$. This is a problem we've seen before. We find that $\sin\alpha = 12/13$.

Example 84 Find $\cos(\arcsin(-3/5))$.

Solution: Let $\alpha = \arcsin(-3/5)$. Then $\sin\alpha = (-3/5)$, $-\pi/2 \leq \alpha \leq \pi/2$, and we seek $\cos\alpha$. This time, α is in quadrant IV, so $\cos\alpha = 4/5$, a positive number.

In summary,

$$y = \arcsin x \text{ means } x = \sin y \text{ and } -\pi/2 \leq x \leq \pi/2$$
$$y = \arccos x \text{ means } x = \cos y \text{ and } 0 \leq x \leq \pi$$
$$y = \arctan x \text{ means } x = \tan y \text{ and } -\pi/2 \leq x \leq \pi/2.$$

[3]The notation $\sin^{-1} 1/3$ looks too much like the notation $\sin^2 1/3$, which of course means $(\sin 1/3)(\sin 1/3)$. By analogy, the symbol $\sin^{-1} 1/3$ "should" mean $1/\sin(1/3) = \csc 1/3$. But it means something completely different. While it remains standard in some texts, and on some calculators, we will not use it.

[4]One way to understand this is to say that there are "just as many numbers" on the whole line as there are on the interval $-\pi/2 \leq x \leq \pi/2$. When mathematicians started talking like this, some people thought this statement strange, since the interval has finite length while the line is infinite in length. What they meant, however, was simply that the notion of "length" is not based on the "number" of points in the segment being measured.

Exercises

1. Find the value of:
   (a) $\arcsin 0.5$ (b) $\arccos 0.5$ (c) $\arctan 1$
   (d) $\arcsin(-\frac{\sqrt{3}}{2})$ (e) $\arccos(-\frac{\sqrt{3}}{2})$ (f) $\arctan(-\sqrt{3})$
   (g) $\arcsin 2$

2. Find the numerical value of the following expressions:
   (a) $\sin(\arcsin 0.5)$ (b) $\cos(\arccos \frac{\sqrt{3}}{2})$ (c) $\tan(\arctan(-1))$
   (d) $\arcsin(\sin \frac{\pi}{3})$ (e) $\arccos(\cos \frac{11\pi}{6})$

3. Show that $\sin(\arccos b) = \pm\sqrt{1-b^2}$. What determines whether we should choose the positive sign or the negative sign?

4. Express $\tan(\arcsin b)$ in terms of b. Will we need an ambiguous sign, as we did in Problem 3?

5. Express $\cos(\arctan b)$ in terms of b.

6. Show that $\arccos(\sin\alpha) = \pi/2-\alpha$, for $0 \leq \alpha \leq \pi/2$. What can you say for values of α outside this set?

7. Find each of the following values:
   (a) $\arcsin(\sin \frac{\pi}{11})$ (b) $\arcsin(\sin \frac{2\pi}{11})$ (c) $\arcsin(\sin \frac{3\pi}{11})$
   (d) $\arcsin(\sin \frac{4\pi}{11})$ (e) $\arcsin(\sin \frac{5\pi}{11})$ (f) $\arcsin(\sin \frac{6\pi}{11})$
   Hint: For most students, Part (f) is much more difficult than the others.

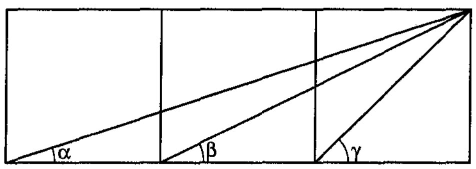
8. Draw the graph of the function $y = \cos(\arccos x)$.

9. Draw the graph of the function $y = \arccos(\cos x)$.

10. Find the numerical value of $\sin(\arcsin 3/5 + \arcsin 5/13)$. (Hint: Let $\alpha = \arcsin 3/5$, $\beta = \arcsin 5/13$, and use the formula for $\sin(\alpha+\beta)$.)

11. Recall that $\tan(\alpha+\beta) = (\tan\alpha + \tan\beta)/(1-\tan\alpha\tan\beta)$. Using this formula, prove that $\arctan a + \arctan b = \arctan \frac{a+b}{1-ab}$.

12. The diagram below shows three equal squares, with angles α, β, γ as marked. Prove that $\alpha + \beta = \gamma$.



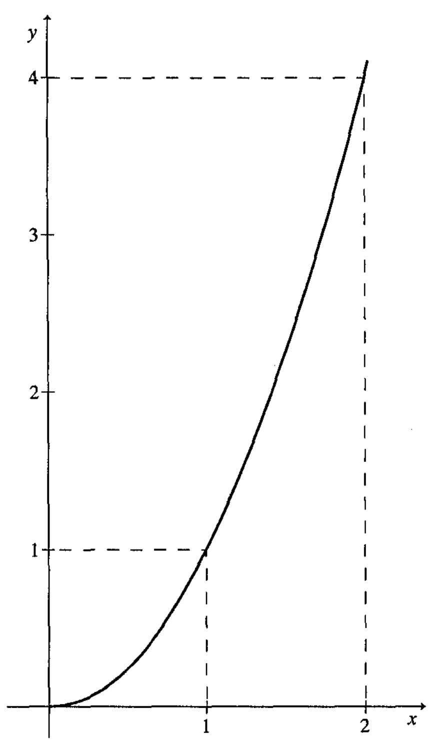
Hint: Note that $\alpha = \arctan 1/3$, $\beta = \arctan 1/2$, and $\gamma = \arctan 1$. Then use the formula from Problem 8.

13. Extra credit: Can you prove the result in Problem 9 without using trigonometry?

3 Graphing inverse functions

How is the graph of a function related to the graph of its inverse function?

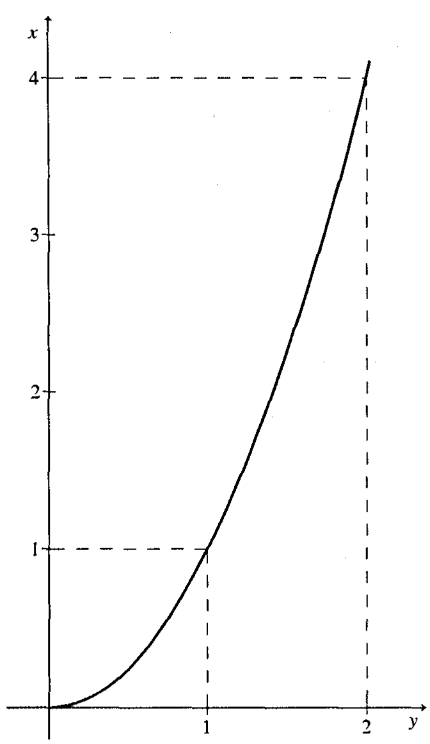
Example 85 Let $y = x^2$, for $x \geq 0$ and $y \geq 0$. As we have seen, it is monotone increasing. Here is its graph:



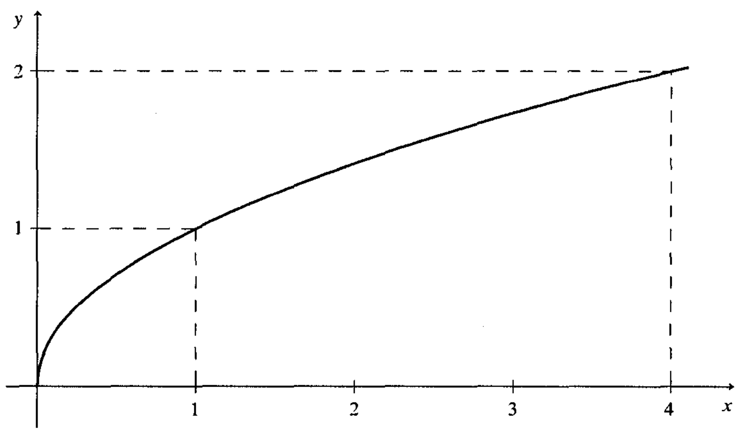
We can read the values of the function from the graph. For example, the diagram shows that $f(2) = 4$, since the x-value 2 corresponds to the y-value 4 on the graph.

The inverse function, as we have seen, is $g(y) = \sqrt{y}$. This graph also contains all the information we need to find values of the inverse function. We just choose our first number on the y-axis, and use the graph to get the corresponding number on the x-axis. For example, if we want $g(4)$, we find the number 4 on the y-axis, and use the graph to find the corresponding number (which is 2) on the x-axis.

However, many people are more comfortable using the letter x to denote the number in set A for which the function is making an assignment, and the letter y for the number in set B to which x is assigned. There are two ways to accommodate this need. We can simply relabel the axes of the original graph:

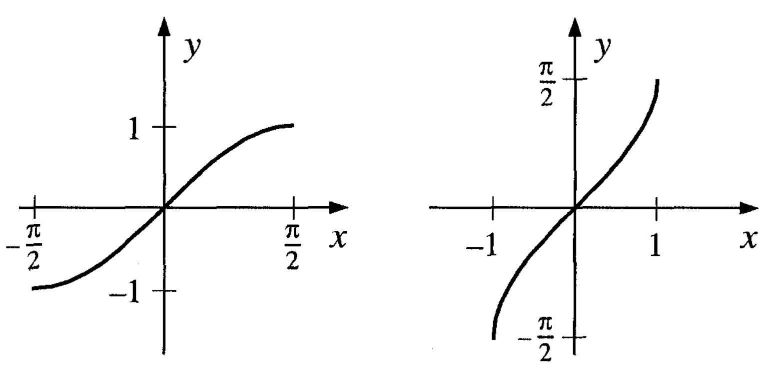


But many people prefer the x-axis to appear horizontal, and the y-axis to appear vertical, on the page. We can accommodate them by reflecting the graph around a diagonal line:

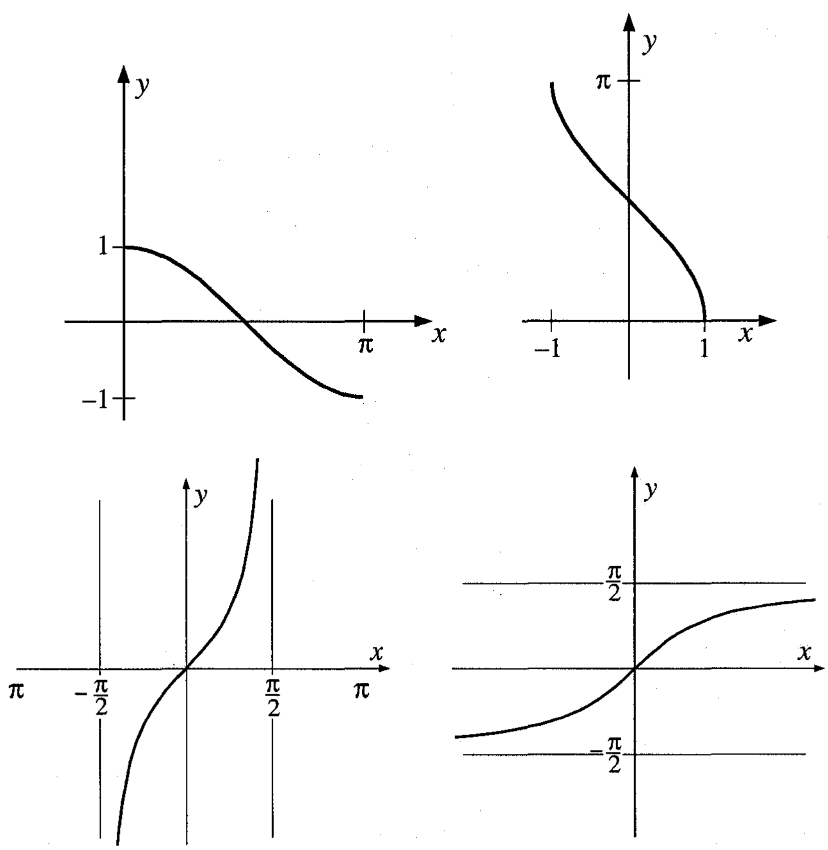


This graph contains the same information as the others, but in a more conventional form.

Here are graphs of the sine function, and its inverse, the arcsine function. The graph of the inverse function is given in the conventional position. Note that the domains are restricted as we discussed above.



And here are graphs of $y = \arccos x$ and $y = \arctan x$:



The graph of $y = \arctan x$ shows clearly how the function maps the entire real line onto a finite interval.

4 Trigonometric equations

We must often solve trigonometric equations: equations in which trigonometric functions of the unknown quantity appear. We can often use the following method to solve these:

1. Reduce them to the form $\sin a = a$, $\cos x = a$, or $\tan x = a$;
2. Locate the solutions to these simple equations between 0 and 2π;
3. Use the periods of the functions $\sin x$, $\cos x$, and $\tan x$ to find all the solutions.

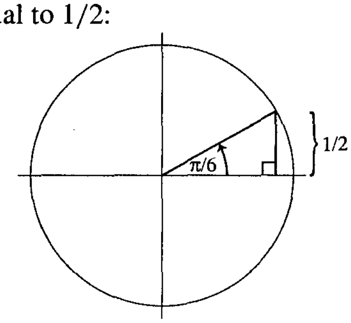
We start with a simple example.

Example 86 Solve the equation $\sin x = 1/2$.

This means that we must find *all* the values of x for which $\sin x = 1/2$.

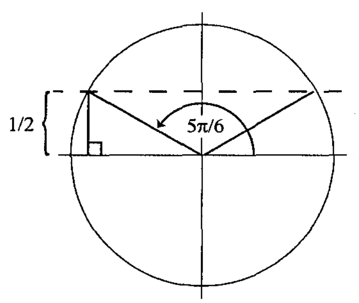
We will describe two ways of finding these values. Our first method uses a circle, and our second uses a graph of the function $y = \sin x$.

Solution 1: We first use a unit circle, centered at the origin. As a first step, we find two particular answers. We recall that $\sin \pi/6 = 1/2$. Let us illustrate this on our circle. We draw an angle of $\pi/6$, and find the line segment which is equal to $1/2$:



This is our first answer.

But if we draw a horizontal line across the circle, we find another angle whose sine is $1/2$:



So we have the answers $x = \pi/6$ and $x = 5\pi/6$. These are all the possible answers in the interval $0 \le x \le 2\pi$.

To find more answers, note that we can make as many complete rotations about the circle as we like (either clockwise or counterclockwise), and we will get back to the same point.

From our first answer, we get the new values $\pi/6 \pm 2\pi$, $\pi/6 \pm 4\pi$, $\pi/6 \pm 6\pi$, and so on. We can write these as $\pi/6 + 2\pi n$ for any integer n.

From our second answer, we get the new values $5\pi/6 \pm 2\pi$, $5\pi/6 \pm 4\pi$, $5\pi/6 \pm 6\pi$, and so on. We can write these as $5\pi/6 + 2\pi n$ for any integer n.

So we have two sequences of answers:

$$\begin{aligned} \pi/6 + 2\pi n \quad & \text{for any integer } n \text{, and} \\ 5\pi/6 + 2\pi n \quad & \text{for any integer } n. \end{aligned}$$

These two sequences contain all the solutions to our equation.

We can express all these solutions more elegantly. We started with the basic answers $x = \pi/6$ and $5\pi/6$. We can write $5\pi/6$ as $\pi - \pi/6$. Then the second set of answers will be $\pi - \pi/6 + 2\pi n$. Then we can write the two sequences of solutions as:

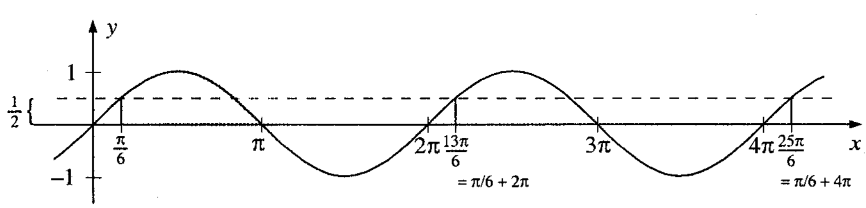
$$\begin{aligned} 2n\pi + \pi/6 \quad & \text{for any integer } n \text{, and} \\ (2n+1)\pi - \pi/6 \quad & \text{for any integer } n. \end{aligned}$$

Now we note that the expression $2n\pi$ represents any even integer multiple of π, and we must add $\pi/6$ to this to get an answer to our equation, while $(2n+1)\pi$ represents any odd integer multiple of π, and we must subtract $\pi/6$ to get an answer. So we can write our solutions elegantly as:

$$\pi k + (-1)^k(\pi/6) \quad \text{for any integer } k.$$

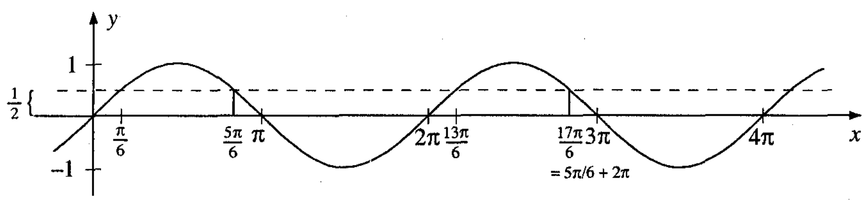
The reader can verify that for $k = 2n$ (that is, for an even integer k), we obtain the first sequence of solutions, and for $k = 2n + 1$ (for odd integers k), we obtain the second sequence.

Solution 2: We can use the graph $y = \sin x$ to solve our equation. Along with the graph of the function $y = \sin x$, we draw the line $y = 1/2$:



This line intersects the graph at a point whose x-coordinate is $\pi/6$. This is our first initial solution. Since the graph of $y = \sin x$ has period 2π, we will find more solutions, whose x-coordinates are $\pi/6 \pm 2\pi$, $\pi/6 \pm 4\pi$, $\pi/6 \pm 6\pi$, and so on.

The line $y = 1/2$ also intersects the graph at the point whose x-coordinate is $5\pi/6$. This is our second initial solution. Again, periodicity gives us more solutions, whose x-coordinates are $5\pi/6 \pm 2\pi$, $5\pi/6 \pm 4\pi$, $5\pi/6 \pm 6\pi$, and so on.



So, as before, we have two sequences of solutions: $\pi/6 + 2\pi n$ and $5\pi/6 + 2\pi n$, and we can express them elegantly as $\pi k + (-1)^k(\pi/6)$, for any integer k.

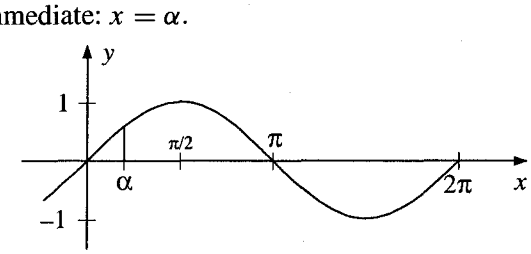
> *In general, if we need to solve a simple trigonometric equation, we can first find all the solutions between* 0 *and* 2π*, then use periodicity to get all the other solutions*

Exercises

1. Using the graph above, find all the points x on the x-axis such that $\sin x > 1/2$.
2. Solve the equation $\sin x = -1/2$.
3. Solve the equation $\cos x = \sqrt{2}/2$.
4. Solve the equation $\tan x = 1$.
5. Solve the equation $\sin x = -1$.

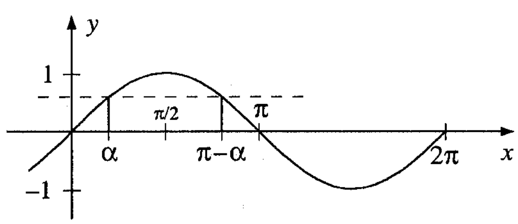
5 A more general trigonometric equation

Take some acute angle α. We wish to solve the equation $\sin x = \sin\alpha$. One solution is immediate: $x = \alpha$.



Periodicity then gives us the sequence of solutions $\alpha + 2n\pi$, for any integer n.

We also have a second immediate solution: $x = \pi - a$.



This solution gives us a second sequence of solutions, which we can write as $\pi - a + 2n\pi$, for any integer n.

So the solutions are given by

$$\alpha + 2n\pi \quad \text{and} \quad \pi - a + 2n\pi$$

for any integer n. As before, we can state this result more elegantly as

$$k\pi + (-1)^k \alpha .$$

Exercises

1. Solve the equation $\sin x = \sin \pi/5$.
2. Solve the equation $\sin x = \sin \pi/2$.
3. Using the graph of the function $y = \cos x$, show that the solutions to the equation $\cos x = \cos \alpha$ (for some acute angle α) are given by $2\pi n + \alpha$ and $2\pi n - \alpha$, for any integer n.
4. Solve the equation $\cos x = \cos \pi/5$.
5. Using the graph of the function $y = \tan x$, show that the solutions to the equation $\tan x = \tan \alpha$ (for some acute angle α) are given by $\alpha + \pi n$, for any integer n.
6. Solve the equation $\tan x = \tan \pi/5$.
7. Suppose α is some fixed angle. Express in terms of α all the solutions to the equation $\sin x = -\sin \alpha$. (Hint: One approach is to recall that $-\sin \alpha = \sin(-\alpha)$.)

8. Suppose α is some fixed angle. Express in terms of α all the solutions to the equation $\cos x = -\cos\alpha$.

9. Check that the formula $x = (-1)^n\alpha + \pi n$ represents all the solutions to the equation $\sin x = \sin\alpha$, as n takes on all integer values.

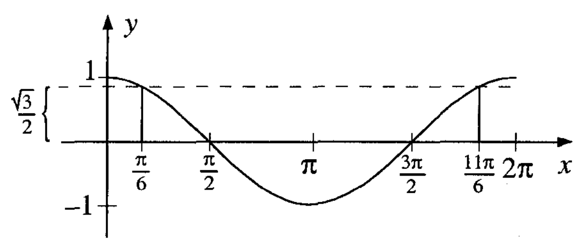
6 More complicated trigonometric equations

Example 87 Solve the equation $\cos^2 x = 3/4$.

Solution 1: This equation is equivalent to the two equations

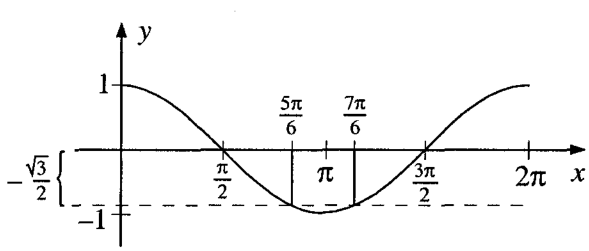
$$\cos x = \frac{\sqrt{3}}{2} \text{and} \quad \cos x = -\frac{\sqrt{3}}{2}.$$

The first equation has two solutions between 0 and 2π. They are $x = \pi/6$ and $x = 11\pi/6$:



Then periodicity gives us two sequences of solutions for our first equation: $x = \pi/6 + 2\pi n$ and $x = 11\pi/6 + 2\pi n$, for any integer n.

Now we turn to our second equation. The equation $\cos x = -\sqrt{3}/2$ has two solutions between 0 and 2π, namely, $5\pi/6$ and $7\pi/6$.



This gives two more sequences of solutions: $x = 5\pi/6 + 2\pi n$ and $x = 7\pi/6 + 2\pi n$, for any integer n.

Altogether, there are four sequences of solutions:

$$x = \pi/6 + 2\pi n \quad \text{for any integer } n$$
$$x = 5\pi/6 + 2\pi n \quad \text{for any integer } n$$
$$x = 7\pi/6 + 2\pi n \quad \text{for any integer } n\text{, and}$$
$$x = 11\pi/6 + 2\pi n \quad \text{for any integer } n.$$

The reader is invited to try to find one elegant formula that will give all these solutions.

Solution 2: [Outline] We can write $\cos 2x = 2\cos^2 x - 1 = 2(3/4) - 1 = 1/2$. Then we solve for $2x$ (for example, by looking at the graph of $y = \cos 2x$) to find the four sequences of values for $2x$. Finally, we divide each value we find by 2, to solve for x.

Example 88 Solve the equation $\sin x = \cos x$.

Solution 1: We can recall that $\cos x = \sin(\pi/2 - x)$, and rewrite the equation in terms of the sine function:

$$\sin x = \sin(\pi/2 - x)\,.$$

But, as we saw earlier, the equation $\sin x = \sin\alpha$ has two sequences of solutions:

(i) $x = \alpha + 2\pi n$ for any integer n, or

(ii) $x = (\pi - \alpha) + 2\pi n$ for any integer n.

We apply this result with $\alpha = \pi/2 - x$. From sequence (i) we get $x = \pi/4 + \pi n$. From sequence (ii) we get the equation $x = \pi/2 + x + 2\pi n$, which has no solutions at all.

Thus the solutions to the equation $\sin x = \cos x$ are given by the sequence

$$x = \pi/4 + \pi n \quad \text{for any integer } n.$$

Solution 2: If we divide both sides of the equation by $\cos x$, we obtain a new equation involving only the tangent function: $\tan x = 1$.

The only solution between 0 and π is $x = \pi/4$. Since the period of the tangent function is π, this initial solution give all the others, which can be written as $\pi/4 + \pi n$, for any integer n.

A fine point: If we divide by $\cos x$, we must check that this expression cannot be equal to 0. In fact, if $\cos x = 0$, we cannot have $\cos x = \sin x$, because the two functions are never 0 for the same value of x.

Example 89 Solve the equation $\sin x = \cos 2x$.

Again, we offer two solutions.

Solution 1: We rewrite the equation in terms of the sine function, and proceed as in the second solution to Example 87. We have $\sin x = \sin(\pi/2 - 2x)$.

We can now distinguish two cases, as we did in Solution 1 to Example Example 88. If $x = (\pi/2 - 2x) + 2\pi n$, then $x = \pi/6 + 2\pi n/3$, which gives one sequence of solutions.

In the second case, we have $x = (\pi - (\pi/2 - 2x)) + 2\pi n$. This leads to $x = -\pi/2 - 2\pi n$. This is a second sequence of solutions.

Solution 2: We know that $\cos 2x = 1 - 2\sin^2 x$ (see Chapter 7). So we can rewrite the given equation as

$$\sin x = 1 - 2\sin^2 x \quad \text{or} \quad 2\sin^2 x + \sin x - 1 = 0\,.$$

Let us try to solve for $\sin x$ by factoring (if this doesn't work, we can always use the quadratic formula). We have $(2\sin x - 1)(\sin x + 1) = 0$, so $\sin x = 1/2$ or $\sin x = -1$.

We can solve these equations separately, using the methods we have already demonstrated.

For $\sin x = 1/2$, we find $x = \pi/6 + 2\pi n$ or $x = 5\pi/6 + 2\pi n$, for any integer n.

For $\sin x = -1$, we have $x = 3\pi/2 + 2\pi n$, for any integer n.

There are three sequences of solutions.

Example 90 Solve the equation $\tan^2 x = 3$.

Solution: The equation is equivalent to the two equations

$$\tan x = \sqrt{3} \text{ and } \tan x = -\sqrt{3}\,.$$

An initial answer to the first equation is $x = \pi/3$, and periodicity gives the answers $\pi/3 + \pi n$, for any integer n.

The second equation has an initial solution $x = -\pi/3$, and periodicity gives the answers $-\pi/3 + \pi n$, for any integer n. These two sequences give the complete solution.

In conclusion, we note that we have already shown (Ch. 7, Appendix I.2; p. 159), that any trigonometric identity can be reduced to an algebraic identity. The same is true for trigonometric equations. However, the algebraic equation that results is often more difficult than the same equation in trigonometric form.

Exercises

1–12. Find the solution sets for the following equations:

1. $\sin 2x = 1$	2. $\sin x/2 = 1/2$
3. $\cos x = \sin 2x$	4. $\sin x = \sin 3x$
5. $\cos x = \sin 4x$	6. $26 \sin^2 x + \cos^2 x = 10$
7. $\cos^2 x - \cos x = \sin^2 x$	8. $3 \tan^2 x = 12$
9. $\cos 2x = 2 \sin^2 x$	10. $\tan^2 x = \cot x$
11. $\dfrac{5}{\cos^2 x} = 7 \tan x + 3$	12. $\sqrt{3} \tan^2 x + 1 = (1 + \sqrt{3}) \tan x$

13. Let us look back at Example 89. Solution 1 gave the general solution as

$$\begin{aligned} x &= \pi/6 + 2\pi n/3 \quad \text{or} \\ x &= -\pi/2 - 2\pi n \quad \text{for any integer } n. \end{aligned}$$

But Solution 2 gave the general solution

$$\begin{aligned} x &= \pi/6 + 2\pi n \quad \text{or} \\ x &= 5\pi/6 + 2\pi n \quad \text{or} \\ x &= 3\pi/2 + \pi n \quad \text{for any integer } n. \end{aligned}$$

Show that these two sets of solutions are actually identical.

Appendix – The Miracles Revealed

In Chapter 5 we discussed two small miracles:

The Miracle of the Tangent

If we draw a tangent to the curve $y = \sin x$ at the point $x = \alpha$, then the distance between d, the point of intersection of this tangent with the x-axis, and the point $(\alpha, 0)$ is $|\tan \alpha|$.

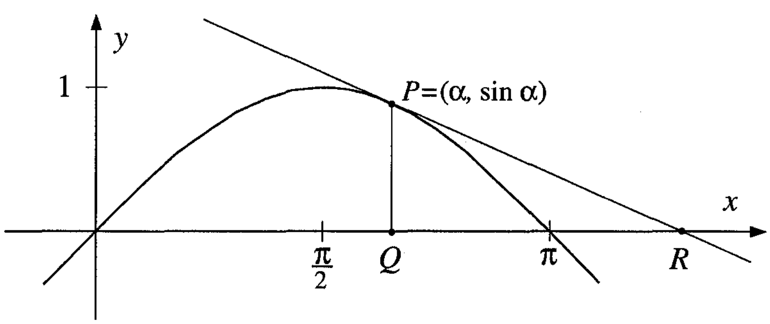
The Miracle of the Arch

The area under one arch of the curve $y = \sin x$ is 2.

We now return to these results and furnish their proofs. Each draws on techniques that are standard in the study of the calculus. In particular, each uses the fact that the quotient $\sin h/h$ approaches 1 as h gets close to 0. We showed why this is true on Chapter 5, p. 118. A more rigorous proof would involve the notion of limit, which is the fundamental notion of the calculus. In this section, we give a sketch of a proof for each miracle that parallels the more formal approach used in a course on calculus.

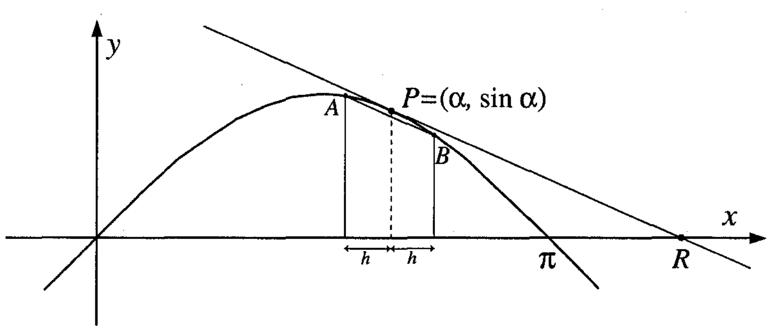
Proof of The Miracle of the Tangent

The diagram shows a point $P(\alpha, \sin\alpha)$ on the curve $y = \sin x$. It intersects the x-axis at point R. We will show that $QR = |\tan\alpha|$, by writing an equation for line PR, then finding the coordinates of point R.



We can write the equation of a line using the coordinates of a point on the line and the line's slope. The point will be P, with coordinates $(\alpha, \sin\alpha)$.

To get the slope of line PR, we use a technique from the calculus. Instead of looking at tangent PR, we look at a secant to the curve $y = \sin x$, which intersects the curve near point P. We take two points, A and B, one just to the left of P and one just to the right, at a small distance h along the x-axis:



The coordinates of point A are $(\alpha - h, \sin(\alpha - h))$, and the coordinates of B are $(\alpha + h, \sin(\alpha + h))$. From these two points we can compute the slope of secant AB:

$$\begin{aligned}&\frac{\sin(\alpha+h)-\sin(\alpha-h)}{2h}\\&\quad=\frac{\sin\alpha\cos h+\cos\alpha\sin h-(\sin\alpha\cos h-\cos\alpha\sin h)}{2h}\\&\quad=\frac{2\cos\alpha\sin h}{2h}=\cos\alpha\left(\frac{\sin h}{h}\right).\end{aligned}$$

Now we take smaller and smaller (positive) values of h, so that points A and B get closer together, and secant AB begins to resemble tangent PR. The expression $\sin h/h$ gets closer and closer to 1 as h approaches 0. And of course $\cos\alpha$ does not change as h approaches 0. So the slope of secant AB, which is looking more and more like tangent PR, gets closer and closer to the value $\cos\alpha$. It is reasonable, then, to expect that the slope of PR is exactly $\cos\alpha$. (In calculus, this technique of finding the slope of a tangent to a curve will receive a full justification. It is related to the notion of the *derivative* of a function.)

Now we can find the equation of line PR, through point $P(\alpha, \sin\alpha)$ and with slope $\cos\alpha$:

$$\frac{y-\sin\alpha}{x-\alpha}=\cos\alpha\ .$$

We need the x-coordinate of point R. Its y-coordinate is 0, so its x-coordinate is obtained by letting $y = 0$ in the equation above. We find that

$$x=\alpha-\frac{\sin\alpha}{\cos\alpha}=\alpha-\tan\alpha\ .$$

Then the length of QR is just $|\alpha - (\alpha - \tan\alpha)| = |\tan\alpha|$.

Exercises

1. The diagrams above show a case where $\alpha > \pi/2$. Take a numerical value of α slightly larger than $\pi/2$ (for example, $\alpha = 1.6$), and follow the argument above. (Note that for such values value of α, $\tan\alpha < 0$.)

2. Take a value of α between 0 and $\pi/2$, and follow the argument again. Note that for such values of α, $\tan\alpha > 0$. Where does point R fall in these cases?

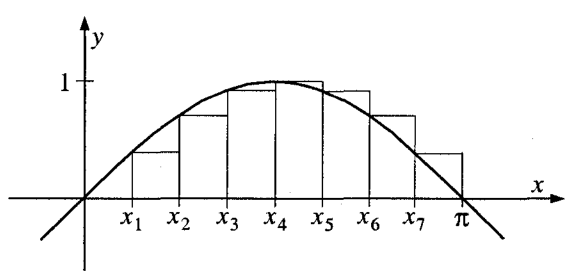
3. When does point R fall on the origin?

4. Where does point R fall when α is very close to $\pi/2$? How does your answer depend on whether α is greater or less than $\pi/2$?

Proof of the Miracle of the Arch

This miracle concerns the area under one arch of the curve $y = \sin x$, which we claim is exactly 2. On p. 117 we showed that this area A satisfies the inequalities $\pi/2 < A < \pi$. We did this by drawing figures bounded by straight lines that approximated the area A. We can improve on this approximation by taking regions closer and closer to the region whose area we want to measure. We will construct these regions out of rectangles.

We take the interval from 0 to π along the x-axis, and divide it into many equal pieces. If there are n of these pieces, then the points of division are $x_0 = 0$, $x_1 = \pi/n$, $x_2 = 2\pi/n$, $\ldots x_{n-1} = (n-1)\pi/n$, and $x_n = n\pi/n = \pi$. For each point x_i, we draw a rectangle by erecting a perpendicular to the x-axis with one endpoint at x_i and the other on the curve $y = \sin x$ (the diagram shows the case $n = 8$):



Note that the rectangles are inscribed in the arch for $0 < x_i < \pi/2$, and they are circumscribed for $\pi/2 < x < \pi$. Also, the widths of the rectangles are all π/n. Let us set $h = \pi/n$. Then as h approaches 0, the rectangles get thinner and more numerous, and the sum of their areas approaches the area A.

Finally, note that the rectangles associated with $x_0 = 0$ and $x_n = \pi$ are "degenerate": their area is 0 (no matter what value we choose for n). It will be convenient for us to ignore the rectangle associated with x_0, but to include the one associated with x_n. Then we can write the sum of the areas of these rectangles as

$$\begin{aligned} &h\sin x_1 + h\sin x_2 + h\sin x_3 + \cdots + h\sin x_n \\ &= h\Big(\sin\frac{\pi}{n} + \sin\frac{2\pi}{n} + \sin\frac{3\pi}{n} + \cdots + \sin\frac{n\pi}{n}\Big) \end{aligned}$$

$$= h\frac{\sin\frac{n+1}{2}\frac{\pi}{n}\sin\frac{n}{2}\frac{\pi}{n}}{\sin\frac{\pi}{2n}} = h\frac{\sin\frac{n+1}{2}\frac{\pi}{n}}{\sin\frac{\pi}{2n}}\cdot\sin\frac{\pi}{2} = h\frac{\sin\frac{n+1}{n}\frac{\pi}{2}}{\sin\frac{\pi}{2n}}\,.$$

If n is very large, our set of rectangles will look more and more like the area A. But as n gets very large, the fraction $(n+1)/n$ approaches the value 1. Hence our expression for A gets close to

$$h\frac{\sin\pi/2}{\sin\pi/2n} = \frac{h}{\sin h/2}\,.$$

Now if we let $k = h/2$ this expression is equal to $2k/\sin k = 2(k/\sin k)$. As h gets close to 0, so does k, and so the expression approaches 1. Its reciprocal, which is $k/\sin k$, also approaches 1. This means that the sum that approximates A gets close to $2\cdot 1 = 2$, a miraculous result.

In calculus, this technique for finding the area under a curve is related to the *integral* of a function.

Exercises

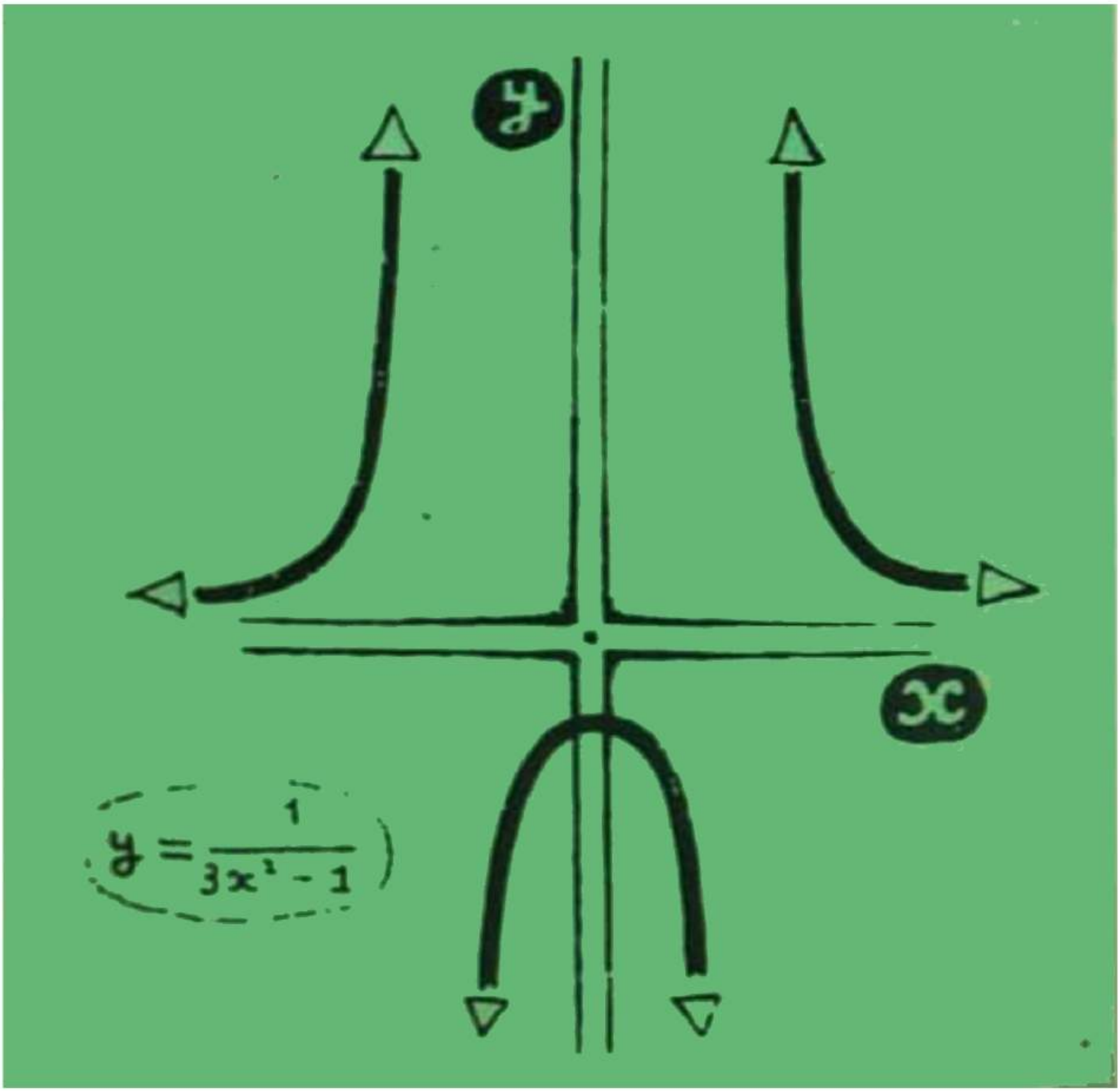
1. Using a calculator, find the approximations to A given by taking $n = 4$ and $n = 8$.

2. What do you think the area under the curve $y = \sin x$ is from $x = 0$ to $x = \pi/2$?

3. Try using the method outlined above to find the area under the curve $y = \sin x$ from $x = 0$ to $x = \pi/2$. Is the result what you might have expected?

I.M. Gel'fand E.G. Glagoleva E.E. Shnol

Functions and Graphs



Cover art by V. Yankelevski

I.M. Gel'fand E.G. Glagoleva E.E. Shnol

Functions and Graphs

Birkhäuser
Boston • Basel • Berlin

I. M. Gelfand
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903

E.G. Glagoleva
University of Moscow
117234 Moscow
USSR

E.E. Shnol
University of Moscow
117234 Moscow
USSR

Functions and Graphs was originally published in 1966 in the Russian language under the title Funktsii i grafiki.

The english language edition was translated and adapted by Leslie Cohn and David Sookne under the auspices of the Survey of Recent East European Mathematical Literature, conducted by the University of Chicago under a grant from the National Science Foundation. It is republished here with permission from the University of Chicago for its content and the MIT Press for its form.

Library of Congress Cataloging-in-Publication Data
Gel'fand, I. M. (Izrail ' Moiseevich)
[Funktsii i grafiki. English]
Functions and graphs / I.M. Gelfand, E.G. Glagoleva, and E.E. Shnol ; translated and adopted from the second Russian edition by Thomas Walsh and Randell Magee.
p. cm.
Translation of Funktsii i grafiki.
ISBN 0-8176-3532-7
1. Functions. 2. Algebra--Graphic methods. I. Glagoleva, E. G. (Elena Georgievna) II. Shnol', E. E. (Emmanuil El'evich) III. Title.
QA331.G3613 1990 90-48213
515--dc20 CIP

© I.M. Gel'fand, 1990

Printed on acid-free paper.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the copyright owner.

ISBN 0-8176-3532-7 SPIN 10882894
ISBN 3-7643-3532-7

Text reproduced with permission of the MIT Press, Cambridge, Massachusetts, from their edition published in 1967.

Printed and bound by Quinn-Woodbine, Woodbine, NJ.

Printed in the U.S.A.

9 8 7 6

Preface

Dear Students,

We are going to publish a series of books for high school students. These books will cover the basics in mathematics. We will begin with algebra, geometry and calculus. In this series we will also include two books which were written 25 years ago for the Mathematical School by Correspondence in the Soviet Union. At that time I had organized this school and I continue to direct it.

These books were quite popular and hundreds of thousands of each were sold. Probably the reason for their success was that they were useful for independent study, having been intended to reach students who lived in remote places of the Soviet Union where there were often very few teachers in mathematics.

I would like to tell you a little bit about the Mathematical School by Correspondence. The Soviet Union, you realize, is a large country and there are simply not enough teachers throughout the country who can show all the students how wonderful, how simple and how beautiful the subject of mathematics is. The fact is that everywhere, in every country and in every part of a country there are students interested in mathematics. Realizing this, we organized the School by Correspondence so that students from 12 to 17 years of age from any place could study. Since the number of students we could take in had to be restricted to about 1000, we chose to enroll those who lived outside of such big cities as Moscow, Leningrad and Kiev and who inhabited small cities and villages in remote areas. The books were written for them. They, in turn, read them, did the problems and sent us their solutions. We never graded their work -- it was forbidden by our rules. If anyone was unable to solve a problem then some personal help was given so that the student could complete the work.

Of course, it was not our intention that all these students who studied from these books or even completed the School should choose mathematics as their future career. Nevertheless, no matter what they would later choose, the results of this training remained with them. For many, this had been their first experience in being able to do something on their own -- completely independently.

I would like to make one comment here. Some of my American colleagues have explained to me that American students are not really accustomed to thinking and working hard, and for this reason we must make the material as attractive as possible. Permit me to not completely agree with this opinion. From my long experience with young students all over the world I know that they are curious and inquisitive and I believe that if they have some clear material presented in a simple form, they will prefer this to all artificial means of attracting their attention -- much as one buys books for their content and not for their dazzling jacket designs that engage only for the moment.

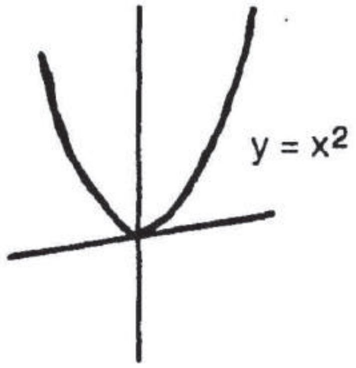
The most important thing a student can get from the study of mathematics is the attainment of a higher intellectual level. In this light I would like to point out as an example the famous American physicist and teacher Richard Feynman who succeeded in writing both his popular books and scientific works in a simple and attractive manner.

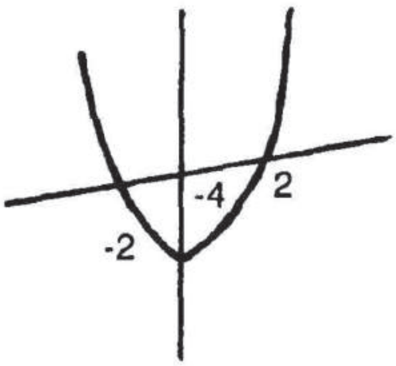
I.M. Gel'fand

Foreword

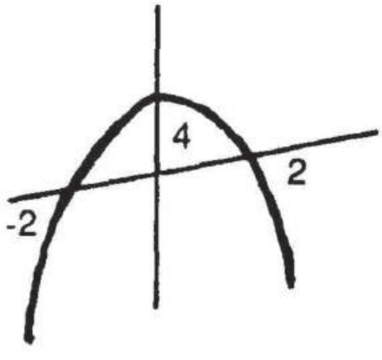
This book is called *Functions and Graphs*. As I mentioned, this book was written 25 years ago. I believe that today we would write in the same way.

The process of drawing graphs is a way of transferring your formulas and your data into geometrical form. Thus, drawing graphs is one of the ways to "see" your formulas and your functions and to observe the way in which this function changes. For example, if it is written $y = x^2$ then you already see a parabola [see picture 1]; if $y = x^2 - 4$ then you see a parabola dropped by four [see picture 2]; and if $y = 4 - x^2$ then you see the previous graph turned upside down.





This skill, to see simultaneously the formula and its geometrical representation, is very important not only for studies in mathematics but in studies of other subjects as well. It will be a skill that will remain with you for the rest of your life, like riding a bicycle, typing, or driving a car. Graphs are widely used in economics, engineering, physics, biology, applied mathematics, and of course in business.



This book, along with the others in this series, is not intended for quick reading. Each section is designed to be studied carefully. You can first scan the section and choose what is interesting for you. Nor is it necessary to solve all of the problems. Choose what you like. And if it is difficult for you, come back to it and try to understand what made it hard for you.

One more remark. This book as well as other books in this series is intended to be compatible with computers. However, do not think for one moment that in your study of mathematics you will be able to rely solely on computers. Computers can help you solve a problem. The computer cannnot -- nor will it ever be able -- to think and understand like you can. But of course it can sometimes inspire you to think and to understand.

Note to Teachers

This series of books includes the following material:

1. *Functions and Graphs*
2. *The Method of Coordinates*
3. *Algebra*
4. *Geometry*
5. *Calculus*
6. *Combinatorics*

Of course, all of the books may be studied independently. We would be very grateful for your comments and suggestions. They are especially valuable because books 3 through 6 are in progress and we can incorporate your remarks. For the book *Functions and Graphs* we plan to write a second part in which we will consider other functions and their graphs, such as cubic polynomials, irrational functions, exponential function, trigonometrical functions and even logarithms and equations.

Contents

Introduction

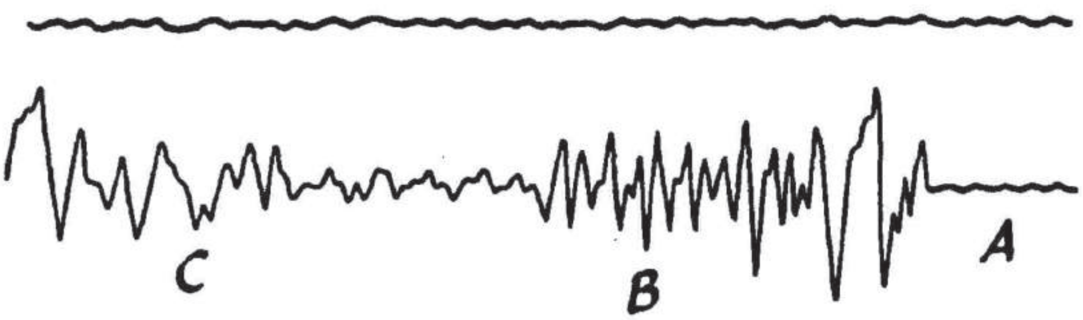


Fig. 1

In Fig. 1 the reader can see two curves traced by a seismograph, an instrument which registers fluctuations of the earth's crust. The upper curve was obtained while the earth's crust was undisturbed, the lower shows the signals of an earthquake.

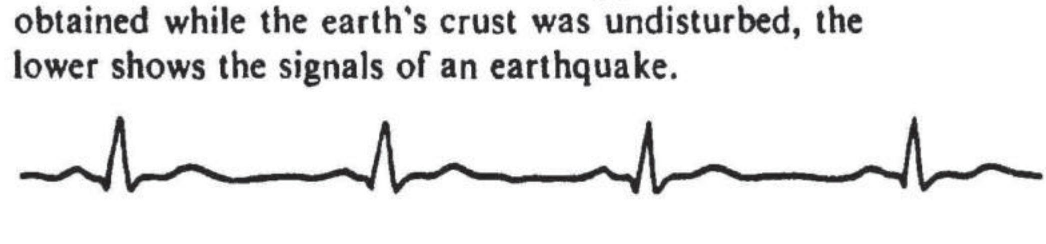


Fig. 2

In Fig. 2 there are two cardiograms. The upper shows normal heartbeat. The lower is taken from a heart patient.

Figure 3 shows the so-called characteristic of a semiconducting element, that is, the curve displaying the relationship between current intensity and voltage.

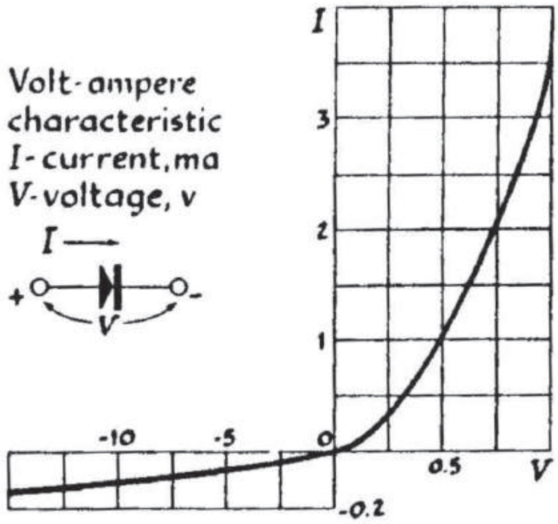


Fig. 3

In analyzing a seismogram, the seismologist finds out when and where an earthquake occurred and determines the intensity and character of the shocks. The physician examining a patient can, with the help of a cardiogram, judge the disturbances in heart activity; a study of the cardiogram helps to diagnose a disease correctly. The characteristic curve of a semiconducting element enables the radio-electrical engineer to choose the most favorable condition for his work.

All of these people study certain functions by the graphs of these functions.

What, then, is a function, and what is the graph of a function?

Before giving a precise definition of a function, let us talk a little about this concept. Descriptively speaking, a function is given when to each value of some quantity, which mathematicians call the argument and usually denote by the letter x, there corresponds the value of another quantity y, called the function.

Thus, for example, the magnitude of the displacement of the earth's surface during an earthquake has a definite value at each instant of time; that is, the amount of displacement is a function of time. The current intensity in a semiconducting element is a function of voltage, since to each value of the voltage there corresponds a definite value of current intensity.

Many such examples can be mentioned: the volume of a sphere is a function of its radius, the altitude to which a stone rises when thrown vertically upward is a function of its initial speed, and so on.

Let us now pass to the precise definitions. To say that the quantity y is a function of the quantity x, one first of all specifies which values x can take. These "allowed" values of the argument x are called *admissible values*, and the set of all admissible values of the quantity or variable x is called the *domain of definition* of the function y.

For instance, if we say that the volume V of the sphere is a function of its radius R, then the domain of definition of the function $V = \frac{4}{3}\pi R^3$ will be all numbers greater than zero, since the value R of the radius of the sphere can be only a positive number.

Whenever a function is given, it is necessary to specify its domain of definition.

Definition 1

We say that y is a *function* of the variable x, if: (1) it is specified which values of x are admissible, i.e., the domain of definition of the function is given, and (2) to each admissible value of x there corresponds exactly one value of the variable y.

Instead of writing "the variable y is a function of the variable x", one writes

$$y = f(x).$$

(Read: "y is equal to f of x.")

The notation $f(a)$ denotes the numerical value of the function $f(x)$ when x is equal to a. For example, if

$$f(x) = \frac{1}{x^2 + 1},$$

then

$$f(2) = \frac{1}{2^2 + 1} = \frac{1}{5},$$

$$f(1) = \frac{1}{1^2 + 1} = \frac{1}{2},$$

$$f(0) = \frac{1}{0^2 + 1} = 1, \text{ etc.}$$

The rule by which for each value of x the corresponding value of y is found can be given in different ways, and no restrictions are imposed on the form in which this rule is expressed. If the reader is told that y is a function of x, then he has only to verify that: (1) the domain of definition is given, that is, the values that x can assume are specified, and (2) a rule is given whereby to each admissible value of x there can be associated a unique value of y.

What kind of rule can this be?

Let us mention some examples.

1. Suppose we are told that x may be any real number and y can be found by the formula

$$y = x^2.$$

The function $y = x^2$ is thus given by a formula.

The rule may also be verbal.

2. The function y is given in the following manner: If x is a positive number, then y is equal to 1; if x is

a negative number, then y is equal to -1; if x is equal to zero, then y is equal to zero.

Let us mention yet another example of a function given by a verbal rule.

3. Every number x can be written in the form $x = y + \alpha$, where α is a nonnegative number less than one, and y is an integer. It is clear that to each number x there corresponds a unique number y; that is, y is a function of x. The domain of definition of this function is the entire real axis. This function is called "the integral part of x" and is denoted thus:

$$y = [x].$$

For example,

$$[3.53] = 3, \quad [4] = 4, \quad [0.3] = 0, \quad [-0.3] = -1.$$

We shall use this function later in our exercises.

4. Let us consider the function $y = f(x)$, defined by the formula

$$y = \frac{\sqrt{x+3}}{x-5}.$$

What can reasonably be considered as its domain of definition?

If a function is given by a formula, then usually its so-called *natural* domain of definition is considered, that is, the set of all numbers for which it is possible to carry out the operations specified by the formula. This means that the domain of definition of our function does not contain the number 5 (since at $x = 5$ the denominator of the fraction vanishes) and values of x less than -3 (since for $x < -3$ the expression under the root sign is negative). Thus, the natural domain of definition of the function

$$y = \frac{\sqrt{x+3}}{x-5}$$

consists of all numbers satisfying the relations:

$$x \geq -3, \qquad x \neq 5.$$

A function can be represented geometrically with the help of a graph. In order to construct a graph of some function, let us consider some admissible value of x and the corresponding value of y. For example, suppose the value of x is the number a, and the corresponding value of y is the number $b = f(a)$. We represent this pair of numbers a and b by the point with the coordinates (a, b) in the plane. Let us construct such points for all admissible values of x. The collection of points obtained in this way is the graph of the function.

Definition II

The *graph of a function* is the set of points whose abscissas are admissible values of the argument x and whose ordinates are the corresponding values of the function y.

For example, Fig. 4 depicts the graph of the function

$$y = [x].$$

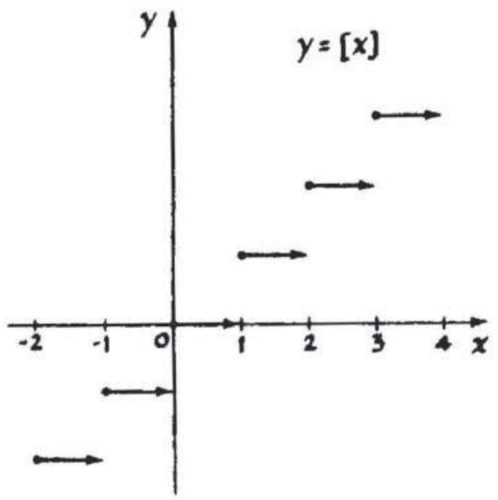


Fig. 4

It consists of an infinite set of horizontal line segments. The arrows indicate that the right end points of these segments do not belong to the graph (whereas the left end points belong to it and therefore are marked by a thick point).

A graph can serve as the rule by which the function is defined. For example, from the characteristic curve of a semiconducting element

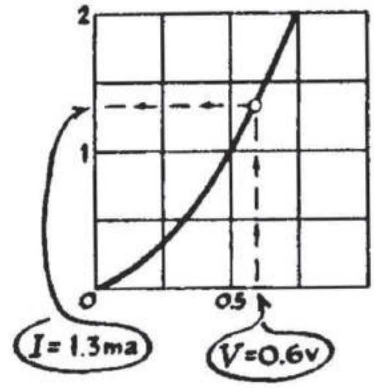


Fig. 5

it is possible to determine (see Fig. 5) that, if the argument V equals 0.6 volt, then the function I is equal to 1.3 milliamperes.

Representing functions by graphs is very convenient because, by looking at graphs, it is at once possible to distinguish one function from the other.

The reader is asked to look once more at the lower curve of Fig. 1. In this graph the most inexperienced person will at once recognize the signals of an earthquake (parts B and C). On closer examination, he will certainly also note the difference in character between the waves in parts B and C (a seismologist could point out that part B represents the so-called P-wave, the wave traveling deep in the earth's crust, while part C represents the S-wave traveling on the surface).

Try to distinguish these two sections by means of the tables of values arranged side by side. (We could not include here a table of values for the whole curve, because it would fill a whole page. In the margin the reader finds tables for small portions of sections B and C.)

Wave P (interval 0.2 sec)	Wave S (interval 0.4 sec)
0.1	0.2
0.1	0.5
−1.6	2.5
−1.7	4.9
−2.4	7.1
−3.0	6.1
−4.5	3.8
−3.8	0.4
−2.9	0.2
−1.1	0.7
0.8	1.5
3.3	2.5
5.1	3.2
3.7	2.8
0.0	0.4
−2.0	−2.2
−4.4	−3.3
−5.8	−4.5
−3.8	−4.8
−1.6	−4.8
	−4.8
	−3.7
	−3.5
	−4.4
	−6.6

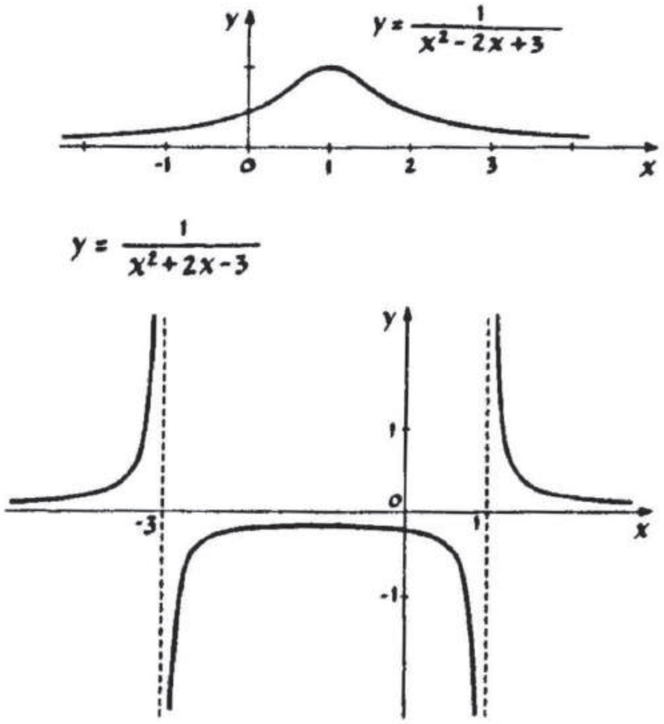


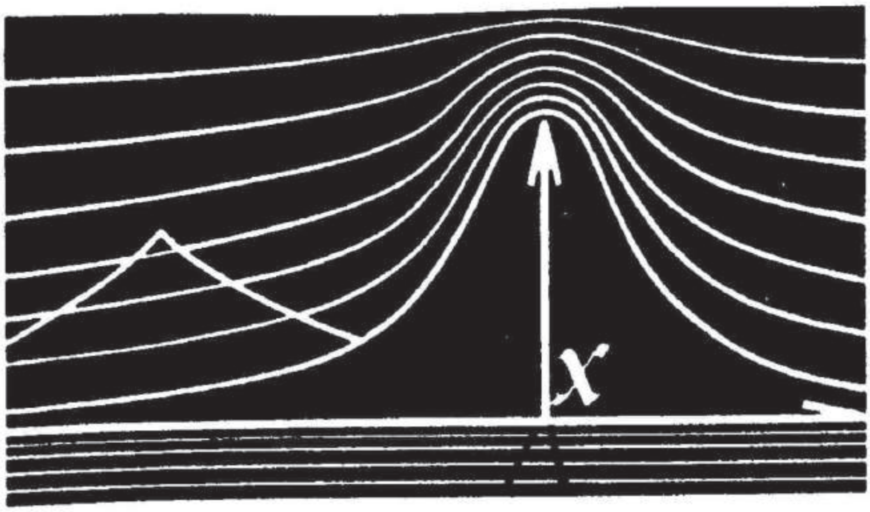
Fig. 6

Figure 6 shows graphs of two functions that are defined by very similar formulas:

$$y = \frac{1}{x^2 - 2x + 3} \quad \text{and} \quad y = \frac{1}{x^2 + 2x - 3}.$$

Of course, the difference in the behavior of these two functions can also be detected from their formulas. But if one looks at their graphs, this difference stands out immediately.

Whenever it is necessary to explain the general behavior of a function, to find its distinctive features, a graph is irreplaceable, by virtue of its visual character. Therefore, once an engineer or scientist has obtained a function in which he is interested, in the form of a formula or a chart, he usually takes a pencil and sketches a graph of this function, to find out how it behaves, what it "looks like."



CHAPTER 1

Examples

1

If the definition is followed literally, then in order to construct a graph of some function, it is necessary to find all pairs of corresponding values of argument and function and to construct all points with these coordinates. In the majority of cases it is practically impossible to do this, since there are infinitely many such points. Therefore, usually a few points belonging to the graph are joined by a smooth curve.

In this way, let us try to construct the graph of the function

$$y = \frac{1}{1 + x^2}. \qquad (1)$$

$y = \frac{1}{1+x^2}$

x	y
0	1
1	$\frac{1}{2}$
2	$\frac{1}{5}$
3	$\frac{1}{10}$

Table 1

Let us choose some values of the argument, find the corresponding values of the function, and write them down in a table (see Table 1). We construct the points with the computed coordinates and join them by a dotted line, for the time being (Fig. 1).

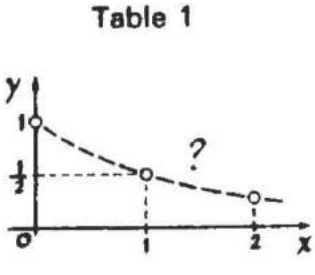


Fig. 1

Let us now verify whether we have drawn the curve correctly between the points found to lie on the graph.

x	y
$1\frac{1}{2}$	$\frac{4}{13}$
$\frac{1}{2}$	$\frac{4}{5}$

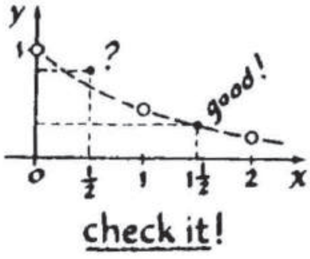
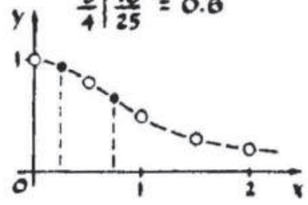


Fig. 2

x	y
$\frac{1}{4}$	$\frac{16}{17} = 1$
$\frac{3}{4}$	$\frac{16}{25} = 0.6$



one more check

x	y
$\frac{1}{3}$	$\frac{9}{10}$
$\frac{2}{3}$	$\frac{9}{13}$

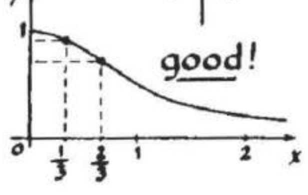


Fig. 3

x	y
1	$\frac{1}{2}$
2	$\frac{1}{5}$
3	$\frac{1}{10}$
⋮	

x	y
-1	$\frac{1}{2}$
-2	$\frac{1}{5}$
-3	$\frac{1}{10}$
⋮	

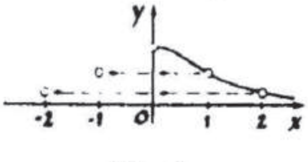


Fig. 4

For this purpose let us take some intermediate value of the argument, say, $x = 1\frac{1}{2}$, and compute the corresponding value of the function $y = \frac{4}{13}$. The point obtained, $(1\frac{1}{2}, \frac{4}{13})$, falls nicely on our curve (Fig. 2), so that we have drawn it quite accurately.

Now we try $x = \frac{1}{2}$. Then $y = \frac{4}{5}$, and the corresponding point lies above the curve we have drawn (Fig. 2). This means that between $x = 0$ and $x = 1$ the graph does not go as we thought. Let us take two more values, $x = \frac{1}{4}$ and $x = \frac{3}{4}$, in this doubtful section. After connecting all these points, we get the more accurate curve represented in Fig. 3. The points $(\frac{1}{3}, \frac{9}{10})$ and $(\frac{2}{3}, \frac{9}{13})$, taken as a check, fit the curve nicely.

2

In order to construct the left half of the graph, it is necessary to fill in one more table for negative values of the argument. This is easy to do. For example,

$$\text{for} \quad x = 2 \quad \text{we have} \quad y = \frac{1}{2^2 + 1} = \frac{1}{5},$$

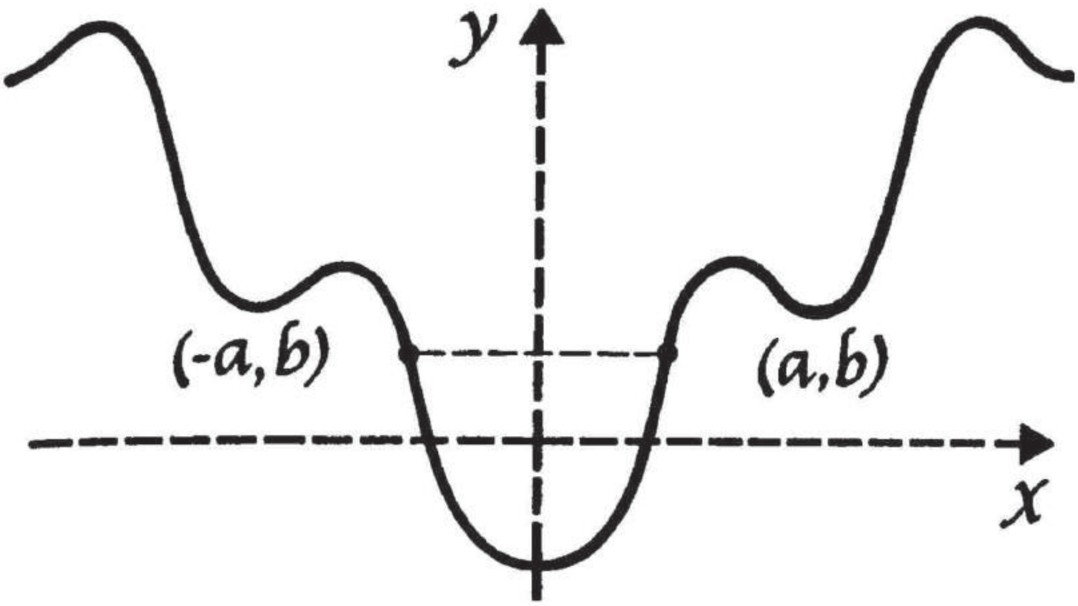
$$\text{for} \quad x = -2 \quad \text{we have} \quad y = \frac{1}{(-2)^2 + 1} = \frac{1}{5}.$$

This means that together with the point $(2, \frac{1}{5})$, the graph also contains $(-2, \frac{1}{5})$, the point symmetric to the first with respect to the y-axis.

In general, if the point (a, b) lies on the right half of our graph, then its left half will contain the point $(-a, b)$ symmetric to (a, b) with respect to the y-axis (Fig. 4). Therefore, in order to obtain the left part of the graph of function (1) corresponding to negative values of x, it is necessary to reflect the right half of this graph in the y-axis.

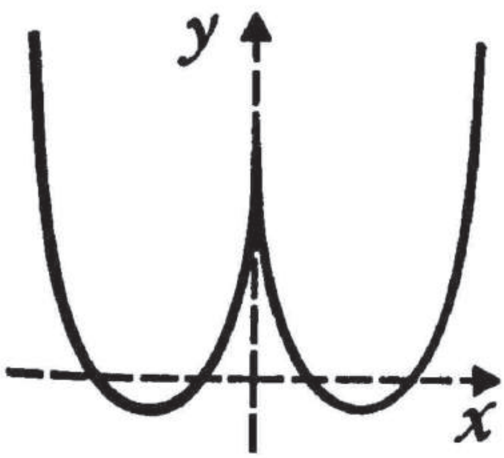
Figure 5 shows the over-all form of the graph.

If we had been hasty and had used our original sketch for the construction of the part of the graph corresponding to negative x (Figs. 1 and 2), then it would have had a "kink" (corner) at $x = 0$. There is

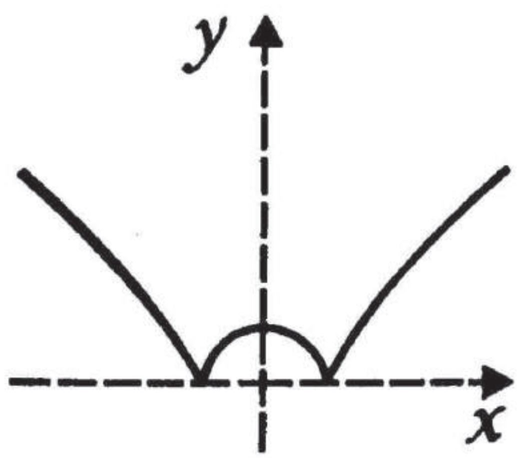


$$f(-a) = f(a)$$

If the values of some function corresponding to any two values of the argument equal in absolute value but with opposite signs (that is, the values a and -a) are equal, then such a function is called even. Any even function has a graph which is symmetric with regard to the y-axis.



$y = x^2 - 3|x| + 2$



$y = \sqrt{|x^2 - 1|}$

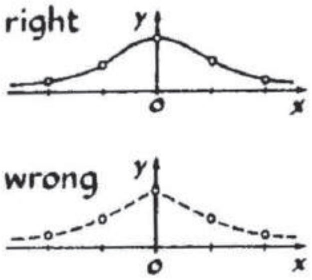


Fig. 5

no such kink in the accurate graph; instead there is a smooth "dome."

EXERCISES

1. The graph of the function

$$y = \frac{1}{3x^2 + 1} \tag{2}$$

is similar to the graph of the function

$$y = \frac{1}{x^2 + 1}.$$

Construct it.

2. Which of the following functions are even (for the definition and the graph of an even function as well as some examples, see p. 11)?

(*a*) $y = 1 - x^2$; (*b*) $y = x^2 + x$;

(*c*) $y = \dfrac{x^2}{1 + x^4}$; (*d*) $y = \dfrac{1}{1 - x} + \dfrac{1}{1 + x}$.

3

Let us now take the function

$$y = \frac{1}{3x^2 - 1}. \tag{3}$$

$$y = \frac{1}{3x^2 - 1}$$

x	y
0	1
1	$\frac{1}{2}$
2	$\frac{1}{11}$
-1	$\frac{1}{2}$
-2	$\frac{1}{11}$

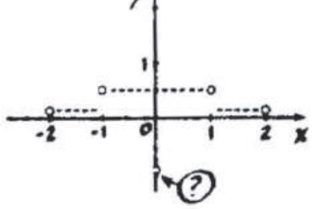


Fig. 6

In form this formula differs little from Formula 2. However, in the construction of this graph by points, troubles immediately arise.

Let us again work out a table and plot the computed points in a diagram. It is not clear how to join these points; it seems that the point $(0, -1)$ does not "fit in" (Fig. 6).

Try to construct a graph of this function yourself. Do not become discouraged if you have to find more points than expected in order to understand how this curve behaves.

Thereafter you are urged to read, on pages 14 to 16 how the authors construct this graph and what useful conclusions can be drawn from the construction.

4

The graph of the polynomial

$$y = x^4 - 2x^3 - x^2 + 2x \qquad (4)$$

will also be constructed first by points.

Taking for the argument values equal to 0, 1, and 2, we obtain values of the function equal to zero. Let us now take $x = -1$. Again we obtain the result $y = 0$. The corresponding points of the graph (0, 0), (1, 0), (2, 0), (−1, 0) lie on the x-axis (Fig. 7).

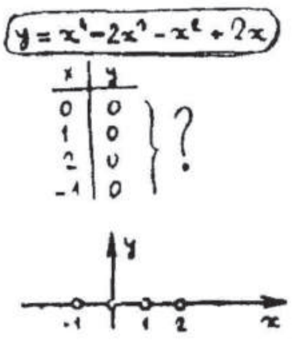


Fig. 7

If we confine ourselves to these four values of the argument, then the x-axis will be a "smooth" curve joining the points obtained. It is clear, however, that the x-axis is not the graph of our function because the polynomial

$$x^4 - 2x^3 - x^2 + 2x$$

cannot equal zero for all values of x.

Let us take two more values of the argument, $x = -2$ and $x = 3$. The corresponding points (−2, 24) and (3, 24) do not lie on the x-axis; on the contrary, they are located very far from it (Fig. 8).

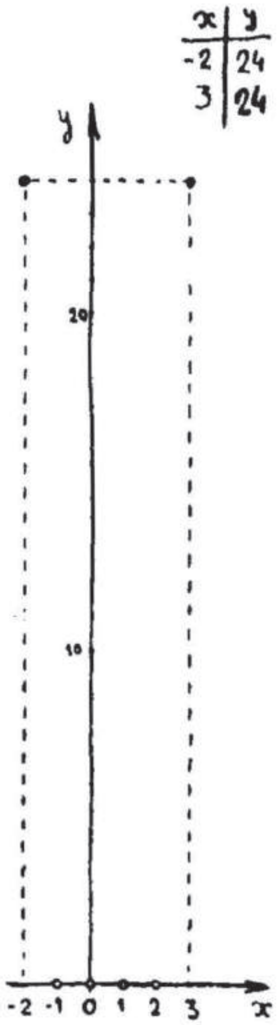


Fig. 8

From the appearance of the graph, its shape is still unclear. Of course, it is possible to find a sufficient number of intermediate points and construct an approximation to the graph, as we did before, but this method is not very reliable.

We shall try to proceed differently.

Let us find out where the function is positive (and, therefore, the graph lies above the x-axis) and where it is negative (that is, the graph lies below the x-axis).

For this purpose, let us factor the polynomial that defines the function:

$$\begin{aligned} x^4 - 2x^3 - x^2 + 2x &= x^3(x-2) - x(x-2) \\ &= (x^3 - x)(x-2) = x(x^2-1)(x-2) \\ &= (x+1)x(x-1)(x-2). \end{aligned}$$

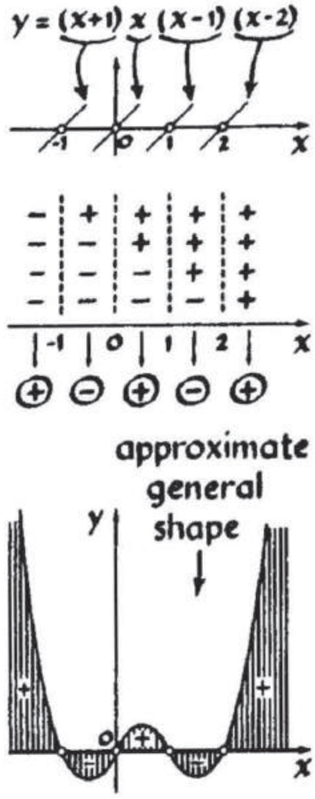


Fig. 9

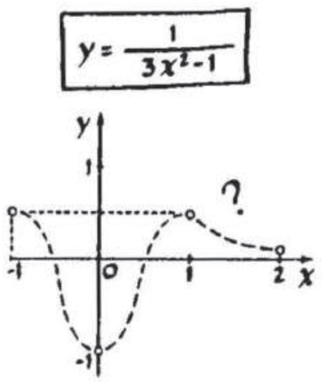


Fig. 10

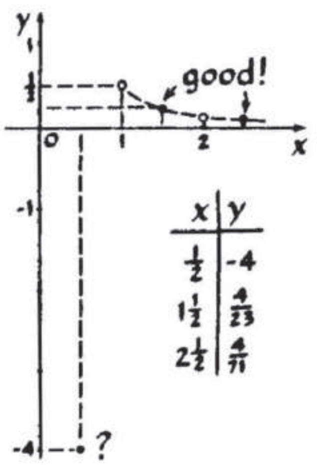


Fig. 11

It is now obvious that our function equals zero only at those four points which we have already plotted on the graph. To the left of the point $x = -1$, all four factors are negative, and the function is positive. Between the points $x = -1$ and $x = 0$ (that is, in the interval $-1 < x < 0$), the factor $x + 1$ becomes positive, while the remaining factors remain negative: the function is negative. In the region $0 < x < 1$, we have two positive and two negative factors: the function is positive. In the next interval the function is again negative. Finally, as the argument passes through the point $x = 2$, the last of the factors become positive: the function becomes positive.

The graph of the function now takes approximately the appearance given in Fig. 9.

5

We now pass to the construction of the graph of the function

$$y = \frac{1}{3x^2 - 1},$$

which we have already discussed on page 12.

Let us mark in the figure the points of the graph corresponding to the values $x = -1, 0, 1, 2$, and let us join them by a curve. As a result we obtain something like Fig. 10.

Let us now take $x = \frac{1}{2}$. We obtain $y = -4$, and the point $(\frac{1}{2}, -4)$ lies considerably below our curve. This means that between $x = 0$ and $x = 1$ the graph runs quite differently!

A more accurate course of the graph is represented in Fig. 11. If we take two more values, $x = 1\frac{1}{2}$ and $x = 2\frac{1}{2}$, the corresponding points fall quite nicely on our curve.

How, then, does the graph proceed between the points $x = 0$ and $x = 1$?

Let us take $x = \frac{1}{4}$ and $x = \frac{3}{4}$. We obtain $y = -\frac{16}{13} \approx -1\frac{1}{4}$ and $y = \frac{16}{11} \approx 1\frac{1}{2}$, respectively. The course of the graph between the points $x = 0$ and

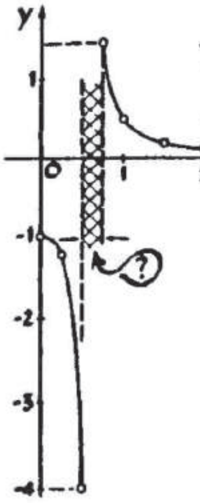
$x = \frac{1}{2}$ thus has become somewhat clearer (Fig. 12), but the behavior of the function between $x = \frac{1}{2}$ and $x = \frac{3}{4}$ remains, as previously, obscure.

If we take a few more intermediate values between $x = \frac{1}{2}$ and $x = \frac{3}{4}$, we see that the corresponding points of the graph lie not on one but on two smooth curves, and the graph assumes approximately the appearance of Fig. 13.

The reader can now well understand that the pointwise construction of a graph is risky and lengthy. If we take few points, then it may turn out that we obtain an altogether false picture of the function. If, on the other hand, we take more points, there will be much superfluous work, and some doubt still remains as to whether we did not omit something significant. How are we to proceed?

Let us recall that when we constructed the graph $y = 1/(x^2 + 1)$ in the interval $2 < x < 3$ and $1 < x < 2$ no additional points were required, while in the interval $0 < x < 1$, it was necessary to find 5 more points. Similarly in the construction of the graph of $y = 1/(3x^2 - 1)$ the most work was required for the interval $0 < x < 1$, where the curve splits into two branches.

Is it not possible to isolate such "dangerous" regions beforehand?



x	y
$\frac{1}{4}$	$-\frac{16}{13}$
$\frac{3}{4}$	$\frac{16}{11}$

Fig. 12

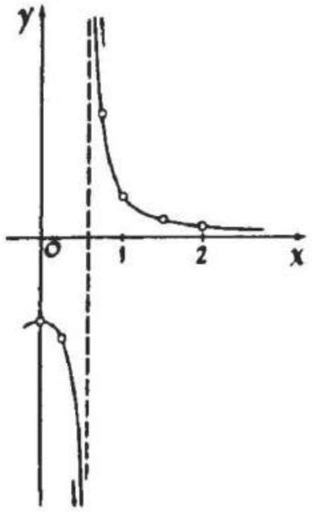


Fig. 13

6

Let us return for the third time to the graph of

$$y = \frac{1}{3x^2 - 1}.$$

If one looks at the expression defining the function, it is immediately obvious that for two values of x the denominator of this expression vanishes. These values are equal to $+\sqrt{\frac{1}{3}}$ and $-\sqrt{\frac{1}{3}}$, that is, approximately ± 0.58. One of them lies in the interval $\frac{1}{2} < x < \frac{3}{4}$, precisely where the function displays unusual behavior, where the graph does not go smoothly. It is now clear why this is the case.

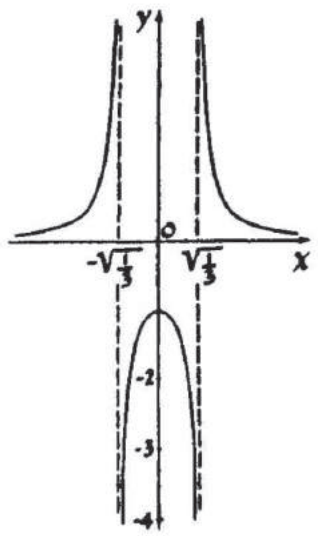


Fig. 14

In fact, for the values $x = \pm\sqrt{\frac{1}{3}}$, the function is not defined (division by zero is impossible); therefore, there can be no points on the graph with these abscissas — the graph does not intersect the straight lines $x = \sqrt{\frac{1}{3}}$ and $x = -\sqrt{\frac{1}{3}}$. Therefore, the graph decomposes into three separate branches. If x approaches one of the "forbidden" values, say, $x = \sqrt{\frac{1}{3}}$, then the fraction $1/(3x^2 - 1)$ increases without bond in absolute value: two branches of the graph approach the vertical straight line $x = \sqrt{\frac{1}{3}}$.

Our (even) function behaves analogously near the point $x = -\sqrt{\frac{1}{3}}$.

The general form of the graph of $y = 1/(3x^2 - 1)$ is shown in Fig. 14.

We now understand that *whenever a function is defined by a formula in the form of a fraction, it is necessary to focus attention on those values of the argument at which the denominator vanishes.*

7

What lesson, then, can be learned from the examples we have considered? In the study of the behavior of a function and in the construction of its graph, not all values of the argument are of equal importance. In the case of the function

$$y = \frac{1}{3x^2 - 1}$$

we saw the importance of those "special" points at which the function is not defined. The character of the graph of

$$y = x^4 - 2x^3 - x^2 + 2x$$

became clear to us when we found the points of intersection of the graph with the x-axis, namely, the roots of the polynomial.

In most cases the main part of work in the construction of graphs consists precisely in finding values of the argument significant for the given function and in investigating its behavior near these values. To complete the construction of the graph after such an analysis, it usually suffices to find some intermediate values of the function between these characteristic points.

EXERCISES

1. Construct the graph of the function

$$y = \frac{1}{3x - 1}.$$

At what points does the graph intersect the coordinate axes?

Imagine that we placed the origin in the very center of a page in an exercise book and took 1 centimeter as the unit of measure (for definiteness we shall consider a page to be a rectangle of size 16 cm $\times$ 20 cm). Find the coordinates of the points at which the graph leaves the page.

2. Construct the graphs of the polynomials*

(*a*) $y = x^3 - x^2 - 2x + 2;$
(*b*) $y = x^3 - 2x^2 + x.$ $\oplus$

(Notice that in case *b* the factorization of the polynomial results in two identical factors.)

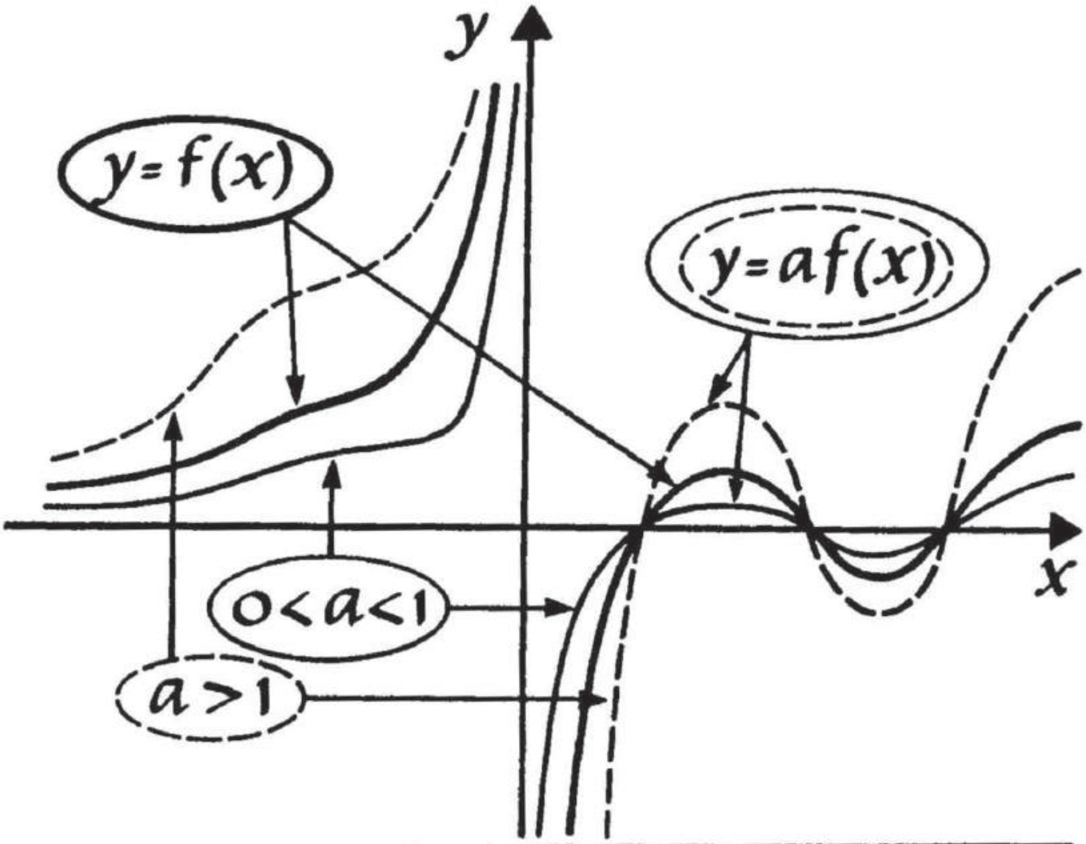
8

Having constructed the graph of some function, we can use various methods to construct the graphs of some "related" functions.†

One of the simplest of these methods is the so-called *stretching along the y-axis.*

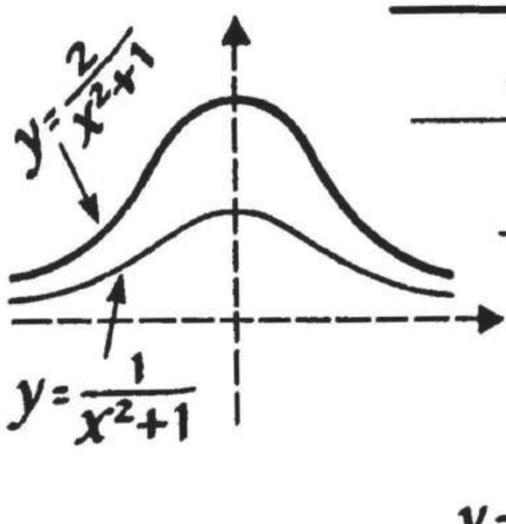
*The sign $\oplus$ marks problems and exercises to which there are answers at the end of the book (see p. 103).

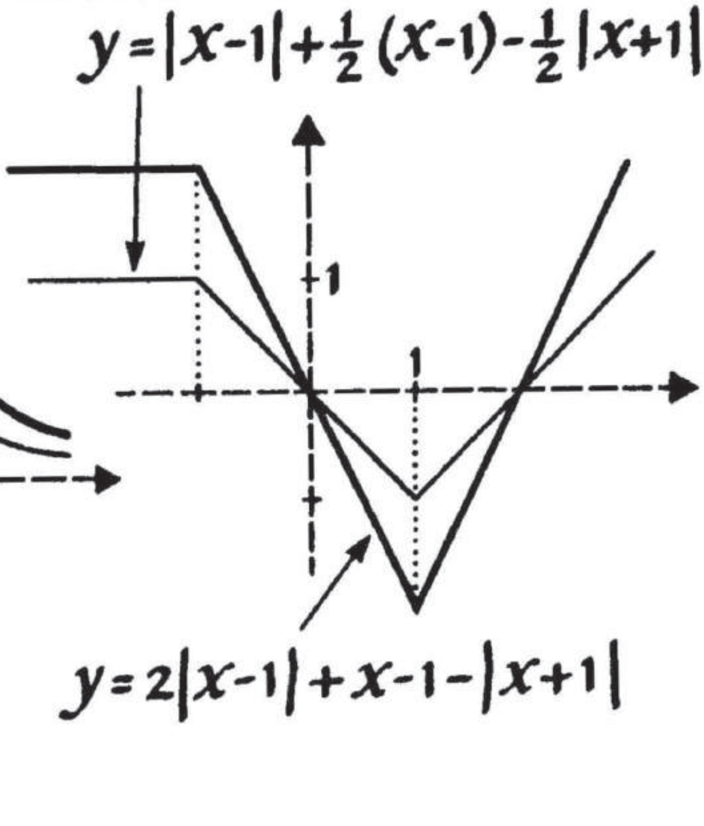
†A method of this kind was already encountered on page 10, when we constructed the graph of $y = 1/(x^2 + 1)$. Having constructed the graph of this function for positive values of x, we were immediately able to construct the graph for negative x as well.



$$f(x) \longrightarrow af(x)$$

The graph of the function $y = af(x)$ is obtained from the graph of the function $y = f(x)$ by stretching it in the ratio a:1 along the y-axis (in case $|a| < 1$ we obtain contraction).





We constructed the graph of the function

$$y = \frac{1}{x^2 + 1}$$

(Fig. 5 on p. 12).

Let us now construct the graph of

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

$y = \frac{1}{x^2+1}$

x	y
0	1
$\frac{1}{2}$	$\frac{4}{5}$
2	$\frac{1}{5}$

$y = \frac{3}{x^2+1}$

x	y
0	3
$\frac{1}{2}$	$\frac{12}{5}$
2	$\frac{3}{5}$

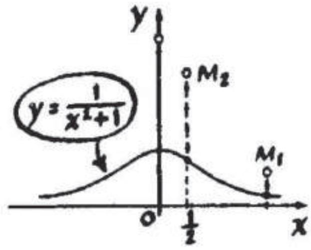


Fig. 15

Let us take any point of the first graph, for example, $x = \frac{1}{2}$, $y = \frac{4}{5}$, that is, the point $M_1(\frac{1}{2}, \frac{4}{5})$. Clearly, we can obtain a point of the second graph by keeping x constant (that is, $x = \frac{1}{2}$) and increasing y in the ratio 3:1. We thus obtain the point $M_2(\frac{1}{2}, \frac{12}{5})$. It can be obtained directly in the diagram (Fig. 15). For this it is necessary to increase the ordinate of the point $M(\frac{1}{2}, \frac{4}{5})$ in the ratio 3:1.

If we carry out such a transformation with each point of the graph of $y = 1/(x^2 + 1)$, then the point $M(a, b)$ is transformed into the point $M'(a, 3b)$ of the graph of $y = 3/(x^2 + 1)$, and the entire graph, stretched in the ratio 3:1 along the y-axis, turns into the graph of the function $y = 3/(x^2 + 1)$ (Fig. 16).

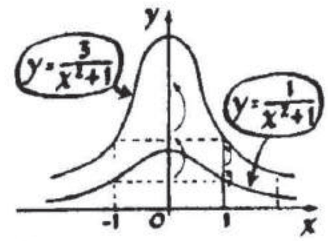


Fig. 16

Thus, the graph of $y = 3/(x^2 + 1)$ represents the graph of $y = 1/(x^2 + 1)$ stretched in the ratio 3:1 along the y-axis.

9

It is still easier to obtain the graph of the function

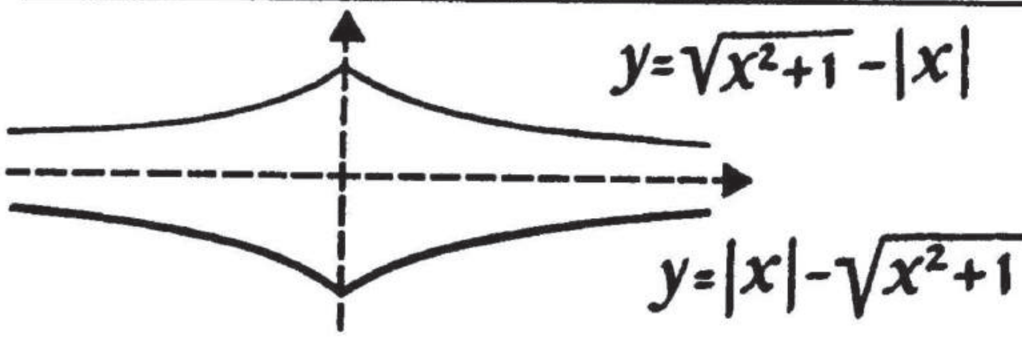
$$y = -\frac{1}{x^2 + 1}$$

from the graph of $y = 1/(x^2 + 1)$. In order to obtain a table for the function $y = -1/(x^2 + 1)$ from Table 1 on page 9 for the function $y = 1/(x^2 + 1)$, it is necessary only to change the sign of each of the numbers in the second column.

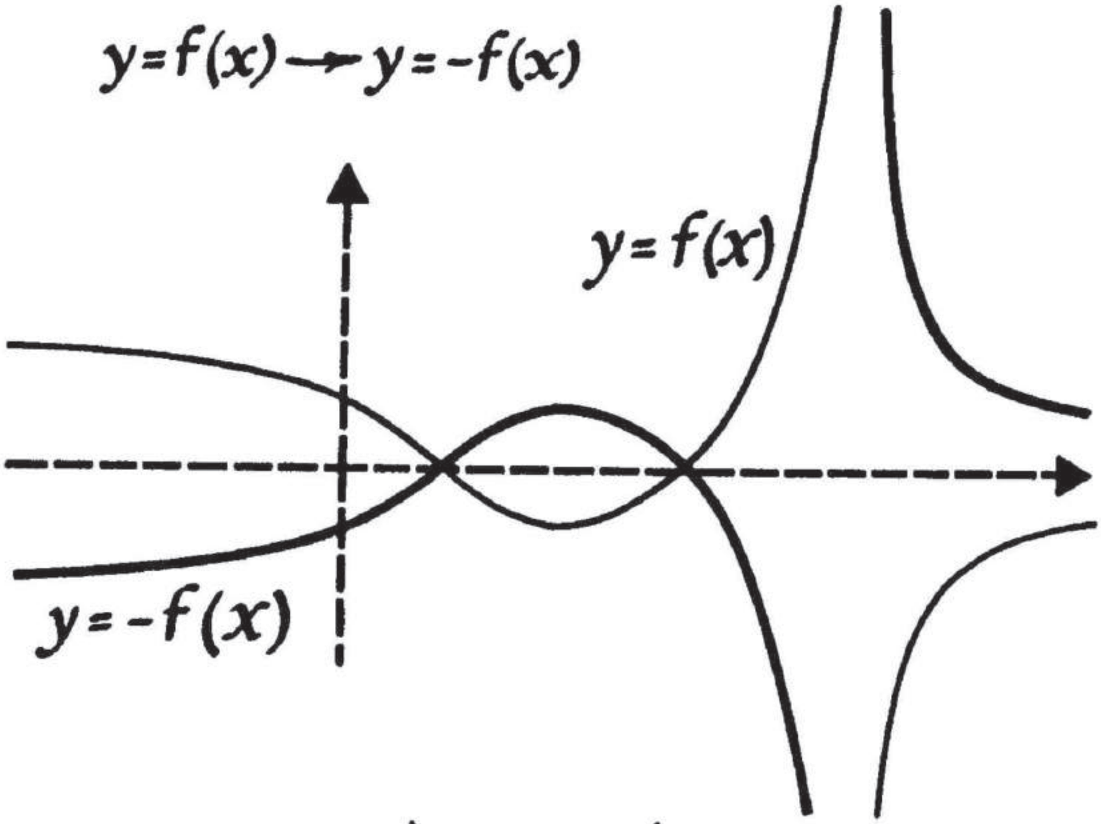
Then each point of the graph of $y = 1/(x^2 + 1)$, for example the point M with abscissa 2 and ordinate

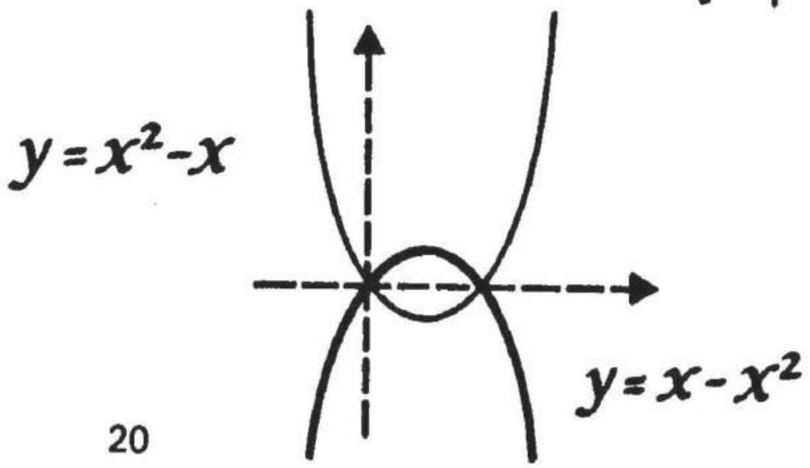
$$f(x) \longrightarrow -f(x)$$

The graph of $y = -f(x)$ can be obtained from the graph of the function $y = f(x)$ by reflection in the x-axis.



$y=f(x) \rightarrow y=-f(x)$





$\frac{1}{5}$, yields a point M' of the graph $y = -1/(x^2 + 1)$ with the same abscissa 2 but whose ordinate has the opposite sign $(-\frac{1}{5})$. Obviously the point $M'(2, -\frac{1}{5})$ is symmetric to the point $M(2, \frac{1}{5})$ with respect to the x-axis. Generally speaking, to the point $N(a, b)$ of the graph of $y = 1/(x^2 + 1)$ there corresponds the point $N'(a, -b)$ of the graph of $y = -1/(x^2 + 1)$.

Thus, the graph of the function $y = -1/(x^2 + 1)$ can be obtained from the graph of $y = 1/(x^2 + 1)$, by finding its mirror image in the x-axis (Fig. 17).

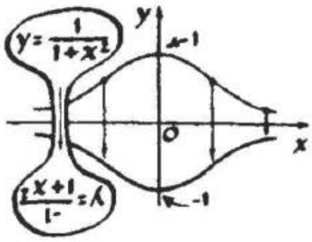


Fig. 17

EXERCISES

1. From the graph of $y = x^4 - 2x^3 - x^2 + 2x$ (Fig. 9, page 14), construct the graphs of $y = 3x^4 - 6x^3 - 3x^2 + 6x$ and $y = -x^4 + 2x^3 + x^2 - 2x$.

2. Construct the graph of $y = 1/(2x^2 + 2)$, using the graph of $y = 1/(x^2 + 1)$.

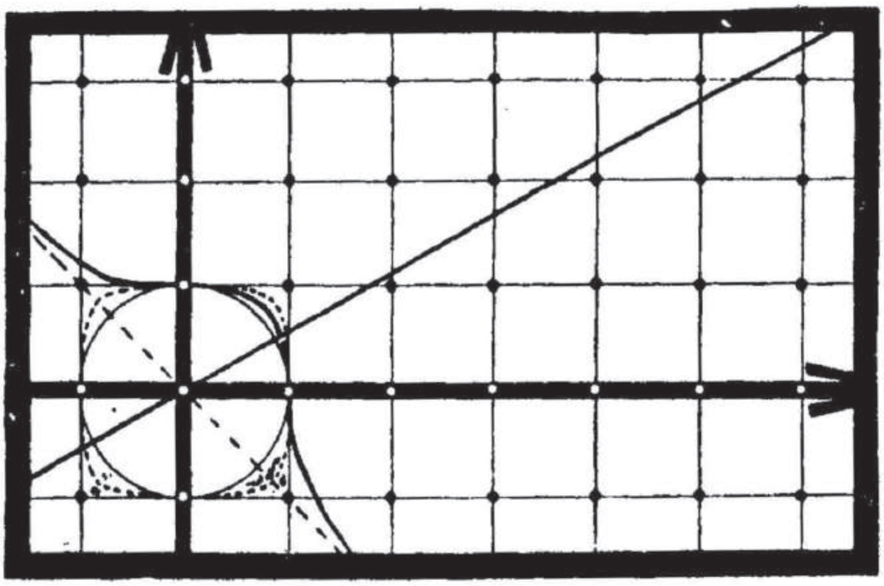
3. Construct the graphs of*

(*a*) $y = \frac{1}{2}[x]$;

(*b*) $y = x - [x]$ and $y = -2(x - [x])$;

(*c*) $y = [2x]$.

*The meaning of the symbol $[x]$, the integral part of the number x, was explained on page 5.



CHAPTER 2

The Linear Function

1

Let us now begin to study systematically the behavior of different functions and to construct their graphs. The characteristic behavior of functions and peculiarities of their graphs will be studied using the simplest examples. When we construct more complicated graphs, we shall try to find familiar elements in them.

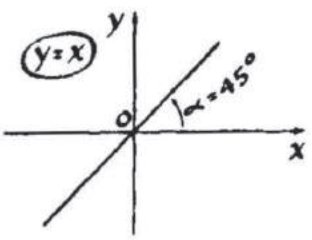


Fig. 1

The simplest function is the function $y = x$. The graph of this function is a straight line, the bisector of the first and third quadrants (Fig. 1).

In general, as you know, the graph of any linear function

$$y = kx + b$$

is some straight line. Conversely, any straight line not parallel to the y-axis is the graph of some linear function. The position of a straight line is completely determined by two of its points. Accordingly, a linear function is completely determined by giving its values for two values of the argument.

EXERCISES

1. Find the linear function

$$y = kx + b$$

that takes the value $y = 41$ at $x = -10$, and the value $y = 9$ at $x = 6$.

2. A straight line passes through the points $A(0, 0)$ and $B(a, c)$. Find the linear function whose graph is this straight line.

3. Draw a straight line through the origin at an angle of 60° with the y-axis. What function has this straight line as its graph?

4. (*a*) In Table 1 of the values of some linear function, two out of its five values are written down incorrectly. Find and correct them.

(*b*) The same question for Table 2.

5. Find the function

$$y = ax + b$$

if its graph is parallel to the graph of $y = x$ and passes through the point $(3, -5)$.

6. Find the linear function whose graph makes an angle of 60° with the x-axis and passes through the point $(3, -5)$.

7. The slope of a straight line* equals k. The straight line passes through the point $(3, -5)$. Find the linear function whose graph is this straight line.

Table 1

x	y
$\vdots$	$\vdots$
-2	-2
-1	3
0	1
1	2
2	-3

Table 2

x	y
-15	-33
-10	-13
0	7
10	17
15	27

2

A characteristic property of a linear function is that if x is increased uniformly, that is, by the same number for all x, then y also changes uniformly. Let us take the function $y = 3x - 2$. Suppose x takes the values

$$1, 3, 5, 7, \cdots,$$

each of which is larger than the preceding one, always

*The coefficient a is called the slope of the straight line $y = ax + b$.

by the same number, 2. The corresponding values of y will be

$$1, 7, 13, 19, \cdots.$$

The reader can see that each of the values of y is larger than the preceding one, also always by the same number, 6.

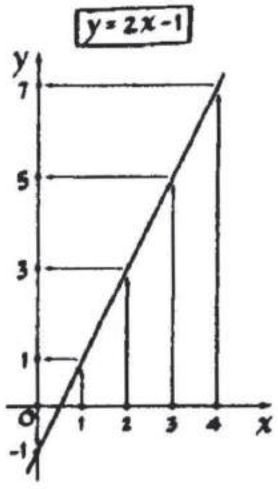


Fig. 2

A sequence of numbers that is obtained from some number by constantly adding the same number forms an *arithmetic progression*. Thus, the characteristic property which we mentioned can be expressed in the following fashion: A linear function converts one arithmetic progression into another arithmetic progression. In our example (page 23), the function $y = 3x - 2$ converts the arithmetic progression 1, 3, 5, 7, ... into the arithmetic progression 1, 7, 13, 19, ... Figure 2 illustrates another example of how the function $y = 2x - 1$ changes the arithmetic progression 0, 1, 2, 3, ... into the arithmetic progression —1, 1, 3, 5, 7, ...

EXERCISES

1. Find the linear function which would convert the arithmetic progression $-3, -1, 1, 3, \cdots$ into the arithmetic progression $-2, -12, -22, \cdots$, etc.

What linear function transforms the second progression into the first?

2. Suppose we are given two arithmetic progressions:

$$a, a + h, a + 2h, \cdots \text{ and } c, c + l, c + 2l, \cdots.$$

Is it always possible to find a linear function $y = kx + b$ which transforms the first progression into the second?

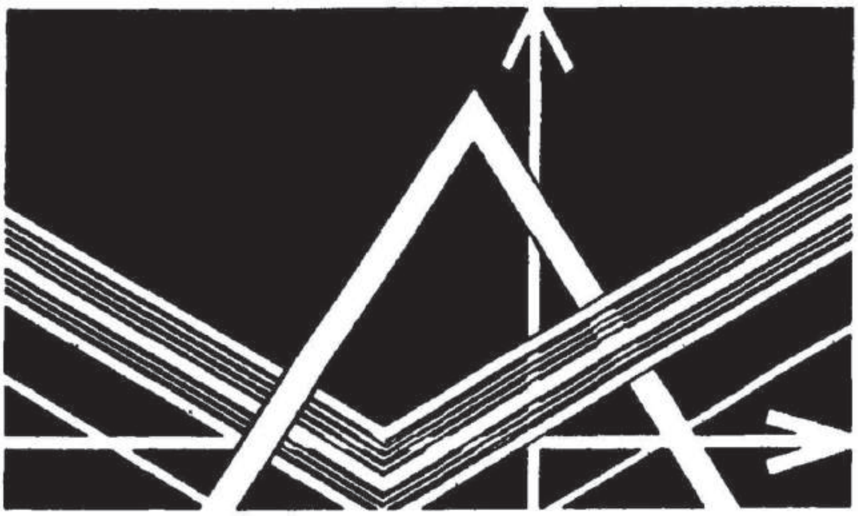
3. (*a*) The straight line $y = \frac{7}{15}x + \frac{1}{3}$ passes through two points with integral coordinates: $A(10, 5)$ and $B(-20, -9)$. Are there other "integral points" (points with integral coordinates) on this straight line?

(*b*) It is known that the straight line $y = kx + b$ passes through two integral points. Are there other integral points on this straight line?

(*c*) It is easy to construct a straight line that does not pass through any integral point. For example, $y = x + \frac{1}{2}$.

Is it possible for some straight line $y = kx + b$ to pass through just one integral point?*

*If you do not find an answer, look at Problem 4 on page 91.



CHAPTER 3

The Function $y = |x|$

1

Let us now consider the function

$$y = |x|,$$

where $|x|$ means the absolute value* or modulus of the number x.

Let us construct its graph, using the definition of absolute value. For positive x we have $|x| = x$; that is, this graph coincides with the graph of $y = x$ and is a half-line leaving the origin at an angle of 45° with the x-axis (Fig. 1). For $x < 0$, we have $|x| = -x$, which means that for negative x the graph of $y = |x|$ coincides with the bisector of the second quadrant (Fig. 2).

But then, the second half of the graph (for negative values of x) can easily be obtained from the first, if it

$y=|x|$

Fig. 1

Fig. 2

*Let us recall: The absolute value of a positive number is equal to this number (if $x > 0$, then $|x| = x$); the absolute value of a negative number is equal to the number provided with the positive sign (if $x < 0$, then $|x| = -x$); the absolute value of zero is equal to zero ($|0| = 0$).

is noted that the function $y = |x|$ is even, since $|-a| = |a|$ (see the definition of an even function on page 11).

This means the graph of this function is symmetric with respect to the y-axis, and the second half of its graph can be obtained by reflection in the y-axis of the part traced out for positive x. This yields the graph represented in Fig. 3.

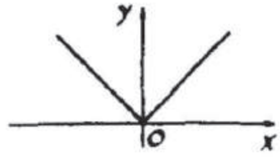


Fig. 3

2

Let us now construct the graph of

$$y = |x| + 1.$$

This graph can easily be constructed at once. We shall obtain it, however, from the graph of the function $y = |x|$. Let us work out a table of values of the function $y = |x| + 1$ and compare it with the same table worked out for $y = |x|$ by writing these tables side by side (Tables a, b). It is obvious that from each point of the first graph, $y = |x|$, a point of the second graph, $y = |x| + 1$, can be obtained by increasing y by 1. (For instance, the point $(-2, 2)$ of the graph of $y = |x|$ goes into the point $(-2, 3)$ of the graph of $y = |x| + 1$, located one above the first — see Fig. 4.) Hence, in order to get the points of the second graph, it is necessary to move each point of the graph up by 1; that is, the entire second graph is obtained from the first by an upward translation of 1 unit (see Fig. 4).

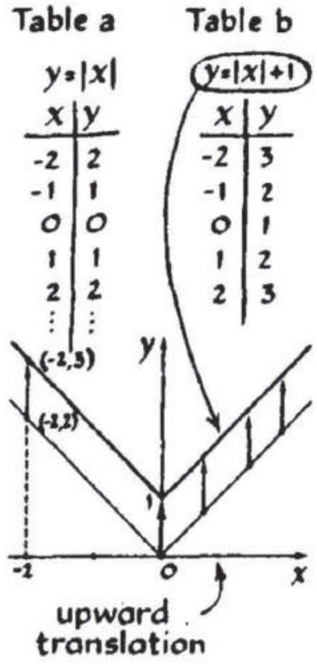


Fig. 4

Problem. Construct the graph of the function

$$y = |x| - 1.$$

Solution. Let us compare this graph with the graph of $y = |x|$. If the point $x = a$, $y = |a|$ lies on the first graph, then the point $x = a$, $y = |a| - 1$ will lie on the second graph. Therefore each point $(a, |a| - 1)$ of the second graph can be obtained from the point $(a, |a|)$ of the first graph by a downward translation of 1 unit, and the whole graph is obtained if the graph $y = |x|$ is moved downward 1 unit (Fig. 5).

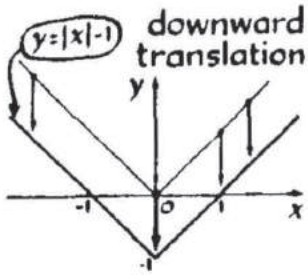


Fig. 5

Such a translation along the y-axis is useful in the construction of many graphs (see p. 29).

Suppose we are to construct the graph of the function

$$y = \frac{x^2 + 2}{x^2 + 1}.$$

Let us represent this function in the form

$$y = \frac{x^2 + 1 + 1}{x^2 + 1},$$

or

$$y = 1 + \frac{1}{x^2 + 1}.$$

It is now obvious that its graph can be obtained from the graph (Fig. 5 on p. 12) of $y = 1/(x^2 + 1)$ by an upward translation of 1 unit along the y-axis.

3

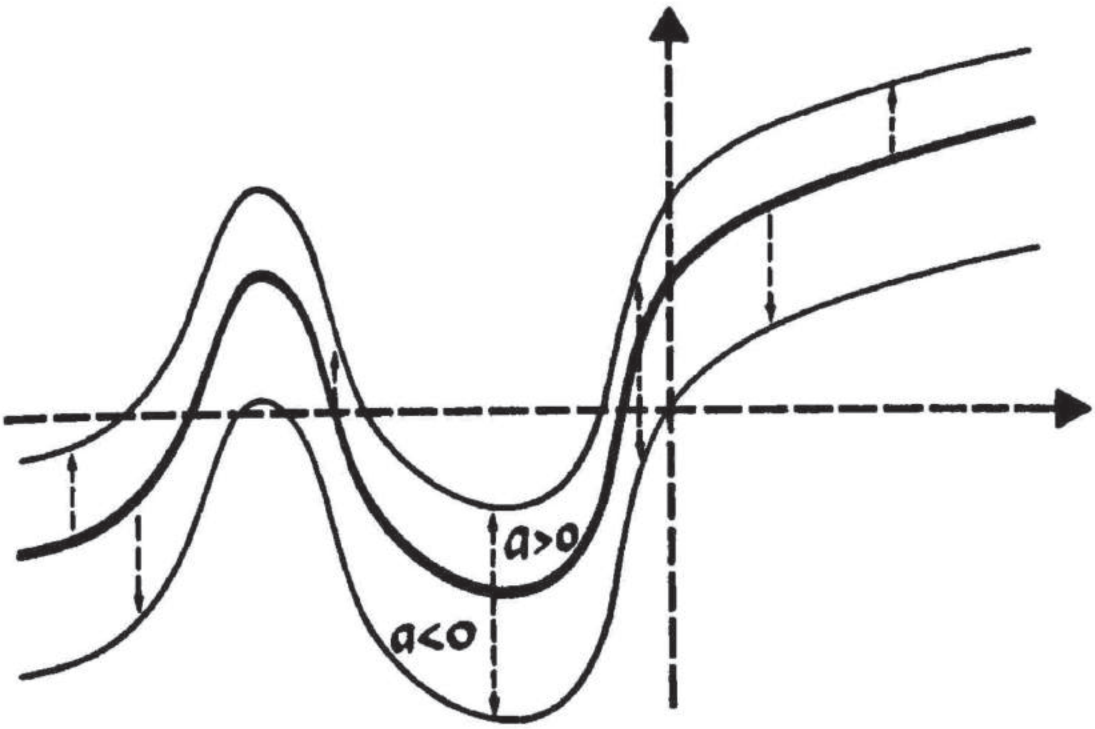
Let us now take the function

$$y = |x + 1|.$$

Fig. 6

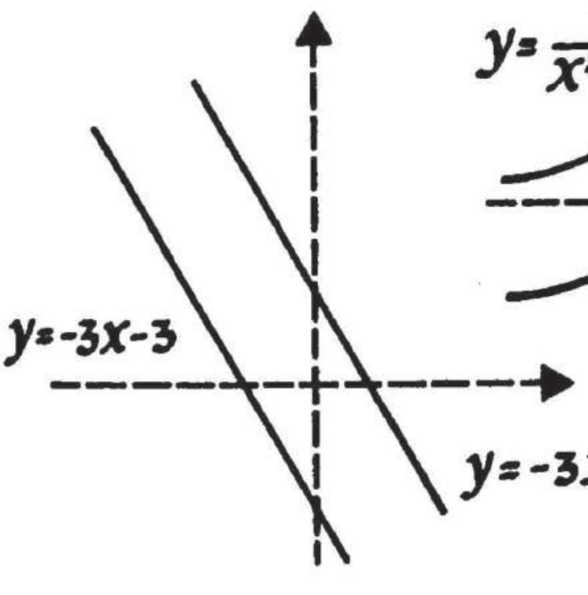
The graph of this function will also be obtained from the graph of $y = |x|$. Let us again write two tables side by side, one for $y = |x|$ and one for $y = |x + 1|$ (Tables a, b). If we compare the values of the functions for the same x, then it turns out that for some x the ordinate of the first graph is larger than that of the second, and for some it is the other way around.

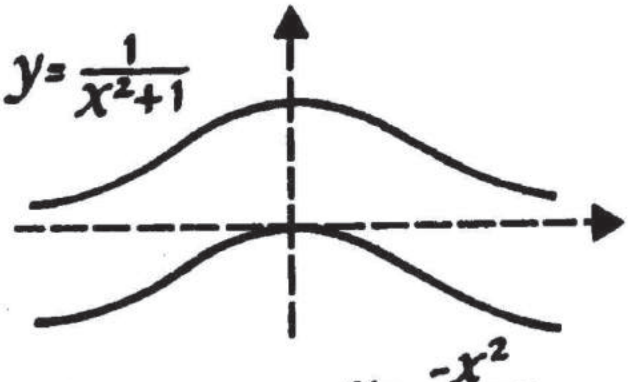
However, if we look closely at the right columns of these two tables, a connection between the tables can be established: The second function assumes the same values as the first, but does so one unit earlier, for smaller values of x. (Why?) Hence each point of the first graph, $y = |x|$, yields a point of the second graph, $y = |x + 1|$, moved one unit to the left; for example, the point $(-1, 1)$ gives rise to the point with



$$f(x) \longrightarrow f(x) + a$$

The graph of the function $y = f(x) + a$ is obtained from the graph of the function $y = f(x)$ by a translation along the y-axis of a units. The direction of the translation is determined by the sign of the number a (if $a > 0$, the graph moves up; if $a < 0$, the graph moves down).





the coordinates $(-2, 1)$ (Fig. 6). Therefore, the whole graph of $y = |x + 1|$ is obtained if the graph of $y = |x|$ is moved to the left by one unit along the x-axis.

Problem. Construct the graph of the function

$$y = |x - 1|.$$

Solution. Let us compare it with the graph of $y = |x|$. If A is a point of the graph of $y = |x|$, with coordinates $(a, |a|)$, then the point $A'(a + 1, |a|)$ will be a point of the second graph with the same value for the ordinate y. (Why?) This point of the second graph can be obtained from the point $A(a, |a|)$ of the first graph by a translation to the right along the x-axis. Hence, the whole graph of $y = |x - 1|$ is obtained from the graph of $y = |x|$ by moving it to the right along the x-axis by 1 unit (Fig. 7). We can say the function $y = |x - 1|$ assumes the same values as the function $y = |x|$, but with a certain delay (a delay of 1 unit).

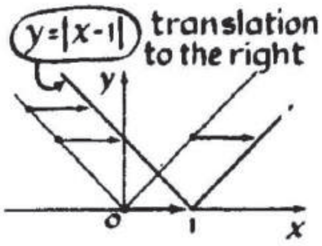


Fig. 7

Such a shift along the x-axis is useful in the construction of many graphs.

EXERCISES

1. Construct a graph of the function

$$y = \frac{1}{x^2 - 2x + 2}.$$

Hint: Represent the denominator of the fraction $1/(x^2 - 2x + 2)$ in the form $(x - 1)^2 + 1$.

2. State rules according to which from the graph of the function $y = f(x)$ it is possible to obtain the graphs of the functions $y = f(x + 5)$ and $y = f(x - 3)$.

3. Construct the graphs of $y = |x| + 3$ and $y = |x + 3|$.

4. Find all linear functions that at $x = 3$ assume the value $y = -5$.

Solution. Geometrically the condition can be formulated thus: find all straight lines passing through the point $(3, -5)$. Any (nonvertical) straight line passing through the origin is the graph of some function $y = kx$. Let us translate this straight line so that it passes through the required point $(3, -5)$, that is, 3 units to the right and 5 units downward (Fig. 8). After the first translation we obtain the equation

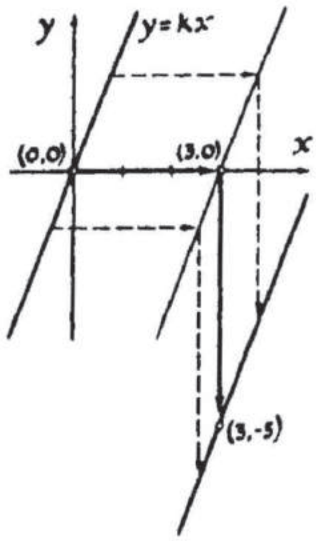


Fig. 8

$$y = k(x - 3),$$

after the second

$$y = k(x - 3) - 5.$$

Answer. All linear functions, which at $x = 3$ assume the value $y = -5$, can be expressed by the formula

$$y = k(x - 3) - 5,$$

where k is any real number. (Compare this problem with Problem 7 on p. 23.)

4

Problem. Construct the graph of

$$y = |x + 1| + |x - 1|.$$

Solution. Let us first construct in one diagram the graphs of each of the terms:

$$y = |x + 1| \text{ and } y = |x - 1|.$$

The ordinate y of the desired graph is obtained by adding the ordinates of the two constructed graphs at the same point. Thus, for example, if $x = 3$, then the ordinate y_1 of the first graph is equal to 4, the ordinate y_2 of the second graph is equal to 2, and the ordinate y of the graph of $y = |x + 1| + |x - 1|$ is equal to 6.

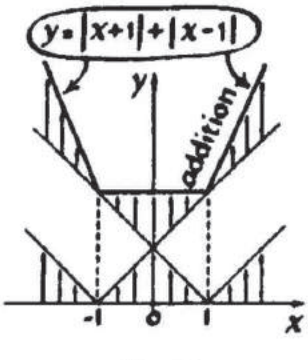


Fig. 9

Let us try to obtain the desired graph by adding at each point (that is, for each x) the ordinates of the two graphs. We thus obtain the drawing given in Fig. 9.

We see that the graph of $y = |x + 1| + |x - 1|$ is a broken line, composed of portions of three straight lines. Hence, in each of three intervals, the function varies linearly.

EXERCISES

1. Write down equations for each part of the broken line

$$y = |x + 1| + |x - 1|.$$

(**Answer.** for $x \leq -1, \quad y = \cdots x + \cdots;$
for $-1 \leq x \leq 1, \quad y = \cdots;$
for $x \geq 1, \quad y = \cdots.$)

2. Where are the breaks in the broken line which is the graph of the function $y = |x| + |x + 1| + |x + 2|$? Find the equations of each of its parts.

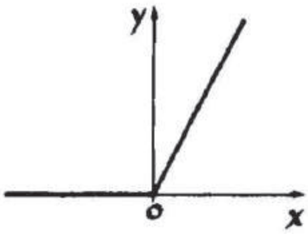


Fig. 10

3. (*a*) The function whose graph is represented in Fig. 10 can be defined by the following conditions:

$$\text{for } x < 0, \quad y = 0$$
$$\text{for } x \geq 0, \quad y = 2x.$$

Try to give this function in terms of one formula.

(*b*) Write down formulas for the functions whose graphs are represented in Figs. 11 and 12, respectively. ⊕

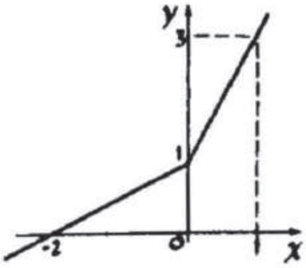


Fig. 11

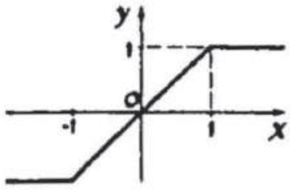


Fig. 12

4. Construct the graph of the function

$$y = |3x - 2|.$$

Hint: Obtain this graph from the graph of $y = |x|$ by two transformations: a translation along the x-axis and stretching along the y-axis. In order to determine the correct value of the translation, it is necessary to pull the coefficient of x in front of the absolute value sign: $|3x - 2| = 3|x - \frac{2}{3}|$.

5

Problem. Construct the graph of

$$y = |2x - 1|.$$

Solution. We shall obtain this graph from the straight line $y = 2x - 1$ (Fig. 13). Wherever the straight line is above the x-axis, y is positive; that is, $2x - 1 > 0$. Hence, in this interval $|2x - 1| = 2x - 1$, and the desired graph coincides with the graph of $y = 2x - 1$. Wherever $2x - 1 < 0$ (that is, the straight line $y = 2x - 1$ is below the x-axis), $|2x - 1| = -(2x - 1)$. Thus, in order to obtain the graph of $y = |2x - 1|$ from this section of the graph of $y = 2x - 1$, it is necessary to change the sign of the ordinate of each point of the straight line $y = 2x - 1$, that is, to reflect this straight line in the x-axis. We thus obtain Fig. 14.

Fig. 13

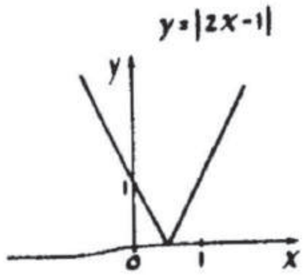
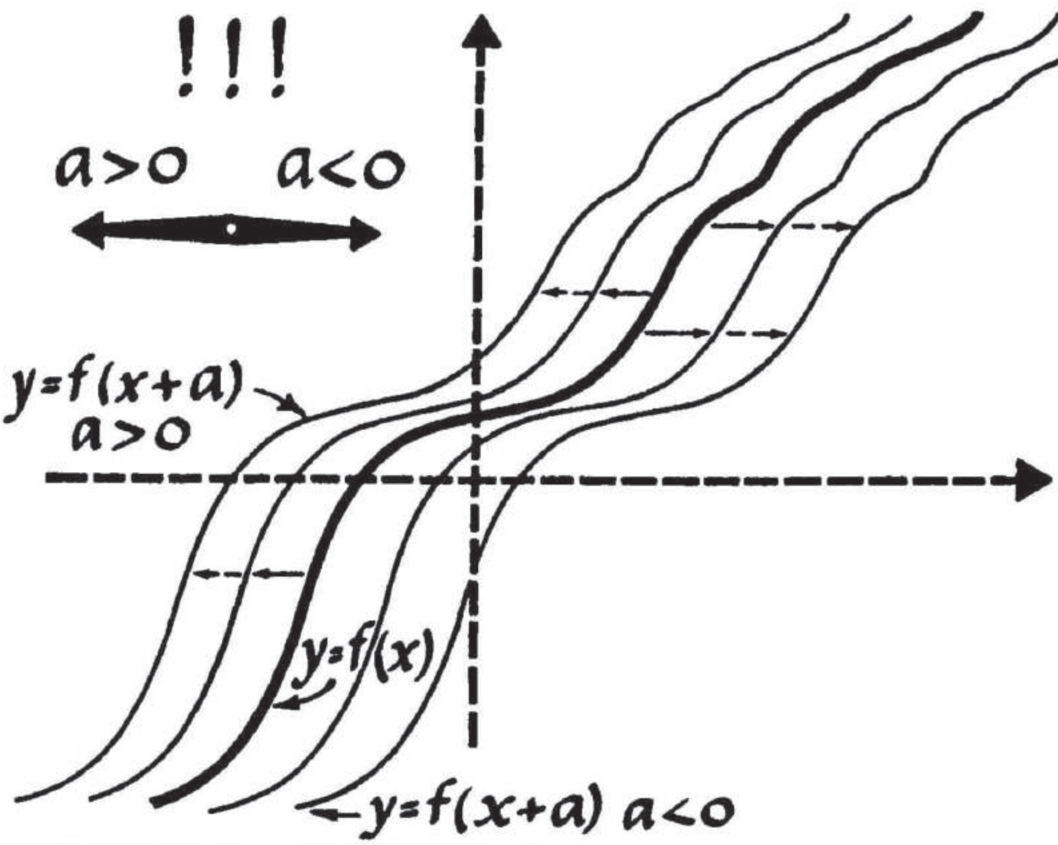
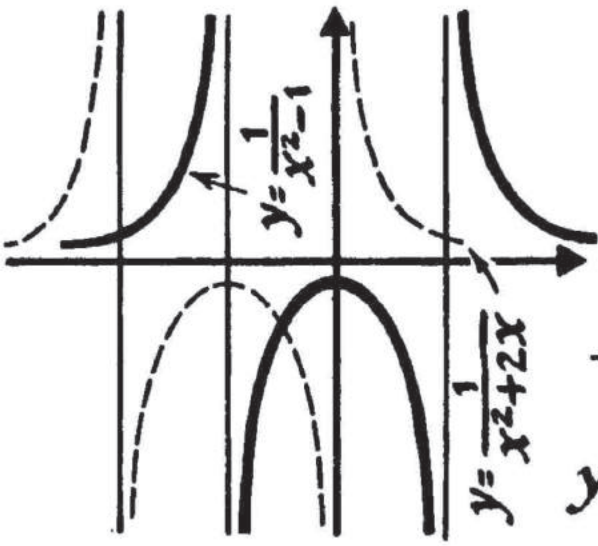


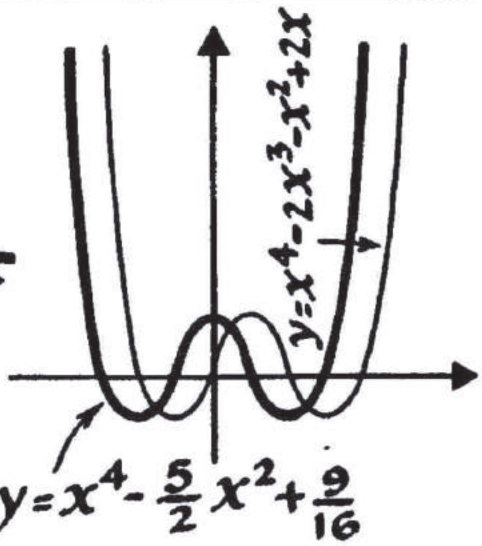
Fig. 14



$$f(x) \longrightarrow f(x+a)$$

The graph of the function $y = f(x+a)$ is obtained from the graph of the function $y = f(x)$ by a translation along the x-axis by -a units. The minus sign means that, if a is postive, the graph moves to the left, and if a is negative, the graph moves to the right.





EXERCISE

From the known graph of

$$y = x^4 - 2x^3 - x^2 + 2x$$

(Fig. 9, p. 14), construct the graph of

$$y = |x^4 - 2x^3 - x^2 + 2x|.$$

6

Problem. From the known graph of

$$y = \frac{1}{x^2 - 2x + 2} \tag{1}$$

(Fig. 15), construct the graph of

$$y = \frac{1}{x^2 - 2|x| + 2}. \tag{2}$$

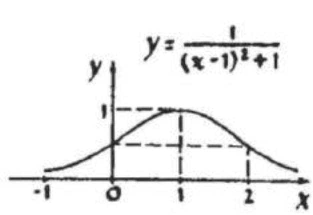


Fig. 15

Solution. Since for positive values of the argument $|x| = x$,

$$\frac{1}{x^2 - 2|x| + 2} = \frac{1}{x^2 - 2x + 2}.$$

Hence, to the right of zero the graph of Eq. 2 coincides with the graph of Eq. 1 (Fig. 16). In order to obtain the left half of the desired graph of Eq. 2, we note that the function $y = 1/(x^2 - 2|x| + 2)$ is even. This means that the left half of the graph of Eq. 2 is obtained from its right half by reflection in the y-axis (Fig. 17). The same is true in the general case: in order to obtain the graph of $y = f(|x|)$ from the graph of $y = f(x)$, it is necessary to reflect the half of the first graph lying to the right of the y-axis in the y-axis.

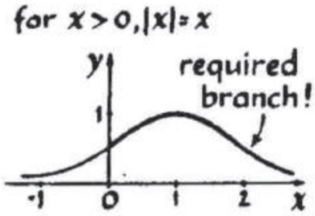


Fig. 16

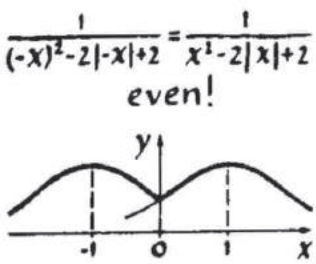
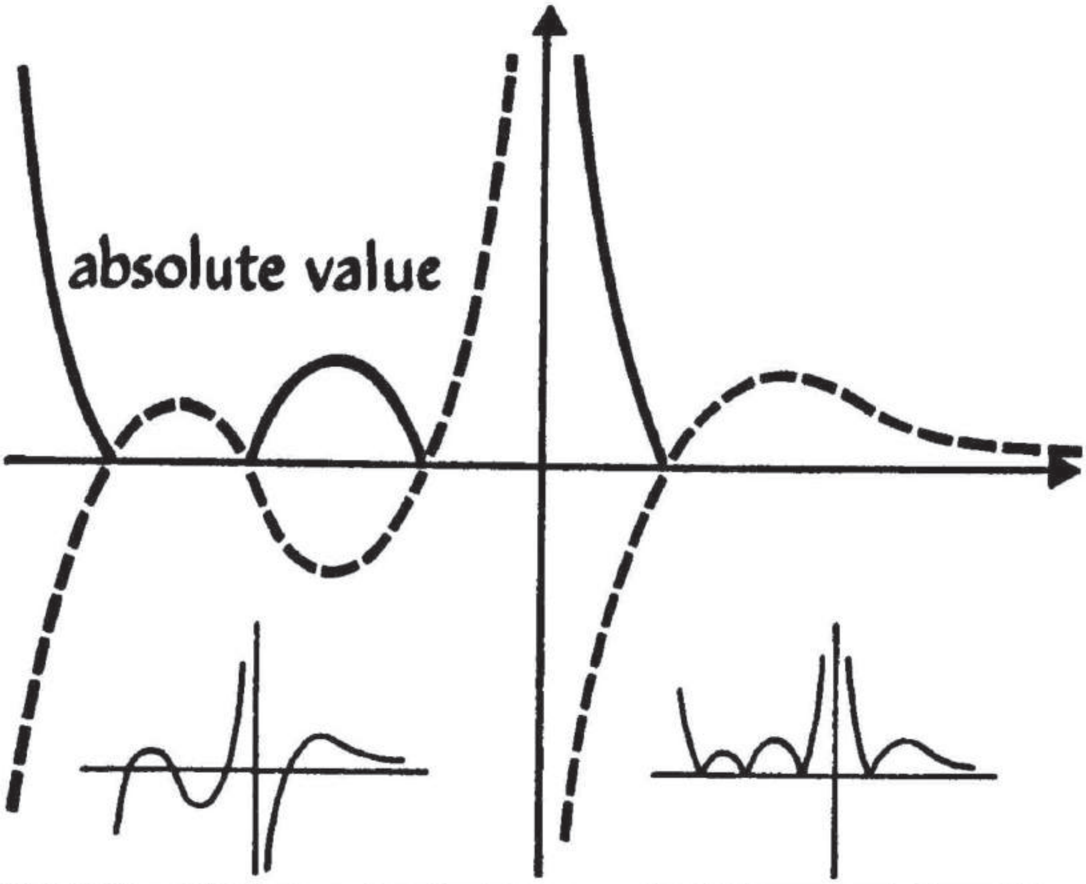


Fig. 17

EXERCISES

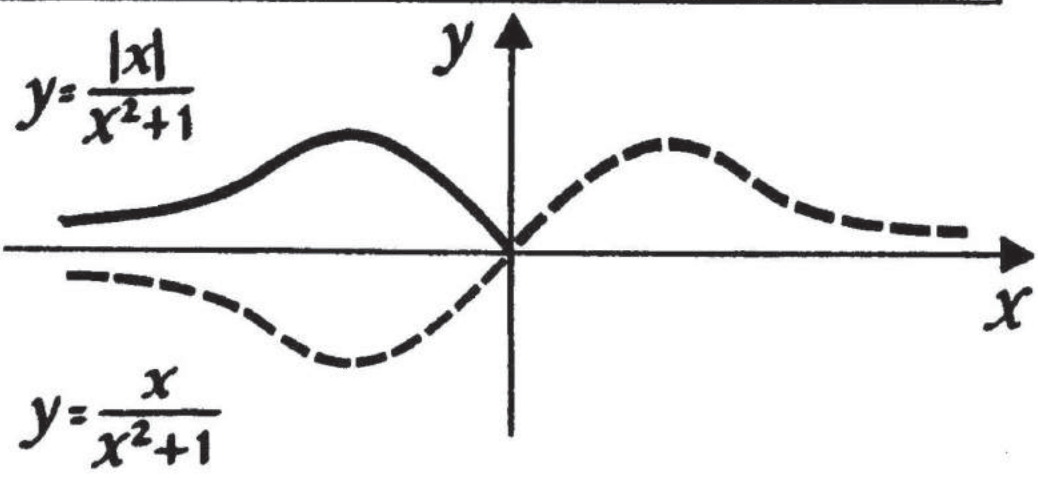
1. Construct the graph of $y = 2|x| - 1$.
2. Construct the graphs of

(*a*) $y = 4 - 2x$;



$$y = f(x) \longrightarrow y = |f(x)|$$

In order to obtain the graph of $y = |f(x)|$ from the graph of $y = f(x)$, it is necessary to leave the parts of the graph lying above the x-axis unchanged, and to reflect the parts lying below the x-axis in the x-axis.



(*b*) $y = |4 - 2x|$;

(*c*) $y = 4 - 2|x|$;

(*d*) $y = |4 - 2|x||$.

3. Find the least value of the function

$$y = |x - 2| + |x| + |x + 2| + |x + 4|. \quad \oplus$$

In concluding this section we suggest that you solve some problems. At first glance they seem to bear no relation whatever to what we have been concerned with in this section, but on reflection you will see that this is not so.

Problems*

1. Seven matchboxes are arranged in a row. The first contains 19 matches, the second 9, and the following ones contain 26, 8, 18, 11, and 14 matches, respectively (Fig. 18). Matches may be taken from any box and put into any adjacent box. The matches must be redistributed so that their number in all boxes becomes the same. How can this be accomplished, shifting as few matches as possible?

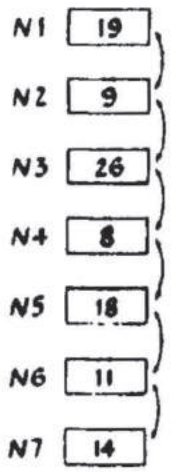


Fig. 18

Solution. There is a total of 105 matches in all the boxes. Hence, if there were the same number of matches in all boxes, then each box would contain 15 matches. With such an arrangement of the boxes the problem has but one solution: Shift 4 matches from the first box to the second. As a result the first box will contain 15 matches and the second 13 matches. Take 2 matches from the third box and place them in the second, so that 24 matches remain in the third box. Shift the excess number of matches from this box into the fourth, and so on.

Problems 2 and 3 are somewhat more difficult. The question will be the same as in Problem 1.

*Problems 1 through 4 and the method of their solution were suggested by M. L. Tsetlin.

2. Seven matchboxes are arranged in a straight line, as previously, but the number of matches in the boxes is different. That is, the first contains 1 match, the second contains 2, the following ones 3, 72, 32, 20, and 10.

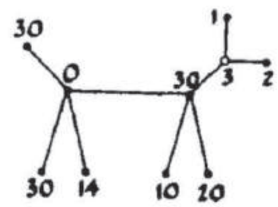


Fig. 19

3. The matchboxes are arranged in the shape of a "dog" (Fig. 19). Matches may be moved only along the lines (routes) joining the boxes.

Graphs can be used to advantage for the following problem.

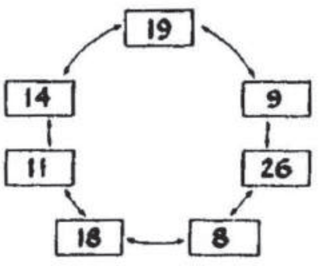
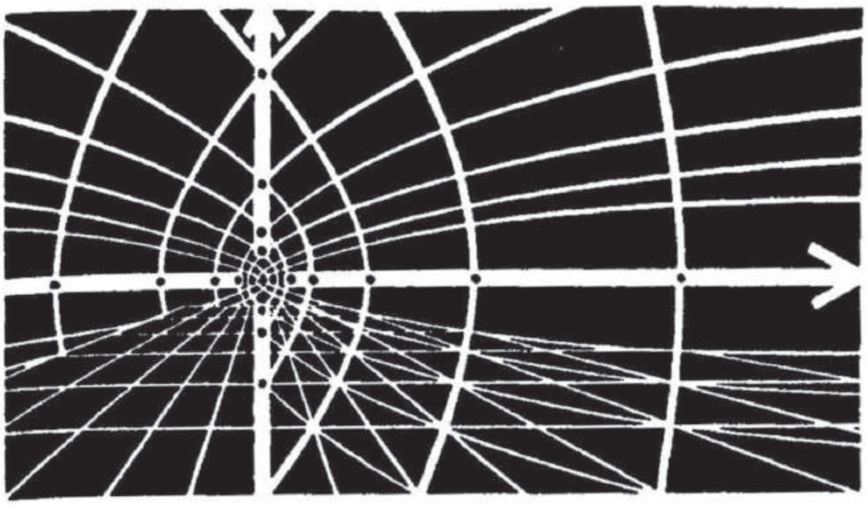


Fig. 20

4. Seven matchboxes are arranged in a circle. The first contains 19 matches, the second 9 matches, and the remaining ones contain 26, 8, 18, 11, and 14 matches, respectively (Fig. 20). It is permissible to move the matches from any box to any adjacent box. The matches must be shifted in such a way that the number of matches in all boxes becomes the same. How can this be done, shifting as few matches as possible? ⊕



CHAPTER 4

The Quadratic Trinomial

1

Let us now pass to the function

$$y = x^2.$$

You have, of course, constructed its graph, and you know that this curve has the special name of *parabola*. Graphs of the functions $y = ax^2$ are obtained from the graph of $y = x^2$ by stretching and are also called parabolas.*

EXERCISE

Figure 1 represents a parabola. (*a*) It is known that it is the graph of the function $y = x^2$. Find the scale of the diagram (the scale is the same for both axes). (*b*) What scale unit must be taken along the axes in order that the same curve serve as graph of the function $y = 5x^2$?

Fig. 1

*It is interesting that all parabolas are similar to one another (see p. 43 as well as p. 95, Problem 16*d*).

2

Let us find out how the values of the function $y = x^2$ will change if the values of the argument constantly change by the same quantity; that is, if they form an arithmetic progression. For simplicity let us consider positive values of x. For example, suppose x takes the values

$$1, 2, 3, 4, \cdots, \text{etc.}$$

Then y will assume the values

$$1, 4, 9, 16, \cdots, \text{etc.}$$

x	y	increment
1	1	$4-1=3$
2	4	$9-4=5$
3	9	$16-9=7$
4	16	$25-16=9$
5	25	
$\vdots$	$\vdots$	$\vdots$

increment of increment = 2 constant!

Fig. 2

You can see that the values of y no longer form an arithmetic progression.

Let us add another column to the table of values of argument and function (Fig. 2). In this column we shall write down by how much the value of y changes when the argument x passes from one of its values to the next. For example, suppose the argument changes from the value $x = 2$ to the value $x = 3$. Then the function changes from the value $y = 4$ to the value $y = 9$. The change, or, as they say, the *increment* of the function* equals the difference between the new value and the previous value of the function, namely,

$$9 - 4 = 5.$$

Thus in the third column of our table we write the increments of the function $y = x^2$. It can now be clearly seen that the function $y = x^2$ varies in such a way that if x increases, so does not only the function itself but also its increment. In the graph this fact is also apparent: the curve $y = x^2$ goes up more and more steeply, while the graph of the linear function, which varies uniformly, always forms the same angle with the x-axis (Fig. 3).

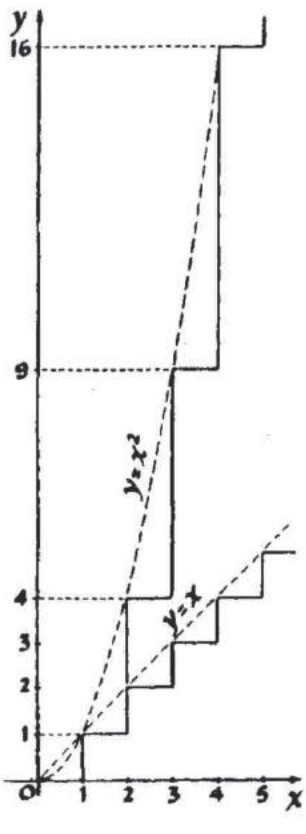


Fig. 3

*An increment of the function $y = f(x)$ is usually denoted by the Greek letter Δ (delta):

$$\Delta y, \text{ or } \Delta f(x).$$

It is interesting to note that the increments of the function $y = x^2$ form an arithmetic progression. The reader is asked to try to prove this fact in general: If the values of the argument x form an arithmetic progression,

$$a, a + d, a + 2d, \cdots, a + bd, \cdots,$$

then the values of the corresponding increments of the quadratic function $y = x^2$ also form an arithmetic progression.

If the argument t is the time and the function s is the distance covered (we change x to t and y to s according to the accepted notation in physics), then the relationship $s = t^2$ corresponds to uniformly accelerated motion (with acceleration equal to 2), while the formula $s = kt + b$ represents uniform motion with velocity k. In uniform motion a body covers equal distances in equal intervals of time; that is, equal increments in the function correspond to equal increments in the argument (a linear function converts every arithmetic progression into an arithmetic progression). In uniformly accelerated motion the distances of the path covered in equal intervals of time increase uniformly. Correspondingly, for a quadratic function (by the way, not only for $y = x^2$, but for any function $y = ax^2 + bx + c$), uniformly increasing increments of the functions correspond to equal increments of the argument.

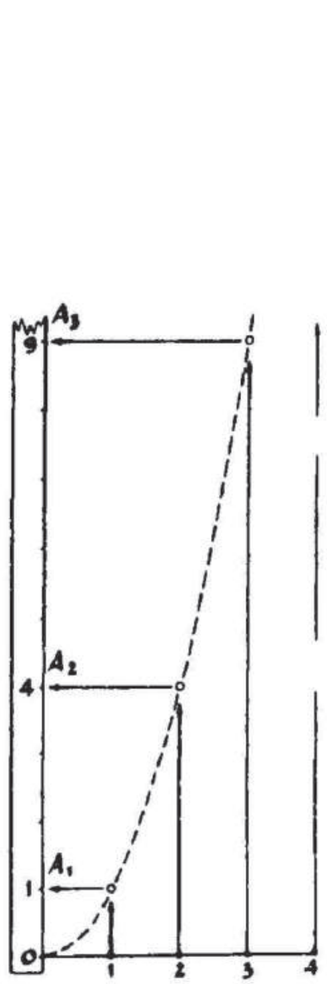


Fig. 4

EXERCISES

1. Work out a table with three columns (values of the argument, values of the function, and values of the increments of the function) for the trinomial $y = x^2 + x - 3$, taking for x the values $1, 0, -1, -2, -3$. Add one more column to the table and enter in it the differences between two consecutive increments.

Now take another trinomial: $x^2 + 3x + 5$. Work out a similar table for it. Compare the last columns

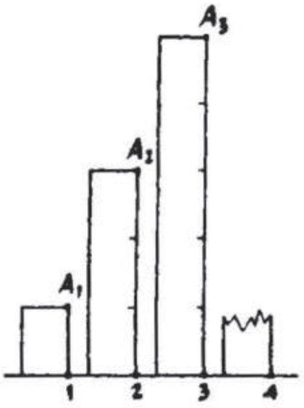


Fig. 5

of these two tables. And what is the result if the trinomial $y = 2x^2 + 3x + 5$ is used?

2. From Fig. 4 it is obvious that if a uniform scale is taken on the positive half of the x-axis, then the graph of $y = x^2$ transforms it into the scale O, A_1, A_2, A_3, etc., located on the y-axis, which is no longer uniform. With this scale the y-axis splits into the segments OA_1, A_1A_2, A_2A_3, etc. Imagine that the y-axis is cut into these segments, which are then placed vertically one after the other along the x-axis at equal distances from each other (with their base points at the points 1, 2, 3, etc.) (Fig. 5). How are the end points of the segments arranged? Explain the result.

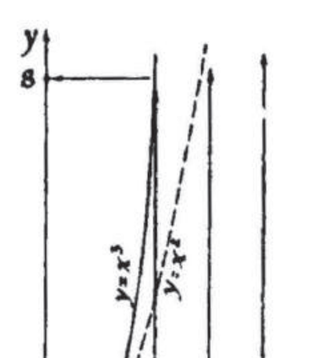


Fig. 6

3. Let us consider the graph of $y = x^3$ for positive values of x (Fig. 6). Do the same with it as with the graph of $y = x^2$ in Exercise 2. In your drawing, show how the end points of the segments are arranged in this case. A more difficult question: Can you find the equation of the curve on which the end points of the segments lie?

Problem

Construct the graph of $y = x^2$. Take a fairly large scale: $1 = 2$ cm (4 squares). Mark the point $F(0, \frac{1}{4})$ on the y-axis.

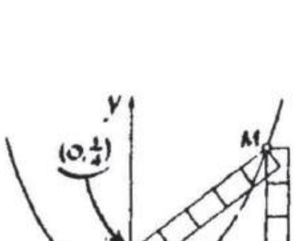


Fig. 7

Measure the distance from the point F to any point M of the parabola with a strip of paper. Then fasten the strip at the point M and turn it around this point so that it becomes vertical. The end of the strip drops somewhat below the x-axis. Mark on the strip how much it goes beyond the x-axis (Fig. 7). Now take another point on the parabola and repeat the measurement. By how much does the edge of the strip now drop below the x-axis? We can tell you the result in advance: Whatever point is taken on the parabola $y = x^2$, the distance from this point to the point $(0, \frac{1}{4})$ will always be larger than the distance from the same point to the x-axis by the same number, $\frac{1}{4}$.

This can be put differently: The distance from any point of the parabola to the point $(0, \frac{1}{4})$ is equal to the distance from the same point of the parabola to the line $y = -\frac{1}{4}$, which is parallel to the x-axis.

This remarkable point, $F(0, \frac{1}{4})$, is called the *focus* of the parabola $y = x^2$, and the straight line $y = -\frac{1}{4}$ is called the *directrix* of this parabola.

Every parabola has a directrix and a focus. (See also Fig. 1 on p. 44.)

3

Let us now consider the graphs of quadratic trinomials of the form

$$y = x^2 + px + q.$$

Let us show that in their form they differ in no way from the parabola $y = x^2$ but are only in a different position with respect to the coordinate axes.

To begin with, let us consider a numerical example. We take the trinomial

$$y = x^2 + 2x + 3.$$

To obtain its graph, let us represent it in the form

$$y = (x + 1)^2 + 2,$$

where we have completed the square.

The graph of $y = (x + 1)^2$ is obtained from the parabola $y = x^2$ by translation along the x-axis. (Explain why the curve $y = (x + 1)^2$ is obtained from $y = x^2$ by a translation to the left.) From the graph of $y = (x + 1)^2$ the graph of $y = (x + 1)^2 + 2$ can be obtained very simply (Fig. 8).

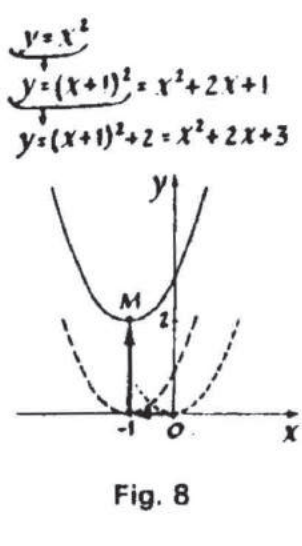


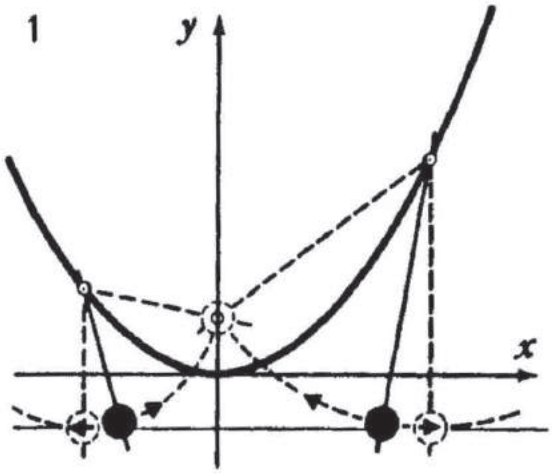
Fig. 8

Thus, the graph of the trinomial

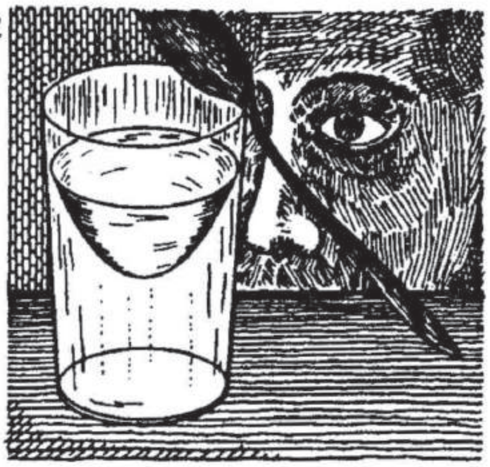
$$y = (x + 1)^2 + 2 = x^2 + 2x + 3$$

is obtained from the parabola $y = x^2$ by translation to the left by one unit and up by two units. In this

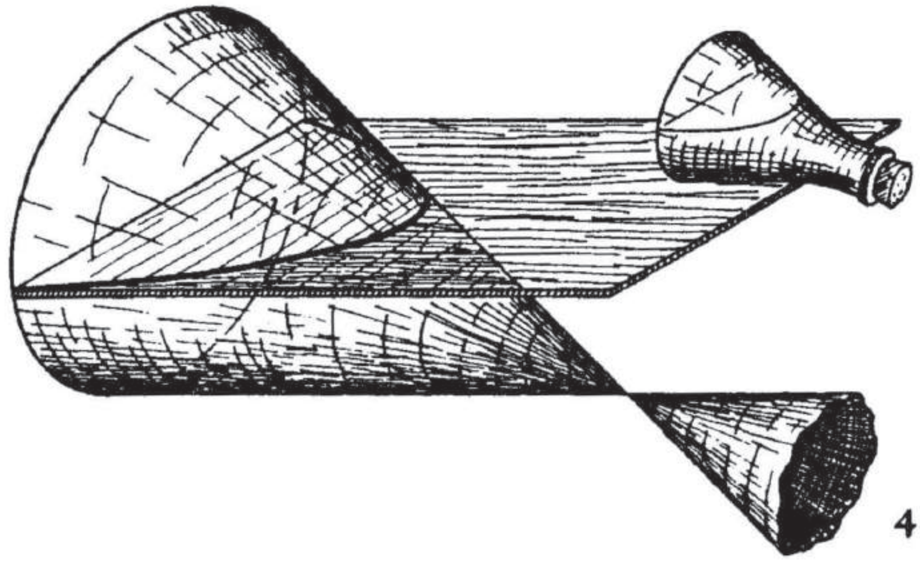
1



2



3



4

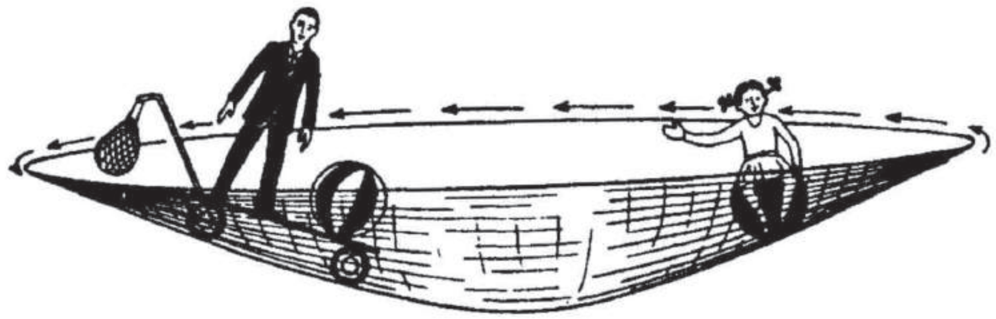
Interesting Properties of the Parabola

1. Any point of the parabola is equidistant from some point, called the focus of the parabola, and some straight line, called the directrix.

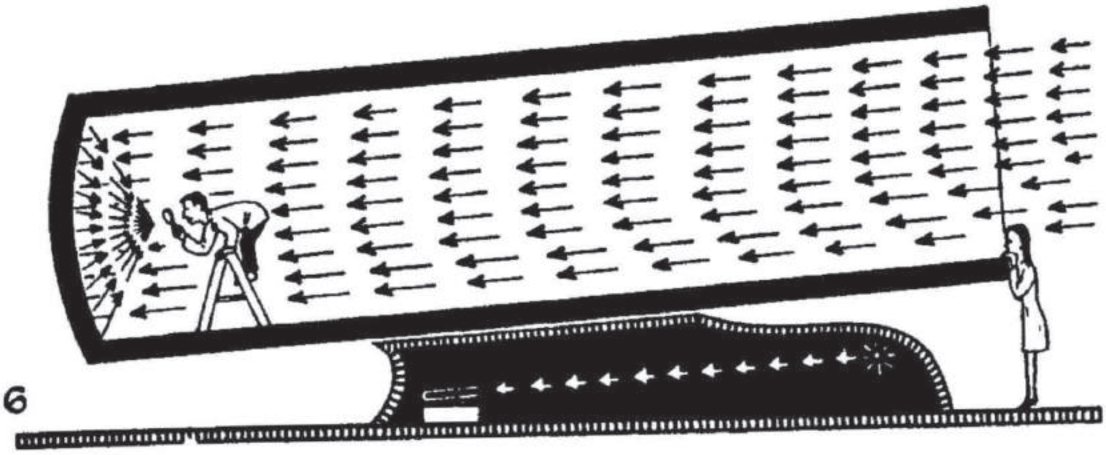
2. The surface of a liquid in a rotating container has the shape of a paraboloid of revolution. You can see this surface if you stir a partially filled glass of water vigorously with a small spoon and then remove the spoon.

3. If a stone is thrown at some angle to the horizon, then it travels along a parabola.

4. If the surface of a cone is intersected with a plane parallel to one of its generators, then a parabola is obtained as section.



5



6

5. In parks an amusement attraction, "the miraculous paraboloid," is occasionally set up. To each of the persons standing in the paraboloid it seems that he is standing on the floor, while the other persons by some miracle stick to the walls.*

6. Parabolic mirrors are also used in reflecting telescopes: the light of the distant star, traveling in a parallel pencil, having fallen on the mirror of the telescope, converges to the focus.

*The experiments described in points 2 and 5 are based on the same property of the paraboloid: is rotated with appropriate speed about its vertically directed axis, then the resultant of the centrifugal force of gravitation in any point of the paraboloid is directed perpendicular to its surface.

transformation the vertex of the parabola, located at the point (0, 0), the origin, goes into the point M with coordinates $(-1, 2)$.

EXERCISES

1. Draw graphs of the functions:

(*a*) $y = (x + 2)^2 + 3$;

(*b*) $y = (x + 2)^2 - 3$;

(*c*) $y = (x - 2)^2 + 3$;

(*d*) $y = (x - 2)^2 - 3$.

2. (*a*) Find the least value of the function

$$y = x^2 + 6x + 5.$$

Solution. The least value of the given function is the ordinate of the vertex of the parabola $y = x^2 + 6x + 5$. To determine the coordinates of the vertex, let us complete the square:

$$x^2 + 6x + 5 = (x + 3)^2 - 4.$$

Thus our parabola was obtained from $y = x^2$ by translation along the x-axis by -3 units and along the y-axis by -4 units; that is, the least value of the function equals -4.

(*b*) The vertex of the parabola $y = x^2 + px + q$ is located at the point $(-1, 2)$. Find p and q.

Now let us prove that by translating the parabola $y = x^2$, we can obtain the graph of any quadratic trinomial of the form

$$y = x^2 + px + q.$$

For this purpose, we complete the square, as before, that is, we represent our trinomial in the form $y = (x + \cdots)^2 + \cdots$, where the second summand in brackets and the constant term must be chosen so as not to depend on x.

After expansion of the expression in parentheses, the term linear in x is the result of doubling a product, and since this term must equal px, the second summand in the parentheses must be taken equal to $p/2$. Thus we have

$$x^2 + px + q = \left(x + \frac{p}{2}\right)^2 + \cdots$$
$$= x^2 + px + \frac{p^2}{4} + \cdots.$$

Since the constant term of the trinomial must equal q, instead of the three dots we have to take $q - p^2/4$.

Thus the trinomial $y = x^2 + px + q$ can be rewritten in the form

$$y = \left(x + \frac{p}{2}\right)^2 + q - \frac{p^2}{4}.$$

We see (Fig. 9) that *the graph of*

$$y = x^2 + px + q$$

represents the parabola $y = x^2$ translated by $-p/2$ along the x-axis and by $q - p^2/4$ along the y-axis.*

The vertex M of this parabola has the abscissa $x_M = -p/2$ and the ordinate $y_M = q - (p^2/4)$.

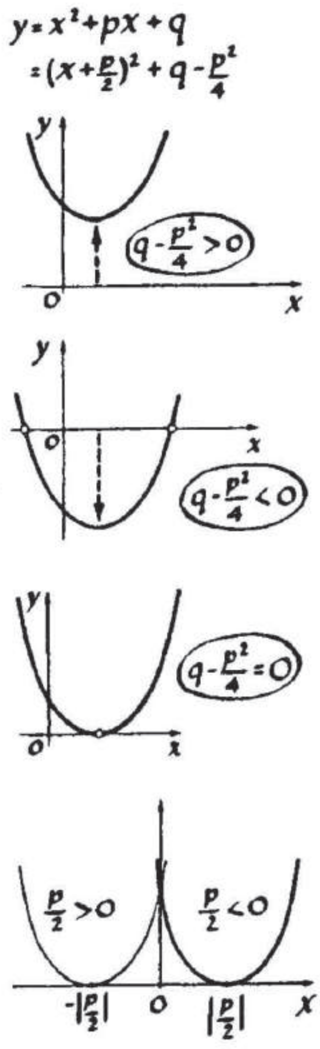


Fig. 9

4

Taking as "base" the graph of $y = ax^2$, in the same way we can obtain the graph of the quadratic trinomial of the more general form,

$$y = ax^2 + bx + c.$$

Let us analyze this, using an example. Let us take the trinomial $y = \frac{1}{2}x^2 - 3x + 6$. Let us put the coefficient of x^2 in front of the parentheses:

$$\tfrac{1}{2}x^2 - 3x + 6 = \tfrac{1}{2}(x^2 - 6x + 12).$$

*Translation by $-p/2$ along the x-axis means a translation to the right if $-p/2 > 0$, and a translation to the left if $-p/2 < 0$.

We complete the square in the expression inside the parentheses:

$$\tfrac{1}{2}(x^2 - 6x + 12) = \tfrac{1}{2}(x^2 - 2 \times 3x + 9 + 3)$$
$$= \tfrac{1}{2}[(x - 3)^2 + 3].$$

Thus, finally,

$$y = \tfrac{1}{2}(x - 3)^2 + \tfrac{3}{2}.$$

We see that the graph of $y = \frac{1}{2}x^2(x - 3)^2 + \frac{3}{2}$ is obtained from the parabola $y = \frac{1}{2}x^2$ by translating 3 units to the right along the x-axis and $\frac{3}{2}$ units up along the y-axis.

Problems

1. Translate the parabola ax^2 along the x-axis and y-axis so that the graph of the trinomial $y = ax^2 + bx + c$ is obtained.

(**Answer.** The parabola $y = ax^2 + bx + c$ is obtained from the parabola $y = ax^2$ by translation of $-b/2a$ along the x-axis and $(4ac - b^2)/4a$ along the y-axis.)

2. Find the least value of the function $y = 2x^2 - 4x + 5$ in the intervals:

(*a*) from $x = 0$ to $x = 5$ $\quad (0 \le x \le 5)$,

(*b*) from $x = -5$ to $x = 0$ $\quad (-5 \le x \le 0)$.

Solution. Let us make use of the results of the preceding problem and construct the graph of the function $y = 2x^2 - 4x + 5$ (Fig. 10).

From the diagram it is obvious that as x varies from the value $x = 0$ to the value $x = 5$, the function $y = 2x^2 - 4x + 5$ initially decreases (to $x = 1$), and then increases. Thus the least value of the function $y = 2x^2 - 4x + 5$ in interval (*a*) is its value at $x = 1$ (Fig. 10a). As x varies from -5 to 0, the function $y = 2x^2 - 4x + 5$ constantly decreases. Hence

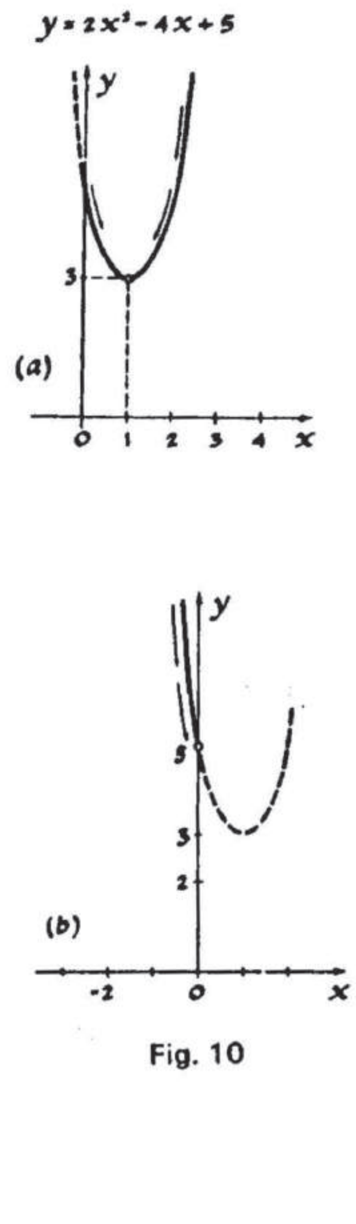


Fig. 10

the least value of the function in interval (*b*) is its value at $x = 0$ (Fig. 10b).

(**Answer.** The least value in the interval (*a*) equals 3; in interval (*b*) it is 5.)

EXERCISES

1. Draw graphs of the following functions, indicating the exact coordinates of the vertex of each of the parabolas and the coordinates of the points of intersection of the graphs with the coordinate axes:

(*a*) $y = x - x^2 - 1$;
(*b*) $y = -3x^2 - 2x + 1$;
(*c*) $y = 10x^2 - 10x + 3$;
(*d*) $y = 0.125x^2 + x + 2$.

2. The graph of what function is obtained if the parabola $y = x^2$ is first stretched in the ratio 2:1 along the y-axis and then translated 3 units down along the same axis? The graph of what function is obtained if these two transformations are carried out in reverse order: first the parabola $y = x^2$ is translated 3 units downward, and then the curve thus obtained is expanded in the ratio 2:1 along the y-axis (Fig. 11)?

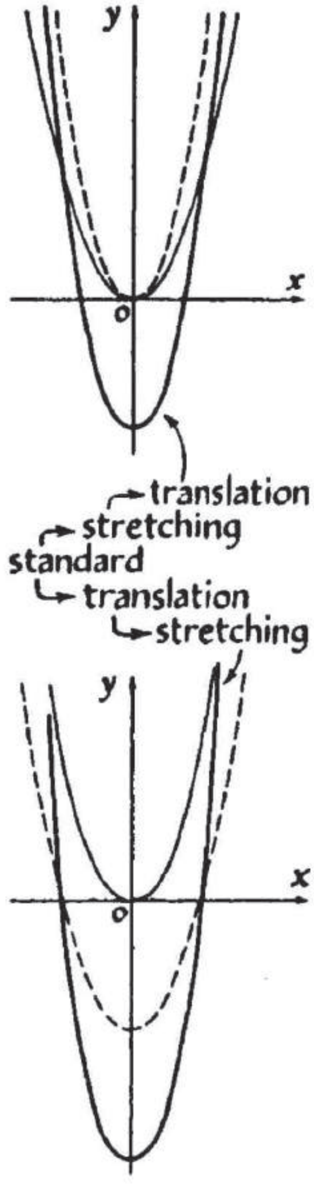


Fig. 11

3. By how much must the parabola

$$y = x^2 - 3x + 2$$

be translated along the x-axis and along the y-axis so that the parabola $y = x^2 + x + 1$ is obtained?

4. Translate the parabola $y = x^2$ along the x-axis so that it will pass through the point (3, 2). The graph of what function is obtained (Fig. 12)?

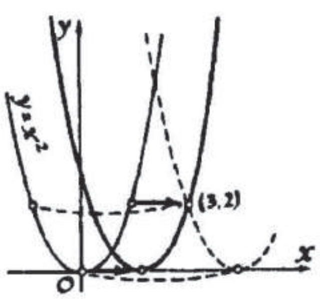


Fig. 12

5

Let us now see what can be said about the solution of the quadratic equation $x^2 + px + q = 0$ using the graph of the function $y = x^2 + px + q$.

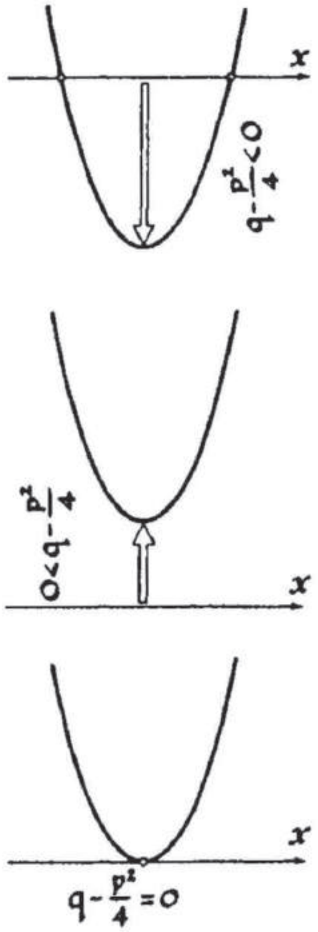


Fig. 13

The roots of this equation are those values of x for which the value of the function $y = x^2 + px + q$ is equal to zero. In the graph these points have ordinates equal to zero; that is, they lie on the x-axis.

From the graph of the quadratic trinomial

$$y = x^2 + px + q,$$

it is immediately obvious that the quadratic equation $x^2 + px + q = 0$ has two real roots if $p^2/4 - q > 0$, and has no roots if $p^2/4 - q < 0$. (Recall that the parabola $y = x^2$ moves down if $q - p^2/4 < 0$, and up if $q - p^2/4 > 0$) (see Fig. 13).

If $p^2/4 - q = 0$, our quadratic equation

$$x^2 + px + q = 0$$

transforms into the equation $(x + p/2)^2 = 0$. This case is particularly interesting. Let us consider it in detail.

The equation $x - 2 = 0$ has one solution, $x = 2$. The equation $(x - 2)^2 = 0$ also has but one solution, $x = 2$; no other number satisfies this equation.

However, in the first case we say that the equation $x - 2 = 0$ has one root, while in the second case we say that the equation $(x - 2)^2 = 0$ has a multiple root or two equal roots: $x_1 = 2$ and $x_2 = 2$.

How can this difference be explained?

There are several methods, and we shall give one of them. Let us alter the first equation a little: let us replace zero in the right-hand term by some small number. The root changes, of course, but it will remain unique, as before; the equation is satisfied by only one number, as before. For example,

$$x - 2 = 0.01, \qquad x = 2.01.$$

Let us now alter the second equation in the same way:

$$(x - 2)^2 = 0.01, \qquad x^2 - 4x + 3.99 = 0.$$

The resulting equation will now have two roots, $x_1 \approx 2.1$ and $x_2 \approx 1.9$. Now we shall again change the right-hand side in the equation $(x - 2)^2 = 0.01$, replacing it by smaller and smaller numbers. As long as this right-hand side does not equal zero, the equation will have two different roots. As the right-hand side diminishes, the roots "approach each other" so that their values will differ from one another by a smaller and smaller quantity. Finally, when the right-hand side becomes equal to zero, the two roots "coincide" — the values of the two roots become equal to each other. Therefore one says that the equation $(x - 2)^2 = 0$ has two roots, merging into a double root.

Geometrically the case of coincident roots corresponds to the parabola $y = (x - 2)^2$ touching the x-axis.

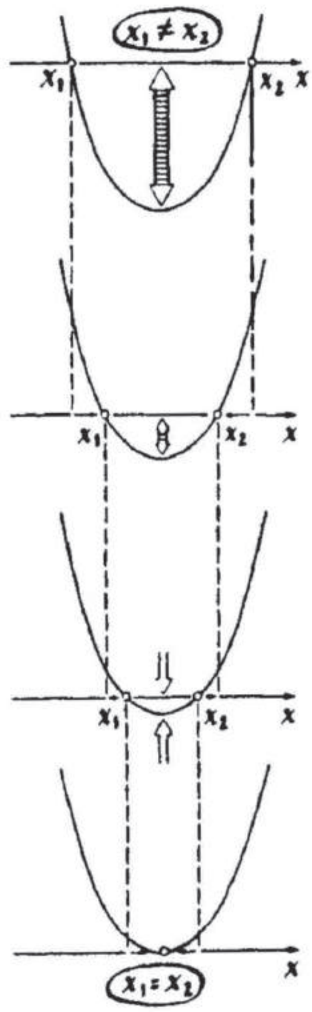


Fig. 14

The general case of a quadratic trinomial $y = x^2 + px + q$ will be analyzed geometrically. Suppose, to begin with, the constant term q is smaller than $p^2/4$ (i.e., $q - p^2/4 < 0$), so that the parabola $y = x^2 + px + q$ has two points of intersection with the x-axis (Fig. 14). We shall increase the constant term: At first the parabola, moving up, will have two points of intersection with the x-axis (the equation $x^2 + px + q$ then has two distinct roots); then as these points of intersection get closer, at a certain moment (when $q - p^2/4 = 0$) they merge into one point. At this moment the parabola

$$y = x^2 + px + q = \left(x + \frac{p}{2}\right)^2$$

touches the x-axis, and the equation

$$x^2 + px + \frac{p^2}{4} = 0$$

has one double root. As the constant then is increased further, the parabola ceases to intersect the x-axis, and the equation $x^2 + px + q = 0$ will not have any real roots.

EXERCISES

1. Find the parabola $y = ax^2 + bx + c$ which intersects the x-axis at the points $x = 3$ and $x = -5$, and the y-axis at the point $y = 30$.

Solution. The quadratic trinomial defining this parabola will have the form

$$a(x - 3)(x + 5).$$

The point of intersection with the y-axis is obtained by setting $x = 0$. Hence, when $x = 0$ our function must equal 30. We obtain

$$a(-3)(+5) = 30, \qquad \text{hence } a = -2.$$

(**Answer.** The parabola $y = -2x^2 - 4x + 30$.)

2. (a) Find the quadratic trinomial of the form $x^2 + px + q$, if its graph intersects the x-axis at the points $x = 2$ and $x = 5$.

(b) Find the cubic polynomial of the form $y = x^3 + px^2 + qx + r$, if it is known that its graph intersects the x-axis at the points $x = 1$, $x = 2$, and $x = 3$.

(c) Can you devise a polynomial whose graph would intersect the x-axis at 101 points: $x_1 = -50$, $x_2 = -49$, $x_3 = -48, \cdots, x_{101} = 50$?

What is the least degree of such polynomials?

3. The trinomial $-x^2 + 6x - 9$ has two identical roots:

(a) Change the constant term by 0.01 so that the resulting trinomial has two distinct roots.

(b) Can the same result be obtained by making a change of 0.01, only this time in the coefficient of x?

4. Figures 15a and b represent graphs of quadratic trinomials $y = x^2 + px + q$. Find p and q. Draw the graph of Fig. 15b, making a more fortunate choice of scale and position of the axes.

Fig. 15

5. Figures 16a, b, and c depict graphs of the quadratic trinomials $y = ax^2 + bx + c$. Find a, b, and c.

6. (*a*) Solve the inequality

$$x^2 - 5x + 4 > 0.$$

Solution. From Fig. 17 it is obvious that the function $y = x^2 - 5x + 4$ is positive in two intervals: for x less than 1 and for x larger than 4.

(**Answer.** $x < 1$ and $x > 4$.)

(*b*) Solve the inequality

$$x - 1 < |x^2 - 5x + 4|.$$

Solution. Let us draw in one diagram the functions on the right-hand and left-hand sides. From Fig. 18 it is apparent that the straight line $y = x - 1$ has three points in common with the graph of

$$y = |x^2 - 5x + 4|: A(x_1, y_1), B(x_2, y_2), C(x_3, y_3).$$

The condition $x - 1 < |x^2 - 5x + 4|$ is satisfied in three intervals: $x < x_1$, $x_1 < x < x_2$, $x > x_3$. The values of x_1 and x_3 can be found from the equation

$$x - 1 = x^2 - 5x + 4.$$

The value of x_2 is found from the equation

$$x - 1 = -(x^2 - 5x + 4).$$

(**Answer.** $x < 1$, $1 < x < 3$ and $x > 5$, that is, all x, except $x = 1$ and $3 \leq x \leq 5$.)

(*c*) Write down the answer for the inequalities

$$x - 1 > |x^2 - 5x + 4|,$$
$$x - 1 \geq |x^2 - 5x + 4|.$$

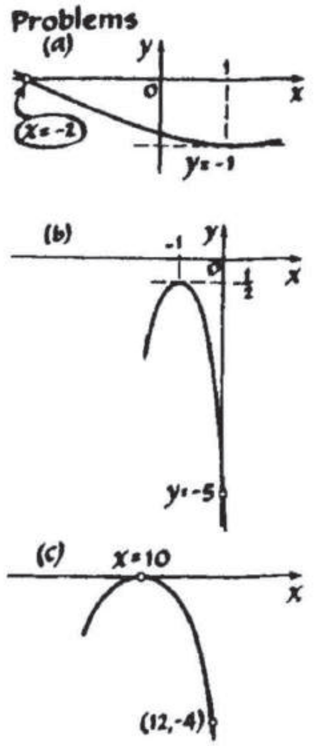


Fig. 16

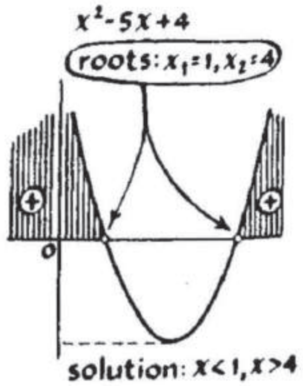


Fig. 17

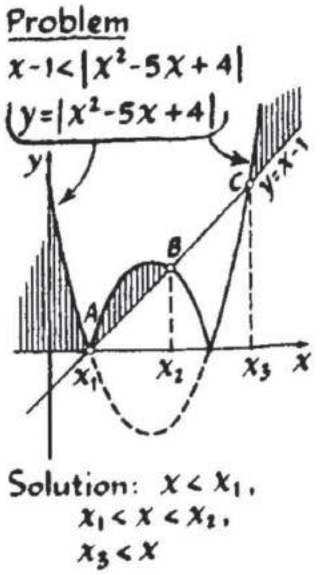


Fig. 18

7. Find the largest value of the function

$$y = x^2 - 5|x| + 4$$

in the interval from -2 to 2.

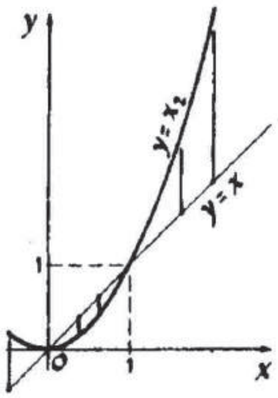


Fig. 19

6

The graph of $y = x^2$ can also be drawn by "squaring" the graph of $y = x$, that is, mentally squaring the value of each ordinate (Fig. 19).

EXERCISES

1. Given the graph (Fig. 20) of $y = x - 1$, draw the graph of $y = (x - 1)^2$ in the same diagram.

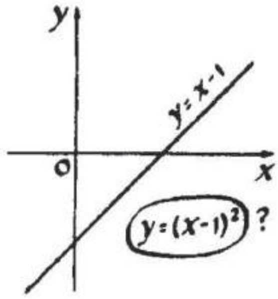


Fig. 20

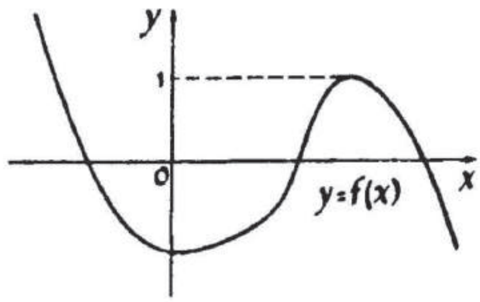


Fig. 21

2. Given the graph (Fig. 21) of $y = f(x)$, draw the graph of $y = \big(f(x)\big)^2$ in the same diagram.

3. Making use of the graph of

$$y = x(x + 1)(x - 1)(x - 2)$$

(Fig. 9 on p. 14), draw the graph of

$$y = x^2(x + 1)^2(x - 1)^2(x - 2)^2.$$

4. Draw graphs of:

(*a*) $y = [x]^2$,

(*b*) $y = (x - [x])^2$.

CHAPTER 5

The Linear Fractional Function

1

Figure 1 represents the "graph" of the function $y = 1/x$ in the form in which it is frequently drawn by persons not initiated into the construction of graphs. They argue like this: "For $x = 1$, $y = 1$. For $x = 2$, $y = \frac{1}{2}$. For $x = 3$, $y = \frac{1}{3}$. For $x = -1$, $y = -1$. For $x = 0, \cdots$? It's unclear. . . . It is not known what $1/0$ means, and therefore we omit $x = 0 \cdots$."

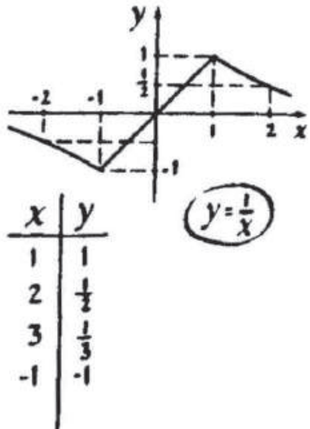


Fig. 1

The reader knows from the preceding text that graphs must not be drawn in this manner. In order to get a correct picture, let us note first that at $x = 0$ the function is not defined. In such cases it is interesting to see how the function behaves near this point. When x, decreasing in absolute value, approaches zero, then y becomes as large in absolute value as we please. If x approaches zero from the right ($x > 0$), then $y = 1/x$ is also positive. Therefore, approaching zero from the right, the curve of the graph moves up (Fig. 2a). If x approaches zero from the left ($x < 0$),

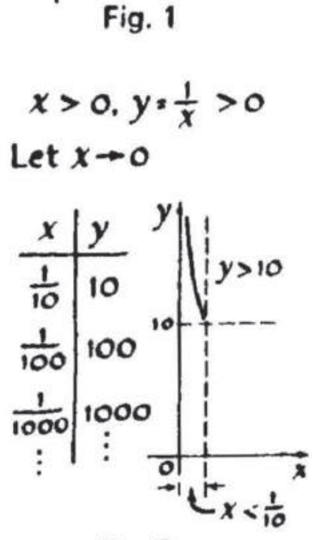
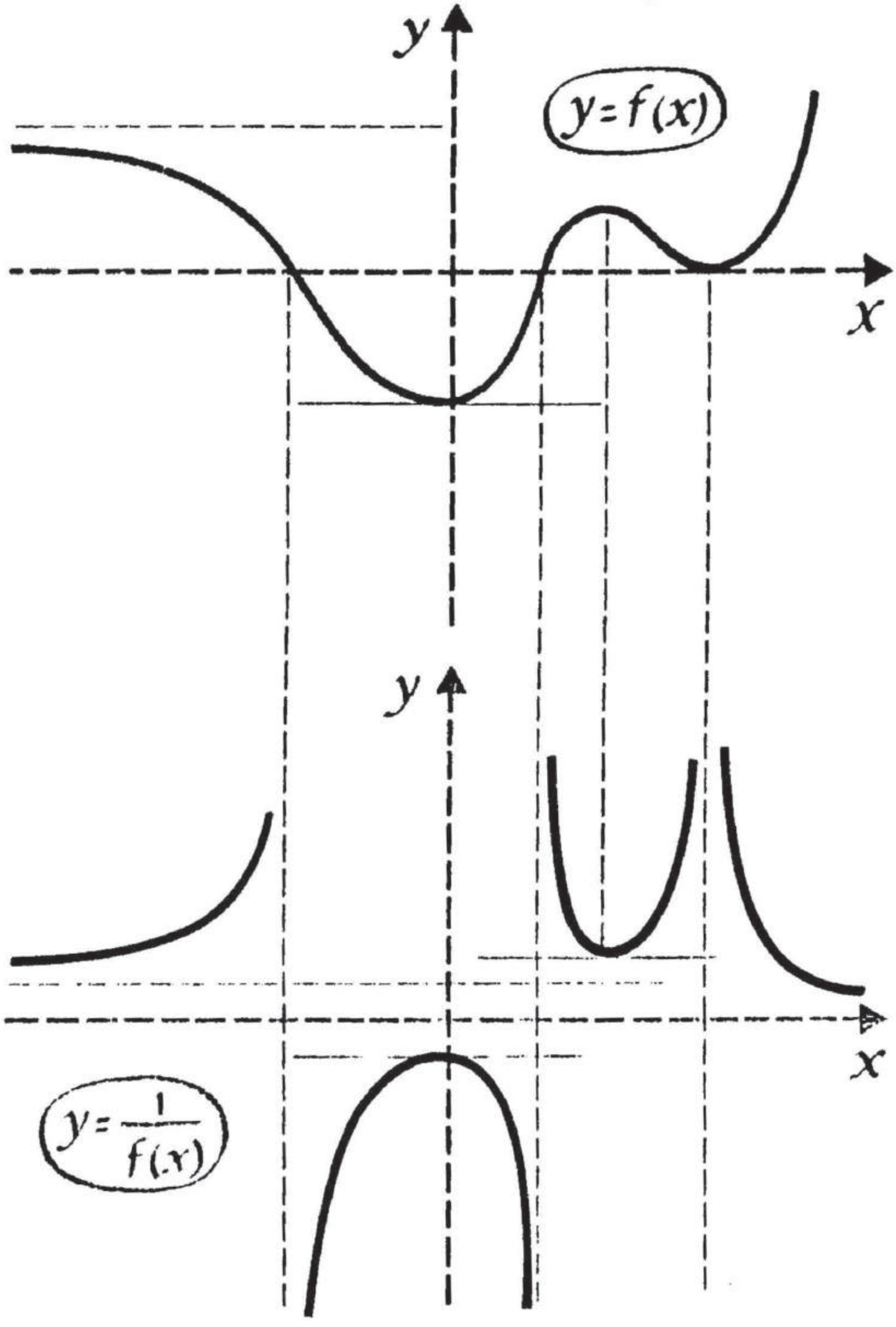
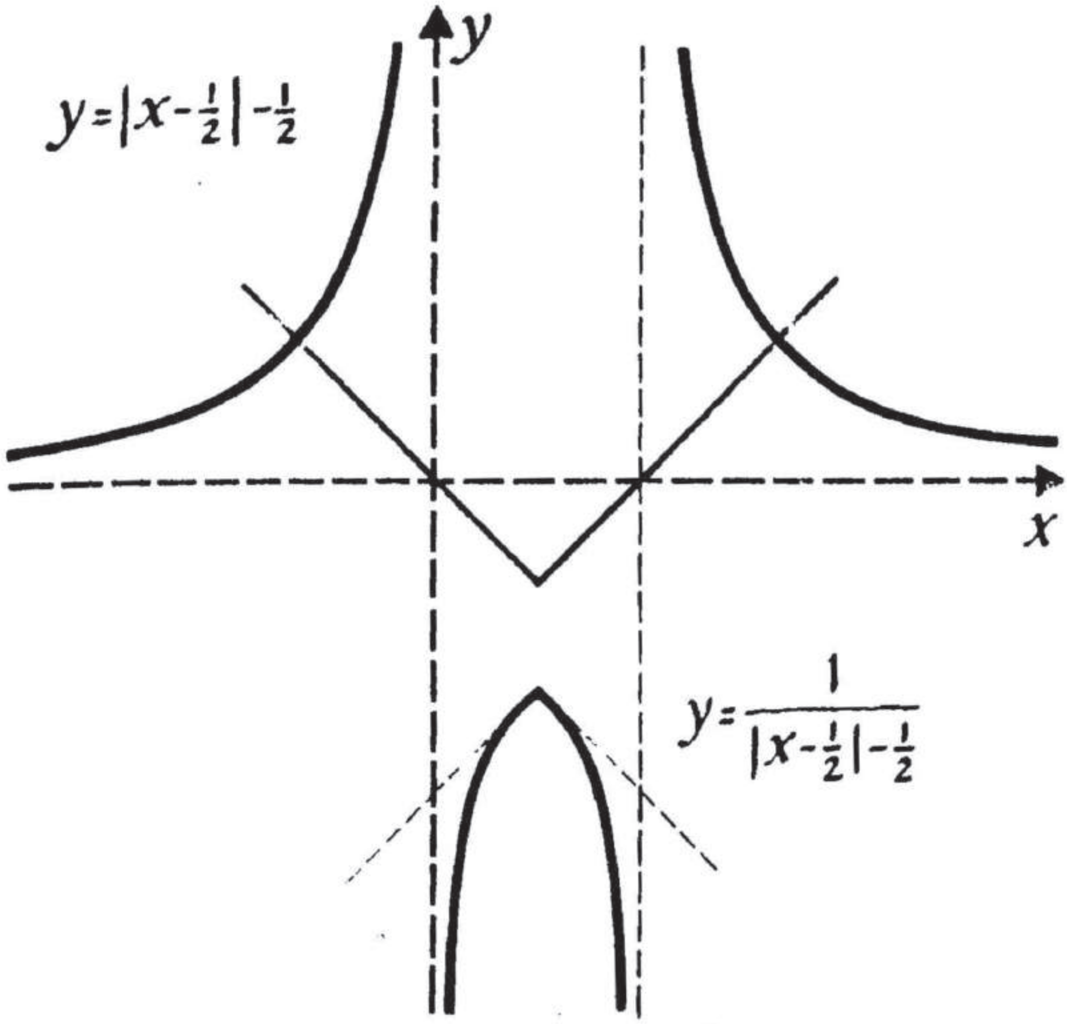


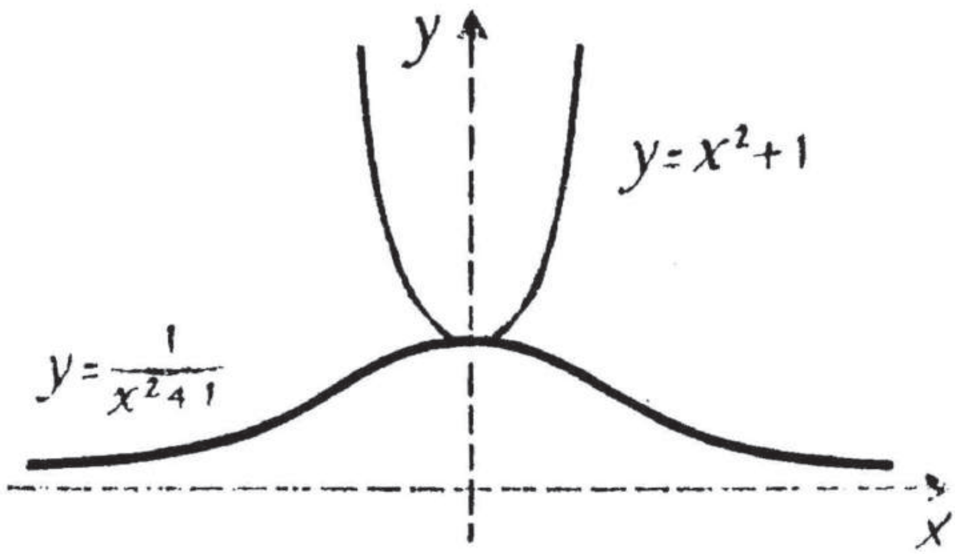
Fig. 2a

$$y = f(x) \rightarrow y = \frac{1}{f(x)}$$





$$y=(x)\to y=\frac{1}{f(x)}$$



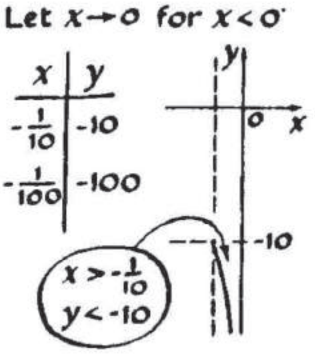


Fig. 2b

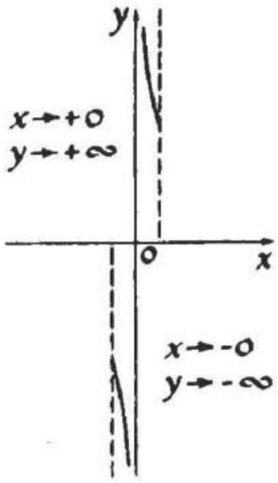


Fig. 3

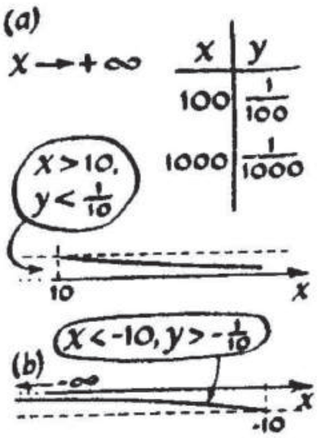


Fig. 4

then y is negative, and therefore from the left the graph moves down (Fig. 2b).

It is now obvious that near the "forbidden" value, $x = 0$ the graph, having separated into two branches, diverges along the y-axis: the right branch goes up, while the left goes down (Fig. 3).

Let us find out now how the function behaves if x increases in absolute value. First let us consider the right branch, that is, values of $x > 0$. For positive x the values of the function y are also positive. This means that the entire right branch is above the x-axis. As x increases, the fraction $1/x$ decreases. Therefore, in moving from zero to the right, the curve $y = 1/x$ drops lower and lower, and it can approach the x-axis to within an arbitrarily small distance (Fig. 4a). For $x < 0$ an analogous picture is obtained (Fig. 4b).

Thus as x increases indefinitely in absolute value, the function $y = 1/x$ decreases indefinitely in absolute value and both branches of the graph approach the x-axis: the right-hand one from above, the left from below (Fig. 5).

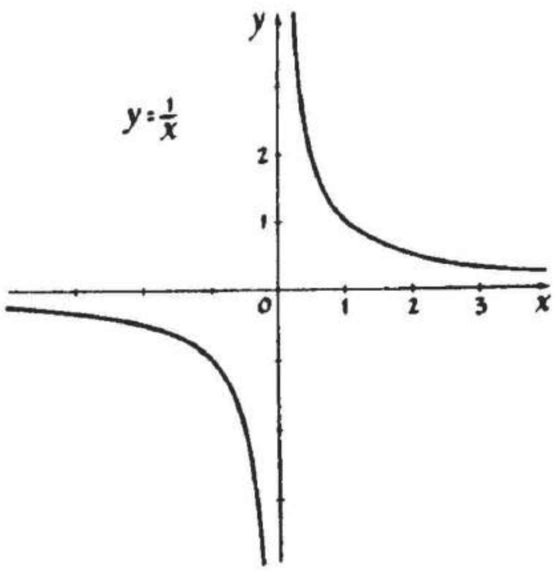


Fig. 5

The curve that is the graph of $y = 1/x$ is called a *hyperbola*. The straight lines that the branches of the hyperbola approach are called its *asymptotes*.

2

The graph of $y = 1/x$ can be constructed somewhat differently.

Let us draw the graph of the function $y = x$ (Fig. 6a). Let us replace each ordinate by its inverse and draw the corresponding points in Fig. 6b. We thus obtain the graph of $y = 1/x$.

The picture we have drawn clearly shows how small ordinates of the first graph transform into large ordinates of the second, and, on the other hand, large ordinates of the first transform into small ordinates of the second.

This method of "dividing" graphs is useful whenever we can construct the graph of $y = f(x)$, and we have to construct the graph of the function $y = 1/f(x)$ (see pp. 56 and 57).

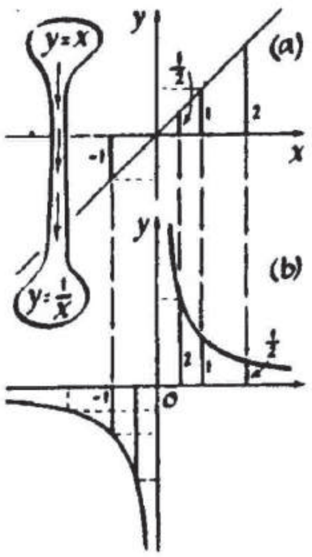


Fig. 6

EXERCISES

1. From the graph of $y = x^2$ construct the graph of $y = 1/x^2$. (Solution in Fig. 7.)

2. Construct the graphs of

$$(a)\ y = \frac{1}{x^2 - 3x - 2}; \quad (b)\ y = \frac{1}{x^2 - 2x + 3}.$$

(It will be seen that these two graphs look quite different.)

3. From the graph of $y = [x]$ (see p. 6) and $y = x - [x]$, construct graphs of

$$(a)\ y = \frac{1}{[x]}; \quad (b)\ y = \frac{1}{x - [x]}.$$

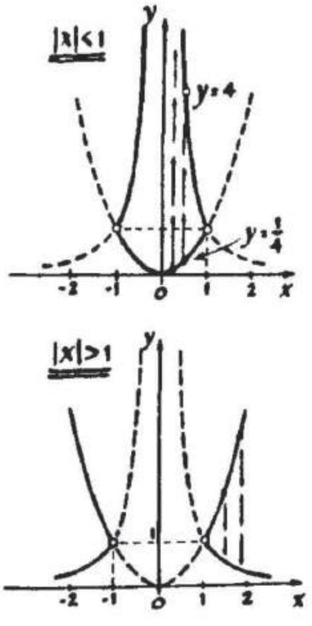


Fig. 7

3

The curves that you are asked to construct in the following exercises are obtained from the hyperbola $y = 1/x$ by transformations with which you are already familiar. All of them are also called hyperbolas.

EXERCISES

1. Draw the graphs of the functions:

(*a*) $y = 1/x + 1$;

(*b*) $y = 1/(x + 1)$;

(*c*) $y = 1/(x - 2) + 1$.

Indicate which are the asymptotes of each of these hyperbolas.

2. (*a*) Prove that the straight lines $y = x$ and $y = -x$ are axes of symmetry of the hyperbola $y = 1/x$.

(*b*) Does the right-hand branch of the graph of $y = 1/x^2$ have an axis of symmetry? $\oplus$

3. From the graph of $y = 1/x$ construct the graph of $y = 4/x$. Does this curve have any axes of symmetry?

The graphs of functions of the form

$$y = \frac{b}{cx + d} \qquad \text{(where } c \neq 0 \text{ and } b \neq 0\text{)}$$

can be obtained from the graph of $y = 1/x$ by translation along the x-axis and stretching along the y-axis. In order to determine the correct value of the translation and the stretching ratio, it is necessary to divide the numerator and denominator of the fraction by c, the coefficient of x:

$$\frac{b}{cx + d} = \frac{b/c}{x + (d/c)}.$$

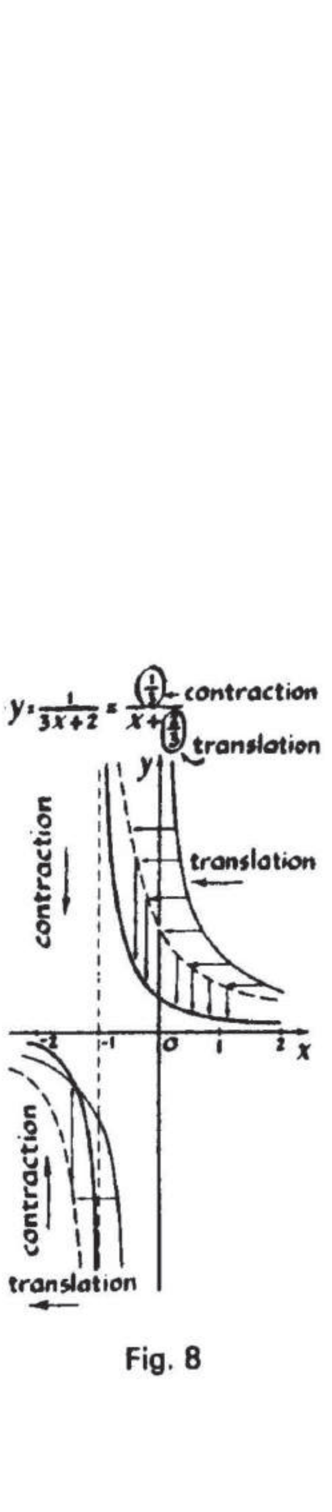


Fig. 8

Let us do this for the example $y = 1/(3x + 2)$. We have (Fig. 8)

$$\frac{1}{3x + 2} = \frac{\frac{1}{3}}{x + \frac{2}{3}}.$$

It is now obvious that the graph of our function $y = 1/(3x + 2)$ is the graph of $(1/x)$, translated by $(-\frac{2}{3})$* along the x-axis and contracted along the y-axis in the ratio 3:1 (Fig. 8).

*Not by (-2), as some students would say rashly.

EXERCISE

Draw the graph of

$$y = \frac{1}{2 - x} + 1.$$

(**Hint.** Transform the fraction $1/(2 - x)$ as already stated: Divide numerator and denominator by the coefficient of x, that is, by (-1), to obtain

$$y = \frac{-1}{x - 2} + 1.)$$

4

The graphs of functions of the form

$$y = \frac{ax + b}{cx + d},$$

called *linear fractional functions*, do not differ in form from the graph of $y = 1/x$. We assume, of course, that $c \neq 0$ (otherwise the linear function $y = (a/d)x + b/d$ is obtained) and that $a/c \neq b/d$; that is, the numerator is not a multiple of the denominator (as with the function

$$y = \frac{4x + 6}{2x + 3}\Big),$$

otherwise the function is constant.

Let us prove this. We first consider the example: $y = (2x + 1)/(x - 3)$. Let us separate the "integral part" of the fraction, dividing the numerator by the denominator (Fig. 9). We obtain

$$\frac{2x + 1}{x - 3} = 2 + \frac{7}{x - 3}.$$

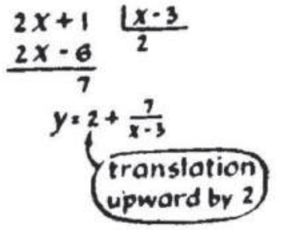
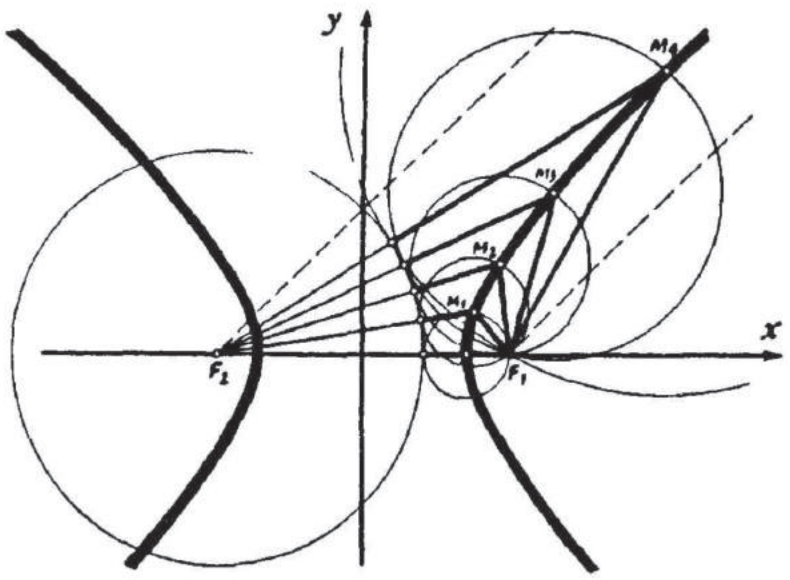
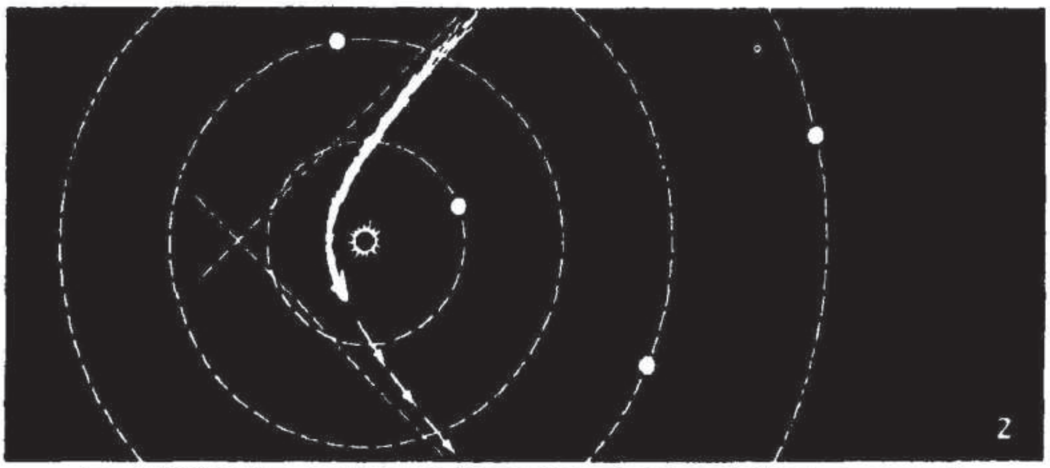
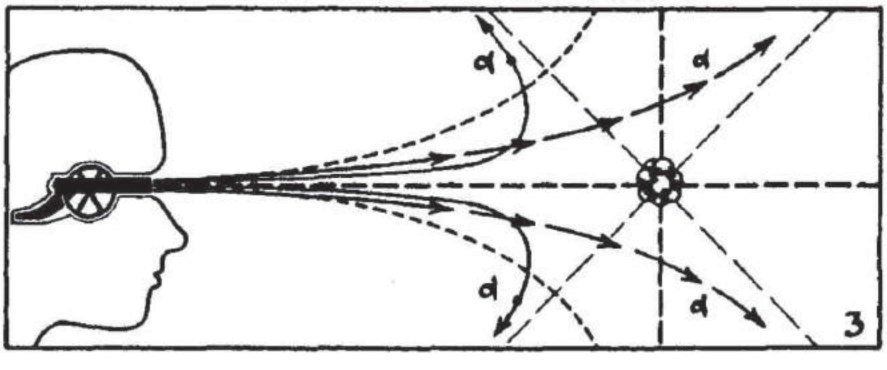


Fig. 9

The graph of this function is evidently obtained from the graph of $y = 1/x$ by the following transformations: a translation by 3 units to the right, a stretching in the ratio 7:1 along the y-axis, and a translation by 2 units upward.







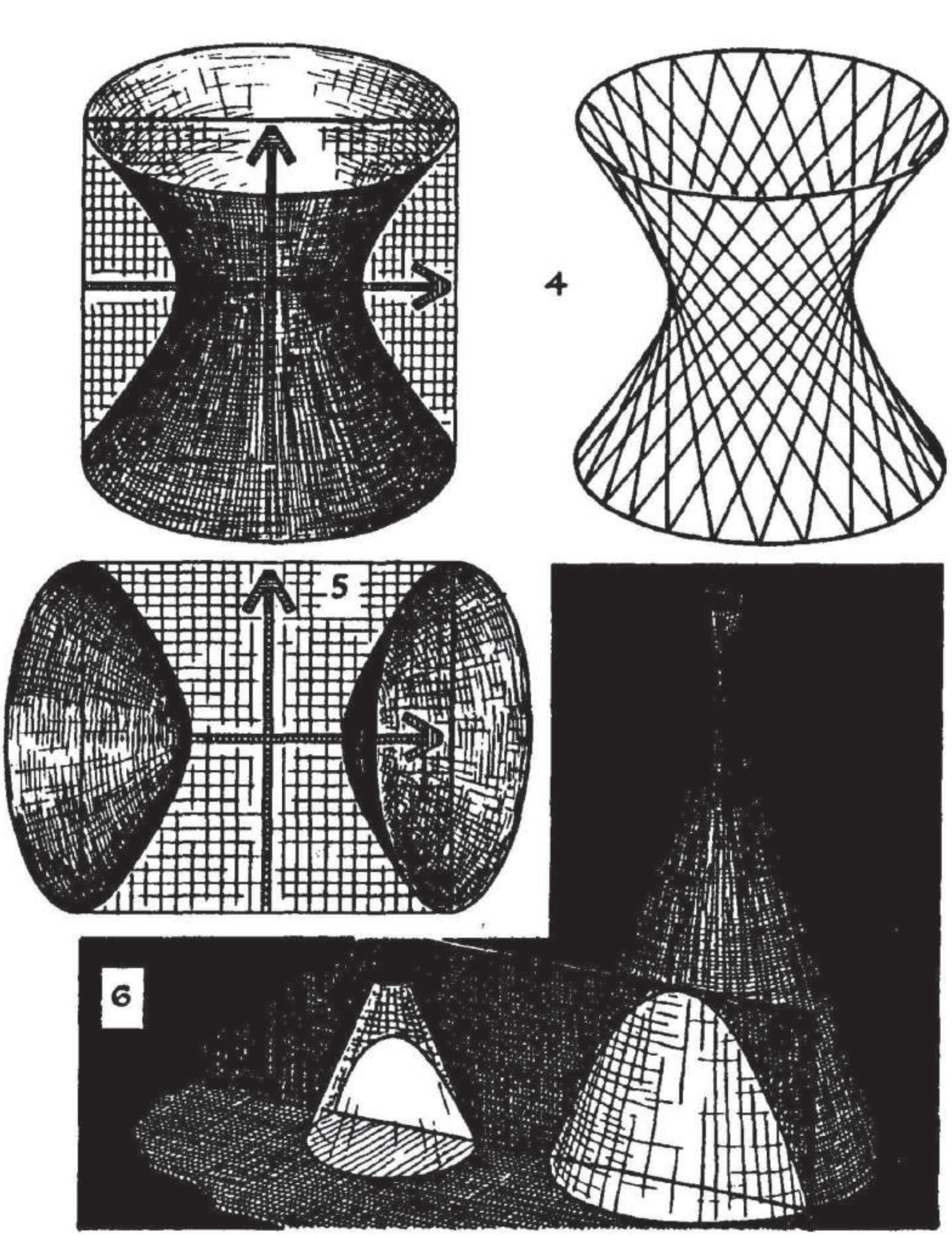
Interesting Properties of the Hyperbola

1. The hyperbola is the locus of all points M, the difference of whose distances from two given points F_1 and F_2, called foci, is equal in absolute value to a given number.

2. A comet or meteorite traveling into the solar system from a great distance moves on a branch of a hyperbola, with the sun in its focus. One asymptote* gives the direction in which the comet approaches, and the second asymptote gives the direction in which it leaves the solar system.

3. In the bombardment of an atomic nucleus, an α-particle, flying past the nucleus, travels in a hyperbola.

*Every hyperbola has two asymptotes. The hyperbolas that are graphs of the linear fractional function $y = (ax + b)/(cx + d)$ have mutually perpendicular asymptotes. Other hyperbolas have asymptotes intersecting each other at a different angle.



4. If a hyperbola is rotated about its axis of symmetry that does not intersect its branches, a surface is obtained called a *hyperboloid of one sheet*. This surface has a striking property: it is "woven" from straight lines. The tower of the Moscow telecenter is composed of "pieces" of such hyperboloids, made entirely from straight steel rods.

5. If a hyperbola is rotated about the other axis of symmetry, a surface consisting of two "pieces," the *hyperboloid of two sheets* is obtained. It was this hyperboloid that A. Tolstoi had in mind in his novel, *The Hyperboloid of Engineer Garin*. But then, the property required by engineer Garin, to collect light rays in a parallel pencil, is actually not possessed by the hyperboloid but the paraboloid, so that it would be more correct to call the book *The Paraboloid of Engineer Garin*.

6. If an infinite cone is appropriately intersected with a plane, then at the intersection a hyperbola is obtained. If the reader has a lamp with a shade in the form of a circular cone, he can satisfy himself of the truth of this: the lamp illuminates a part of the wall, which is bounded by a piece of a hyperbola.

Any fraction $y = (ax + b)/(cx + d)$ can be written in an analogous way, separating its "integral part." Consequently, the graphs of all linear fractional functions $y = (ax + b)/(cx + d)$ are hyperbolas (translated different distances along the coordinate axes and stretched in different ratios along the y-axis).

Remark. To construct the graph of some linear fractional function, the fraction defining this function does not have to be transformed. Since we know that the graph is a hyperbola, it suffices to find the straight lines which its branches approach (the asymptotes of the hyperbola) and a few more points.

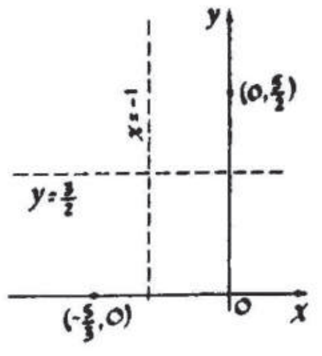


Fig. 10

Example. Let us construct the graph of the function

$$y = \frac{3x + 5}{2x + 2}.$$

Let us first find the asymptotes of this hyperbola. The function is not defined where $2x + 2 = 0$, that is, at $x = -1$ (Fig. 10). Consequently, the straight line $x = -1$ serves as vertical asymptote.

In order to find the horizontal asymptote, let us find what number the values of the function approach as the argument increases in absolute value. For large (in absolute value) values of x,

$$y = \frac{3x + 5}{2x + 2} \approx \frac{3x}{2x} = \frac{3}{2}.$$

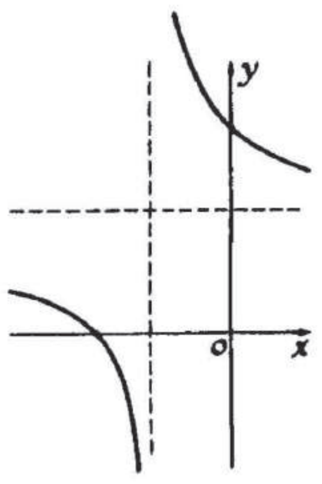


Fig. 11

Consequently, the horizontal asymptote is the straight line $y = \frac{3}{2}$.

Let us determine the points of intersection of our hyperbola with the coordinate axes. At $x = 0$ we have $y = \frac{5}{2}$. The function equals zero when $3x + 5 = 0$, that is, when $x = -\frac{5}{3}$.

Having marked the points $(-\frac{5}{3}, 0)$ and $(0, \frac{5}{2})$ in the diagram, we construct the graph (Fig. 11).

EXERCISES

1. Construct graphs of the functions:

$$(a)\ y = \frac{1}{1-2x}; \qquad (b)\ y = \frac{3+x}{3-x};$$

$$(c)\ y = \left|\frac{2x+1}{x+1}\right|.$$

2. Figures 12a and b represent the graphs of linear fractional functions $y = (px+q)/(x+r)$. Find these functions (determine p, q, and r).

3. (*a*) How many solutions has the equation

$$\frac{x}{1-x} = x^2 + 4x + 2?$$

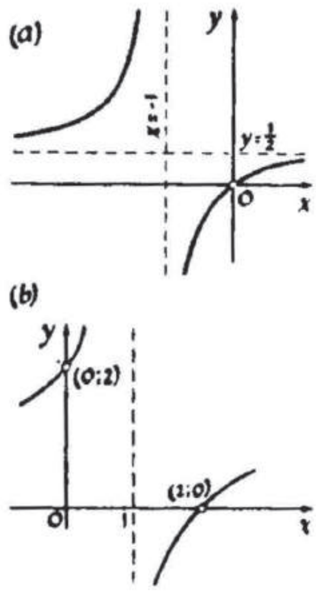


Fig. 12

Solution. Let us construct in one diagram the graphs of the functions

$$y = \frac{x}{1-x} \qquad \text{and} \qquad y = x^2 + 4x + 2.$$

In Fig. 13 two points of intersection of these graphs can be seen. Evidently, there is also a third point, since the parabola intersects the asymptote of the hyperbola. The abscissas of the points of intersection of the graphs are the solutions of the equation.

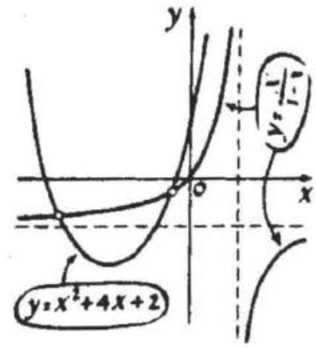
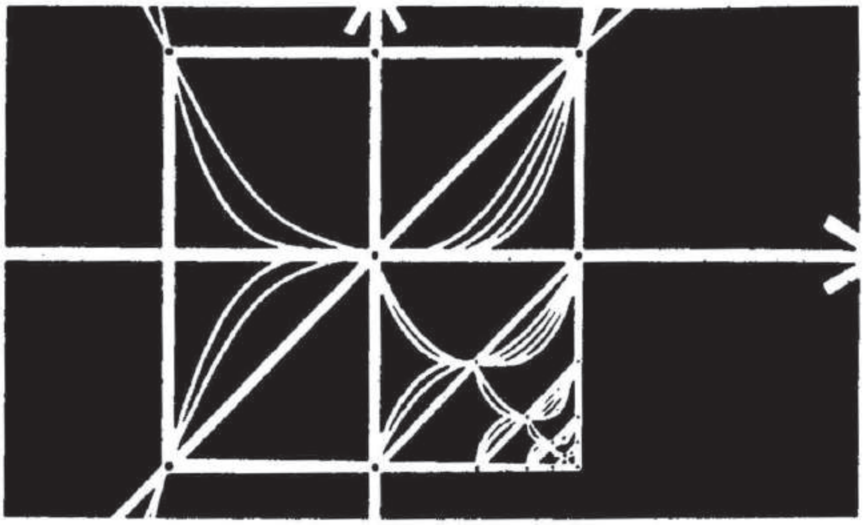


Fig. 13

(**Answer.** Three solutions.)



CHAPTER 6

Power Functions

1

Power functions are functions of the form $y = x^n$. We have already constructed graphs of the power functions for $n = 1$ and $n = 2$. For $n = 1$ we obtain the function $y = x$; the graph of this function is a straight line (Fig. 1a). For $n = 2$ the function $y = x^2$ is obtained; the graph of this function is a parabola (Fig. 1b).

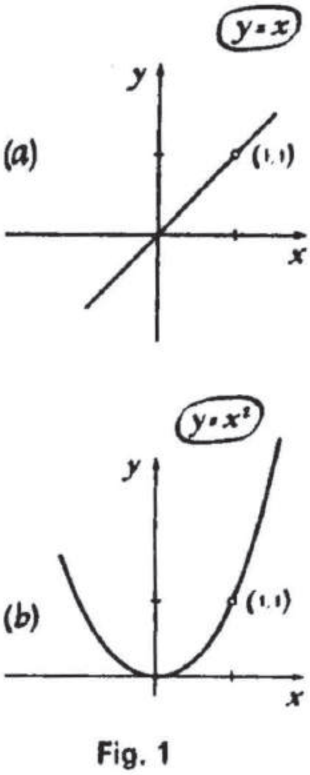


Fig. 1

The graph of the function $y = x^3$ ($n = 3$) is also called a parabola — a *parabola of the third degree* or *cubic parabola*. For positive values of the argument the cubic parabola $y = x^3$ is similar to the parabola of the second degree $y = x^2$. In fact, at $x = 0$ the function $y = x^2$ equals zero and the function $y = x^3$ also equals zero; both graphs pass through the origin; at $x = 1$ the value of x^2 equals 1 and that of x^3 also equals 1; and both graphs pass through the point (1, 1).

As x increases (if x is positive), the values of the function $y = x^2$ as well as the values of the function $y = x^3$ increase. To the right of the origin the cubic parabola $y = x^3$, like the ordinary parabola $y = x^2$, steadily rises (Fig. 2).

For negative values of x the behavior of the curve $y = x^3$ is different from that of the curve $y = x^2$: for negative x the values of x^3 are also negative, therefore the curve heads downward (Fig. 3). Thus, on the whole, the cubic parabola is altogether different from the quadratic one.

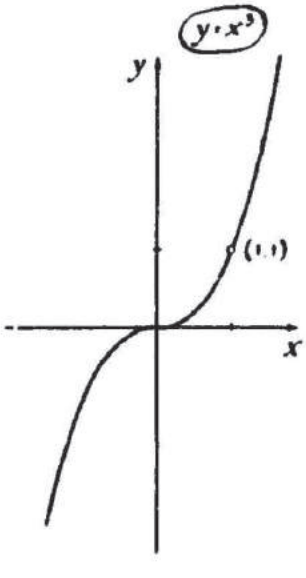


Fig. 2

The left half of the graph of $y = x^3$ can be obtained from its right half by using symmetry, of a different kind though from that which we considered on pages 10 and 11. Let us take any point M on the right half of the graph of $y = x^3$ (Fig. 3). If a is the abscissa of this point, then its ordinate b equals $a^3 (b = f(a) = a^3)$. Let us now find the point of the graph that corresponds to the value of the abscissa with opposite sign, $x = -a$. The ordinate of such a point equals $(-a)^3$, that is, $-a^3$, or $-b$. Thus for each point $M(a, b)$ on the right half of the graph of $y = x^3$ there is a point $M'(-a, -b)$ on its left half. Obviously (Fig. 3), the point M' is symmetric to the point M with respect to the origin. Hence the entire left half of the graph can be obtained from the right half by symmetric reflection with respect to the origin of coordinates.

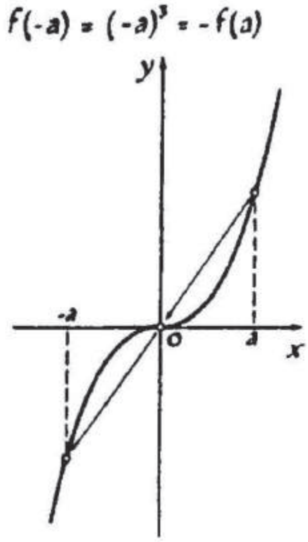


Fig. 3

EXERCISES

1. Which of the graphs of the following functions have a center of symmetry, which have an axis of symmetry:

$$y = x^4; \qquad y = x^5; \qquad y = x^7; \qquad y = x^{16}?$$

2. Prove that the graph of the function $y = 1/x^3$ is symmetric with respect to the origin.

Solution. Let us consider the two points of the curve $y = 1/x^3$ with abscissas $x = a$ and $x = -a$. The ordinate of the first equals $1/a^3$, the ordinate of the second $1/(-a)^3 = -1/a^3$.

Consequently, for each point $M(a, 1/a^3)$ of our curve there is a point $M(-a, -1/a^3)$ symmetric to the first with respect to the origin. Therefore, the whole curve $y = 1/x^3$ is symmetric with respect to the origin.

3. Which of the following functions are even and which are odd:*

$$y = x^3|x|; \qquad y = |x^3| + x; \qquad y = \frac{x}{|x|};$$

$$y = |x - x^2|; \qquad y = (2x + 1)^4 + (2x - 1)^4;$$

$$y = \frac{1}{|2x - x^2|} - \frac{1}{|2x + x^2|};$$

$$y = (x^3 + 1)^2; \qquad y = (x^2 + 1)^3;$$

$$y = \frac{1}{x + \dfrac{1}{x + (1/x)}};$$

$$y = (3 - x)^5 - (3 + x)^5?$$

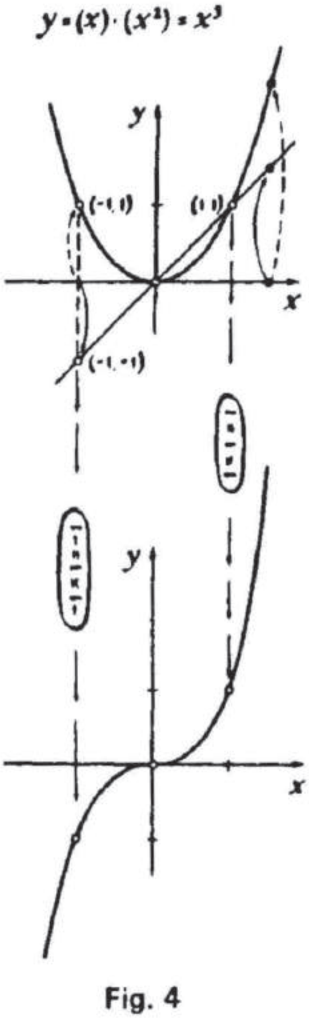


Fig. 4

2

Let us now see how the graphs of $y = x^3$ and $y = x^2$ differ from each other for positive values of x. To this end let us represent x^3 as $x^2 \cdot x$ and obtain the graph of $y = x^3$, by "multiplying" the graph of $y = x^2$ by the graph of $y = x$ (Fig. 4). At $x = 1$ the value of x^3 equals the value of x^2: the point $(1, 1)$ is common to both graphs. Let us now go to the right and to the left from this point.

To the right of the point $(1, 1)$ the values of the function $y = x^3$ are obtained from the values of the function $y = x^2$ by multiplication by numbers larger than one. Therefore for $x > 1$ the values of x^3 are larger than x^2 — to the right of the point $(1, 1)$ the graph of the cubic parabola $y = x^3$ lies above the graph of the parabola $y = x^2$, and as the value of x

*The definition of an even function is given on p. 11, the definition of an odd function is on p. 72.

increases, so does the difference between the two functions.

If one goes from the point (1, 1) to the left, toward $x = 0$, the values of x^3 will be obtained from the values of x^2 by multiplication by a number less than one. Therefore to the left of the point (1, 1) the cubic parabola lies below the parabola $y = x^2$, approaching the x-axis at the origin more quickly (Fig. 5).

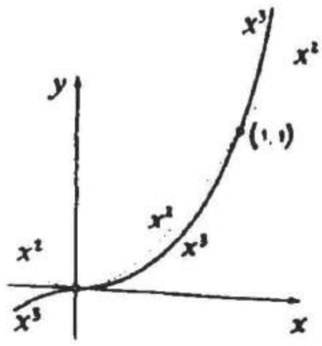


Fig. 5

QUESTIONS

For what x will the value of x^3 be 100 times larger than the value of x^2? 1000 times? How much above the quadratic parabola will the cubic parabola be at these x? Will the cubic parabola rise at any x 100 units above the quadratic parabola? 1,000,000?

If the thickness of a pencil line is considered to be equal to 0.1 mm and the unit of measure is taken to be 1 cm, at $x = 0.1$ the parabola $y = x^2$ can no longer be distinguished from the x-axis. How many times closer to the x-axis is the cubic parabola at this value of x?

3

Obtaining x^3 from x^2 by multiplication by x, *how many times* is the ordinate of $y = x^3$ larger (or smaller) than the ordinate of $y = x^2$? Let us now try to represent graphically the difference between the values of the functions $y = x^3$ and $y = x^2$. To this end let us draw the graph of the function

$$y = x^3 - x^2,$$

whose ordinates can be obtained by subtracting from the ordinates of the graph of $y = x^3$ the ordinates of the graph of $y = x^2$ (Fig. 6).

Let us construct
$y = x^3 - x^2$

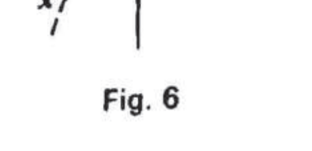


Fig. 6

At $x = 0$ both x^3 and x^2 vanish; consequently the graph of the function $y = x^3 - x^2$ passes through the origin. To the left of the origin the positive x^2 is subtracted from the negative x^3; the difference between $x^3 - x^2$ is negative so that the

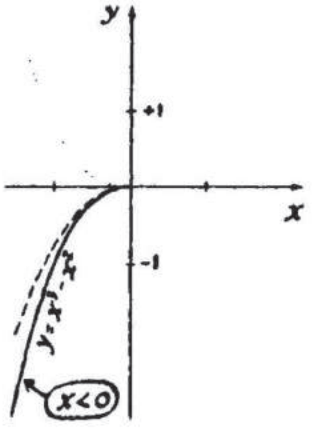


Fig. 7

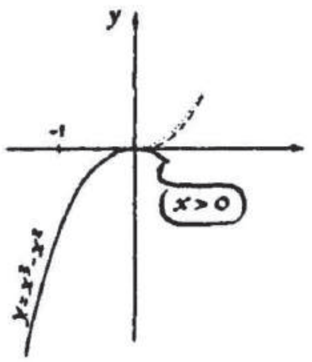


Fig. 8

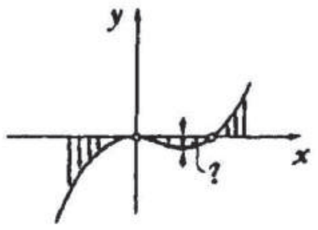


Fig. 9

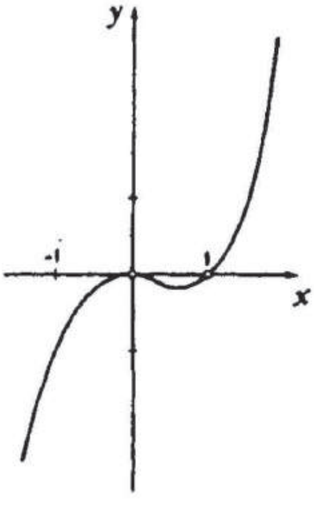


Fig. 10

graph of $y = x^3 - x^2$ runs below the x-axis (and even below the graph of $y = x^3$ (see Fig. 7).

To the right of the origin things are more complicated; both functions are positive, and the result depends on which is larger in absolute value: x^3 or x^2. At first x^2 is larger than x^3; therefore the curve $y = x^3 - x^2$ is situated below the x-axis near the origin (Fig. 8). Gradually x^3 starts to grow faster and faster and at $x = 1$ catches up with the function $y = x^2$ in value. Therefore (somewhere between $x = 0$ and $x = 1$) the curve $y = x^3 - x^2$ starts to rise and at $x = 1$ intersects the x-axis (Fig. 9).

Further, after $x = 1$ the values of the function $y = x^3 - x^2$ increase, the graph moves up and for large x, when x^2 is small in comparison with x^3, is almost indistinguishable in its shape from the graph of $y = x^3$ (Fig. 10).

EXERCISES

1. It is possible to determine approximately at which x the values of the function $y = x^3 - x^2$ start to increase. Try to find this value of x with an accuracy of at least 0.1 (that is, one digit after the decimal point). We shall later be able to find the exact value. Exactly at this spot is also located the lowest point of the dip of the graph.
2. Solve the inequalities:

$$x^3 - x^2 > 0; \qquad x^3 - x^2 \leq 0.$$

Let us now compare the behavior of the functions $y = x^3$ and $y = cx^2$ and construct the graph of the function $y = x^3 - cx^2$. Let us first take a small value of c, for example, $c = 0.3$. The appearance of the graph depends on how the graph of $y = x^3$ and the graph of $y = 0.3x^2$ are situated with respect to each other. In Fig. 11 it is easy to understand that the construction of the graph far away from the origin, that is, for values of x of large absolute value, does not cause any difficulties. However, from the diagram it cannot be determined which of the parabolas $y = x^3$ and $y = 0.3x^2$ lies below the other near the origin, yet the answer to this question determines whether there will be a dip in the resulting graph or not.

In order to clarify this question, let us solve the inequality:

$$x^3 > 0.3x^2, \qquad \text{or} \qquad x^2(x - 0.3) > 0.$$

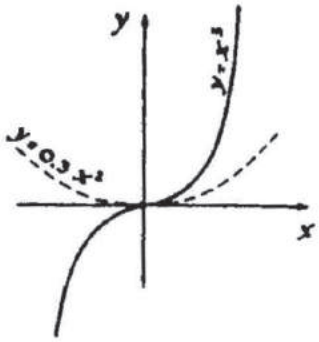


Fig. 11

It is now clear that close to the origin, namely for positive values of x less than 0.3, the cubic parabola lies below the parabola $y = 0.3x^2$ (Fig. 12a). Therefore we can now draw the part of Fig. 11 which was not clear to us before in magnified form and construct the graph of the difference

$$y = x^3 - 0.3x^2.$$

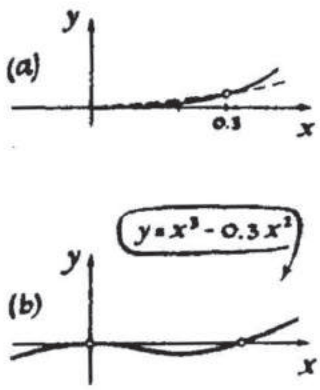


Fig. 12

In this graph, as in the graph of

$$y = x^3 - x^2,$$

there will be a dip, but a narrower one (Fig. 12b).

EXERCISES

1. Find the width of the "dip" for the graphs of the functions:

(*a*) $y = x^3 - 0.01x^2$; (*b*) $y = x^3 - 1000x^2$.

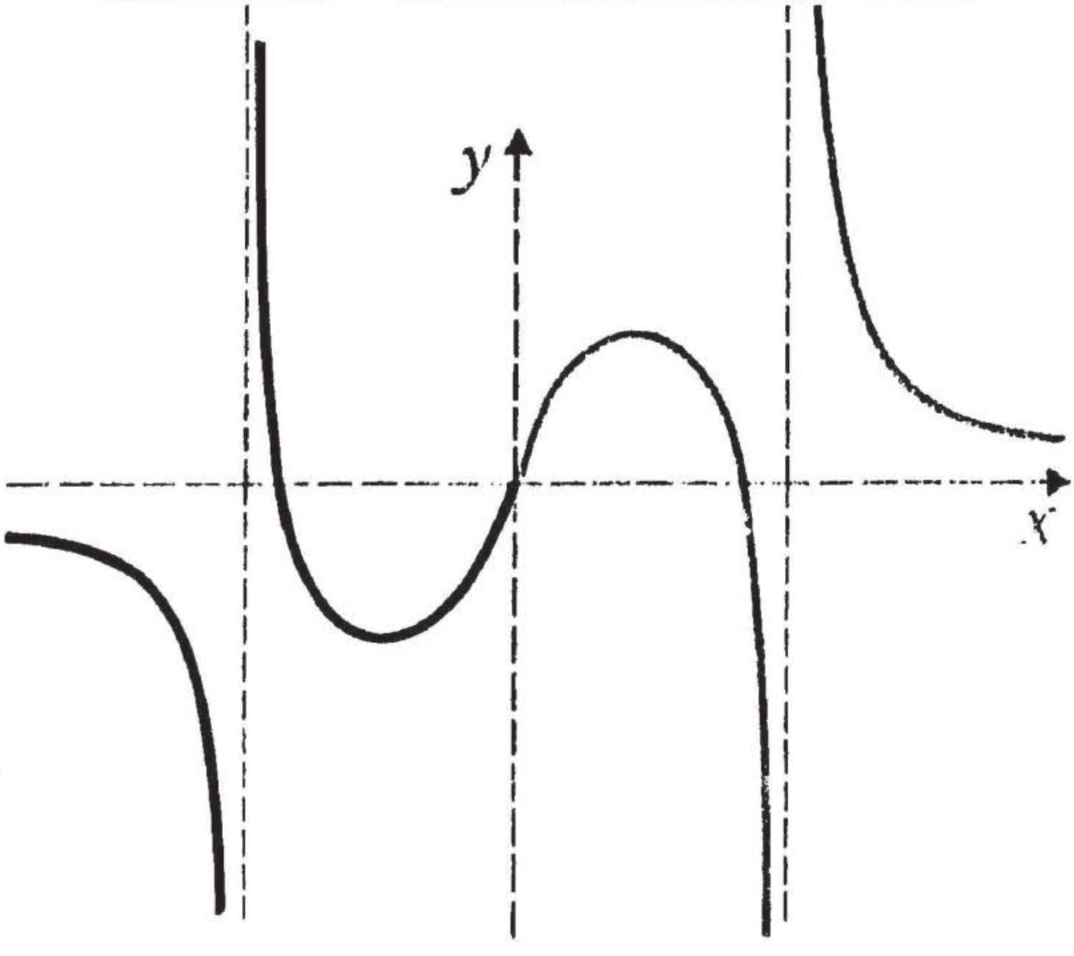
2. Will there be a dip in the graph of

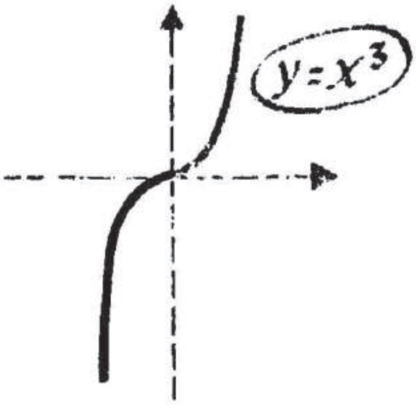
$$y = x^3 + 0.001x^2?$$

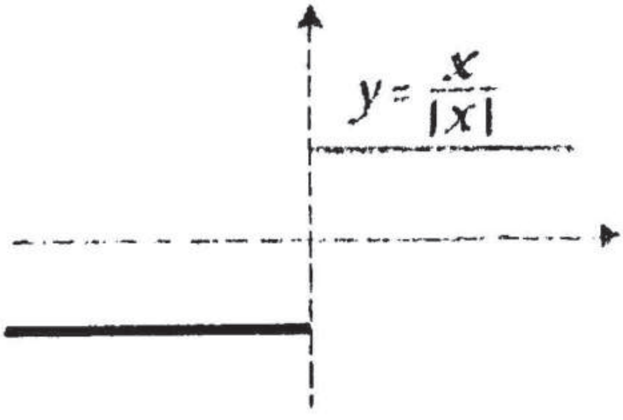
3. Show after which x the parabola $y = x^3$ will lie above the parabola $y = 50x^2$; the parabola $y = 10{,}000x^2$.

Having done these exercises, you will understand that the graphs of the functions $y = x^3 - cx^2$ for any $c > 0$ have the same character, the same form: to the left of the origin the graph goes downward,

$f(-a) =$ $= -f(a)$	*The function $y = f(x)$ is called odd if for each a the equation $f(-a) = -f(a)$ is satisfied.* *The graph of an odd function is symmetric with respect to the orgin.*







at the origin it touches the x-axis, farther on it turns downward again and then up; the dip in the graph that is obtained as a result becomes more pronounced as c becomes larger (Fig. 13a).

If we gradually diminish c, the dip will gradually flatten out and finally, when c becomes equal to zero, will disappear and the graph will turn into the ordinary cubic parabola $y = x^3$ (Fig. 13b).

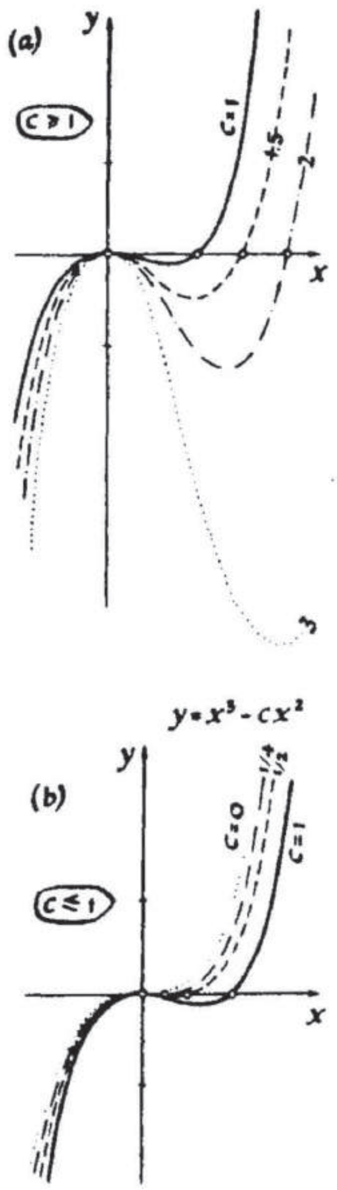


Fig. 13

4

We can now draw a general conclusion as to how the function $y = x^3$ behaves for positive values of x in comparison with any function of the form $y = cx^2$. For x close to zero, the function $y = x^3$ will be less than any function $y = cx^2$, even if the coefficient c is very small. For larger values of x, on the other hand, the function $y = x^3$ will be larger than any function of the form $y = cx^2$, even if the coefficient c is very large.

This can be expressed differently: the parabola of the third degree approaches the x-axis at the origin so quickly that between the parabola and the x-axis there can pass not only no straight line but also not one parabola $y = cx^2$, however small the coefficient c. On the other hand, for large values of x (for $x > 0$) the parabola $y = x^3$ "outruns" any parabola $y = cx^2$ for any coefficient c, no matter how large.

EXERCISES

1. Draw graphs of the functions:

$$y = -x^3, \quad y = |x^3|; \qquad y = 1 + x^3;$$
$$y = (2 + x)^3;$$
$$y = (2 - x)^3; \qquad y = x^3 + 3x^2 + 3x.$$

2. Figure 14 represents the parabolas $y = 5x^3$ and $y = x^2$; because of the small scale of the diagram the relative position of the graphs near zero is not clear.

Look at this drawing "under a microscope" and draw what you see there: draw on a larger scale the region marked off by a small circle.

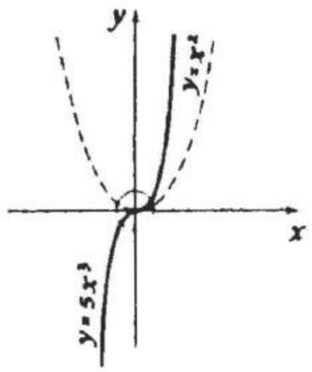


Fig. 14

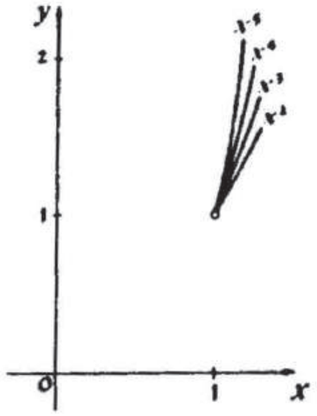


Fig. 15

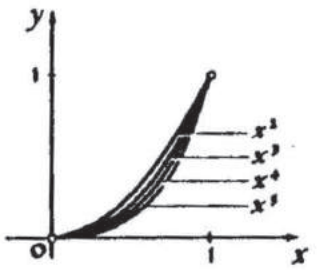


Fig. 16

5

We shall not analyze the functions $y = x^n$ for $n > 3$ as fully as $y = x^3$. The graphs of these functions remind one in their outward appearance either of the parabola $y = x^2$ (for n even), or of the parabola $y = x^3$ (for n odd).

It is understood that the function $y = x^4$ for large values (large in absolute value) of x grows still faster than $y = x^3$. Generally, the larger n is, the faster will grow the power function $y = x^n$ for large values of x (Fig. 15).

When x approaches zero, the values of all power functions $y = x^n$ also approach zero; the larger n is, the faster they will do so. The graphs of all power functions $y = x^n$ (starting with $n = 2$) touch the x-axis at the origin of coordinates, approaching it more rapidly the larger n is (Fig. 16).

For large n it is practically impossible to draw the graph of the function $y = x^n$ according to scale. On almost the whole segment from 0 to 1 the values of the function are very small and the graph of $y = x^n$ is indistinguishable from the x-axis. In a small region around $x = 1$, the function grows to 1 and then quickly grows at such a rate that the graph goes beyond the edges of any sheet of paper.* For example, let $n = 100$. Let us try to draw the graph of $y = x^{100}$, beginning with $x = 1$. At $x = 2$ we get $y = 2^{100}$. This is too big! Let us take $x = 1.1$. Then $y = (1.1)^{100}$. This is still a large number. In fact, $(1.1)^{100} = [(1.1)^{10}]^{10}$.

Let us now use the inequality $(1 + \alpha)^n > 1 + n\alpha$ (valid for $\alpha > 0$).† We obtain $(1.1)^{10} > 1 + 10 \times 0.1 = 2$. Thus, $(1.1)^{100} > 2^{10} > 1000$.

*For negative values of x the situation is analogous.

$$†(1 + \alpha)^n = (1 + \alpha)(1 + \alpha)\cdots(1 + \alpha)$$
$$= 1 + \underbrace{1.1\cdots 1}_{n-1}\cdot\alpha + \underbrace{1.1\cdots 1}_{n-1}\cdot\alpha + \cdots.$$

The terms not written down and indicated by $\cdots$ are all positive.

Thus the region from 1 to 1.1 is still too large for the construction of the graph of $y = x^{100}$. It is only in a region of length 0.01–0.02 that the values of the function do not differ from each other too greatly.

Let us now take the scale on the x-axis 100 times larger than on the y-axis. The graph of $y = x^{100}$ then is expanded in horizontal direction 100 times and will have the shape represented in Fig. 17a.

If n is still larger, it will be necessary to choose the region for accurate construction of the graph still smaller. In Fig. 17b, the reader can see the graph of $y = x^{1000}$ stretched 1000 times along the x-axis. It is remarkable that two identical curves were obtained!*

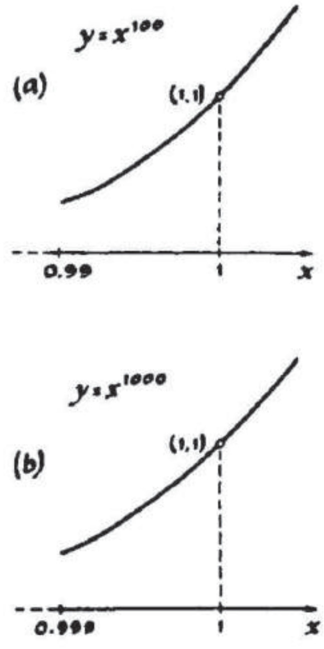


Fig. 17

EXERCISES

1. Construct the graph of the function $y = x^2 - x^4$ by two methods:

(a) subtraction of the graphs of $y = x^2$ and $y = x^4$;

(b) factoring the polynomial $x^2 - x^4$.

2. (a) There are given two increasing sequences

$$a_n: 0.001; 0.004; 0.009; \cdots,$$
$$b_n: 100; 300; 500; \cdots.$$

Can the first sequence catch up with the second (that is, can the inequality $a_n > b_n$ be satisfied for some n)?

(b) Answer the same question for the sequences

$$a_n: 0.001; 0.008; 0.027; \cdots,$$
$$b_n: 100; 400; 900; \cdots.$$

3. Determine how many solutions there are to the following equations:

(a) $x^3 = x^2 + 1$; (b) $x^3 = x + 1$;

(c) $x^3 + 0.1 = 10x$; (d) $x^5 - x - 1 = 0$.

4. (a) Draw graphs of the functions

$$y = x^3 - x; \qquad y = x^3 + x.$$

*Of course a difference between the curves exists; it is too small, however, to be observed in the graph.

(*b*) Select a and b so that in the graph of the function $y = ax^3 + bx$ there will be a dip 10 wide and not less than 100 deep.

6

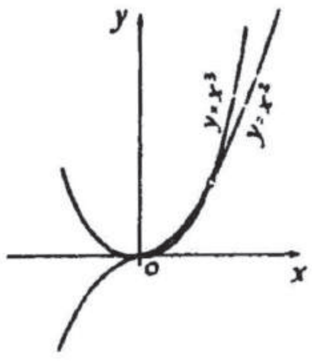


Fig. 18

While talking about the behavior of power functions $y = x^n$ for small values of the argument x, we noted that the graphs of these functions touch the x-axis at the origin. Let us dwell a little on this fact to clarify the exact meaning of the expression, a straight line touches a curve.

Actually, why do we say that the x-axis touches both the parabola $y = x^2$ and the cubic parabola $y = x^3$ (although the parabola $y = x^3$ is "pierced through" by the x-axis; Fig. 18)?

Why do we not consider the y-axis a tangent to the curve $y = x^2$, although it has only one point in common with it (see Fig. 18)?

Everything depends on what meaning is given to the expression "a straight line touches a curve," what is to be considered the basic defining property of a tangent, and what definition is to be given this concept. Up to now in the school geometry course we have become acquainted with only a single curve, the circle, and have learned only the definition of a tangent to a circle.

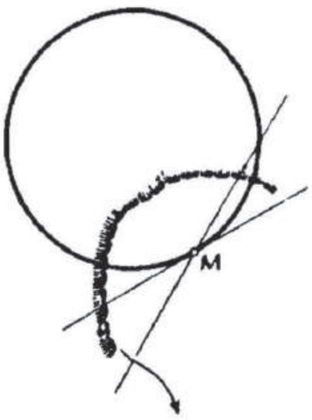


Fig. 19

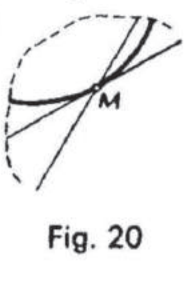


Fig. 20

Let us try to understand in what respects the tangent to a circle differs from a secant. The following circumstances immediately become apparent: (1) a tangent has only one point in common with the circle while a secant has two (Fig. 19); (2) in the neighborhood of the point of tangency M the tangent line "lies nearer" the curve than any secant line through M. (Therefore, even if only part of the circle is drawn in the figure and we do not see whether in fact there is a second common point, we could still distinguish a tangent from a secant, Fig. 20).

Which of these two conditions is to be considered the more important and upon which should be based the definition of a tangent to any curve in general and not only to a circle?

The first is simpler. It is possible to try to call a straight line having but one point in common with the curve a tangent.

For instance, through the vertex of the parabola $y = x^2$, beside the x-axis, there passes still another straight line having but one point in common with the parabola — this is the y-axis; the y-axis, however, is not called a tangent to the parabola $y = x^2$ (Fig. 18).

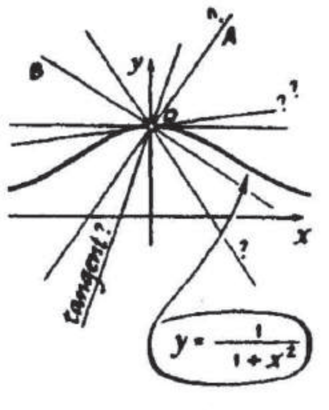
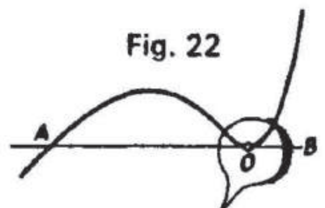


Fig. 21

In Fig. 21 the situation is still worse. All straight lines lying within the angle AOB, intersect the curve only once! On the other hand, in Fig. 22 the straight line AB has two common points with the curve, yet, obviously has every right to be called a tangent. In fact, if having "clipped" the drawing, one looks at the region close to the point of contact O, the situation of the curve relative to the straight line will be of precisely the same character as the situation of the parabola relative to the x-axis (Fig. 23).

Fig. 22


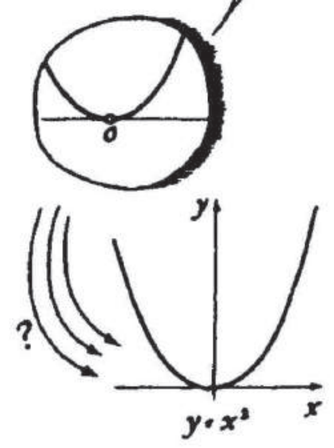


Fig. 23

Consequently, it is reasonable to choose as a basic defining property of a tangent the fact that the tangent attaches itself closely to the curve.

Thus, for instance, it is natural to consider that the cubic parabola $y = x^3$ is tangent to the x-axis at the origin: for the parabola $y = x^3$ attaches itself closely to the x-axis at the origin (still more closely than the parabola $y = x^2$).

In order to give the definition of a tangent we must formulate what is meant by a straight line "closely attaching itself" to a curve. Let us examine again the parabola $y = x^2$. Between the x-axis (the straight line $y = 0$) and the parabola there cannot pass a single straight line: any straight line $y = kx$ (for $k > 0$) in some region runs above the parabola and intersects it once more.

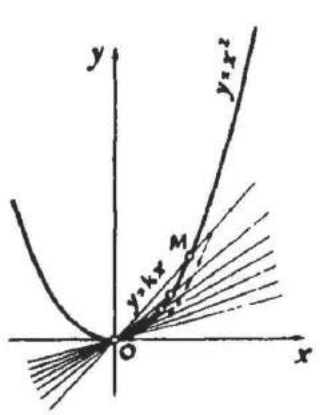


Fig. 24

If, rotating the straight line clockwise, we diminish k, the second point (the point M in Fig. 24) approaches the first (the point O) and finally coincides with it. At this moment the straight line turns from a secant into a tangent.

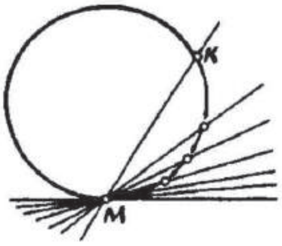


Fig. 25

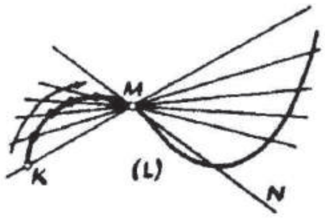


Fig. 26

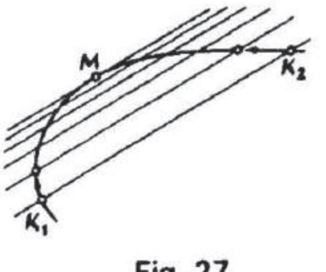


Fig. 27

Fig. 28

The same can be seen in any case of tangency. In Fig. 25 the secant MK is drawn through the point M of the circle. If the point K is drawn nearer to M, the secant will turn about the point M and, finally when the point K coincides with the point M, will turn into a tangent to the circle at the point M. Then it will not have any other points in common with the circle. This circumstance, however, is nonessential and of secondary importance.

Thus we adopt the following definition of a tangent.

Definition. Suppose there is some curve (L) and a point M on this curve (Fig. 26). Let us draw a straight line MK through M and any other point K of the curve (L) (such a straight line, passing through any two points of the curve, is called a secant; it may also intersect the curve in some other points). If the point K is now moved along the curve (L) so as to approach the point M, the secant MK will turn about the point M. If finally, when the point K coincides with the point M, the straight line coincides with some definite straight line MN (Fig. 26), then this straight line MN is called a tangent to the curve (L) at the point M.

Thus the essential difference of a straight line tangent to some curve at the point M from other straight lines passing through the same point is that for the tangent its common point with the curve — the point of tangency — is a double point resulting from the merging of two approaching points of intersection.

It is not necessary here that one of these points remain fixed: both points of intersection may move toward each other and coincide at the point of tangency (Fig. 27). Occasionally not two but three points merge in the point of tangency (Fig. 28).

Remark 1. In the definition nothing is stated about the number of common points of the curve and the tangent to it. This number can be arbitrary. In Fig. 29a the reader sees a straight line touching the curve (G) at the point M and intersecting it in two more points, and in Fig. 29b a straight line MN, which touches the curve at several points at a time.

Remark 2. In the definition of tangent it is assumed that the point K can approach the point M in any manner. In all cases the secant must tend to the same straight line, which is then called the tangent. If in different methods of approach of K to M the secant line tends to different straight lines the curve is said to have no tangent at this point.

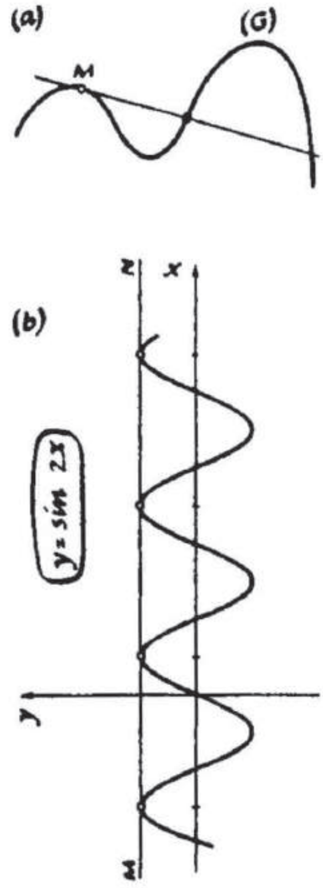


Fig. 29

EXERCISES

1. Find the tangent at the point $O(0, 0)$ to the parabola $y = x^2 + x$.

Solution. Let us take some point M on the parabola with coordinates (a, b). Obviously, $b = a^2 + a$. Let us draw a straight line through the points O and M.

The equation of this straight line has the form $y = kx$. At $x = a$ we have $y = a^2 + a$, hence $k = a + 1$, and the equation of the secant is $y = (a + 1)x$. We shall now make the point $M(a, b)$ approach the point $O(0, 0)$. When the point M coincides with the point O, its abscissa a vanishes, and the secant $y = (1 + a)x$ becomes the tangent $y = x$.

Answer. The equation of the tangent is $y = x$.

2. Find the tangent to the parabola $y = x^2 + x$ at the point $A(1, 2)$. ⊕

3. Which of the straight lines parallel to the straight line $y = x$ is tangent to the parabola $y = -x^2 + 1$? ⊕

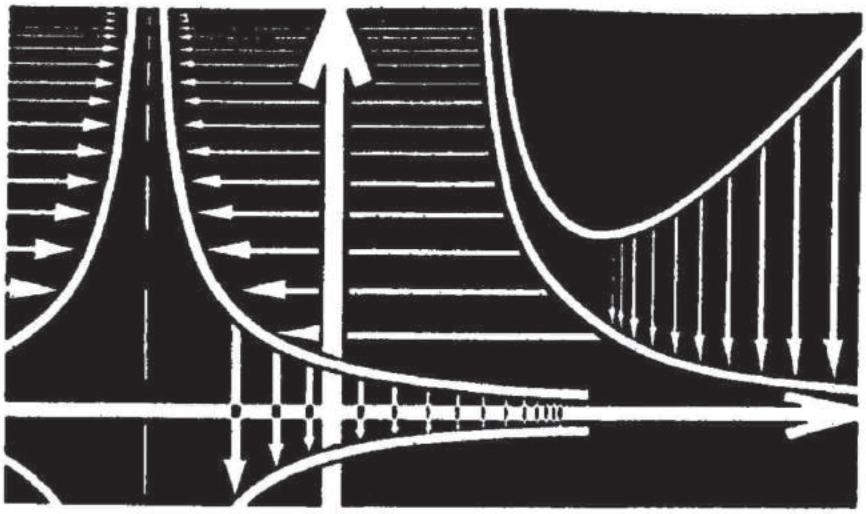
4. (*a*) Prove that the straight line $y = 0$ is the tangent to the curve

$$y = x^3 + x^2 \text{ at the origin.}$$

(*b*) Find the tangent at the point $O(0, 0)$ to the curve

$$y = x^3 - 2x.$$

5. For what cubic polynomials $y = ax^3 + bx^2 + cx$ does the x-axis serve as tangent at the origin?



CHAPTER 7

Rational Functions

1

Rational functions are functions that can be represented in the form of a quotient of two polynomials.

Examples of rational functions are

$$y = \frac{x^3 - 5x + 3}{x^6 + 1}, \qquad y = \frac{(x-1)^2(x+1)}{x^2+3},$$
$$y = x^2 + 3 - \frac{1}{x-1}.^*$$

The linear fractional function

$$y = \frac{ax+b}{cx+d},$$

analyzed in Chapter 5, is rational. It is the quotient

*$y = x^2 + 3 - 1/(x-1)$ is a rational function, since it can be written in the form of a ratio of two polynomials:

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

of two linear functions — polynomials of the first degree.

If the function $y = f(x)$ is the quotient of two polynomials of a degree higher than the first, then its graph will, as a rule, be more complicated, and constructing it accurately with all its details will sometimes be difficult. Frequently, however, it is sufficient to use methods analogous to those with which we have already become familiar.

2

Let us analyze some examples. We construct the graph of the function

$$y = \frac{x-1}{x^2+2x+1}.$$

First, let us notice that at $x = -1$ the function is not defined (since the denominator of the fraction $x^2 + 2x + 1$ equals zero at $x = -1$). For x close to -1, the numerator of the fraction $x - 1$ is approximately equal to -2, while the denominator $(x + 1)^2$ is positive and small in absolute value. Hence, the whole fraction $(x - 1)/(x + 1)^2$ will be negative and large in absolute value (and the more so, the closer x is to the value $x = -1$).

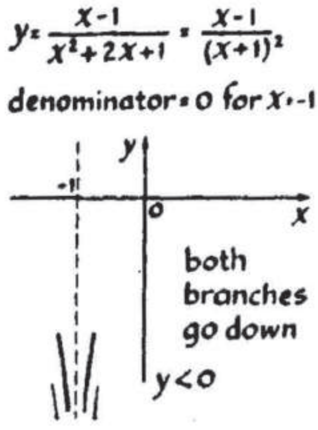


Fig. 1

Conclusion. The graph splits into two branches (since there is no point on it with abscissa equal to -1); both branches go down as x approaches -1 (Fig. 1).

Let us consider the numerator. It vanishes at $x = 1$. Hence at the point $x = 1$ the graph intersects the x-axis. Having also drawn the point of intersection with the y-axis (at $x = 0, y = -1$), we can get an approximate idea of how the graph behaves in its central part (Fig. 2).

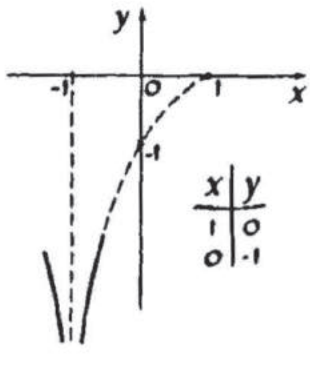


Fig. 2

We still must find out what happens to the function at values of x large in absolute value.

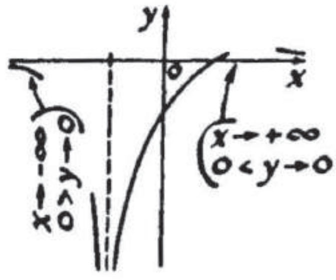


Fig. 3

If x is positive and increases, the numerator and denominator of the fraction increase. But since the numerator is of the first degree, while the denominator is of the second degree, the denominator, for large x, increases considerably faster than the numerator. Therefore, as x increases to infinity, the function $y = (x - 1)/(x^2 + 2x + 1)$ gets closer and closer to zero. In this way the right branch of the graph, to the right of the point $x = 1$, rises a little above the x-axis (Fig. 3), and then again starts to drop and will approach the x-axis.

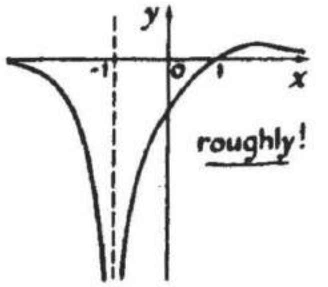


Fig. 4

Analogous considerations show that the left branch of the curve also approaches the x-axis as x increases in absolute value, except not from above but from below (Fig. 3). Later we shall show (see page 84) how to find the exact point where the right branch of the curve reaches its highest point.

From the above-mentioned details the general form of the graph can be found (Fig. 4).

3

Let us construct the graph of the function

$$y = \frac{x}{x^2 + 1}.$$

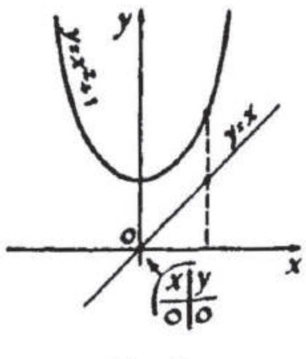


Fig. 5

For convenience let us first draw the graphs of the numerator $y = x$ and the denominator $y = x^2 + 1$ (Fig. 5). For the construction of the graph of our function it is necessary to divide the values of the numerator by the values of the denominator.

At $x = 0$ the numerator is equal to zero — the graph passes through the origin. Let us go to the right (that is, we consider positive values of the argument). Since, for very small x, the value of x^2 is much smaller than x, as the graph leaves the origin, the denominator will for some time be almost equal to 1 (somewhat larger than 1); therefore the whole function will be approximately equal to the numerator x (slightly smaller than the numerator): the graph runs alongside the straight line $y = x$, gradually falling below it (Fig. 6).

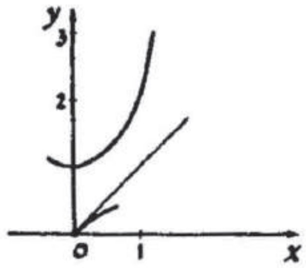


Fig. 6

Soon, however, $x^2 + 1$ starts to grow faster than x, the denominator leaves the numerator behind, and the fraction starts to decrease: the graph turns downward (Fig. 7).

Since the numerator is of the first degree, and the denominator contains an x^2-term, for large x the denominator grows faster than the numerator. Therefore, as x increases, the fraction becomes smaller and smaller — the graph approaches the x-axis (Fig. 8).

The left half of the graph can be obtained in an analogous way if it is observed that the given function is odd. The general form of the graph is shown in Fig. 9.

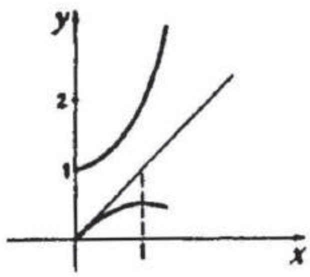


Fig. 7

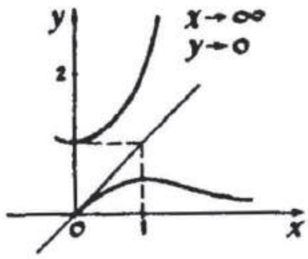


Fig. 8

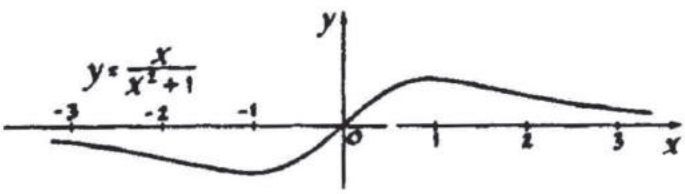


Fig. 9

4

Let us return to the graph of the function that we have constructed,

$$y = \frac{x}{x^2 + 1},$$

and analyze another interesting question, using this example. Let us try to find the highest point of the right half of the graph, exactly (and, hence, also the lowest point of the left half).

Obviously, our curve cannot rise very high, because the denominator $x^2 + 1$ starts quite rapidly to outgrow the numerator x. Let us find out whether the curve can reach a height equal to 1, that is, whether for some x, the value of y can be equal to 1.

Since $y = x/(x^2 + 1)$, it is necessary to solve the equation $x/(x^2 + 1)$ or the equation $x^2 - x + 1 = 0$. This equation does not have any real roots. (Check it!) Hence, on the graph there are no points with the ordi-

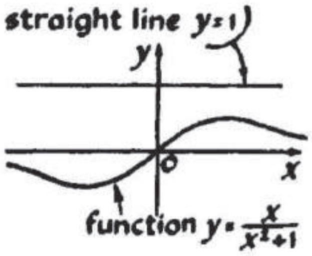


Fig. 10

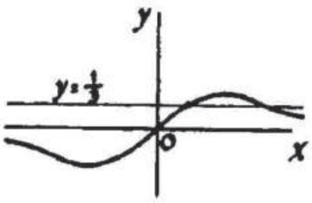


Fig. 11

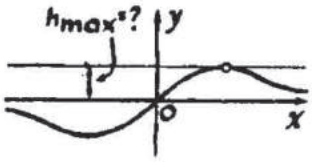


Fig. 12

nate $y = 1$ — the graph does not intersect the straight line $y = 1$ (Fig. 10).

Let us find out whether the curve reaches a height equal to $\frac{1}{3}$. For this purpose it is necessary that $x/(x^2 + 1) = \frac{1}{3}$, or $x^2 - 3x + 1 = 0$. This equation has two real roots (check it!), and therefore our graph has two points with ordinates equal to $\frac{1}{3}$; that is, it intersects the straight line $y = \frac{1}{3}$ at two points (Fig. 11).

In order to find the highest point, one must know for which largest h the equation $x/(x^2 + 1) = h$ will have a solution (Fig. 12).

Let us replace $x/(x^2 + 1) = h$ by the quadratic equation

$$hx^2 - x + h = 0.$$

This equation has a solution when

$$1 - 4h^2 \geq 0.$$

Hence, we can find the greatest height to which our graph can rise:

$$h = \tfrac{1}{2}.$$

Let us find at which x this largest value of y is obtained. Since $y = x/(x^2 + 1)$, $x/(x^2 + 1) = \frac{1}{2}$, $x^2 - 2x + 1 = 0$,* hence $x = 1$.

Thus the highest point of the graph is the point $(1, \frac{1}{2})$.

EXERCISE

Find the largest ordinate of the graph of $y = (x - 1)/(x + 1)^2$ (see Chapter 2).

5

Let us construct the graph of the function

$$y = \frac{x^2 + 1}{x}.$$

*Was it mere coincidence that a complete square was obtained?

Its general form can easily be drawn if it is noticed that

$$\frac{x^2+1}{x} = \frac{1}{x/(x^2+1)},$$

and, consequently, we have arrived at a problem that we have already solved: constructing the graph of $y = 1/f(x)$ from the graph of $y = f(x)$ (see pp. 56 and 57).

We get approximately the picture in Fig. 13.

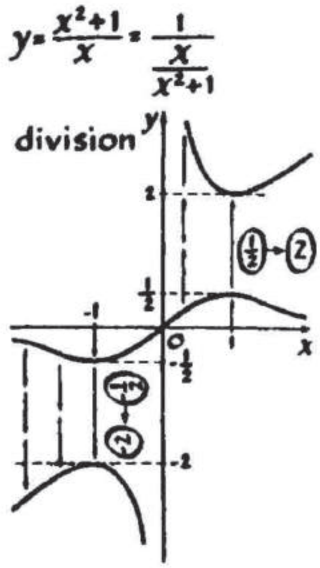


Fig. 13

Let us construct the graph by a different method. In doing so, we shall be able to explain another interesting peculiarity of this curve.

Let us divide the numerator by the denominator:

$$\frac{x^2+1}{x} = x + \frac{1}{x}.$$

Now we shall construct the graph of $y = x + (1/x)$ by "addition" of the known graphs of $y = x$ and $y = 1/x$ (Fig. 14).

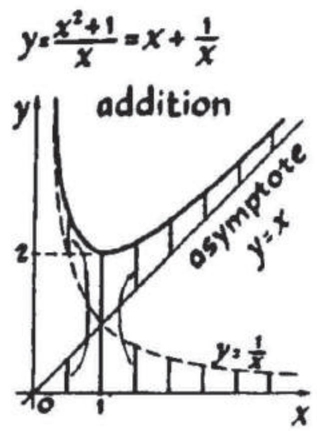


Fig. 14

We saw that the graph of $y = (x^2 + 1)/x$ has the y-axis as a vertical asymptote, which the graph approaches when x decreases in absolute value. It is now obvious that this graph also has an inclined asymptote, the straight line $y = x$ (this straight line is approached when x increases without bound).

EXERCISES

1. Check that the graph of $y = (x^2 + 1)/x$ is symmetric with respect to the origin.

2. Find the coordinates of the lowest point of the right-hand branch of the graph of $y = (x^2 + 1)/x$.

The answer to the second exercise is clear from the first method by which this graph was constructed (see Fig. 13): The lowest point of the graph of $y = (x^2 + 1)/x$ is obtained for the x for which the graph of $y = x/(x^2 + 1)$ reaches its highest point, that is, for $x = 1$. The smallest ordinate value of the graph of $y = (x^2 + 1)/x$ is thus equal to 2.

We have obtained an interesting inequality: *for positive x* (the right-hand part of the graph was considered) we always have

$$x + \frac{1}{x} \geq 2.$$

Problems

1. Prove the inequality

$$x + \frac{1}{x} \geq 2 \text{ for } x > 0 \qquad (1)$$

directly.

2. Prove the inequality

$$\frac{a+b}{2} \geq \sqrt{ab}. \qquad (2)$$

It expresses this fact: "The arithmetic mean of two positive numbers a and b is always greater than or equal to the geometric mean of these numbers."

Inequality 1 is a particular case of Inequality 2. For what a and b?

3. The inequality $x + (1/x) \geq 2$ is used in solving the well-known problem of the "honest merchant." An honest merchant knew that the scales on which he was weighing his merchandise were inaccurate, because one beam was somewhat longer than the other (at that time they were still using scales as shown in Fig. 15). What was he to do? Cheating his customers would be dishonest, but he did not want to hurt himself either. The merchant decided that he would weigh out half the merchandise to each buyer on one scale and the second half on the other scale.

The question is: Did the merchant gain or lose, as a result?

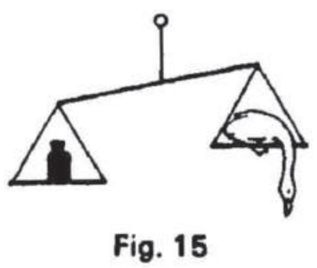


Fig. 15

6

In the following example let us take the function

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

It is already in the form of the sum of two functions, and of course it is possible to construct its graph by adding the graphs of $y = 1/(x+1)$ and $y = 1/(x-1)$. In this case, however, it is perhaps possible to find the general form of the graph from the following considerations:

(*a*) The function is not defined at $x = 1$ and $x = -1$, and therefore the curve separates into three branches (Fig. 16).

(*b*) As x approaches one, the second term, and hence the whole function as well, increases in absolute value, and therefore the branches of the graph move away from the x-axis, approaching the straight line $x = 1$. On the right of $x = 1$ the curve goes up, and on the left it goes down (Fig. 17).

An analogous picture results near the straight line $x = -1$:

(*c*) $y = 0$ at $x = 0$; the curve passes through the origin (Fig. 18).

(*d*) For numbers large in absolute value, both terms are small in absolute value, and both extreme branches of the graph approach the x-axis: the right from above, and the left from below (Fig. 19).

Combining all this information, we can obtain the general form of the graph (Fig. 20).

Show that this graph is symmetric with respect to the origin.

7

The examples discussed show that even for the construction of the same graph, different methods can be employed. Therefore, we shall now give a few more examples for the construction of graphs and

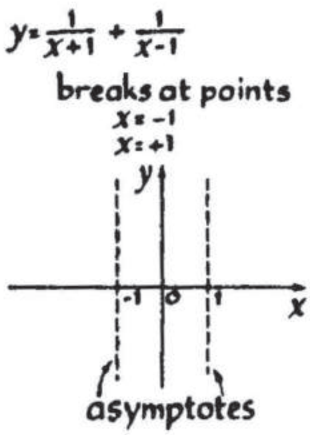


Fig. 16

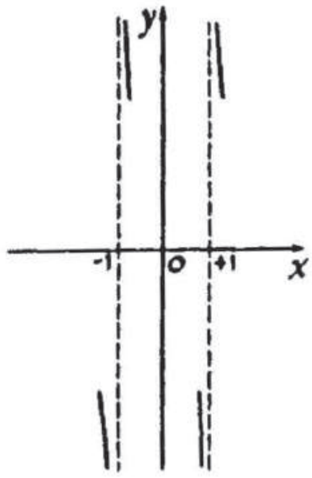


Fig. 17

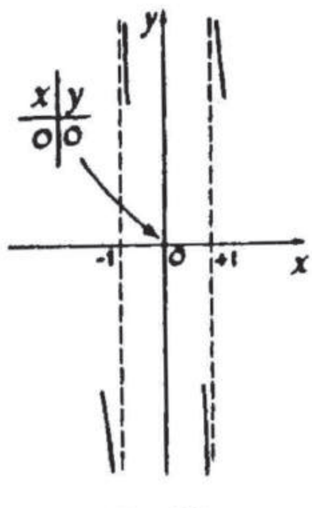


Fig. 18

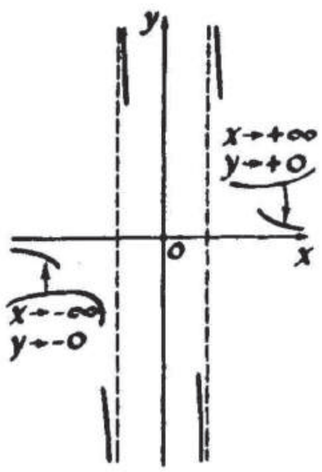


Fig. 19

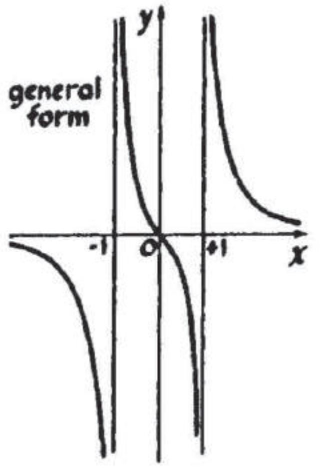


Fig. 20

shall not deprive the reader of the pleasure of selecting for himself the most suitable methods for constructing them.

EXERCISES

1. Construct the graph of

$$y = \frac{1}{x} + \frac{1}{x+1} + \frac{1}{x+2}.$$

Into how many pieces does the curve split?

2. (*a*) Construct the graph of

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

(*b*) Construct the graph of

$$y = \frac{1}{x} - \frac{1}{x+2}.$$

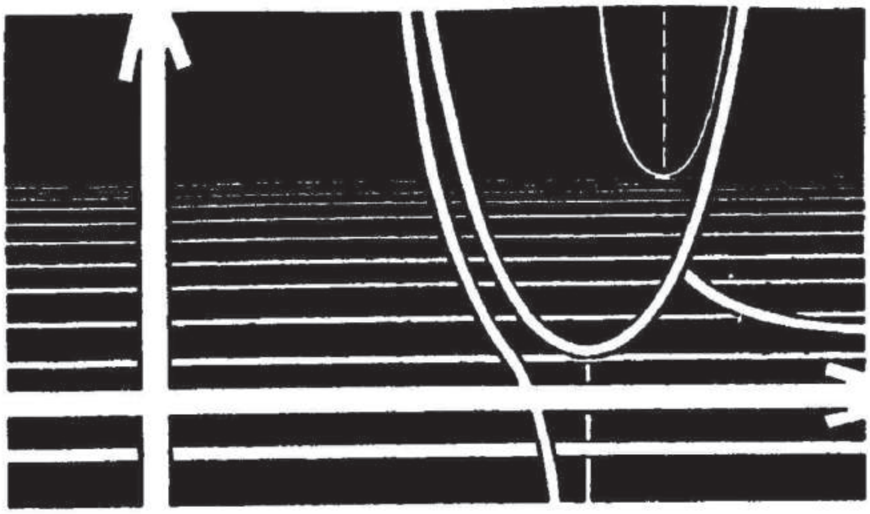
Indicate the axis of symmetry of this curve.

3. Construct the graphs of

(*a*) $y = \dfrac{1}{(x-1)(x-2)}$;

(*b*) $y = \dfrac{1}{(x-1)(x-2)(x+1)}$.

(*c*) $v = x + \dfrac{1}{x^2}$.



Problems for Independent Solution

1. Construct the graphs of the functions:

(a) $y = x(1 - x) - 2$;

(b) $y = x(1 - x)(x - 2)$;

(c) $y = \dfrac{4 - x}{x^3 - 4x}$;

(d) $y = \dfrac{2|x| - 3}{3|x| - 2}$;

(e) $y = \dfrac{1}{4x^2 - 8x - 5}$;

(f) $y = \dfrac{1}{x^3 - 5x}$;

(g) $y = \dfrac{1}{x^2} + \dfrac{1}{x - 1}$;

(h) $y = \dfrac{1}{x^2} + \dfrac{1}{x^3}$; $\oplus$

(i) $y = (2x^2 + x - 1)^2$;

(j) $y = |x| + \dfrac{1}{1 + x^2}$;

(k) $y = \dfrac{x^2 + 2x}{x^2 + 4x + 3}$;

(l) $y = \dfrac{x^2 - 2x + 4}{x^2 + x - 2}$;

(m) $y = (x - 3)|x + 1|$;

(n) $y = |x - 2| + 2|x| + |x + 2|$;

(o) $y = \left[\dfrac{1}{x}\right]$;

(p) $y = \dfrac{|x + 1| - x}{|x - 2| + 3}$;

(q) $y = \dfrac{x}{[x]}$ ⊕

(r) Construct graphs of the linear fractional functions of the form

$$y = \frac{3x + a}{2x + 2}$$

for different values of a.

2. The function $y = f(x)$ is defined by the following rule:

$$f(x) = \begin{cases} 1 \text{ for } x > 0, \\ 0 \text{ for } x = 0, \\ -1 \text{ for } x < 0. \end{cases}$$

This function is frequently encountered, and therefore there is a special notation for it:

$$y = \text{sign } x \text{ (read "signum } x\text{")}.$$

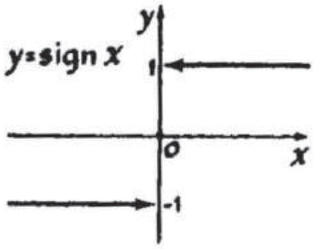


Fig. 1

The graph of this function is represented in Fig. 1. For $x \neq 0$ the function sign x can be defined by the formula $y = x/|x|$.*

Draw graphs of the functions:

$$y = \text{sign}^2 x; \quad y = (x - 1)\,\text{sign } x; \quad y = x^2 \,\text{sign } x.$$

3. The general form of the graph of the function that is the quotient obtained by dividing one quadratic

*Why only for $x \neq 0$?

trinomial by another,

$$y = \frac{ax^2 + bx + c}{x^2 + px + q},$$

depends on how many and what roots the numerator and denominator have.

(*a*) Construct the graphs of the functions:

$$y = \frac{4x^2 - 8x + 3}{x - x^2},$$

$$y = \frac{x^2 - 2x + 1}{x^2 + 2}, \qquad y = \frac{3x^2 - 10x + 3}{x^2 - x - 6}.$$

(*b*) What is the form of the graph of the function

$$y = \frac{ax^2 + bx + c}{x^2 + px + q},$$

if both roots of the numerator are less than the roots of the denominator? ⊕

(*c*) Analyze all possible cases and draw the possible types of graphs for functions of the form

$$y = \frac{ax^2 + bx + c}{x^2 + px + q}.$$

Try not to omit a single case, and give one example for each type.

4. Construct the graph of the function $y = \sqrt{3}\,x$.

(*a*) Prove that it cannot pass through any point whose coordinates are integers, except the point (0, 0).

If one square is taken as scale unit, then the vertices of the squares will be these "integral" points. Take the origin close to the lower left corner of a notebook and draw the straight line $y = \sqrt{3}\,x$ accurately. (At what angle with the x-axis must it be drawn?) Some of the integral points turn out to be very close to this straight line. Use this to find approximate values for $\sqrt{3}$ in the form of ordinary fractions. Compare the values obtained with the tabular one: $\sqrt{3} \approx 1.7321$.