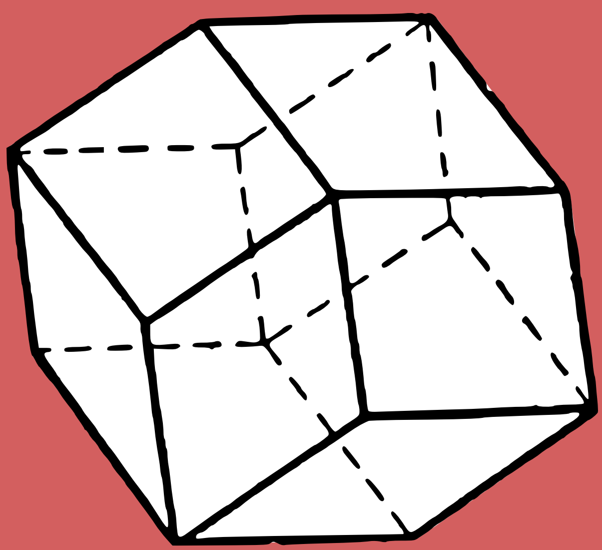
G. Dorofeev, M. Potapov,
N. Rozov

ELEMENTARY MATHEMATICS

Selected Topics & Problem Solving



MIR PUBLISHERS • MOSCOW

Г. В. Дорофеев, М. К. Потапов
Н. Х. Розов

Пособие по математике
для поступающих в вузы

Избранные вопросы
элементарной
математики

Издательство
«Наука»
Москва

ELEMENTARY MATHEMATICS

SELECTED TOPICS AND PROBLEM SOLVING

G. DOROFEEV

M. POTAPOV

N. ROZOV

TRANSLATED FROM THE RUSSIAN
BY
GEORGE YANKOVSKY

MIR PUBLISHERS
MOSCOW

First published 1973
Second printing 1980
Third printing 1982

На английском языке

CONTENTS

the most natural way for the average student. The main thing is that such a solution is carried through with extreme logical care and is made as rigorous as possible.

The reader may find that certain simple examples are analyzed in far too great an amount of detail. But do not hurry to criticize this approach, for what appears simple may merely be something that has not been studied in sufficient depth. Also, not all solutions are given with full details. It is the hope of the authors that this text will not only be read but studied with pencil and paper in hand. A good deal is left up to the student to think through by himself. This pertains to some parts of the theory and certain stages of problem solving.

It must be emphasized that this book is not an ordinary textbook but one in which certain carefully selected topics of theory and an abundant amount of problem solving will enable the student to expand and deepen his knowledge of the school course of elementary mathematics and enable him better to begin the study of higher mathematics in higher educational institutions. The topics chosen here for detailed discussion are those that usually cause the most trouble or do not, for a variety of reasons, receive the attention they deserve. The most complicated and important parts of elementary mathematics are analyzed and illustrated in detailed problem solving and subsequent discussion. Particular attention is paid to analyzing typical mistakes of the student.

Another point to bear in mind is that the authors consider only the more traditional topics of elementary mathematics. They do not use methods of analytic geometry or differential and integral calculus; in geometry, axiomatics is not dwelt on, nor is the terminology of set theory made much use of.

This textbook is supplied with a large number of problems in the form of exercises appended to each section. The answers are given at the end of the book.

This book is aimed at a broad range of readers, from students of secondary school to students of teachers' colleges and universities, and mathematics instructors in secondary and higher educational institutions. It can also be used in self-instruction as a supplement to any standard textbook.

G. Dorofeev, M. Potapov, N. Rozov

Chapter 1 ARITHMETIC AND ALGEBRA

1.1 General remarks on arithmetic and algebra

Of fundamental importance to the student is the fact that all the concepts he employs in his mathematical discourse must be rigorously *defined*, the only exception being, of course, such starting terms as natural number, equation, point, line, plane, and the like. The requisite definitions are of course given in any textbook, but the student becomes accustomed so soon to using these concepts in solving problems that he feels more and more inclined (without always realizing it) to regard the initial notions as intuitively clear and not in any need of being defined.

The student of mathematics must at all times have a clear-cut understanding of all fundamental mathematical concepts (we will return to this subject in Secs. 2.1 and 3.1).

Also important, besides definitions, are mathematical conventions involving the formation of an entity or of a relation between entities (indicated by a special symbol). These conventions serve essentially as a *definition of the symbol* and must be memorized. For example, the plus (+) sign is used to indicate the sum of two numbers, the symbol a^2 stands for the square of the number a, which is to say the product $a \cdot a$; the fact that a is less than b, that is, the number $a-b$ is negative, is written conventionally with the aid of the $<$ sign as $a<b$.

The student will also recall the signs of weak inequalities: $\leqslant$ (less than or equal to) and $\geqslant$ (greater than or equal to). The student usually finds no difficulty when using them in formal transformations, but examinations have shown that many students do not fully comprehend their meaning.

To illustrate, a frequent answer to: "*Is the inequality* $2 \leqslant 3$ *true*?" is "No, since the number 2 is less than 3". Or, say, "*Is the inequality* $3 \leqslant 3$ *true*?" the answer is often "No, since 3 is equal to 3". Nevertheless, students who answer in this fashion are often found to write the result of a problem as $x \leqslant 3$. Yet their understanding of the sign $\leqslant$ between concrete numbers signifies that not a single specific number can be substituted in place of x in the inequality $x \leqslant 3$, which is to say that the sign $\leqslant$ cannot be used to relate any numbers whatsoever.

Actually, the situation is this: by *definition* of the sign $\leqslant$, the *inequality* $a \leqslant b$ *is considered to be true when* $a < b$ *and also when* $a = b$. Thus, the inequality $2 \leqslant 3$ *is true* because 2 *is less* than 3, and the inequality $3 \leqslant 3$ *is true* because 3 *is equal* to 3.

From this definition of the sign $\leqslant$ it follows that the inequality $a \leqslant b$ is not true only when $a > b$. For this reason, the sign $\leqslant$ may be read not only as "less than or equal to" but also as "not greater than". Thus, the inequalities $2 \leqslant 3$ and $3 \leqslant 3$ are read, respectively, as "2 is not greater than 3" and "3 is not greater than 3".

The same applies to the sign $\geqslant$, which can be read both as "greater than or equal to" and as "not less than". By *definition* of the sign $\geqslant$, the *inequality* $a \geqslant b$ *is valid if* $a > b$ *or if* $a = b$; it is not valid only if $a < b$.

Almost every student knows that the function $y = 2^x$ is defined for all real x and can readily draw the graph of the function. However, $2^{\sqrt{3}}$ is often a riddle to the student. The best he can usually do is to indicate how one should give an approximate *computation* of the number. But where is the logic? How can you expect to give an approximate computation of a number without knowing its definition?

To be able to state what the number $2^{\sqrt{3}}$ represents, one has to recall the special definition for a number raised to an irrational power, and of course it is necessary to recall the other definitions of powers having natural exponents (a zero, rational or negative exponent). Note that the general definition of a power with a natural exponent n is inapplicable when $n = 1$ since a product involving a single factor is meaningless. For this reason, the equation $a^1 = a$ is the *definition* of the first power of a number. In the very same fashion, the zero power ($a^0 = 1$) is introduced as a *definition*.

Now let us find out *why the equation*

$$(\sqrt[3]{a})^3 = a \tag{1}$$

holds true. Students often prove this by manipulating the left-hand member. This is of course permissible, but it simply indicates that the rules for handling radicals have displaced in the mind of the student the definition of a radical. Indeed, how does one define the cube root of a number? By convention, the cube root of a number a is that number whose cube is equal to a. The cube root of a number a is conventionally denoted by the symbol $\sqrt[3]{a}$. Thus, equation (1) is merely the formula for the definition of a cube root with regard for the convention concerning the meaning of the symbol $\sqrt[3]{\ }$.

The course of algebra includes a considerable number of propositions (assertions). The view is rather widely held that in geometry one has to reason rigorously and there are theorems which require careful proofs with the use of definitions but that in algebra there is only one theorem (Viète's theorem),* and the rest is just verbal formulations and formulas. This is not so in the least. Even the formula for the square of a sum is a *theorem*. The properties of the logarithmic function constitute several theorems. As in geometry, every theorem of algebra must be proved, and all the initial concepts must be defined.

Experience shows that the more ordinary an algebraic statement is and the more often it is used in problem solving, the more frequently the student forgets that he should be able not only to state it properly and employ it, but also to *prove* it. At all times, particular attention must be paid to the ability of the student to justify (substantiate) statements, particularly those which appear to be "self-evident".

All students are familiar with the formula for solving quadratic equations, but not so many know its derivation. The same difficulties are encountered when dealing with *theorems involving the solution of quadratic inequalities*. Even if the student obtains correct solutions of such inequalities, he is frequently not able to explain *why*, for instance, a quadratic trinomial with positive leading coefficient is positive outside the interval between the roots if the latter are real, and is positive for arbitrary x if the roots are imaginary.

Yet rigorous proofs of the theorems dealing with the sign of a quadratic trinomial are simple in the extreme.

If the quadratic ax^2+bx+c, $a\neq 0$, has real roots x_1 and x_2 (which means its discriminant is positive), then it can be factored:

$$ax^2+bx+c=a(x-x_1)(x-x_2) \tag{2}$$

It is thus evident that for any x exceeding the larger root, both factors in parentheses, that is $(x-x_1)$ and $(x-x_2)$, are positive, and for any x less than the smaller root, they are negative, which means that in both cases their product $(x-x_1)(x-x_2)$ is positive and therefore the right member of (2) has the same sign as the number a. However, if x lies in the interval between the roots x_1 and x_2, then one of the parentheses in (2) is positive and the other one is negative. And so the sign of the product in the right member of (2) is opposite that of a.

We have thus proved the following **theorem**: *the value of a quadratic trinomial ax^2+bx+c with positive discriminant $(b^2-4ac>0)$ has for any x outside the interval between the roots of the quadratic a sign that coincides with the sign of the coefficient a, and is of opposite sign for any x inside the interval between the roots.* **

* Viète's theorem states that the sum of the roots of a quadratic equation is equal to the coefficient (with sign reversed) of the unknown to the first power, and the product of the roots is equal to the constant term.

** The student himself can state and prove the theorem referring to the case when the quadratic trinomial ax^2+bx+c has equal roots, i.e. when its discriminant is zero: $b^2-4ac=0$.

There is also another **theorem** that is valid: *the value of a quadratic ax^2+bx+c with negative discriminant* ($b^2-4ac<0$) *has for any x a sign coincident with the sign of the coefficient a.*

To prove this theorem, isolate a perfect square:

$$ax^2+bx+c=a\left[\left(x+\frac{b}{2a}\right)^2+\frac{4ac-b^2}{4a^2}\right] \tag{3}$$

Since the discriminant $b^2-4ac<0$ (it will be recalled that in this case the quadratic has imaginary roots), it is evident that the expression in square brackets is positive for any value of x, and the product in the right-hand member of (3) is, for any x, of the same sign as the number a.

The student is often surprised to encounter difficulties when dealing with **biquadratic equations**. There would seem to be no difficulties, since *any biquadratic equation $ax^4+bx^2+c=0$ can be reduced to a quadratic equation by the standard substitution $x^2=y$.* But suppose that the resulting quadratic has imaginary roots y_1 and y_2. Then determining x requires taking the square root of a complex number. In itself this is not so complicated and appropriate formulas are given in the standard textbooks. However, this may be avoided altogether if one does not resort to the standard substitution but factors the left-hand member by means of a special transformation.

This transformation consists in isolating a perfect square in the trinomial ax^4+bx^2+c and gives a valid result only when the quadratic equation $ay^2+by+c=0$ has **imaginary** roots.

However, in this case the perfect square is isolated in a somewhat different fashion than ordinarily: namely, group together the highest-degree term and the constant term, and then take their sum and complete the square.

Suppose we have an equation like $x^4+bx^2+c=0$ (for the sake of simplicity, we set $a=1$, which can always be done readily), and the equation $y^2+by+c=0$ has imaginary roots. This condition means that the discriminant $D=b^2-4c<0$, that is, $b<4c$, whence it is clear that $c>0$ and $|b|<2\sqrt{c}$, that is $b<2\sqrt{c}$. We can therefore perfom the following manipulations:

$$\begin{aligned} x^4+bx^2+c &= (x^4+c)+bx^2=(x^4+2\sqrt{c}\,x^2+c)-(2\sqrt{c}-b)\,x^2 \\ &= (x^2+\sqrt{c})^2-(2\sqrt{c}-b)\,x^2 \\ &= \left(x^2+x\sqrt{2\sqrt{c}-b}+\sqrt{c}\right)\left(x^2-x\sqrt{2\sqrt{c}-b}+\sqrt{c}\right) \end{aligned}$$

The solution of the given biquadratic equation now reduces to that of two quadratic equations with real coefficients.

These rather involved formulas need not be memorized of course; it is much better to isolate a perfect square in each given instance. To illustrate, let us solve the equation

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

We first reduce the equation to $x^4+x^2+{}^3/_2=0$. Its discriminant is equal to $1^2-4\cdot{}^3/_2=-5<0$, and so, applying the foregoing method, we obtain

$$\begin{aligned} x^4+x^2+{}^3/_2 &= \left(x^4+2\sqrt{{}^3/_2}\,x^2+{}^3/_2\right)-\left(2\sqrt{{}^3/_2}-1\right)x^2 \\ &= \left(x^2+\sqrt{{}^3/_2}\right)^2-\left(\sqrt{6}-1\right)x^2 \\ &= \left(x^2+x\sqrt{\sqrt{6}-1}+\sqrt{{}^3/_2}\right)\left(x^2-x\sqrt{\sqrt{6}-1}+\sqrt{{}^3/_2}\right) \end{aligned}$$

We can now solve the quadratics without any fear of complicated radicals. The first equation

$$x^2+x\sqrt{\sqrt{6}-1}+\sqrt{{}^3/_2}=0$$

has a negative discriminant: $D=\left(\sqrt{\sqrt{6}-1}\right)^2-4\sqrt{{}^3/_2}=-1-\sqrt{6}$, and, consequently, its roots

$$x_{1,2}=-\frac{\sqrt{\sqrt{6}-1}}{2}\pm i\,\frac{\sqrt{\sqrt{6}+1}}{2}$$

Similarly we find the roots of the second equation:

$$x^2-x\sqrt{\sqrt{6}-1}+\sqrt{{}^3/_2}=0$$

They are

$$x_{3,4}=\frac{\sqrt{\sqrt{6}-1}}{2}\pm i\,\frac{\sqrt{\sqrt{6}+1}}{2}$$

Two-term equations of the sixth degree ($x^6+a^6=0$) likewise reduce to the solution of this type of biquadratic equation (expand the left-hand member as a sum of cubes and apply the technique described above).

A few words are in order concerning the statements of a number of definitions and theorems. Textbooks frequently state definitions and theorems verbally without much use of convenient literal notation. Occasionally, this is justified, but very often it simply makes for hard-to-digest formulations. For instance, instead of writing "the square of the sum of any two numbers is equal to the sum of the squares of the numbers plus two times their product," one could more simply write: "for any numbers a and b we have $(a+b)^2=a^2+2ab+b^2$." A logarithm is conveniently defined as "a number x is the logarithm of a number N to the base a ($a>0$, $a\neq 1$) if $a^x=N$."

It is important to develop the habit of converting verbal statements into formula statements, and vice versa, for this is precisely what is ordinarily required when proving theorems. For example, to prove that "the logarithms of numbers exceeding unity to a base exceeding unity are positive," we must first introduce the designations: let the base be $a>1$, the number $x>1$, and let $y=\log_a x$; then establish that the number $y>0$. Rephrasing of this nature can also involve the necessity of using a definition. Thus, before proving the assertion that "for $a>1$ the function $y=\log_a x$ increases," one has to recall what an increasing function is, and then the proof begins thus: "Let $a>1$, and let x_1 and x_2 be positive numbers, $x_1<x_2$; we will prove that $\log_a x_1<$ $<\log_a x_2$."

It is not always properly understood that certain formula-type statements make use of symbols of certain concepts.

It is precisely this that explains why formula (1) is not readily recognized as the definition of a cube root written as the symbol $\sqrt[3]{\ }$, and that the equation $a^{\log_a N}=N\,(N>0,\ a>0,\ a\neq 1)$ is a symbolic notation of the "customary" verbal definition of a logarithm which employs the convention of denoting the logarithm of a number N to a base a in the form of $\log_a N$.

Exercises

1. What is (a) a periodic decimal fraction, (b) $a^{2/3}$, (c) a quadratic equation, (d) $\sqrt{11}$, (e) the modulus (absolute value) of a complex number, (f) $a>b$, (g) the sum of a nonterminating decreasing geometric progression?

2. State which of the following is a definition, an axiom or a theorem: (a) an equation is unaltered if both members are multiplied by the same number, (b) the modulus of any number is nonnegative, (c) $a^{1/3}=\sqrt[3]{a}$, (d) the graph of the function $y=-3x$ passes through the origin of coordinates.

3. Is the following equation always valid: $\sqrt{a}\sqrt{b}=\sqrt{ab}$?

4. If the discriminant of a quadratic equation is positive, then the equation has two distinct real roots. State the converse theorem, the inverse and the contrapositive. Which of these theorems are valid?

5. Prove that if the roots of a quadratic equation are imaginary, then the discriminant is negative.

6. Using formulas, state the condition that at least one of the numbers $a_1, \ldots, a_n$ is equal to the number α.

7. Use a single equation to denote that at least two of the numbers a, b, c are equal to zero.

8. What can be said about the numbers a and b if $1/a<1/b$? From what properties of the function $y=1/x$ can we obtain an answer to this question?

9. Using mathematical relations, state the assertion that the function $y=3x-$ $-x^2$ increases when the argument varies in the interval from -1 to $+1$.

10. Is the condition that the sum of the digits of a number is divisible by 3 a necessary, sufficient or necessary and sufficient condition for the number to be divisible by 12?

1.2 Integers, rational numbers, irrational numbers

Problems involving various parts of arithmetic often give trouble. This is frequently due to the fact that arithmetic is studied in the junior forms where many results are given without proof, and the material is actually never taken up again. Yet this does not in the least diminish the significance of such sections of arithmetic as the divisibility of the natural numbers, the properties of fractions, the theory of proportions, etc.

The senior student must know the statements of these results and should also be able to prove them (say, to derive a given criterion of divisibility).

To illustrate, let us prove the *criterion for divisibility by* 9. Given a natural number $N=\overline{a_n a_{n-1} \ldots a_2 a_1 a_0}$. Here, the symbol $\overline{a_n a_{n-1} \ldots a_1 a_0}$ (where the bar on top means that the digits are not to be thought of as a product of the numbers $a_n, \ldots, a_0$) denotes an $(n+1)$-digit number, where $a_n, a_{n-1}, \ldots, a_1, a_0$ are the digits of the appropriate orders of the number* (so that $1 \leqslant a_n \leqslant 9$, $0 \leqslant a_{n-1} \leqslant 9, \ldots, 0 \leqslant a_1 \leqslant 9$, $0 \leqslant a_0 \leqslant 9$). We have to prove two assertions: (a) if the sum of the digits $a_n+a_{n-1}+\ldots+a_1+a_0$ of the number N is divisible by 9, then the number N itself is divisible by 9; (b) if the number N is divisible by 9, then the sum of its digits is divisible by 9.

In accord with the positional principle of the decimal number system, we have

$$\overline{a_n a_{n-1} \ldots a_2 a_1 a_0}=a_n \cdot 10^n+a_{n-1} \cdot 10^{n-1}+\ldots+a_2 \cdot 10^2+a_1 \cdot 10+a_0$$

Since $10^k=\underbrace{\overline{99 \ldots 9}}_{k \text{ times}}+1$ for any natural $k \geqslant 1$, we get

$$N=[a_n \underbrace{\overline{99 \ldots 9}}_{n \text{ times}}+a_{n-1} \underbrace{\overline{99 \ldots 9}}_{n-1 \text{ times}}+\ldots+a_2 \cdot \overline{99}+a_1 \cdot 9] \\ +(a_n+a_{n-1}+\ldots+a_2+a_1+a_0) \quad (1)$$

It is obvious that the number in square brackets is divisible by 9, for it is a sum of n terms, each of which is divisible by 9. If the sum $a_n+\ldots+a_1+a_0$ is divisible by 9, then from (1) it is clear that the number N is also divisible by 9. The proof of Assertion (a) is complete. Assertion (b) likewise follows from a consideration of (1): if the left member (the number N) is divisible by 9 and since the first summand of the right member (the number in square brackets) is divisible by 9, it follows that the second summand (the sum of the digits of N) must be divisible by 9.

In the solution of problems, various arithmetical facts are sometimes useful. We shall now review a number of them using literal symbolism.

If we have two integers** a and b, $b>0$, then there is a unique integer q and a unique integer r, $0 \leqslant r<b$, such that

$$a=bq+r \quad (2)$$

* It is natural to regard the highest-order digit as nonzero.

** Recall that the numbers 1, 2, 3, ... are called *natural numbers* (the positive integers), and the numbers —2, —1, 0, 1, 2, ... are the *integers* (whole numbers). It is convenient to write the set of integers as 0, ± 1, ± 2, ...

Equation (2) is simply the *division of a number a by a number b with a remainder*. In particular, from (2) it is clear that any even number is of the form $2k$, where k is an integer, and any odd number may be represented as $2n+1$, where n is an integer.

If we have a natural number N exceeding unity, and if $N=n_1^{\alpha_1} \dots n_k^{\alpha_k}$ is the decomposition of this number into prime factors (here, $n_1, \dots, n_k$ are distinct prime divisors of N, and $\alpha_1, \dots, \alpha_k$ represent the number of their repetitions in the decomposition of N), then any divisor of N is of the form $D=n_1^{\beta_1} \dots n_k^{\beta_k}$ where $0 \leqslant \beta_1 \leqslant \alpha_1, \dots, 0 \leqslant \beta_k \leqslant \alpha_k$.

If we have natural numbers $a_1, \dots, a_n$, then their *common divisor* is a natural number which exactly divides each of the numbers $a_1, \dots, a_n$. The largest of these common divisors of the numbers $a_1, \dots, a_n$ is termed the greatest common divisor. If the greatest common divisor is equal to 1, then the numbers $a_1, \dots, a_n$ are *relatively prime* (coprime).

If a natural number N is divisible by each of two relatively prime integers a_1, a_2, then N is also divisible by the product a_1a_2 of these integers.* Furthermore, if the product NM of natural numbers N and M is exactly divisible by a natural number D and if M and D are relatively prime, then N is divisible by D.

Finally, it is well to recall the following property: one of a sequence of n integers $k+1, k+2, \dots, k+n$, where k is an arbitrary integer, is definitely divisible by n.

Let us consider some examples of the use of the properties of integers in solving problems which involve divisibility.

1. *Prove that for an arbitrary even n the number $N=n^3+20n$ is divisible by* 48.

Quite naturally, a direct verification of the fact that the assertion holds true for $n=2, 4, 6, \dots$ does not solve the problem since we are not able to run through all the even numbers. Hence, we have to give a proof that will hold true for any even n.

An even number n can be written in the form $n=2k$, where k is an integer; therefore $N=8k\,(k^2+5)$. If we demonstrate that for *any* integer k the number $k\,(k^2+5)$ is divisible by 6, it will be clear that N is divisible by 48.

We perform the following obvious transformation:

$$k\,(k^2+5)=k\,(k^2-1+6)=(k-1)\,k\,(k+1)+6k \qquad (3)$$

We see that the second summand in the right member of (3) is divisible by 6. Now the first summand on the right is a product of *three* successive integers, and for this reason one of them is definitely divisible

* It is easy to see that if a_1 and a_2 are not relatively prime, then the number N is not necessarily divisible by the product a_1a_2 (give an example!).

by 3. What is more, of two successive integers (and all the more so, of three) one must definitely be even. Since 2 and 3 are relatively prime, it follows that $k(k^2+5)$ is indeed divisible by 6 for any integer k.

2. *Prove that no matter what the natural number n, the number $N=n^2+1$ is not divisible by* 3.

The only possible remainders upon the division of a natural number by 3 are 0, 1, 2 (see (2)). In solving this problem we therefore find it expedient to partition all the natural numbers into three classes: $3k$, where k is a natural number; $3k+1$, where k is a natural number or 0; $3k+2$, k natural or 0.*

For an arbitrary natural number n which is exactly divisible by 3, that is, such as can be represented in the form $n=3k$ for some positive integer k, we have $n^2+1=9k^2+1$. Since the first summand in the right member is divisible by 3 and the second one is not, the number N is not divisible by 3 for these values of n.

If $n=3k+1$ for a certain natural number k (or for $k=0$), then $n^2+1=9k^2+6k+2$. It is obvious in this case that when N is divided by 3 there is a remainder of 2.

The case of $n=3k+2$ is considered in a similar manner.

3. *How many zeros are there at the end of the product of all natural numbers from* 1 *to* 1962 *inclusive?*

This problem appears to be very difficult due to its rather unusual statement, yet in reality the underlying idea of the solution is simple. If the number $N=1\cdot 2\ldots 1961\cdot 1962$ is factored into primes:

$$N=2^{\alpha_1}\cdot 3^{\alpha_2}\cdot 5^{\alpha_3}\ldots p^{\alpha_q} \tag{4}$$

it is clear that *each pair* of prime factors 2 and 5 will generate one zero in the number N because $10=2\cdot 5$. Now to obtain the representation (4) it suffices to factor separately into primes each of the factors of the product N, and then collect identical prime factors. Since we are interested only in the numbers α_1 and α_3 in the expansion (4), we have to find out how many twos and fives appear in the expansion of each factor of the product N.

To illustrate, let us determine the number α_3. It is clear that every factor of N divisible by 5 yields one five when decomposed into prime factors. Altogether, there are $[1962/5]=392$ such factors in the product N (the symbol $[a]$ denotes the largest integer in a). However, there will be, among the factors of the product N, such factors as are divisible by 25 and these will yield an *additional* five in the prime decomposition. There will be a total of $[1962/25]=78$ such factors. Yet another five will be obtained from all factors of N which are mul-

* Whereas in the representation $3k$ we must take, for k, one of the numbers 1, 2, ..., in the representations $3k+1$, $3k+2$ we also have to take $k=0$. To this value of k correspond the natural numbers 1 and 2, respectively.

tiples of 125. There will be $[1962/125]=15$. Finally, there are three factors divisible by 625; they will vield one five each. We thus have $392+78+15+3=488$ fives, or $\alpha_3=488$, in a decomposition of the number N into prime factors.

A similar computation shows that in the formula (4), the number $\alpha_1=1955$, whence it is evident that there are only 488 pairs of the primes 2 and 5 and therefore the number N ends in 488 zeros.

Ideas involving divisibility of numbers are very often employed in the solution of problems from other sections of algebra.

4. *Find numbers which are common terms of the two following arithmetic progressions:*

$$3,\ 7,\ 11,\ \ldots,\ 407 \text{ and } 2,\ 9,\ 16,\ \ldots,\ 709 \tag{5}$$

It is clear that the general term of the first arithmetic progression is of the form $a_n=3+4(n-1)$; the indicated terms of the progression are associated with the values $n=1, 2, \ldots, 102$. Similarly, the terms of the second progression are obtained from the formula $b_k=2+7(k-1)$, $k=1, 2, \ldots, 102$. The problem thus consists in finding all numbers n and k, $1\leqslant n\leqslant 102$, $1\leqslant k\leqslant 102$, for which $a_n=b_k$, that is, $4n+4=$ $=7k$.

From the equation $4(n+1)=7k$ it is evident that it is valid provided k is a multiple of 4, that is, if $k=4s$; it is clear that s can take on the values $1, 2, \ldots, 25$ (since $1\leqslant k\leqslant 102$). But if $k=4s$, then $4(n+1)=7\cdot 4s$, or $n=7s-1$. Since $1\leqslant n\leqslant 102$, only the numbers $1, 2, \ldots, 14$ are permissible values of s.

We thus have 14 numbers that are common to both progressions (5). The numbers themselves are readily found either from the formula for a_n when $n=7s-1$, $s=1, 2, \ldots, 14$, or from the formula for b_k for $k=4s$, $s=1, 2, \ldots, 14$.

It will be recalled that *rational* numbers are numbers of the form p/q, where p is an integer and q is a natural number. If p/q is a positive number, then $p>0$, if p/q is negative, then $p<0$. It is obvious that the fraction p/q is always considered to be in lowest terms, that is, it is always possible to consider the numbers $|p|$ and q as relatively prime. The number 0 is associated with the representation p/q for $p=0$ (and any q).

A rigorous and complete theory of *irrational* numbers (a substantiation of their properties and of operations involving them) is given in the course of higher mathematics. Nevertheless, what the school curriculum offers must be mastered thoroughly.

A typical mistake of students is that they judge the rationality or irrationality of a number solely on the basis of its outward appearance and readily assume that an involved combination of irrational numbers is all the more so an irrational number. Yet this is not always the case. For instance, the number $[(\sqrt{3}+\sqrt{2})/(\sqrt{3}-\sqrt{2})]-$

$-2\sqrt{6}$ is not irrational: a simple computation shows that it is equal to 5. Likewise the number $\sqrt[3]{7+\sqrt{50}}+\sqrt[3]{7-5\sqrt{2}}$, despite its intricate and "irrational" aspect, is actually a rational number equal to 2 (this is quite evident if one notes that the radicands are perfect cubes).

It is therefore necessary to thoroughly justify one's stand when stating that a given number is rational or irrational. Some problems will help to illustrate this point.

5. *Prove that* $\log_4 18$ *is an irrational number.*

Since $\log_4 18=\frac{1}{2}+\log_2 3$, it suffices to demonstrate that the number $\log_2 3$ is irrational. Acsume the contrary, that this number is rational. This means that $\log_2 3=p/q$. Since $\log_2 3>0$, both numbers p and q may be regarded as natural. Taking advantage of the definition of a logarithm, we rewrite the equation $\log_2 3=p/q$ as $2^p=3^q$. But this latter equation is not possible for any natural p and q whatsoever; on the left is an *even* number (since $p>0$), while the right member is *odd*. The resulting contradiction completes the proof.

6. *Prove that the numbers* $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$ *cannot be the terms of a single arithmetic progression.*

To many this statement appears to be almost obvious. Some state right off that the irrational numbers $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$ "cannot be separated from one another by the same number". Others even "justify" this idea by the following computations: $\sqrt{3}-\sqrt{2}\approx 1.732-1.414=0.318$, and $\sqrt{5}-\sqrt{3}\approx 2.236-1.732=0.514$.

Note first of all that *approximate computations without an estimate of their accuracy cannot be used as proof in mathematics.* But even if one estimates the accuracy of the computations (this is easy to do), the proof is not correct since it demonstrates that the given numbers cannot be three *successive* terms of an arithmetic progression. But it is not prooved that they cannot be three, generally speaking, nonadjacent terms of a single arithmetic progression.

We carry out the proper proof by contradiction. Suppose, in a certain arithmetic progression with first term a_1 and common difference d, we have

$$\sqrt{2}=a_k=a_1+(k-1)d,\quad \sqrt{3}=a_m=a_1+(m-1)d,$$
$$\sqrt{5}=a_n=a_1+(n-1)d$$

Subtracting the first equation from the second and the second from the third, and then dividing one of the resulting relations by the other, we get

$$\frac{\sqrt{3}-\sqrt{2}}{\sqrt{5}-\sqrt{3}}=\frac{m-k}{n-m} \qquad (6)$$

The right member of this equation is a rational number because m, k and n are integers. For brevity, denote this number by r and rewrite (6) as $\sqrt{3}-\sqrt{2}=r(\sqrt{5}-\sqrt{3})$, whence, squaring, we get $r^2\sqrt{15}-\sqrt{6}=(8r^2-5)/2$. The right side of this equation is again a rational number. Denote it by s. Squaring both sides of $r^2\sqrt{15}-\sqrt{6}=s$, we get $\sqrt{10}=(15r^4-s^2+6)/(6r^2)$. This equation shows that $\sqrt{10}$ is a rational number, which is not true (the irrationality of $\sqrt{10}$ is proved in the same way as the irrationality of $\sqrt{2}$ is). This contradiction demonstrates that equation (6) is impossible, which is to say that the numbers $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$ cannot be terms of a single arithmetic progression.

7. *Determine all such integers a and b for which one of the roots of the equation* $3x^3+ax^2+bx+12=0$ *is equal to* $1+\sqrt{3}$.

The number $1+\sqrt{3}$ is, by definition, a root of the equation $3x^3+ax^2+bx+12=0$ if the following is valid:

$$3(1+\sqrt{3})^3+a(1+\sqrt{3})^2+b(1+\sqrt{3})+12=0$$

or, after simplification and regrouping,

$$(4a+b+42)+(2a+b+18)\sqrt{3}=0$$

We are interested only in the integers a and b; in this case the numbers $p=4a+b+42$ and $q=2a+b+18$ will also be integers.

We thus have to determine the integers a and b for which $p+q\sqrt{3}=0$. At this point many students make a logical mistake by considering as "quite obvious" that the latter equation is possible only for the case of $p=q=0$. However, not many can give a convincing justification of this fact. Let us prove it.

Indeed, suppose that $p+q\sqrt{3}=0$ holds true for a certain integer $q\neq 0$. Then it would follow immediately that $\sqrt{3}=-p/q$, which contradicts the irrationality of the number $\sqrt{3}$. Thus, $q=0$. But if $q=0$, then from $p+q\sqrt{3}=0$ it follows that p too is equal to zero.

Consequently, the number $1+\sqrt{3}$ is a root of the equation $3x^3+ax^2+bx+12=0$ if and only if

$$4a+b+42=0$$
$$2a+b+18=0$$

This system has the unique solution $a=-12$, $b=6$.

Exercises

1. State and prove the criterion of divisibility by 11.

2. Prove that there is not a single natural number N with sum of digits equal to 15 that is the square of an integer.

3. Let $p\geqslant 5$ be a prime number. Prove that the number p^2-1 is divisible by 24.

4. It is given that the numbers p, $p+2$ and $p+4$ are primes. Find p.

5. Demonstrate that if a natural number ends in the digit 7, it cannot be the square of an integer.

6. How many factors 2 are there in the product of all integers from 1 to 500 inclusive?

7. Find numbers which are simultaneously the terms of two arithmetic progressions: 2, 5, 8, ..., 332 and 7, 12, 17, ..., 157.

8. For what natural numbers n is the fraction $(3n+4)/5$ an integer?

9. For what natural numbers n is the fraction $(2n+3)/(5n+7)$ reducible to lower terms?

10. Find a four-digit number that satisfies the following conditions: the sum of the squares of the extreme digits is equal to 13; the sum of the squares of the middle digits is equal to 85; if we subtract 1089 from the desired number we obtain a number containing the same digits as the desired number but in reverse order.

11. Find a three-digit number $\overline{abc}$ such that the four-digit numbers $\overline{abc1}$ and $\overline{2abc}$ satisfy the equation $\overline{abc1}=3\cdot\overline{2abc}$.

12. Find all five-digit numbers of the form $\overline{34x5y}$ (x and y are digits) divisible by 36.

13. Determine for what natural numbers n the number n^4+4 is a composite number.

14. Prove that if the sum $k+m+n$ of three natural numbers is divisible by 6, then $k^3+m^3+n^3$ is also divisible by 6.

15. Demonstrate that for any natural n the number $\underbrace{\overline{1\ldots1}}_{2n \text{ times}}-\underbrace{\overline{2\ldots2}}_{n \text{ times}}$ is the square of an integer.

16. Find all the integral solutions to the equation $2x^2-3xy-2y^2=7$.

17. Find all the integral solutions to the equation $2x^2y^2+y^2-6x^2-12=0$.

18. Prove that between any two unequal rational numbers a and b there is at least one rational number and at least one irrational number.

19. Can the numbers 10, 11, 12 be terms of a single geometric progression?

20. Prove the irrationality of the number $\tan 5°$.

1.3 The method of mathematical induction

The method of mathematical (or complete) induction is a very strong tool in mathematical proofs. Unfortunately, in secondary school it does not receive the attention which it deserves. Most students have a rather hazy idea concerning this important method. What is more, many proofs in the school course of mathematics are logically unsound due precisely to the fact that the induction method was not invoked. Recall for instance the *derivation of the formula for the general term of an arithmetic progression.* We write down a number of equations:

$$\left.\begin{aligned} a_1 &= a_1+d(1-1), \\ a_2 &= a_1+d=a_1+d(2-1), \\ a_3 &= a_2+d=a_1+2d=a_1+d(3-1), \\ a_4 &= a_3+d=a_1+3d=a_1+d(4-1) \end{aligned}\right\} \quad (1)$$

and so forth; consequently, for any n the formula $a_n=a_1+d(n-1)$ is valid. The incompleteness of this proof is obvious. We establish the formula for a few values of n and then draw the conclusion that it is true for any integer n. With that approach, it is possible to "prove" the following assertion: *for an arbitrary integer n*, n^2+n+41 *is prime.* Indeed, for $n=1, 2, 3, 4$ we have 43, 47, 53, 61—all primes. "Consequently", the assertion is proved, though it is clear that, for example, when $n=41$ the number n^2+n+41 is divisible by 41.

The foregoing example might seem to suggest that one should take a much larger number of values of n. Suppose, however, that in deriving the formula for the general term of an arithmetic progression we verified it for one million terms. Are we entitled to conclude that the formula will hold for all n? Not in the least. We may know a million steps, but we do not know what might happen on the million and first step. It may very well be that the formula will break down on that next step.*

Therefore, the defect of all proofs of that nature does not lie in the number of special cases examined but in the absence of a "look to the future", the lack of any knowledge of what will happen on the next step. This is all taken care of in the method of mathematical induction.

The essence of this method consists in the following.

Let an assertion be *verified for a single special case*, say for $n=1$. Let us suppose that we can demonstrate that *from the validity of this assertion for* $n=k$ *it will always follow that it is valid for the next value of* n *as well, that is, for* $n=k+1$. We can then reason as follows: we have verified our assertion for $n=1$, but then, by what has been proved, it will be true for $n=1+1=2$. Now, since it is valid for $n=2$, it will also be valid for $n=2+1=3$, and so forth, which means *it will hold true for all values of* n.

It might, at first glance, seem that this "and so forth" is just as lacking in legitimacy as were the earlier examples. This is not so, however, for we are fully confident that each time we can take the next step, and it is therefore **obvious** that the assertion will hold true for any value of n: *because any integer can be reached in a finite number of steps beginning with* $n=1$.**

Thus, in order to prove the validity of an assertion for an arbitrary natural*** n, we have to prove two things: firstly, that it is true for $n=1$, secondly, that from its validity for $n=k$ follows its validity for $n=k+1$. That is the gist of the *method of mathematical induction*: we prove that our assertion holds true for $n=1$ (this is the *basis* of the induction), then we assume that it is valid for a certain $n=k$ (the *hypothesis* of the induction), and we prove that in that case it is valid for $n=k+1$ (the *induction step*).

* Incidentally, the number n^2+n+41 proves to be prime for all n from 1 to 39 inclusive, and only the fortieth step reveals that the assertion no longer holds. What guarantee have we that the formula for the general term of an arithmetic progression will not behave similarly, say, at the million and first step?

** In the rigorous theory of natural numbers, this assertion is accepted as an axiom.

*** Ordinarily, the *natural numbers* are the positive integers 1, 2, 3, etc., the number 0 not being included.

Let us apply this method in proving the formula for the general term of an arithmetic progression. The assertion we are about to prove is of the form

$$a_n = a_1 + d(n-1)$$

For $n=1$ the assertion holds since the left-hand member contains a_1 and the right-hand member $a_1+d(1-1)=a_1$. Assume that it is true for $n=k$, that is, $a_k=a_1+d(k-1)$. By the definition of an arithmetic progression, $a_{k+1}=a_k+d$, whence, using the induction hypothesis, we obtain

$$a_{k+1} = a_1 + d(k-1) + d = a_1 + dk = a_1 + d[(k+1)-1]$$

which is to say the assertion holds true for $n=k+1$. Hence, the formula for the general term is valid for any n.

It must be stressed that the induction method is a method of proof of specified assertions and does not serve as a derivation of these assertions. For instance, this method cannot be used to obtain the formula of the general term; however, if we have found the formula in some way, say by trial and error, then the proof of it can be carried out by the induction method. That is exactly what we did above: equation (1) suggested what the formula for the general term might be like, and then we demonstrated this in rigorous fashion. In this process, of course, the method of trial and error, the mode of obtaining a formula or an assertion is not a necessary element of the proof. On the basis of some kind of reasoning or guessing we conjecture an assertion, then we can proceed to proof by induction.

Let us consider some examples of induction proofs.

1. *Prove that the sum of n terms of a geometric progression is*

$$S_n = \frac{a_1(q^n-1)}{q-1} \tag{2}$$

For $n=1$ the equation holds true since

$$S_1 = a_1 = \frac{a_1(q-1)}{q-1}$$

Assume that (2) is valid for $n=k$, that is, $S_k=\frac{a_1(q^k-1)}{q-1}$. Then

$$S_{k+1} = S_k + a_{k+1} = \frac{a_1(q^k-1)}{q-1} + a_1q^k = \frac{a_1(q^{k+1}-1)}{q-1}$$

or equation (2) is true for $n=k+1$. Therefore it holds true for any n.

2. *Prove that if n is a natural number, then* $4^n+15n-1$ *is divisible by* 9.

For $n=1$, the number $4^n+15n-1$ is 18, which is divisible by 9. Assume that $4^k+15k-1$ is divisible by 9 and take $n=k+1$. Then

$$\begin{aligned}4^{k+1}+15(k+1)-1 &= 4(4^k+15k-1)-45k+18\\ &= 4(4^k+15k-1)-9(5k-2)\end{aligned}$$

But by the induction hypothesis $4^k+15k-1$ is divisible by 9, and therefore the right member, and so too the left member, of the equation is divisible by 9, which completes the proof.*

3. *Demonstrate that if a and b are positive numbers and $a<b$, then for any natural n the inequality $a^n<b^n$ holds true.*

The assertion is obvious for $n=1$. Suppose that $a^k<b^k$; multiplying the inequality by a positive number a, we get $a^{k+1}<ab^k$. But b is a positive number and so $b^k a<b^k b$, or

$$a^{k+1}<b^{k+1}$$

which is what we sought to prove.

4. *Prove the formula for the number of permutations*

$$P_m^n=m(m-1)\dots(m-n+1)$$

We will assume that m is a certain fixed integer and we will carry out the proof by induction with respect to n. For $n=1$, the left member is equal to P_m^1, which is equal to m, and so the formula holds. Assume that

$$P_m^k=m(m-1)\dots(m-k+1)$$

In order to take the induction step, we establish the relation

$$P_m^{k+1}=(m-k)P_m^k$$

To do this, write out all permutations of m elements taken k at a time and affix to each one each of all the elements that did not appear in the permutation. Thus from each permutation of m elements taken k at a time we obtain $m-k$ permutations of m elements taken $k+1$ at a time. There will consequently be a total of $(m-k)P_m^k$ such permutations. It is easy to see however that among the permutations thus obtained there will be all the permutations of m elements taken $k+1$ at a time, each one appearing exactly once. Therefore $P_m^{k+1}=(m-k)\,P_m^k$

Utilizing the relation just proved and the induction hypothesis, we get

$$P_m^{k+1}=P_m^k(m-k)=m(m-1)\dots(m-k+1)(m-k)$$
$$=m(m-1)\dots(m-k+1)[m-(k+1)+1]$$

which completes the proof.

Note here that the method of induction need not begin with $n=1$. It is of course possible to demonstrate the proposition for some $n=n_0$,

---

* This assertion can be proved without induction. Indeed, using the binomial theorem we have, for $n\geqslant 2$,

$$4^n+15n-1=(3+1)^n+15n-1$$
$$=3^n+n\cdot 3^{n-1}+\dots+C_n^2\cdot 3^2+n\cdot 3+1+15n-1$$
$$=9(3^{n-2}+n\cdot 3^{n-3}+\dots+C_n^2+2n)$$

whence it follows that the given number is divisible by 9.

take the induction step and obtain the result that the proposition holds true for all integral values of n which are greater than or equal to the initial number n_0. Naturally, the induction hypothesis is then of a modified form: namely, we assume that the proposition being demonstrated holds true for $n=k\geqslant n_0$. Finally, it is important to bear in mind that for values of $n<n_0$ the assertion may be either true or not true. At any rate, the proof via the method of mathematical induction does not allow us to draw any conclusions concerning the validity of the assertion for $1\leqslant n<n_0$.

Both stages in the induction proof—choice of the basis of induction and substantiation of the induction step—are equally important and quite independent of each other. To take an example, *let us try to find out whether the inequality* $2^n>n^2$, *where* n *is a positive integer, is valid*. It clearly holds true for $n=1$. We verify the possibility of taking the induction step. Suppose that for $n=k$ we have the inequality $2^k>k^2$. Then it is obvious that $2^{k+1}=2\cdot 2^k>2k^2$, and to justify the induction step it suffices to establish the inequality $2k^2\geqslant(k+1)^2$ or $k^2-2k-1\geqslant 0$. However, this latter inequality is only valid if $k\geqslant 1+\sqrt{2}$, that is, if $k\geqslant 3$. We cannot therefore take $n_0=1$ for the basis of induction because we will not be able to take the first induction step. It is then natural to try $n_0=3$ for the basis. In this case, we can take the induction step, but it is verified directly that for $n=3$ the inequality $2^n>n^2$ is not true, and so we cannot start the induction process. It is only for $n=5$ that the inequality holds, and so we can take $n_0=5$ for the induction basis; for $n\geqslant n_0$ the induction step is also valid. And so the *inequality* $2^n>n^2$ *is true for all integral values of* $n\geqslant 5$. For some values of n less than 5 the inequality is also true ($n=1$) but for others it is not true ($n=2, 3, 4$).

5. *Prove that for* $n>1$ *the following inequality is true*:

$$n!<\left(\frac{n+1}{2}\right)^n \qquad (n!=1\cdot 2\cdot 3\ldots n)$$

For $n=2$ we obtain the true inequality $2<9/4$. Suppose that $k!<\left(\frac{k+1}{2}\right)^k$. Then by the induction hypothesis $(k+1)!=k!\,(k+1)<\left(\frac{k+1}{2}\right)^k(k+1)$. If we now prove that

$$\left(\frac{k+1}{2}\right)^k(k+1)<\left(\frac{k+2}{2}\right)^{k+1} \tag{3}$$

the theorem is proved because then

$$(k+1)!<\left(\frac{k+1}{2}\right)^k(k+1)<\left(\frac{k+2}{2}\right)^{k+1}$$

that is our inequality holds true for $n=k+1$.

Inequality (3) can clearly be rewritten as

$$2<\left(1+\frac{1}{k+1}\right)^{k+1}$$

But the binomial theorem yields

$$\left(1+\frac{1}{k+1}\right)^{k+1}=1+(k+1)\frac{1}{k+1}+\ldots>2$$

so that inequality (3) holds and the original inequality is thus proved.*

6. *Prove the theorem: if a product of $n\geqslant 2$ positive numbers is equal to* 1, *then their sum is greater than or equal to n, that is, if $x_1x_2\ldots x_n=1$, $x_1>0$, $x_2>0$, $\ldots$, $x_n>0$, then $x_1+x_2+\ldots+x_n\geqslant n$.*

If $n=2$, we have to prove the assertion: if $x_1x_2=1$, then $x_1+x_2\geqslant 2$. But this is obvious since the arithmetic mean $\frac{x_1+x_2}{2}$ of two positive numbers is greater than or equal to the geometric mean $\sqrt{x_1x_2}=1$, that is, $x_1+x_2\geqslant 2$. Besides, equality, $x_1+x_2=2$, is achieved only when $x_1=x_2=1$.

Making the induction hypothesis, let us take any positive numbers $x_1, \ldots, x_k, x_{k+1}$ that satisfy the condition $x_1\ldots x_{k-1}x_kx_{k+1}=1$. If each of these numbers is equal to 1, then the sum $x_1+\ldots+x_k+x_{k+1}=$ $=k+1$, so that in this case the inequality at hand is true.

But if this is not so, there will be, among them, a number less than 1 and a number exceeding 1. Suppose that $x_k>1$, $x_{k+1}<1$. We have

$$x_1\ldots x_{k-1}(x_kx_{k+1})=1$$

This is a product of k numbers and so the induction hypothesis is applicable and we can assert that

$$x_1+\ldots+x_{k-1}+x_kx_{k+1}\geqslant k$$

But then

$$\begin{aligned} x_1+\ldots+x_{k-1}+x_k+x_{k+1} &\geqslant k-x_kx_{k+1}+x_k+x_{k+1}\\ &=k+1+(x_k-1)(1-x_{k+1})>k+1 \end{aligned}$$

since $x_k-1>0$ and $1-x_{k+1}>0$, which completes the proof.

Note that we also established the fact that, in the relation being demonstrated, equality is only possible when all $x_i=1$; however, if not all x_i are equal to 1, then we have the sign of strict inequality in this relation.

From this theorem follows a *generalized inequality between the arithmetic mean and the geometric mean for $n\geqslant 2$ positive integers*:

$$\frac{x_1+\ldots+x_n}{n}\geqslant\sqrt[n]{x_1\ldots x_n},\quad x_1>0,\ \ldots,\ x_n>0$$

---

* For a different proof of this inequality that does not make use of the method of mathematical induction see Problem 15 of Sec. 1.8.

Indeed, denote $\sqrt[n]{x_1 \ldots x_n}$ by c and x_i/c by y. Then $y_1 \ldots y_n = \frac{x_1 \ldots x_n}{c^n} = 1$. By what has been proved, $y_1 + \ldots + y_n \geqslant n$, that is, $\frac{x_1 + \ldots + x_n}{c} \geqslant n$ or $\frac{x_1 + \ldots + x_n}{n} \geqslant c$. This completes the proof.

This inequality is widely used in proving other inequalities. For example, if we apply it to the integers 1, 2, ..., n, we immediately get the inequality

$$\sqrt[n]{1 \cdot 2 \ldots n} < \frac{1+2+\ldots+n}{n}$$

or $\sqrt[n]{n!} < \frac{n+1}{2}$, whence $n! < \left(\frac{n+1}{2}\right)^n$. In Problem 5 this inequality was proved by induction. The second proof is clearly simpler.

The method of mathematical induction finds extensive use outside of algebra too, such as in the proof of trigonometric relations and geometric propositions.

7. *Prove that for every positive integer n the inequality $|\sin nx| \leqslant n\,|\sin x|$ holds true.*

The inequality is obviously true for $n=1$. Assuming that $|\sin kx| \leqslant k \cdot |\sin x|$, we prove that $|\sin (k+1)x| \leqslant (k+1) \cdot |\sin x|$. Indeed, taking advantage of the inequality $|\cos kx| \leqslant 1$, we have

$$\begin{aligned}|\sin (k+1)x| &= |\sin kx \cdot \cos x + \sin x \cdot \cos kx| \\ &\leqslant |\sin kx| \cdot |\cos x| + |\sin x| \cdot |\cos kx| \\ &\leqslant |\sin kx| + |\sin x| \leqslant k\,|\sin x| + |\sin x| = (k+1) \cdot |\sin x|\end{aligned}$$

The required inequality is thus valid.

8. *Prove that for every natural number n the following equation holds true:*

$$\sin\frac{\pi}{3} + \sin\frac{2\pi}{3} + \ldots + \sin\frac{n\pi}{3} = 2\sin\frac{n\pi}{6}\sin\frac{n+1}{6}\pi \qquad (4)$$

For $n=1$ we obtain the true equality

$$\sin\frac{\pi}{3} = 2\sin\frac{\pi}{6}\sin\frac{\pi}{3}$$

Making the induction hypothesis, we consider the sum in the left-hand member of (4) for $n=k+1$:

$$\begin{aligned}&\sin\frac{\pi}{3} + \sin\frac{2\pi}{3} + \ldots + \sin\frac{k\pi}{3} + \sin\frac{(k+1)\pi}{3} \\ &= 2\sin\frac{k\pi}{6}\sin\frac{(k+1)\pi}{6} + \sin\frac{(k+1)\pi}{3} \\ &= 2\sin\frac{k\pi}{6}\sin\frac{(k+1)\pi}{6} + 2\sin\frac{(k+1)\pi}{6}\cos\frac{(k+1)\pi}{6} \\ &= 2\sin\frac{(k+1)\pi}{6}\left[\sin\frac{k\pi}{6} + \cos\frac{(k+1)\pi}{6}\right] \\ &= 2\sin\frac{(k+1)\pi}{6} \cdot 2\sin\frac{\pi}{6} \cdot \cos\frac{(k-1)\pi}{6} \\ &= 2\sin\frac{(k+1)\pi}{6}\cos\frac{(k-1)\pi}{6}\end{aligned}$$

To complete the proof it suffices to note that

$$\cos\frac{(k-1)\pi}{6}=\sin\left[\frac{\pi}{2}-\frac{(k-1)\pi}{6}\right]$$
$$=\sin\frac{(4-k)\pi}{6}=\sin\left(\pi-\frac{(4-k)\pi}{6}\right)=\sin\frac{(k+2)\pi}{6}$$

The proof of formula (4) is thus complete.*

9. *In a plane draw n straight lines such that no two are parallel and no three pass through a single point. Into how many parts do these lines divide the plane?*

After making appropriate drawings, we can write down the following correspondence between the number n of straight lines satisfying the condition of the problem and the number a_n of parts into which these lines partition the plane:

$$n=1,\ 2,\ 3,\ 4,\ 5,\ \ldots$$
$$a_n=2,\ 4,\ 7,\ 11,\ 16,\ \ldots$$

It is easy to see** that the following expression will serve as the general term of the sequence a_n:

$$a_n=1+\frac{n(n+1)}{2} \tag{5}$$

Formula (5) can readily be verified for the first few values of n, but this of course does not mean that the answer to the problem is at hand. The assertion requires additional proof by the method of mathematical induction.

For the moment, let us forget the foregoing and prove that n straight lines (no two of which are parallel and no three pass through a single point) divide the plane into a_n parts, where a_n is computed from formula (5).

It is obvious that (5) is true for $n=1$. Making the induction hypothesis, we consider $k+1$ straight lines which satisfy the conditions of the problem. Separating out k lines in arbitrary fashion, we can say that they divide the plane into $1+\frac{k(k+1)}{2}$ parts. Now adjoin the $(k+1)$th line. Since it is not parallel to any of the others, it will intersect all k straight lines. Since it does not pass through any of the intersection points of the other lines, it will pass through the $k+1$ pieces into which the plane is divided and will cut each of these pieces

* See Problem 7, Sec. 2.2, for a different proof of (4) that dispenses with mathematical induction.

** First note that, judging by the first terms, the sequence a_n is such that the differences a_2-a_1, a_3-a_2, a_4-a_3, ... form an arithmetic progression, and then take advantage of Problem 8 of Sec. 1.7.

into two pieces, thus adding another $k+1$ pieces. Thus, the total number of pieces into which the plane is divided by the $k+1$ straight lines is

$$1+\frac{k(k+1)}{2}+k+1=1+\frac{(k+1)[(k+1)+1]}{2}=a_{k+1}$$

The proof is complete.

It is evident from the foregoing that the method of mathematical induction is applied in the proof of propositions involving a natural number n. However, many propositions that do not involve n at all can be replaced by equivalent ones that depend explicitly on n. The following is an illustration.

10. *At any time, the total number of persons on the earth who shake hands an odd number of times is even.*

To prove this, assign to each handshake a number in natural (chronological) order. Then our assertion is equivalent to the following: for every n, after a handshake with number n, the number of people who have made an odd number of handshakes is even.

This assertion depends on n and we will prove it by induction. For brevity, call "bad" the people who have made an odd number of handshakes, and "good" the rest.

After the handshake with number 1, we have two bad people, an even number. After the kth handshake, the number of bad people is even, and let handshake number $k+1$ take place. Three cases are possible: it will occur between (a) two good people, (b) two bad people, (c) a good and a bad person.

In the first case, two good persons add one handshake to their even number and become bad, in the second case, two bad persons become good, and in the third, a good person becomes a bad one and a bad one goes into a good one. Thus, the number of bad people either increases by two or decreases by two, or remains unchanged; in any case the number remains even. The proof is complete.

The examples we have examined show how successfully the principle of mathematical induction can be used in diverse problems. At the same time, the strength of this method should not be exaggerated, for there are numerous problems that would seem to be specially made for the induction method, yet attempts to apply it fail.

As an instance, let us try proving by mathematical induction the inequality

$$\frac{1}{9}+\frac{1}{25}+\ldots+\frac{1}{(2n+1)^2}<\frac{1}{4}$$

For $n=1$ this inequality is of the form $1/9<1/4$, which means it is valid. Assume that it holds for $n=k$:

$$\frac{1}{9}+\frac{1}{25}+\ldots+\frac{1}{(2k+1)^2}<\frac{1}{4}$$

For $n=k+1$ the left member is equal to

$$\frac{1}{9}+\frac{1}{25}+\ldots+\frac{1}{(2k+3)^2}=\left[\frac{1}{9}+\frac{1}{25}+\ldots+\frac{1}{(2k+1)^2}\right]+\frac{1}{(2k+3)^2}$$

By the induction hypothesis, the sum in the square brackets is less than 1/4 and so

$$\frac{1}{9}+\frac{1}{25}+\ldots+\frac{1}{(2k+3)^2}<\frac{1}{4}+\frac{1}{(2k+3)^2}$$

It is clear that from the above inequality we cannot in the least conclude that the left member is less than 1/4. Thus, our proof by induction has come to a halt. Yet this inequality is easily proved by a different method (see Problem 13 of Sec. 1.8).

Exercises

Prove the following formulas by mathematical induction:

1. (a) $1^2+2^2+3^2+\ldots+n^2=\frac{n(n+1)(2n+1)}{6}$.

   (b) $1^2-2^2+3^2-4^2+\ldots+(-1)^{n-1}n^2=(-1)^{n-1}\frac{n(n+1)}{2}$.

2. $1^3+2^3+3^3+\ldots+n^3=\left[\frac{n(n+1)}{2}\right]^2$.

3. The number of permutations P_n of n elements, $n>1$, is
$$P_n=n!=1\cdot 2\cdot 3\ldots n$$

4. De Moivre's formula $[r(\cos\varphi+i\sin\varphi)]^n=r^n(\cos n\varphi+i\sin n\varphi)$.

5. (a) $\sin x+\sin 2x+\ldots+\sin nx=\dfrac{\sin\frac{n+1}{2}x\cdot\sin\frac{nx}{2}}{\sin\frac{x}{2}}$, $x\neq 2k\pi$.

   (b) $\sin x+\sin(x+h)+\ldots+\sin(x+nh)=\dfrac{\sin\left(x+\frac{nh}{2}\right)\cdot\sin\frac{n+1}{2}h}{\sin\frac{h}{2}}$,

   $h\neq 2k\pi$.

6. $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots+\frac{1}{2n-1}-\frac{1}{2n}=\frac{1}{n+1}+\frac{1}{n+2}+\ldots+\frac{1}{2n-1}+\frac{1}{2n}$.

Prove the following inequalities:

7. $n!>2^{n-1}$ if $n>2$.
8. $2^n\cdot n!<n^n$ if $n>2$.
9. $(n!)^2<\left[\frac{(n+1)(2n+1)}{6}\right]^n$.
10. $(2n)!<\left(\frac{2n+1}{2}\right)^{2n}$.
11. $\frac{a_1}{a_2}+\frac{a_2}{a_3}+\ldots+\frac{a_{n-1}}{a_n}+\frac{a_n}{a_1}\geqslant n$ if $a_1>0,\ \ldots,\ a_n>0$.

12. $\left(\frac{a^2+b^2+c^2}{a+b+c}\right)^{a+b+c} > a^a b^b c^c$ if a, b, c are distinct positive integers.

13. Prove that for every integer n, n^7-n is divisible by 7.

14. Prove that for every integer n the number $11^{n+2}+12^{2n+1}$ is divisible by 133.

15. Prove

$$\frac{1}{\log_x 2\cdot\log_x 4}+\frac{1}{\log_x 4\cdot\log_x 8}+\ldots+\frac{1}{\log_x 2^{n-1}\cdot\log_x 2^n}=\left(1-\frac{1}{n}\right)\cdot\frac{1}{\log_x^2 2}.$$

16. Prove that for every positive a and b and all natural values of n the inequality $(a+b)^n < 2^n(a^n+b^n)$ is true.

17. Prove that for every $a>0$ the inequality $\sqrt{a+\sqrt{a+\sqrt{a+\ldots+\sqrt{a}}}} < \frac{1+\sqrt{4a+1}}{2}$ (the left member contains an arbitrary number of radicals) holds true.

18. Prove that for all positive integers n and k the equation $C_{n+1}^0-C_{n+1}^1+C_{n+1}^2-\ldots+(-1)^k C_{n+1}^k=(-1)^k C_n^k$ is valid. (By definition, $C_m^0=1$ for every m and $C_p^q=0$ if $q>p$.)

1.4 Real numbers

Two things will engage us in this section: the absolute value (also called modulus) of a real number and the principal square root.

In most cases, students state the correct absolute values of given real numbers, but when it comes to giving the *definition* of the absolute value of a number, we often hear such meaningless phrases as: the absolute value of a number is the number without a sign or the number with the plus sign, or the positive value of the number. A clear-cut definition of this important concept is required at all times. The absolute value of a number a (denoted by $|a|$) is defined as follows:

$$|a|=\begin{cases} a & \text{if } a>0 \\ 0 & \text{if } a=0 \\ -a & \text{if } a<0 \end{cases}$$

This definition enables one to compute the absolute value of any real number. One only needs to use the first, second or third line of the definition, depending on whether the given number is positive, negative or zero.

For example, *what is the absolute value of the number* -3? A full answer is: $-3<0$, therefore, by the third line in the definition, the absolute value of the number -3 is $-(-3)=3$, that is, $|-3|=3$.

Noting that for $a=0$ the equality $|a|=a$ is true, we can write the definition of the absolute value more compactly (also see Exercise 1):

$$|a|=\begin{cases} a & \text{if } a\geqslant 0 \\ -a & \text{if } a<0 \end{cases}. \qquad (1)$$

From the definition of absolute value it follows that $|a| \geqslant 0$ for any a. This is a *theorem*, although it is a very simple one. In the proof we consider two cases:

(a) $a \geqslant 0$. Then $|a| = a \geqslant 0$, which completes the proof.

(b) $a < 0$. Then $|a| = -a$. But $-a > 0$ since $a < 0$, that is $|a| > 0$, which is what we set out to prove.

Bear in mind that the fact that the expression $|a|$ is always positive or zero is not a definition of absolute value but a corollary; the definition does not in any way involve the sign of the expression $|a|$.

It is easy to see that, geometrically, $|a|$ denotes a distance, that is, the length of a line segment of the number line (a positive number or 0) from the point a to zero. It can also be proved (by examining separate cases) that $|b-a|$ is the distance between the points a and b (see Exercise 5). These geometric representations are very useful in solving problems, and in elementary cases permit finding the answer immediately without resorting to the standard method which we now consider.

For example, the equation $|x-1| = 2$ is solved geometrically as follows: its solution set consists of points at a distance 2 from point 1; that is $x_1 = 3$, $x_2 = -1$. The same goes for the solution set of the inequality $|x+2| \leqslant 5$, which consists of points lying at a distance from -2 not exceeding 5, that is, of points in the interval $-7 \leqslant x \leqslant 3$.

The following absolute-value properties are very useful in problem solving:

For any real numbers a and b,

I. $|a+b| \leqslant |a| + |b|$,

II. $|a-b| \geqslant ||a| - |b||$,

III. $|ab| = |a||b|$,

IV. $\left|\frac{a}{b}\right| = \frac{|a|}{|b|}$ $(b \neq 0)$.

The easiest to prove are Properties III and IV. This is done simply by taking all possible combinations of the signs of a and b and is left to the reader as an exercise. Note the following important corollary to Property III: $|a|^2 = a^2$ for every a (indeed, putting $a = b$, we get $|a|^2 = |a^2|$, which is equal to a^2, since $a^2 \geqslant 0$).

To prove Property I note that

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

and

$$(|a| + |b|)^2 = |a|^2 + 2|a| \cdot |b| + |b|^2 = a^2 + 2|ab| + b^2$$

but $ab \leqslant |ab|$, so that

$$|a+b|^2 \leqslant (|a| + |b|)^2$$

But of two nonnegative numbers $|a+b|$ and $|a| + |b|$, the smaller one is that whose square is smaller. This proves Property I. A different proof may be given based on examining possible cases.

Property II may be proved in similar fashion or it may be derived from Property I. Namely, by Property I

$$|a| = |(a-b)+b| \leqslant |a-b| + |b|$$

whence $|a| - |b| \leqslant |a-b|$. Similarly we prove that $|b| - |a| \leqslant |a-b|$. But one of these two expressions, either $|a| - |b|$ or $|b| - |a|$ is nonnegative and consequently coincides with its absolute value so that $||a| - |b|| \leqslant |a-b|$, which is what we set out to prove.

Problems involving absolute values are, as a rule, worked by the standard procedure of getting rid of the modulus (absolute-value) sign: the definition is used in considering all cases of the distribution of signs of the expressions under the absolute-value sign, and in each of these cases the modulus is replaced either by the expression itself or by the expression with opposite sign. This gives us a problem in which there is no absolute-value sign.* This procedure is rather familiar to the student but two grave mistakes are commonly made when applying it.

The first mistake is connected with a misunderstanding (or improper handling) of the definition of absolute value: when in place of x we have some kind of expression $f(x)$ under the modulus sign (this occurs rather often), then in place of the proper equation

$$|f(x)| = \begin{cases} f(x) & \text{if } f(x) \geqslant 0 \\ -f(x) & \text{if } f(x) < 0 \end{cases} \tag{2}$$

we all too often find the student writing

$$|f(x)| = \begin{cases} f(x) & \text{if } x \geqslant 0 \\ -f(x) & \text{if } x < 0 \end{cases}$$

which is obviously quite incorrect.

The second mistake follows from an insufficient grasp of the logical essence of the procedure itself. Indeed, a consideration of specific instances, say in the solution of an equation or inequality means that in each case we see the solution only in some narrow region, namely in the region defined by the conditions of the specific case at hand. This forces us, after finding the solutions, to select those which appear in the indicated region, that is, such as satisfy the conditions defining the specific case. What frequently happens, however, is that the student correctly isolates the individual cases, solves the equation for each case, but leaves untouched the conditions of the cases, regarding them as something unnecessary.

True, both types of mistakes are sometimes merely the result of carelessness, but they still remain mistakes that have to be rectified.

---

* It is well to bear in mind however that this procedure is not in itself a solution to the problem at hand. Serious difficulties may arise after its application. The purpose of this technique is to remove difficulties connected with moduli and to reshape the problem so that there is no absolute-value sign.

Generally speaking, in problems which involve a consideration of a large number of cases, care, accuracy and attentiveness on the part of the student are of prime importance.

Let us consider a number of examples.

1. *Solve the equation* $x^2-2|x|-3=0$.

To get rid of the absolute-value sign, we consider two cases: (a) $x \geqslant 0$ and (b) $x<0$.

In Case (a) we obtain the equation $x^2-2x-3=0$, whose roots are $x_1=3$, $x_2=-1$. But for (a) we need only $x \geqslant 0$ so that only $x=3$ is a root of the original equation.

In Case (b) the equation becomes $x^2+2x-3=0$, whose roots are $x_1=1$, $x_2=-3$. But in this case, by Condition (b), we are only interested in negative roots, $x=-3$.

Thus the original equation has the roots $x_{1,2}=\pm 3$.*

2. *Solve the equation* $|x^2-x-6|=x+2$.

Consider two cases in succession:

(a) $x^2-x-6<0$. In this case we have the equation $-x^2+x+6=$ $=x+2$ with roots $x_1=2$; $x_2=-2$. Now check to see whether x_1 and x_2 satisfy Condition (a). To do this, substitute these values into the left member of the inequality $x^2-x-6<0$. We then obtain the numerical inequalities $-4<0$ and $0<0$. The first is valid, the second is not; and so only 2 is a root of the original equation.

(b) $x^2-x-6 \geqslant 0$. In this case we have the equation $x^2-x-6=x+2$ whose roots are $x_1=4$, $x_2=-2$. Since both of these values of x satisfy Condition (b), both 4 and -2 are roots of the original equation.

Thus, the original equation has three roots: -2, 2, 4.

Let us examine this solution more carefully: first we rejected the value $x=-2$ and then found it again, so that in the end this value was a root of the original equation. How is this to be explained? The point is that in the first case we rejected $x=-2$ but we did not assert that it is not a root of the original equation. The only thing we maintained was that this value is discarded due to the restrictions imposed on x by the condition of Case (a). Quite naturally, there is nothing to stop this value from satisfying the condition of another case and thus to become a root of the original equation.

We now examine a problem in which the first of the two blunders mentioned above have frequently been made at examinations.

3. *Solve the inequality*

$$|x^2+3x|+x^2-2 \geqslant 0$$

In accordance with the definition of absolute value we have to consider two cases:

$$\text{(a) } x^2+3x \geqslant 0, \qquad \text{(b) } x^2+3x<0$$

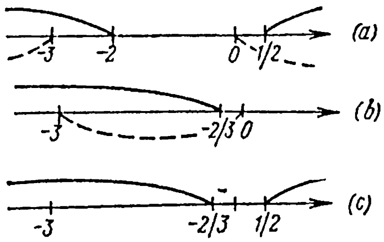
---

* Note that by the substitution $y=|x|$ we can reduce the given equation to a quadratic equation.

yet some students regarded the cases $x \geqslant 0$ and $x < 0$. In the first instance, the absolute-value sign can indeed be removed; for $x \geqslant 0$ the inequality $x^2+3x \geqslant 0$ is also true, but for $x<0$ we cannot say anything about the sign of x^2+3x, yet this did not deter these students from writing, for $x<0$, $|x^2+3x| = -x^2-3x$ or even $|x^2+3x| = x^2-3x$.

In the correct solution, for Case (a) we get the inequality $2x^2+3x-2 \geqslant 0$ the solutions of which are $x \leqslant -2$ and $x \geqslant 1/2$. Condition (a) is satisfied for $x \leqslant -3$ and for $x \geqslant 0$. We now have to choose from these solutions those which satisfy Condition (a), or $x \leqslant -3$ and $x \geqslant 1/2$. This is most easily done in Fig. 1*a*. We get the solution [for Case (a)]: $x \leqslant -3$ and $x \geqslant 1/2$.

Fig. 1



In Case (b) we have the inequality $-3x-2 \geqslant 0$ or $x \leqslant -2/3$. Condition (b) is satisfied for $-3<x<0$, so that out of all the $x \leqslant -2/3$ there only remain the values of x which lie in the interval $-3<x \leqslant -2/3$ (Fig. 1*b*).

Combining the solutions found in (a) and (b) (Fig. 1*c*), we get the answer: $x \leqslant -2/3$ and $x \geqslant 1/2$.

4. *Solve the inequality* $2|3+5x-2x^2|<1-x$.

Consider two cases:

(a) $3+5x-2x^2 \geqslant 0$. In this case the given inequality can be rewritten as $2(3+5x-2x^2)<1-x$ or, after simplification, as $4x^2-11x-5>0$. This inequality holds true for $x>(11+\sqrt{201})/8$ and for $x<(11-\sqrt{201})/8$. But of these values of x only those that satisfy also the condition of the case at hand, that is, the inequality $3+5x-2x^2 \geqslant 0$, can be retained as satisfying the original inequality. Solving this inequality, we find that it is satisfied for $-1/2 \leqslant x \leqslant 3$.

Now we have to choose from the intervals $x>(11+\sqrt{201})/8$ and $x<(11-\sqrt{201})/8$ those values of x which simultaneously lie in the interval $-1/2 \leqslant x \leqslant 3$. This is easily accomplished on the number axis. Mark points $(11-\sqrt{201})/8$, $(11+\sqrt{201})/8$, $-1/2$ and 3 (Fig. 2). From the figure it is clear that not a single value of x satisfying the inequality $x>(11+\sqrt{201})/8$ lies in the interval $-1/2 \leqslant x \leqslant 3$, which is to say that there is not a single solution of the original inequality among these values of x. Among the values of $x<(11-\sqrt{201})/8$ there will be some which appear in this interval; these are all values of x

in the interval

$$-\frac{1}{2} \leqslant x < \frac{11-\sqrt{201}}{8}$$

They form the solution set of the original inequality in the case at hand.

(b) $3+5x-2x^2<0$. In this case we have the inequality $2(2x^2-5x-3)<1-x$ or $4x^2-9x-7<0$. Solving this inequality, we get

$$\frac{9-\sqrt{193}}{8} < x < \frac{9+\sqrt{193}}{8}$$

However, from these values of x we have to choose only such as at the same time satisfy the inequality $3+5x-2x^2<0$ whose solu-

Fig. 2

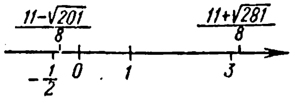
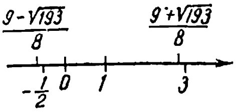


Fig. 3



tion set constitutes two regions: $x<-1/2$ and $x>3$. From Fig. 3 it is evident that the interval

$$\frac{9-\sqrt{193}}{8} < x < -\frac{1}{2}$$

is the solution set of the original inequality in the case at hand.

Thus, the solution set of the original inequality consists of two intervals

$$\frac{9-\sqrt{193}}{8} < x < -\frac{1}{2} \quad \text{and} \quad -\frac{1}{2} \leqslant x < \frac{11-\sqrt{201}}{8}$$

It is easy to see that these two intervals combine into one so that the final solution of the given inequality is the interval

$$\frac{9-\sqrt{193}}{8} < x < \frac{11-\sqrt{201}}{8}$$

A few remarks are in order on how to make drawings like those shown in Figs. 2 and 3. The most important thing is to be very careful in plotting on the number line the points which correspond to the given numbers and see that the proper sequential order of the points is maintained. For example, if the numbers are almost alike, do not bunch them together but spread them out even if the scale becomes somewhat exaggerated. In some cases plotting the points in the required order is so complicated that one has to resort to approximate computations and even, occasionally, to proving numerical inequalities.

For example, in Fig. 2 the number $(11-\sqrt{201})/8$ is to the right of $-1/2$. This follows immediately from the easily proved inequality $-1/2<(11-\sqrt{201})/8$. In the same way, the number $(11+\sqrt{201})/8$

lies to the right of the number 3 since $3<(11+\sqrt{201})/8$ (because $\sqrt{201}>14$ and, hence, the numerator of the fraction in the right member exceeds 25).

Let us now consider some examples in which there are several expressions under the absolute-value sign. To get rid of the absolute-value sign in these examples, the procedure calls for examining all possible combinations of these expressions. That is what we will do in the first example. In the next two we will show how this can be circumvented.

5. *Solve the system of equations*

$$|x^2-2x|+y=1$$
$$x^2+|y|=1$$

There are four possible distinct combinations of signs in the expressions under the absolute-value sign:

$$\text{(a) } x^2-2x\geqslant 0,\ y\geqslant 0, \quad \text{(b) } x^2-2x\geqslant 0,\ y<0,$$
$$\text{(c) } x^2-2x<0,\ y\geqslant 0, \quad \text{(d) } x^2-2x<0,\ y<0$$

We consider each one in succession.

(a) In this case we have the system

$$x^2-2x+y=1$$
$$x^2+y=1$$

whence we readily get $x=0$, $y=1$. This pair satisfies Condition (a) and therefore is a solution of the original system.

(b) In this case the system is of the form

$$x^2-2x+y=1$$
$$x^2-y=1$$

whence, adding the equations we get $x^2-x=1$, that is, $x_{1,2}=(1\pm\sqrt{5})/2$. The appropriate values of y may be computed from the second equation, but this can be done more simply as follows: indeed, x_1 and x_2 satisfy $x^2-x=1$ and, comparing it with the second equation, we get $y=x$, or $y_{1,2}=(1\pm\sqrt{5})/2$.

We check Condition (b). Since $y_1<0$, the pair x_1, y_1 does not satisfy it and must be rejected. Now, since y_2 satisfies (b), and for x_2 the inequality $x^2-2x\geqslant 0$ is true (because $x_2<0$), in this case we have the solution $x=(1-\sqrt{5})/2$, $y=(1-\sqrt{5})/2$.

(c) In this case we get the system

$$-x^2+2x+y=1$$
$$x^2+y=1$$

whence, subtracting the first equation from the second, we get $x^2-x=0$, or $x_1=0$, $x_2=1$, and $y_1=1$, $y_2=0$. The pair x_1, y_1 does not satisfy (c) and x_2, y_2 satisfies (c); hence, the pair $x=1$, $y=0$ is the solution set in this case.

(d) Here we have the system

$$-x^2+2x+y=1$$
$$x^2-y=1$$

whence, adding, we get $x=1$ and so $y=0$. But this pair does not satisfy Condition (d) and so must be rejected, although it is a solution of the original system. The situation here is the same as in Example 2 where this was explained.

Thus, the system of equations has the following three solutions:

$$x_1=0,\ y_1=1;\ x_2=(1-\sqrt{5})/2,\ y_2=(1-\sqrt{5})/2;$$
$$x_3=1,\ y_3=0$$

6. *Solve the inequality*

$$|x-1|-|x|+|2x+3|>2x+4$$

This problem requires considering a total of 8 combinations of signs, but we can manage things so as to consider only four. This is achieved by a special technique called the "method of intervals".

Mark on the number line those values of x for which each of the expressions under the absolute-value sign vanishes: the points $-3/2$, 0 and 1. Thus, the entire number line is divided into four intervals:*

$$x<-\frac{3}{2},\ -\frac{3}{2}\leqslant x<0,\ 0\leqslant x<1,\ 1\leqslant x$$

Let us consider each of these regions.

(a) $x<-\frac{3}{2}$. In this case, $2x+3<0$, $x<0$ and $x-1<0$, i.e. the initial inequality takes the form $-x+1+x-2x-3>2x+4$. It is satisfied for $x<-3/2$; in conjunction with Condition (a) we find that $x<-3/2$ is a solution of the original inequality.

(b) $-3/2\leqslant x<0$. In this case, $2x+3\geqslant 0$, $x<0$ and $x-1<0$, and therefore the original inequality takes the form $-x+1+x+2x+3>$ $>2x+4$, i.e. $0>0$.

This inequality is usually a stumbling block. How is it to be resolved? Actually, of course, there is nothing to solve: it is simply that for every x in the interval $-3/2\leqslant x<0$, the original inequality turns into the invalid inequality $0>0$ and therefore has no solution for Case (b).

(c) $0\leqslant x<1$. In this case, $2x+3\geqslant 0$, $x\geqslant 0$ and $x-1<0$; consequently the original inequality reduces to the inequality $-x+1-$

* Note that the intervals may also be written thus: $x\leqslant -3/2$, $-3/2<x<0$, $0\leqslant x\leqslant 1$, $1<x$. It is easy to see that this does not change anything in the solution. Our choice was made in accordance with the definition of absolute value in the form (1).

$-x+2x+3>2x+4$. It is satisfied for $x<0$. But this relation is inconsistent with Condition (c): there are no solutions of the original inequality among the values of x in the interval $0\leqslant x<1$.

(d) $1\leqslant x$. In this case the inequality takes the form $x-1-x+2x+3>2x+4$, or $2>4$; in other words, there are no values among $x\geqslant 1$ that satisfy the original inequality.

Hence the proposed inequality holds true for $x<-3/2$.

It is quite clear from the foregoing that the function

$$y=|x-1|-|x|+|2x+3|$$

may be written in the following form dispensing with the absolute-value sign altogether:

$$y=\begin{cases} -2x-2 & \text{f} \quad x<-\frac{3}{2}, \\ 2x+4 & \text{if} \quad -\frac{3}{2}\leqslant x<0, \\ 4 & \text{if} \quad 0\leqslant x<1, \\ 2x+2 & \text{if} \quad 1\leqslant x \end{cases}$$

This form of representing a function involving the absolute-value sign can be very useful in problem solving.

In the following problem there are difficulties besides the two moduli. True, it is more the nature of the difficulties than the degree which deter the student. Incidentally, Problem 6 (Case (b)) is of this kind.

7. *Solve the equation*

$$|x^2-9|+|x^2-4|=5$$

Following the method of Problem 6, we consider three cases:* (a) $x^2<4$, (b) $4\leqslant x^2\leqslant 9$, (c) $9<x^2$.

In the first case, $|x^2-9|=9-x^2$, $|x^2-4|=4-x^2$ or

$$9-x^2+4-x^2=5,\ x^2=4,\ x_{1,2}=\pm 2$$

But x^2 must be less than 4 in the first case, so the values $x_{1,2}=\pm 2$ are unsuitable and the given equation has no roots in this case.

In the second case, $|x^2-9|=9-x^2$, $|x^2-4|=x^2-4$ or $9-x^2+x^2-4=5$, or $5=5$. At this point some students think the equation has "disappeared". Actually, nothing serious has occurred; simply the original equation is equivalent to the identity $5=5$ when $4\leqslant x^2\leqslant 9$, which is to say it is satisfied for all values of x. This means that any value of x which satisfies the condition $4\leqslant x^2\leqslant 9$ is a solution of the equation. It now remains to solve this double inequality. We then get $-3\leqslant x\leqslant -2$, $2\leqslant x\leqslant 3$.

Actually, at this examination, some students did just the opposite; they wrote: "the equation becomes an identity and therefore has no

* Here again we could write $x^2<4$, $4\leqslant x^2<9$, $9\leqslant x^2$ instead, and the final answer would naturally be the same.

solutions for $4 \leqslant x^2 \leqslant 9$." This is a bad misunderstanding which consists in regarding an identity as something quite different from an equation. Actually an identity is a special case of an equation.

The third case is considered in a manner similar to the first; no new solutions appear.

Finally, the roots of the original equation fill two intervals of the number line, which is rather unusual for equations (unlike the situation for inequalities).

No less interesting in this respect is the following example, where the solutions constitute an infinite interval and one more point.

8. *Find the solution to the equation*

$$2^{|x+2|} - |2^{x+1} - 1| = 2^{x+1} + 1$$

We consider two cases.

(a) $x+2 \geqslant 0$. In this case, since $2^{|x+2|} = 2^{x+2}$ and $2^{x+2} - 2^{x+1} = 2^{x+1}$, we get the equation $2^{x+1} - 1 = |2^{x+1} - 1|$. This equation is obviously satisfied for $2^{x+1} - 1 \geqslant 0$, or, what is the same thing, $x+1 \geqslant 0$, that is $x \geqslant -1$. These values of x satisfy Condition (a) and are therefore roots of our equation.

(b) $x+2 < 0$. Here, simple manipulations and the substitution of y for 2^{x+1} yields

$$2y^2 + 2y + 2y\,|y-1| = 1$$

This equation can be solved by considering, as in the previous examples, two cases for getting rid of the moduli. But it can also be seen at a glance that for $y \geqslant 1$ the left side exceeds 1 and therefore all we need to do is seek the roots $y < 1$; but when $y < 1$ we get $4y = 1$, whence $y = 1/4$ so that $x = -3$.

Combining the solutions obtained in Cases (a) and (b), we have the answer: $x = -3$ and $x \geqslant -1$.

The foregoing examples show clearly enough that the concept of absolute value does not put up insurmountable barriers since the absolute-value sign can always be eliminated by the standard procedure of considering separate cases. Quite naturally, running through individual cases is not the only way to solve problems involving moduli.

Very often the peculiarities of a specific problem permit finding other, shorter and more elegant solutions. This suggests that the student should not start out immediately with separate cases as soon as he sees an absolute-value sign. This approach will always be there if nothing else avails. It is best first to take a hard look and examine the problem for other approaches.

Sometimes a unique device is found that leads to a solution directly, as witness the following simple problem.

9. *Solve the inequality* $x^2 + x + |x| + 1 \leqslant 0$.

The standard technique can be invoked of course, but if we rewrite the inequality as $|x| \leqslant -(x^2 + x + 1)$, it will be seen immediately that it does not have a solution. Indeed, $|x| \geqslant 0$ for all values of x, and the right-hand member of the latter inequality is always strictly negative because $x^2+x+1=(x+1/2)^2+3/4>0$.

In the next problem the advantage of a special technique is conspicuous since the standard procedure requires tedious computations involving irrational numbers, while this simple device produces the answer very quickly.

10. *Solve the inequality*

$$|x^2-3x-3|>|x^2+7x-13|$$

It will be recalled that, when squared, an inequality with nonnegative members is replaced by an equivalent inequality (see Sec. 1.10). Our inequality is equivalent to the following one:

$$|x^2-3x-3|^2>|x^2+7x-13|^2$$

But $|a|^2=a^2$ so that this inequality can be rewritten as

$$(x^2-3x-3)^2>(x^2+7x-13)^2$$

Now, transposing all terms to the right side and using the formula for the difference of squares, we get

$$2(x^2+2x-8)\cdot 10(x-1)<0$$

or, what is the same thing,

$$(x+4)(x-2)(x-1)<0$$

This inequality is readily solved by the so-called method of intervals (see Sec. 1.10). Its solutions, and consequently the solutions of the original inequality are $x<-4$ and $1<x<2$.

In concluding this examination of the concept of absolute value, we give a problem whose complexity lies in the presence of a *parameter*. However, as we will see, the factor of a parameter already makes the problem rather involved, requiring both a knowledge of method and a proper technique of solution, and also considerable accuracy.

11. *Solve the equation* $x|x+1|+a=0$ *for every real number* a.

We consider two cases: $x<-1$ and $x\geqslant-1$. In the former case, the equation takes the form $x(-x-1)+a=0$ or $x^2+x-a=0$. This is a quadratic equation with parameter a. We are interested only in those real roots of the equation that satisfy the condition $x<-1$. Naturally the roots depend on the parameter a: for certain values of a the roots may be real, for others, imaginary. For this reason we must first indicate the values of a for each of which the roots of the equation $x^2+x-a=0$ are real. The condition for the reality of the roots is the nonnegativity of the discriminant: $D=1+4a\geqslant 0$. In other words, the roots of the equation are real for $a\geqslant -1/4$:

$$x_1=\frac{-1+\sqrt{1+4a}}{2},\quad x_2=\frac{-1-\sqrt{1+4a}}{2}$$

For the remaining values of a, that is, for $a<-1/4$, the roots of this equation are imaginary. Consequently, for $a<-1/4$ (in the first case now under consideration!) the original equation has no solutions.

Thus it remains to seek the solutions of the original equation in this case for $a\geqslant-1/4$. Then, of the numbers x_1 and x_2 thus found we will have to take those which satisfy the condition $x<-1$.

To do this we have to solve the inequalities

$$\frac{-1+\sqrt{1+4a}}{2}<-1 \quad\text{and}\quad \frac{-1-\sqrt{1+4a}}{2}<-1$$

The first one can readily be reduced to the form $1+\sqrt{1+4a}<0$, which means it is not valid for any values of a. The second inequality is reduced to $1<\sqrt{1+4a}$ and is valid for $a>0$, as can readily be seen.

Thus, for $a>0$ the original equation has one real root $x=(-1-\sqrt{1+4a})/2$ that satisfies the condition of the case at hand, $x<-1$, and for $a\leqslant 0$ does not have any such root.

In the latter case, we have the equation $x^2+x+a=0$. The condition for reality of the roots, $D=1-4a\geqslant 0$, shows that this equation has real roots only for $a\leqslant 1/4$; for $a>1/4$ (in this second case) the original equation does not have any solution. It remains to find, among $a\leqslant 1/4$, those values of a for which the roots of the equation $x^2+x+a=0$ satisfy the condition of the case $x\geqslant -1$, that is, to solve the inequalities

$$\frac{-1+\sqrt{1-4a}}{2}\geqslant-1 \quad\text{and}\quad \frac{-1-\sqrt{1-4a}}{2}\geqslant-1$$

The first inequality is reducible to the form $1+\sqrt{1-4a}\geqslant 0$ and hence is valid for all admissible values of a, that is for $a\leqslant 1/4$. The second inequality can be reduced to $\sqrt{1-4a}\leqslant 1$ and is thus valid for all admissible positive values of a, that is, for $0\leqslant a\leqslant 1/4$.

Thus, in the domain $x\geqslant -1$ the original equation has two real roots for $0\leqslant a\leqslant 1/4$:

$$x'=\frac{-1+\sqrt{1-4a}}{2},\quad x''=\frac{-1-\sqrt{1-4a}}{2}$$

(for $a=1/4$ these roots are coincident), and for $a<0$, only one real root $x=(-1+\sqrt{1-4a})/2$. There are no roots of the original equation for $a>1/4$ in the domain $x\geqslant -1$.

To summarize, then, the results of these two cases are

$$x=\frac{-1+\sqrt{1-4a}}{2} \quad \text{for } a<0,$$

$$x_1=\frac{-1-\sqrt{1+4a}}{2},\quad x_2=\frac{-1+\sqrt{1-4a}}{2},\quad x_3=\frac{-1-\sqrt{1-4a}}{2}$$

$$\text{for } 0\leqslant a\leqslant\frac{1}{4}$$

(for $a=0$ we have $x_1=x_3$; for $a=1/4$ we have $x_2=x_3$);

$$x=\frac{-1-\sqrt{1+4a}}{2} \quad \text{for } a>\frac{1}{4}$$

In place of remarking that the roots coincide, we could have allotted the cases $a=0$ and $a=1/4$ separate lines:

$$x_1=-1, \ x_2=0 \ \text{ for } a=0,$$
$$x_1=\frac{-1-\sqrt{2}}{2}, \ x_2=-\frac{1}{2} \text{ for } a=\frac{1}{4}$$

Closely connected with the notion of the absolute value of a real number is the concept of the principal square root. Let us ask the question: *What is $\sqrt{x^2}$ equal to*? In other words, how can this expression be written without the square-root sign? The most common answer is: $\sqrt{x^2}=x$ for every x. It is easy to see that this is not the correct answer. Indeed, for $x=-2$ for instance we have

$$\sqrt{(-2)^2}=\sqrt{4}=2\neq-2$$

Another answer one often hears is that $\sqrt{x^2}=\pm x$. This is also wrong because, by definition, $\sqrt{x^2}$ is a definite number and not two numbers: $+x$ and $-x$.

To get at the heart of the problem, let us recall the basic definitions and facts relating to the concept of a root.

Definition 1. *A number b is called a square root of a number a if $b^2=a$.*

According to this definition the two assertions "b is a square root of a" and "$b^2=a$" are equivalent.

To stress one essential peculiarity of this definition, let us compare it, say, with the definition of the square of a number: the square of a number b is the product of this number by itself. This definition has the merit that it gives a rule for finding the number b^2. In contrast, the definition of a square root is not so good in that it not only fails to indicate a rule for computing the square root, but it does not even follow therefrom whether the square root of a given number a can always be extracted, and how many roots can be extracted, that is how many distinct numbers b can be found that satisfy the equality $b^2=a$. Hence, the first thing we have to do is investigate the problem of the existence and number of square roots of the given number a.

The complete result is given by three assertions:

(1) *If a is positive, then there are exactly two square roots of a, one of them being positive, the other negative.*

(2) *If $a=0$, there is one square root of a; it is equal to zero.*

(3) *If a is negative, the square root of a cannot be extracted (note that in this section we consider only real numbers).*

The mathematics curriculum of secondary school does not include the proof of the existence of a positive root of a positive number.* The remaining assertions in (1) to (3) are easy to prove and are left to the reader as an exercise.

Consider the positive number a. We can extract two square roots of this number. In order to distinguish between them, we introduce the concept of principal square root.

Definition 2. *The positive square root of a positive number is called the principal square root of that number.*

The principal square root of a is denoted by the symbol $\sqrt{a}$. The expression $\sqrt{0}$ is always understood to mean the unique root of zero, that is, 0.

Thus, the statement "b is the principal square root of a" is equivalent to the following two assertions taken together: "$b^2=a$" and "$b \geqslant 0$"; it is assumed here that a is a positive number or zero. If b is the principal square root of a, then the other root is equal to $-b$.

To summarize, then, the expression $\sqrt{x^2}$ of which we spoke at the beginning is not just any number which, when squared, yields x^2, but is definitely a *positive number or zero*.

Then what is $\sqrt{x^2}$ equal to?

For the sake of convenience we will assume that $x \neq 0$ since for $x=0$ we clearly have $\sqrt{0^2}=\sqrt{0}=0$. By Definition 2 we know that $\sqrt{x^2}$ represents a positive number which, squared, yields x^2. It is easy to see that the numbers x and $-x$ (and only these numbers) have this property. But there is only one positive number among them, and it is precisely this positive number that is equal to $\sqrt{x^2}$.

Thus, *if x is positive, then $\sqrt{x^2}=x$, if $-x$ is positive* (that is, *x is negative*), *then $\sqrt{x^2}=-x$.* And so we can form the following table:

$$\sqrt{x^2}=\begin{cases} x & \text{if } x>0 \\ 0 & \text{if } x=0 \\ -x & \text{if } x<0 \end{cases}$$

Using absolute-value notation, we can write compactly

$$\sqrt{x^2}=|x| \tag{3}$$

for an arbitrary (real) x.

The foregoing is very important with respect to algebraic and trigonometric manipulations. Disregard of this property can lead to serious mistakes (see Problem 1, Sec. 2.2).

---

* A procedure is also indicated that permits computing a square root to any preassigned degree of accuracy. The student should know and be able to use this procedure for computing the square root of a given number. Note, however, that the school curriculum does not include proof of the fact that this procedure does indeed permit computing a square root.

If not all letters in algebraic expressions involving radicals denote nonnegative numbers, always use formula (3) when performing identity transformations.

12. *Simplify the expression* ($a>0$, $a\neq 1$)

$$\frac{a^{-x}}{\sqrt{5}}\left[2a^{2x}-a^{x}(2a^{x}-1)\right]\left[1-\left(\frac{\sqrt{5}a^{x}}{2a^{x}-1}\right)^{-2}\right]^{-\frac{1}{2}}$$
$$\times\sqrt{(a^{x}+2)^{2}-5}-(a^{2x}+4)\left[a^{2x}+4(1-a^{x})\right]^{-\frac{1}{2}}$$
$$+4a^{x}\left[1+(a^{x}+2)(a^{2x}-4a^{x}+4)^{-\frac{1}{2}}\right]$$
$$\times\left[a^{x}+2+(a^{2x}-4a^{x}+4)^{\frac{1}{2}}\right]^{-1}$$

and determine for which values of x this expression is equal to unity.

Manipulating algebraically, reduce this expression to the simplest possible form. Taking advantage of the definitions of fractional and negative powers, we can transform the first summand to the form a^x and the third to

$$\frac{4a^{x}}{\sqrt{a^{2x}-4a^{x}+4}}$$

Hence, we can rewrite the given expression as

$$a^{x}-\frac{a^{2x}-4a^{x}+4}{\sqrt{a^{2x}-4a^{x}+4}}=a^{x}-\sqrt{(a^{x}-2)^{2}}$$

We leave it to the reader to carry out all the formal transformations.

We already know that the latter expression cannot be written as $a^x-(a^x-2)=2$: since the difference a^x-2 is not necessarily positive, an answer such as "the proposed expression for all values of x is equal to 2" is erroneous. The true answer is: "The original expression can be transformed to the form $a^x-|a^x-2|$."

It remains to find those values of x for which

$$a^{x}-|a^{x}-2|=1$$

If $a^x\geqslant 2$, then this equation quite obviously has no solution. But if $a^x<2$, then we have the equation $a^x-(2-a^x)=1$, for x, that is, $a^x=3/2$. Noting likewise that the condition $a^x<2$ is also valid for this value of x, we find the desired value to be $x=\log_a 3/2$.

Exercises

1. Are the following equations valid?

(a) $|a|=\begin{cases} a & \text{if } a>0,\\ -a & \text{if } a\leqslant 0\end{cases}$ (b) $|a|=\begin{cases} a & \text{if } a\geqslant 0,\\ -a & \text{if } a\leqslant 0\end{cases}$

2. Prove that: (a) $|x|=|-x|$, (b) $x\leqslant|x|$.

3. Prove that if $|a|=0$, then $a=0$.

4. What can be said about the numbers $a_1, \ldots, a_n$, if it is known that $|a_1|+\ldots+|a_n|=0$?

5. Prove that the distance between the points a and b on the number line is equal to $|b-a|$.

6. Solve the inequalities (a) $|x-a|<b$, (b) $|x-a|\geqslant b$, where a and $b>0$ are given numbers, and give a geometric interpretation of the solutions.

Solve the following equations and inequalities:

7. $|3x-4|=1/2$.
8. $|x+1|+2=2$.
9. $||x-1|+2|=1$.
10. $|x-3|>-1$.
11. $|34+21x-x^2|\leqslant -1$.
12. $|4-3x|\leqslant 1/2$.
13. $|x|+x^3=0$.
14. $|x^2-1|+x+1=0$.
15. $|2x-x^2-3|=1$.
16. $(1+x)^2\geqslant |1-x^2|$.
17. $|x^2-6x+8|\leqslant 4-x$.
18. $|x^2+4x+3|>x+3$.
19. $|x-1-x^2|\leqslant |x^2-3x+4|$.
20. $|x^3-1|\leqslant x^2+x+1$.
21. $|x^2-4x+2|=(5x-4)/3$.
22. $(x+1)(|x|-1)=-1/2$.
23. $|x|-2|x+1|+3|x+2|=0$.
24. $9^{-|x|}=(1/2)^{|x+1|+|x-1|}$.
25. $(x+4)\,3^{1-|x-1|}-x=(x+1)|3^x-1|+3^{x+1}+1$.
26. $|x-1|\geqslant (x+1)/2$.
27. $|x-2|<x/2$.
28. $|x^2-2x-3|<3x-3$.
29. $x^2-|3x+2|+x\geqslant 0$.
30. $x^2+2|x+3|-10\leqslant 0$.
31. Solve the system of equations

$$\begin{gathered} y-2|x|+3=0 \\ |y|+x-3=0 \end{gathered}$$

Solve the following equations for every positive number a:

32. $x^2+|x|+a=0$.
33. $144^{|x|}-2\cdot 12^{|x|}+a=0$.
34. $9^{-|x-2|}-4\cdot 3^{-|x-2|}-a=0$.
35. Prove that if the numbers x, y are of one sign, then

$$\left|\frac{x+y}{2}-\sqrt{xy}\right|+\left|\frac{x+y}{2}+\sqrt{xy}\right|=|x|+|y|$$

36. Simplify the expressions

$$\sqrt[4]{x^2},\ \sqrt{x^4},\ \sqrt[6]{x^6},\ \sqrt[5]{x^5},\ \sqrt{x^6y},\ \sqrt[5]{x^{15}y^{10}}$$

37. Is the equation $a\sqrt{b}=\sqrt{a^2b}$ always true?
38. Simplify the expression $\sqrt{9-6a+a^2}+\sqrt{9+6a+a^2}$ if $a<-3$.
39. Simplify

$$\frac{1}{\sqrt{x+2\sqrt{x-1}}}+\frac{1}{\sqrt{x-2\sqrt{x-1}}}$$

for $1<x<2$.

40. Simplify the expression

$$\sqrt{(1-\cos\alpha\cos\beta)^2-\sin^2\alpha\sin^2\beta}$$

1.5 Complex numbers

One of the chief difficulties of this topic involves the definition of a complex number and the rules for handling complex numbers.

Students sometimes define complex numbers this way: "a complex number is a number of the form $a+bi$, where a and b are real numbers and $i=\sqrt{-1}$". This definition is not clear at all. First of all (see Sec. 1.4) the radical sign is used to denote the principal square root of a positive real number, so what is $\sqrt{-1}$?

The basic definitions of complex numbers may be given as follows.

A complex number is an expression $a+bi$, where a and b are real numbers and i is a symbol. Two complex numbers $a+bi$ and $c+di$ are defined to be equal if $a=c$ and $b=d$.

Algebraic operations involving complex numbers are, by definition, performed according to the same rules as are operations involving real numbers with the convention that i^2 is replaced by -1.

One can then list the formulas for addition, subtraction, multiplication and division of complex numbers that follow from this definition.

The definition of a complex number given above is, strictly speaking, not rigorous. We give here one of several possible logically rigorous constructions of the theory of complex numbers.

In order to construct the complex numbers we will consider *formal expressions* of the type $a+bi$, where a and b are real numbers. The term "formal" indicates that we do not attribute any meaning to them, we do not ask what they might signify and we do not try to relate such expressions to any real thing. We regard them in a strictly formal fashion: to obtain such an expression we have to take two equal and distinct real numbers a and b, and, using the auxiliary symbols $+$ and i, construct expressions of the kind indicated above. For instance,

$$2+3i,\ 2+(-3i),\ 2+0i,\ 0+1i,\ (-\pi)+\sqrt{3}i$$

Likewise, we say nothing about the meaning of the auxiliary symbols $+$ and i. The $+$ sign here is not the sign of addition that we are used to, for we have always added only real numbers! So we regard this $+$ sign as a formal symbol (a straight cross, as it were). Its sole purpose is to help in the construction of the formal expressions which we desire.

Since these expressions are absolutely new entities as far as we are concerned, we will have to agree on how to distinguish them one from another, and in what cases they are to be regarded as equivalent (equal). We stress the fact that we have *to come to an agreement* on this point and to give them a definition, and not derive them from any kind of earlier stated axioms or theorems. The point is that we have just introduced them ourselves and so, naturally, there can be no theorems involving them so far, and we can distinguish them in any way we wish to.

So we start with the following definition.

Definition 1. *The expressions $a+bi$ and $c+di$ will be considered equal if and only if $a=c$ and $b=d$ at the same time. The equality of the expressions $a+bi$ and $c+di$ will be written $a+bi=c+di$.*

We can now state in which cases the two expressions will differ: $a+bi$ and $c+di$ *are distinct* if at least one of the two inequalities $a\neq c$, $b\neq d$ is true.

Our next task is to learn how to manipulate these expressions (via addition, multiplication, etc.). Again, this is something that is left up to us.

Let us proceed from the idea of arithmetic operations: to add or multiply two numbers means to apply a rule by means of which a third number is constructed that is termed the sum (or product). Thus, in order to be able to add and multiply these expressions, we must state certain rules for doing so.

Definition 2. *The sum of two expressions $a+bi$ and $c+di$ is an expression $(a+c)+(b+d)i$. We will denote the sum of the expressions $a+bi$ and $c+di$ by*

$$(a+bi)+(c+di)$$

Note that here, the sign $+$ between the two parentheses has a new meaning, it is the sign of addition of formal expressions.

Definition 3. *The product of the expressions $a+bi$ and $c+di$ is the expression $(ac-bd)+(ad+bc)i$. We denote the product of the expressions $a+bi$ and $c+di$ by*

$$(a+bi)(c+di)$$

We now introduce some generally accepted terminology.

Expressions of the type $a+bi$ which differ in accordance with the rule following from Definition 1, which are added in accordance with Definition 2, and which are multiplied in accordance with Definition 3, are called complex numbers.

A natural question may arise at this point. Why wasn't the term "complex number" introduced at the start, but only after bringing in three definitions? This would not have been the proper approach since it is possible, on the basis of expressions of the type $a+bi$, to construct other theories that differ fundamentally from the theory of complex numbers. The kind of theory to be constructed depends precisely on the rules which we agree upon for handling these expressions. For this reason, when we speak about complex numbers it is not that we have in mind simply the set of expressions of the type $a+bi$, but that we always presuppose that they are to be added and multiplied in accord with Definitions 2 and 3.

So much for the definition of complex numbers. Now let us take up the theory. For instance, we can define the *difference* between two complex numbers $a+bi$ and $c+di$ as a complex number such that when it is combined with $c+di$ yields $a+bi$, and prove that this difference $(a+bi)-(c+di)$ is equal to $(a-c)+(b-d)i$. Next, we can define the *quotient* obtained by the division of $a+bi$ by $c+di$ as a complex number, the product of which by $c+di$ is equal to $a+bi$, and prove that for $c+di \neq 0+0i$ this quotient $\frac{a+bi}{c+di}$ is equal to

$$\frac{ac+bd}{c^2+d^2}+\frac{-ad+bc}{c^2+d^2}i$$

A geometric interpretation of complex numbers can also be given and so on.

Now for the fundamental question of how the real and complex numbers are connected. Let us examine formal expressions of the form $a+0i$. Using Definitions 2 and 3, let us compute the sum and product of two such numbers:

$$(a+0i)+(b+0i)=(a+b)+(0+0)i=(a+b)+0i.$$
$$(a+0i)(b+0i)=(ab-0\cdot 0)+(a\cdot 0+0\cdot b)i=ab+0i$$

We see that finding the sum of the complex numbers $a+0i$ and $b+0i$ amounts to adding the real numbers a and b, and then adjoining $0i$ to the result, or setting up an expression which looks like this: $(a+$ $+b)+0i$. The same goes for a product.

Thus, operations involving complex numbers having the form $a+0i$ are actually performed in the same way as operations involving real numbers. It is therefore quite natural to identify the complex number $a+0i$ with the real number a.

By this identification, we find that the set of real numbers has become a subset (a part) of the set of complex numbers, which is to say that every real number is at the same time a complex number, and so the numbers a and $a+0i$ will not be distinguished in the future.

Now consider the complex number $0+1i$. This number plays so fundamental a role in the theory that for brevity we denote it simply as i. It then turns out that we can attribute the following meaning to the complex number $a+bi$ which up to now has been regarded as a formal expression: it is the sum of the complex number a (that is, $a+0i$) and the product of the complex number b (or, $b+0i$) by the complex number i (or, $0+1i$). Indeed,

$$(a+0i)+(b+0i)(0+1i)=(a+0i)+[(b\cdot 0-0\cdot 1)$$
$$+(b\cdot 1+0\cdot 0)i]=(a+0i)+(0+bi)=(a+0)+(0+b)i=a+bi$$

The foregoing reasoning has enabled us to invest with meaning the sign (+) in the formal expression $a+bi$; it can be regarded as the sign of addition of complex numbers.

It now remains to figure out the main property of the complex number i. It can readily be shown that

$$i^2 = i \cdot i = (0+1i)(0+1i) = -1+0i = -1$$

Now to summarize our findings: operations involving complex numbers are performed by the same rules used with regard to real numbers with the convention that i^2 is replaced by -1.

The equality $i^2=-1$ may be interpreted thus: the number i is a root of the equation $x^2+1=0$. It was precisely the problem of solving the equation $x^2+1=0$, which does not have any real roots, that lead to the construction of the theory of complex numbers.

To summarize, we give here a series of definitions.

The complex number $a+bi$ is called a real number if $b=0$. Examples of real numbers are 1, -3, 0.

A complex number $a+bi$ is called an imaginary number if $b \neq 0$. Examples of imaginary numbers are $2i$, $1-i$, $\sqrt{7}-i\sqrt{3}$.

The complex number $a+bi$ is called a pure imaginary if $a=0$. Examples of pure imaginary numbers are $-2i$, πi, 0.

Note that the number 0 is both a real and a pure imaginary number, but is not an imaginary number.

The number a is called the real part of the complex number $a+bi$.

The number b is called the imaginary part of the complex number $a+bi$.

In the solution of many problems, one requires a geometric interpretation of complex numbers as points in a plane. Here, an essential role is played by the *concept of the absolute value* (or *modulus*) *of a complex number $z=a+bi$, which is defined by*

$$|z| = \sqrt{a^2+b^2} \tag{1}$$

The modulus is clearly a nonnegative real number defined uniquely by this formula for every complex number $z=a+bi$. The modulus is endowed with a simple geometric meaning: $|z|$ is clearly the distance from the origin of coordinates to the point associated with the number z. This geometric interpretation is revealed in many problems.

1. *Given, in a plane, a certain point representing the complex number $z=a+bi$. Locate the points* (a) $z+1$, (b) $z-2+i$.

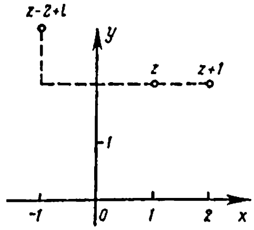
(a) Since the number $z+1=(a+1)+bi$, it follows that the coordinates of the point representing the complex number $z+1$ will be $(a+1, b)$, which is to say the ordinate remained unchanged and the abscissa was increased by 1. And so the point $z+1$ is obtained from the point z by a rightward shift of one unit (Fig. 4).

(b) Since the number $z-2+i=(a-2)+(b+1)i$, it follows that the coordinates of the point representing the complex number $z-2+i$ are $(a-2, b+1)$, or the abscissa is reduced by 2 and the ordinate is increased by 1. And so the point $z-2+i$ is obtained from z by moving 2 units to the left and 1 unit up (Fig. 4).

2. *On a plane, locate points for which* $|z|=1$.

According to the geometric interpretation of the modulus of a complex number, all the points representing complex numbers for which $|z|=1$ lie at one and the same distance, equal to unity, from the origin. In other words, by definition, they lie on a circle of radius 1 with centre at the origin of coordinates.

Fig. 4



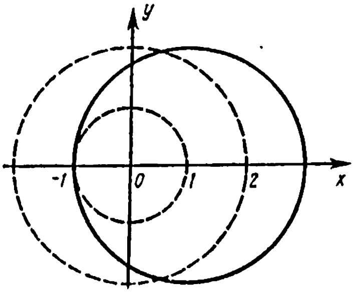
3. *Let* $|z|=2$; *where are the points* $3z$ *located?*

The points z which satisfy the condition $|z|=2$ are located on a circle of radius 2 with centre at the origin (see preceding problem). The point $3z$ is located on the same ray as the point z, but is distant from the origin three times that of point z. (Why? Make a drawing.) For this reason, the points $3z$, where $|z|=2$, are located on a circle of radius 6 with centre at the origin.

4. *Let* $|z|=1$; *where are the points* $1+2z$ *located?*

The points z which satisfy the condition $|z|=1$ lie on a circle of radius 1 with centre at the origin. All points $2z$, where $|z|=1$, are located on a circle of radius 2 with centre at the origin. The point $2z+1$ is obtained from point $2z$ by a rightward shift of 1 unit (see Problem 1). And so the points $1+2z$, where $|z|=1$, are located on a circle of radius 2 with centre at the point (1, 0) (Fig. 5).

Fig. 5



5. *Locate the points for which* $2<|z|<3$.

We know that points satisfying the condition $|z|=2$ are located on a circle of radius 2 with centre at the origin. Now points for which $|z|>2$ are located farther from the origin than the points of this circle, that is, outside the circle. Similarly, points satisfying the condition $|z|<3$ are located inside a circle of radius 3 centred at the origin.

Hence, points which satisfy the condition $2<|z|<3$ lie inside an annulus bounded by concentric circles centred at the origin having radii $r_1=2$ and $r_2=3$ (Fig. 6).

A complex number may also be regarded as a vector, the origin of which lies at the origin of the system of coordinates in the plane and the terminus of which represents the given number. This makes for simple geometric interpretations of the operations of addition and subtraction.

If a vector $\overrightarrow{OM_1}$ is the vector associated with the number $z_1=a+bi$ and $\overrightarrow{OM_2}$ is the vector associated with the number $z_2=c+di$, then the

Fig. 6

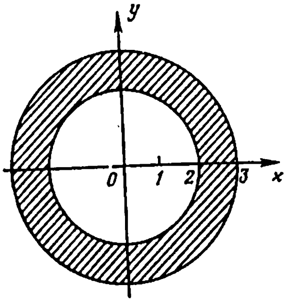
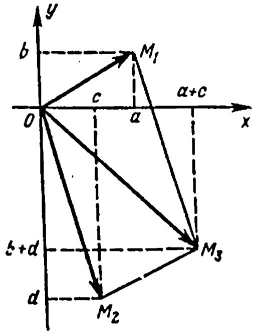


Fig. 7



sum of these vectors, $\overrightarrow{OM_1}+\overrightarrow{OM_2}$, is the diagonal $\overrightarrow{OM_3}$ of a parallelogram $OM_1M_3M_2$. The endpoint of this diagonal, the point M_3, obviously has the coordinates $(a+c, b+d)$ (Fig. 7). Hence, $\overrightarrow{OM_3}$ is a vector that depicts the complex number $z_3=z_1+z_2=(a+c)+(b+d)i$. If the vector $\overrightarrow{OM}$ represents the number z, then the number $-z$ will be represented by the vector $\overrightarrow{ON}$, the terminus of which is a point symmetric to the point M about the origin. Thus the operation of subtraction of complex numbers also admits a simple geometric interpretation. Namely, since $z_1-z_2=z_1+(-z_2)$, then in place of the vector $\overrightarrow{OM_2}$ representing the number z_2, we consider the vector $\overrightarrow{OM_4}$ symmetric to it about the origin (Fig. 8). Adding, as before, the vector $\overrightarrow{OM_1}$, associated with the number z_1, with the vector $\overrightarrow{OM_4}$, we get the vector $\overrightarrow{OM_5}$, which represents the difference z_1-z_2. It is clear that the length of the vector $\overrightarrow{OM_5}$ is equal to the length of the vector $\overrightarrow{M_2M_1}$, which is the length of the diagonal M_1M_2 of the parallelogram $OM_1M_3M_2$. Since the length of the vector $\overrightarrow{OM_5}$ is equal to the modulus of the difference z_1-z_2, the length of the diagonal M_1M_2 is also equal to $|z_1-z_2|$. We obtain a simple geometric interpretation of the modulus of the difference of two complex numbers: $|z_1-z_2|$ is the distance between the points M_1 and M_2 representing the complex numbers z_1 and z_2. This interpretation is frequently used in problem solving.

6. *Where are the points associated with the complex numbers z for which* $|z-1|=2$ *located?*

If z is the desired point, then the distance between z and 1 is equal to 2. But the points distant 2 from 1 lie on a circle. Hence, points representing numbers for which $|z-1|=2$ lie on a circle of radius 2 centred in the point (1, 0).

We can reason differently. Denote $z-1=w$. Then we have the equation $|w|=2$. Consequently, the points w lie on a circle of radius 2 with centre at the origin. But $z=w+1$ so that the points z are obtained from the points w by a rightward shift of 1 unit. Hence, the desired points lie on a circle of radius 2 centred in the point (1, 0).

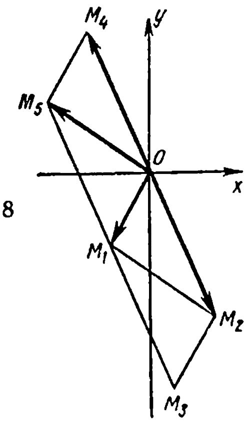


Fig. 8

7. *Locate the points that represent the complex numbers z for which* $|z+2i|\leqslant 1$.

We rewrite this condition as follows: $|z-(-2i)|\leqslant 1$. Hence, the distance from the points z to the point $-2i$ does not exceed 1. In other words, all the points which satisfy this condition will lie inside or on the circumference of a circle of radius 1 with centre in the point (0, —2) representing the complex number $-2i$.

8. *The complex numbers z satisfy the condition* $1<|z+2-3i|<2$. *Locate the points representing these numbers.*

Rewrite the condition as $1<|z-(-2+3i)|<2$. All the points which satisfy this condition lie inside an annulus bounded by concentric circles having radii $r_1=1$ and $r_2=2$ and centred in the point (—2, 3).

9. *Complex numbers z satisfy the condition*

$$|z-i|=|z+2|$$

Give the location of points representing these numbers.

The modulus $|z-i|$ is the distance from the points z to a fixed point representing the number i. The modulus $|z+2|=|z-(-2)|$ is the distance from the points z to a fixed point representing the number -2.

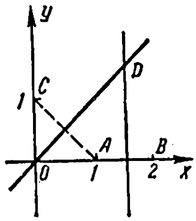
It is required to find the points for which these distances are equal. The solution will thus be a locus equidistant from two fixed points in the plane: the point representing the complex number i, that is, the point (0, 1), and the point representing the number -2, or the point (—2, 0).

From geometry we know that this locus is a straight line perpendicular to a line segment connecting the two indicated points and passing through its midpoint. This means that the points representing the complex numbers z which satisfy the condition $|z-i|=|z+2|$ lie on a straight line perpendicular to the line segment joining points with coordinates $(-2, 0)$ and $(0, 1)$ and passing through the midpoint of that segment.

10. *Locate the points representing the complex numbers z for which* $|z-1|=|z-2|=|z-i|$.

The set of points satisfying the condition $|z-1|=|z-2|$ is a straight line passing through the midpoint of the line segment AB, where A is $(1, 0)$ and B $(2, 0)$, perpendicular to this line segment. The set of points satisfying the condition $|z-1|=|z-i|$ is a straight line passing through the midpoint of AC, where A is $(1, 0)$ and C is $(0, 1)$, and perpendicular to this segment (see Fig. 9).

Fig. 9



It is clear now that the condition

$$|z-1|=|z-2|=|z-i|$$

is only satisfied by the single point D lying at the intersection of these two straight lines. It is easy to compute the coordinates of this point: $x=y=3/2$. In other words, the condition of the problem is satisfied by only one complex number $z=3/2+(3/2)i$.

Complex numbers different from zero are often conveniently written in a form called the *trigonometric* (or *polar*) *form*.

Let us first introduce for these numbers the concept of an argument: *the argument of a number $z=a+bi\neq 0$ is any one of the numbers φ which are solutions of the system of equations*

$$\begin{aligned}\cos\varphi&=\frac{a}{\sqrt{a^2+b^2}}\\ \sin\varphi&=\frac{b}{\sqrt{a^2+b^2}}\end{aligned} \tag{2}$$

The argument is not defined for $z=0$.

From trigonometry it will be recalled that this system of equations has an infinity of solutions, and if φ_0 is one of its solutions, then all the other solutions are obtained from it by the formula

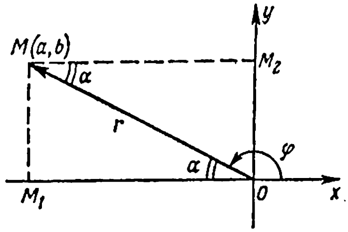
$$\varphi=\varphi_0+2k\pi,\ k \text{ any integer} \tag{3}$$

Thus, any complex number $z \neq 0$ has infinitely many arguments and all of them can be obtained from one using formula (3).

Note that the arguments of a complex number z always include one that satisfies the inequalities: $0 \leqslant \varphi < 2\pi$; it is precisely this value of φ that is called the argument of the number z. However, this restriction is often too confining. We will adhere to the definition given above and use the term *principal argument* for values of φ lying in the interval from 0 to 2π. Accordingly, wherever, in the sequel, it is required to find the argument of some complex number z, we will confine ourselves to finding one of its arguments (and not necessarily the principal argument). We use the symbol arg z for this argument.

The argument of a complex number z has the following geometric significance. If a complex number $z=a+bi \neq 0$ is regarded as a vector $\overrightarrow{OM}$, then the magnitude of the angle φ, through which the x-axis

Fig. 10



must be turned counterclockwise to first coincidence with the vector $\overrightarrow{OM}$ is the principal argument of z (Fig. 10). The magnitude of any angle which differs from φ by an integral number of round angles is an argument of that number z.*

11. *Locate the points which satisfy the condition* arg $z=\pi/3$.

This condition is satisfied by all points lying on a ray emanating from the origin at an angle of $\pi/3$ to the x-axis. It must be stressed that this condition is not satisfied by the entire straight line but only by the ray, and also minus the origin! (Why?)

Now let $z=a+bi \neq 0$ be some complex number. Denote by r its modulus computed from formula (1) and by φ one of its arguments. Then we can write it in the form

$$z = r(\cos\varphi + i\sin\varphi) \tag{4}$$

The right-hand member of this equation is termed the *trigonometric form* of the number z. The trigonometric form of $z=0$ is not defined.

The trigonometric form of complex numbers is closely associated with their geometric interpretation: formula (4) is a natural consequence of geometric reasoning (Fig. 10).

* From this geometric interpretation it is evident that it is not possible, in any reazonable fashion, to introduce the argument of the number $z=0$. That is precisely why it is not done.

Given our definition of the modulus and argument of a complex number, its trigonometric form (4) is obtained automatically. A different approach is also frequently used. The modulus and argument of a complex number are introduced via geometric reasoning and then formula (4) is proved. The formula itself is sometimes derived on the basis of a drawing where point $M(a, b)$ lies in the first quadrant. However, this formula is of course valid for any position of the point M and the student should be able to demonstrate its validity in every case.

For instance, let the point $M(a, b)$ lie in the second quadrant as shown in Fig. 10. Then $OM_1 = r\cos\alpha$, $OM_2 = r\sin\alpha$ and $\alpha = \pi - \varphi$. Since for the points of the second quadrant $a < 0$ and $b > 0$, it follows that $OM_1 = -a$, $OM_2 = b$ and, consequently,

$$-a = r\cos\alpha = r\cos(\pi - \varphi) = -r\cos\varphi,$$

$$b = r\sin\alpha = r\sin(\pi - \varphi) = r\sin\varphi$$

That is,

$$a + bi = r\cos\varphi + ir\sin\varphi = r(\cos\varphi + i\sin\varphi)$$

The formula holds if the point M lies in the second quadrant. It is easy to see, incidentally, that there is no necessity to run through all the quadrants: the formulas $a = r\cos\varphi$ and $b = r\sin\varphi$ can be derived directly from the definitions of the cosine and the sine of the angle φ, whence the validity of formula (4) follows immediately for any position of M.

The formulas

$$a = r\cos\varphi$$

$$b = r\sin\varphi$$

are formulas for passing from the trigonometric form of a complex number to the algebraic form, because it is easy to find a and b if we know r and φ.

More often one encounters the converse problem: knowing a and b, find r and φ. The modulus r is determined [see formula (1)] very simply: $r = \sqrt{a^2 + b^2}$. But many mistakes are made in determining the argument φ. The most typical being this: from system (2) it is clear that $\tan\varphi = b/a$, whence $\varphi = \arctan b/a$. Indeed, although $\tan\varphi = b/a$, it does not yet follow that $\varphi = \arctan b/a$. In order to determine correctly the argument of a complex number z, one has to know in which quadrant the point z lies, and this is best done by applying the geometric interpretation of a complex number, as is done in the following example.

12. *Find the trigonometric form of the complex number* $z = -6 - 8i$.

Clearly, $|z| = 10$ and $\tan\varphi = b/a = 4/3$. As can be seen from Fig. 11, $\arg z = \pi + \alpha$, where α is an acute angle such that $\tan\varphi = 4/3$. For this reason $\alpha = \arctan 4/3$ or $\varphi = \pi + \arctan 4/3$, and so the trigonometric form is

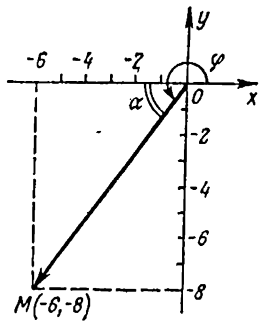
$$z=-6-8i=10(\cos\varphi+i\sin\varphi),\quad \varphi=\pi+\arctan 4/3$$

Note that the concept of the trigonometric form of a complex number different from zero is defined very precisely: namely, it is the complex number $z\neq 0$ written as

$$z=r(\cos\varphi+i\sin\varphi)$$

where r is the modulus of z and is positive, while the cosine and sine are taken of the same angle φ, which is the argument of z; they are

Fig. 11



connected by the + sign. For example, the following complex numbers are not written in trigonometric form:

$$z_1=\cos\frac{\pi}{4}+i\sin\left(-\frac{\pi}{4}\right),\quad z_2=-2\left(\cos\frac{\pi}{3}+i\sin\frac{\pi}{3}\right),$$
$$z_3=\cos\frac{\alpha}{2}-i\sin\frac{\alpha}{2},\quad z_4=\sin 30^\circ+i\cos 30^\circ$$

The trigonometric forms of these complex numbers are, respectively,

$$z_1=\cos\frac{7\pi}{4}+i\sin\frac{7\pi}{4},$$
$$z_2=2\left(\cos\frac{4}{3}\pi+i\sin\frac{4}{3}\pi\right),$$
$$z_3=\cos\left(2\pi-\frac{\alpha}{2}\right)+i\sin\left(2\pi-\frac{\alpha}{2}\right),$$
$$z_4=\cos 60^\circ+i\sin 60^\circ$$

It is well to know how to handle complex numbers written in trigonometric form; besides, certain important properties of complex numbers which are extremely useful follow from a consideration of operations involving complex numbers written in trigonometric form.

Actually, there is nothing complicated in the rules for multiplying and dividing complex numbers in trigonometric form. The familiar formulas of trigonometry are made use of.

From these rules follow such properties of the moduli of complex numbers as:

I. $|z_1 \cdot z_2| = |z_1| \cdot |z_2|$

II. $\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}$

These properties are used rather often in the solution of many problems. Another useful formula is

III. $|z^n| = |z|^n$ *where n is any integer.* It is obtained for any integral value of $n \neq 0$ as a consequence of Properties I and II by means of mathematical induction (see Sec. 1.3). The validity of this property for $n=0$ follows from the generally accepted definition that *every complex number different from zero is, to the zeroth power, equal to unity.*

Finally, two more formulas are valid that express the properties of the modulus of a sum and a difference:

IV. $|z_1 + z_2| \leqslant |z_1| + |z_2|$

V. $|z_1 - z_2| \geqslant ||z_1| - |z_2||$

To prove Property IV, put

$$z_1 = r_1(\cos\varphi_1 + i\sin\varphi_1), \quad z_2 = r_2(\cos\varphi_2 + i\sin\varphi_2)$$

Then, using formula (1), we get

$$|z_1 + z_2| = |(r_1\cos\varphi_1 + r_2\cos\varphi_2) + i(r_1\sin\varphi_1 + r_2\sin\varphi_2)|$$
$$= \sqrt{(r_1\cos\varphi_1 + r_2\cos\varphi_2)^2 + (r_1\sin\varphi_1 + r_2\sin\varphi_2)^2}$$
$$= \sqrt{r_1^2 + r_2^2 + 2r_1r_2\cos(\varphi_1 - \varphi_2)}$$

Noting that $\cos(\varphi_1 - \varphi_2) \leqslant 1$, we get

$$|z_1 + z_2| = \sqrt{r_1^2 + r_2^2 + 2r_1r_2\cos(\varphi_1 - \varphi_2)}$$
$$\leqslant \sqrt{r_1^2 + r_2^2 + 2r_1r_2} = r_1 + r_2 = |z_1| + |z_2|$$

Property V is proved in similar fashion:*

$$|z_1 - z_2| = \sqrt{r_1^2 + r_2^2 - 2r_1r_2\cos(\varphi_1 - \varphi_2)}$$
$$\geqslant \sqrt{r_1^2 + r_2^2 - 2r_1r_2} = |r_1 - r_2| = ||z_1| - |z_2||$$

It is interesting to give a geometric interpretation of Properties IV and V. Let vector $\overrightarrow{OM_1}$ represent the number z_1 and vector $\overrightarrow{OM_2}$ the number z_2 (Fig. 7). Then the vector $\overrightarrow{OM_3}$ represents the sum $z_1 + z_2$. Property IV signifies that the length of the diagonal OM_3 of the parallelogram $OM_1M_3M_2$ does not exceed the sum of the lengths of its

* Property V may be derived from Property IV. Indeed, the equation $z_2 = z_1 + (z_2 - z_1)$ is obvious, whence $|z_2| = |z_1 + (z_2 - z_1)| \leqslant |z_1| + |z_2 - z_1|$ or $|z_2 - z_1| \geqslant |z_2| - |z_1|$. At the same time, from $z_1 = z_2 + (z_1 - z_2)$ we get $|z_1| = |z_2 + (z_1 - z_2)| \leqslant |z_2| + |z_1 - z_2|$ or $|z_1 - z_2| \geqslant |z_1| - |z_2|$. Property V results from combining the two resulting inequalities into a single formula.

sides OM_1 and OM_2. Property V states that the length of the diagonal M_1M_2 is not less than the absolute value of the difference of the sides OM_1 and OM_2.

Note that properties similar to the formulas I to V were stated in the preceding section for the absolute value of a real number. It is quite clear that the modulus of a real number, which is regarded as a particular case of complex numbers, coincides with the absolute value of this real number. Therefore, the above-indicated properties I to V are a generalization of the properties of the absolute value of a number. The proof however is conducted quite differently.

Let us recall another definition from the theory of complex numbers, that of the conjugate complex number. *The conjugate complex number of a complex number $a+bi$ is the complex number $a-bi$.* The conjugate complex of z is denoted by $\bar{z}$.

Quite obviously, $\overline{(\bar{z})}=z$, which is to say that not only is the number $\bar{z}$ the conjugate of z, but, contrariwise, z is the conjugate of $\bar{z}$. Thus, z and $\bar{z}$ are *conjugates of each other.*

It will be useful to remember the following two properties of conjugate numbers:

I. $z\bar{z}=|z|^2$,

II. $|\bar{z}|=|z|$,

which follow directly from the definitions.

Let us now examine some problems that have appeared in examination papers.

13. *Locate the complex numbers $z=a+bi$ for which*

$$\log_{1/2}|z-2|>\log_{1/2}|z|$$

First of all, note that the left member of this inequality is meaningful for all complex numbers z, except $z=2$, and the right member, for all $z\neq 0$. Therefore, the expressions of this inequality have meaning simultaneously for all complex numbers z except $z=0$ and $z=2$. It is among these numbers that we have to seek the solution of this inequality.

By the rules of logarithms (see Sec. 1.6), for all these numbers our inequality is equivalent to the following: $|z-2|<|z|$.

We know (see Example 9 above) that the equation $|z-2|=|z|$ is satisfied by all complex numbers lying on a straight line l parallel to the y-axis and passing through the point A (1, 0) because all the points of this line are equidistant from two points: O (0, 0) and B (2, 0) (Fig. 12). But our job is to find all those points in the plane which are closer to the point B (2, 0) than to the point O (0, 0).

Quite obviously, these are the points of the plane to the right of l in the domain that includes point B. Thus all the points of the half-plane to the right of l satisfy the condition $|z-2|<|z|$ (Fig. 12). The points of the straight line l are excluded.

Recall also that the point $B(2, 0)$ located in the half-plane to the right of l must also be excluded.

To summarize, then, all points of the plane located to the right of the straight line parallel to the y-axis and passing through the point $(1, 0)$, with the exception of the point $(2, 0)$, satisfy the conditions of the problem.

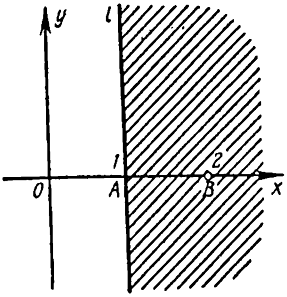


Fig. 12

14. *Suppose the complex number* $z\neq-1$. *Prove that*
(a) *if* $|z|=1$, *then the number* $\frac{z-1}{z+1}$ *is a pure imaginary*;
(b) *if the number* $\frac{z-1}{z+1}$ *is a pure imaginary, then* $|z|=1$.

Let $z=a+bi$ and $z\neq-1$. Then it is clear that $z+1\neq 0$, and the expression $\frac{z-1}{z+1}$ is meaningful.

The number $\frac{z-1}{z+1}$ is the quotient obtained by the division of two complex numbers, and so its $a+bi$ (real-imaginary, or rectangular) form is

$$\frac{z-1}{z+1}=\frac{(a-1)+bi}{(a+1)+bi}=\frac{[(a-1)+bi][(a+1)-bi]}{(a+1)^2+b^2}=\frac{a^2+b^2-1}{(a+1)^2+b^2}+i\frac{2b}{(a+1)^2+b^2}$$

It is then clear that if $|z|=\sqrt{a^2+b^2}=1$, then $a^2+b^2-1=0$, or the number $\frac{z-1}{z+1}$ is a pure imaginary,* and Assertion (a) is proved.

Let us prove Assertion (b). Let the number $\frac{z-1}{z+1}$ be a pure imaginary. Then $\frac{a^2+b^2-1}{(a+1)^2+b^2}=0$, whence $a^2+b^2-1=0$, that is $|z|=$ $=\sqrt{a^2+b^2}=1$ and the proof is complete.

---

* Note that for $z=1$, which means for $a=1$ and $b=0$, the number $\frac{z-1}{z+1}$ is equal to zero Recall that, according to the definition given above, the number 0 is a pure imaginary.

15. *Find the argument of the complex number* $z_1 = z^2 - z$ *if* $z = \cos\varphi + i\sin\varphi$.

Simple manipulations show that

$$\begin{aligned} z_1 &= (\cos\varphi + i\sin\varphi)^2 - (\cos\varphi + i\sin\varphi) \\ &= \cos^2\varphi - \sin^2\varphi + 2i\cos\varphi\sin\varphi - \cos\varphi - i\sin\varphi \\ &= (\cos 2\varphi - \cos\varphi) + i(\sin 2\varphi - \sin\varphi) \\ &= 2\sin\frac{\varphi}{2}\left[-\sin\frac{3\varphi}{2} + i\cos\frac{3\varphi}{2}\right] \end{aligned}$$

Thus

$$|z_1| = \sqrt{4\sin^2\frac{\varphi}{2}\left(\sin^2\frac{3\varphi}{2} + \cos^2\frac{3\varphi}{2}\right)} = 2\left|\sin\frac{\varphi}{2}\right|$$

In accordance with the definition of absolute value we have to consider three cases:

(a) If $\sin\frac{\varphi}{2} = 0$, that is, $\varphi = 2k\pi$, k any integer, then $|z_1| = 0$, and for this reason, $z_1 = 0$ as well. Thus, for $\varphi = 2k\pi$, k any integer, the argument of the number z_1 is not defined.

(b) If $\sin\frac{\varphi}{2} > 0$, which occurs when $2k\pi < \frac{\varphi}{2} < (2k+1)\pi$, that is, for

$$4k\pi < \varphi < (4k+2)\pi, \ k \text{ any integer} \tag{5}$$

then $|z_1| = 2\sin\frac{\varphi}{2}$ and the trigonometric form of the complex number z_1 is as follows:

$$z_1 = 2\sin\frac{\varphi}{2}\left[\cos\frac{\pi+3\varphi}{2} + i\sin\frac{\pi+3\varphi}{2}\right]$$

Consequently, if φ satisfies Condition (5), then

$$\arg z_1 = \frac{\pi+3\varphi}{2}$$

(c) If $\sin\frac{\varphi}{2} < 0$, that is,

$$(4k+2)\pi < \varphi < (4k+4)\pi, \ k \text{ any integer} \tag{6}$$

then $|z_1| = -2\sin\frac{\varphi}{2}$ and the trigonometric form of the complex number z_1 is

$$z_1 = -2\sin\frac{\varphi}{2}\left[\cos\frac{3\pi+3\varphi}{2} + i\sin\frac{3\pi+3\varphi}{2}\right]$$

Hence, if φ satisfies (6), then

$$\arg z_1 = \frac{3\pi+3\varphi}{2}$$

It will be interesting to give a geometric interpretation of the solution, which we will do only for the case of $0 < \varphi < \pi$. The number $z_1 = z^2 - z = z^2 + (-z)$ is the sum of two complex numbers

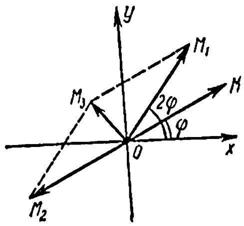
$$z^2 = \cos 2\varphi + i \sin 2\varphi$$

and

$$-z = -\cos\varphi + i\sin\varphi = \cos(\pi+\varphi) + i\sin(\pi+\varphi)$$

which have the same moduli (equal to unity). To find the sum, we have to find the diagonal of the parallelogram constructed on the vectors $\overrightarrow{OM}_1$ and $\overrightarrow{OM}_2$ representing the numbers z^2 and $-z$, respectively (Fig. 13). But this parallelogram is a rhombus, and so the desired

Fig. 13



diagonal OM_3 is the bisector of the angle between the vectors $\overrightarrow{OM}_1$ and $\overrightarrow{OM}_2$ and therefore the angle which is formed by the vector $\overrightarrow{OM}_3$ and the positive x-axis is the half-sum of the angles formed by the vectors $\overrightarrow{OM}_1$ and $\overrightarrow{OM}_2$ with this positive direction; that is,

$$\arg z_1 = \arg(z^2 - z) = \frac{2\varphi + \pi + \varphi}{2} = \frac{\pi + 3\varphi}{2}$$

16. *Find the trigonometric form of the complex number*

$$z = 1 + i \tan\alpha$$

where $-\pi < \alpha < \pi$, $\alpha \neq \pm\pi/2$.

It is natural to rewrite the given number as

$$z = 1 + i\tan\alpha = \frac{1}{\cos\alpha}(\cos\alpha + i\sin\alpha)$$

At this point, many students at the examination made the serious mistake of asserting that this was the trigonometric form of the given number. But this is true only for $1/\cos\alpha > 0$, that is, when $-\pi/2 < \alpha < \pi/2$ (it is given that only values of α are considered which lie in the interval from $-\pi$ to $+\pi$). Now if $1/\cos\alpha < 0$, which is the case when $-\pi < \alpha < -\pi/2$ and $\pi/2 < \alpha < \pi$, then the above equation can be given as

$$z = -\frac{1}{\cos\alpha}(-\cos\alpha - i\sin\alpha) = -\frac{1}{\cos\alpha}[\cos(\pi+\alpha) + i\sin(\pi+\alpha)]$$

This latter expression is the trigonometric form of the number z for $-\pi<\alpha<-\pi/2$ and for $\pi/2<\alpha<\pi$.

This problem can also be solved by the general rule for finding a trigonometric form; to do this we have to find the modulus and the argument of the number z. The modulus of z is found from formula (1):

$$r=|z|=\sqrt{1+\tan^2\alpha}=\frac{1}{|\cos\alpha|}$$

and the argument is any solution of the system [see (2)]

$$\begin{aligned}\cos\varphi&=|\cos\alpha|\\ \sin\varphi&=\tan\alpha\cdot|\cos\alpha|\end{aligned} \qquad (7)$$

To solve this system we have to consider two cases:

(a) $\cos\alpha>0$, that is, α lies in the interval $-\pi/2<\alpha<\pi/2$. In this case $|\cos\alpha|=\cos\alpha$, and system (7) takes the form

$$\begin{aligned}\cos\varphi&=\cos\alpha\\ \sin\varphi&=\sin\alpha\end{aligned}$$

Clearly, one of the solutions of the system is $\varphi=\alpha$ and, consequently, for $-\pi/2<\alpha<\pi/2$ the trigonometric form is

$$z=\frac{1}{\cos\alpha}(\cos\alpha+i\sin\alpha)$$

(b) $\cos\alpha<0$, that is, α lies in the interval $-\pi<\alpha<-\pi/2$ or in the interval $\pi/2<\alpha<\pi$. In this case, $|\cos\alpha|=-\cos\alpha$ and system (7) assumes the form

$$\begin{aligned}\cos\varphi&=-\cos\alpha\\ \sin\varphi&=-\sin\alpha\end{aligned} \quad\text{or}\quad \begin{aligned}\cos\varphi&=\cos(\pi+\alpha)\\ \sin\varphi&=\sin(\pi+\alpha)\end{aligned}$$

One solution is $\varphi=\pi+\alpha$, and so for $-\pi<\alpha<-\pi/2$ and for $\pi/2<$ $<\alpha<\pi$ the trigonometric form is

$$z=\frac{1}{|\cos\alpha|}[\cos(\pi+\alpha)+i\sin(\pi+\alpha)]$$

17. *Find the integral solutions of the equation* $(1-i)^x=2^x$.

Suppose a certain integer k is a solution of this equation. Then from the equality of the complex numbers $(1-i)^k=2^k$ follows the equality of their moduli, or $|(1-i)^k|=2^k$. Noting that $|1-i|=\sqrt{2}$, we have, by the property of a modulus,

$$|(1-i)^k|=|1-i|^k=(\sqrt{2})^k=2^{\frac{k}{2}}$$

Thus, if k is a solution of the original equation, then $2^{\frac{k}{2}}=2^k$, which is only possible when $k=0$.

Now, by substitution, check to see if the number 0 is a solution of the original equation. Recalling that a nonzero complex number to the zeroth power is, by definition, equal to unity, we see that $x=0$ is a root of the original equation.

18. *For every real number $a \geqslant 0$ find all the complex numbers z that satisfy the equation*

$$|z|^2 - 2iz + 2a(1+i) = 0$$

Represent z in real-imaginary form: $z=x+iy$. Then $|z|^2=x^2+y^2$ and the equation becomes

$$x^2 + y^2 - 2ix + 2y + 2a + 2ai = 0$$

Equating the real and imaginary parts to zero, we obtain the following system of equations:

$$x^2 + y^2 + 2y + 2a = 0$$
$$-2x + 2a = 0$$

Whence it follows that $x=a$, and for y we have the quadratic equation

$$y^2 + 2y + a^2 + 2a = 0$$

with parameter a. We seek the real roots of this equation.

It will be recalled that the roots of a quadratic equation are real if its discriminant is nonnegative, and therefore our equation has real roots only for values of a for which $D=1-a^2-2a \geqslant 0$. For these values of a we get

$$y_{1,2} = -1 \pm \sqrt{1-a^2-2a}$$

Thus, if the number a satisfies the inequality $1-a^2-2a \geqslant 0$, then the original equation has two solutions

$$z_{1,2} = a + (-1 \pm \sqrt{1-a^2-2a})\, i$$

(For $1-a^2-2a=0$ these two solutions are the same, which is to say that strictly speaking there is only one solution for certain values of a.) For the remaining values of a, the original equation has no solutions.

It remains to indicate the range of a over which there are solutions. By hypothesis $a \geqslant 0$, and, besides, we found that a must satisfy the inequality $1-a^2-2a \geqslant 0$ or $a^2+2a-1 \leqslant 0$, which is the same thing. The solution of the latter inequality is the interval $-1-\sqrt{2} \leqslant a \leqslant -1+\sqrt{2}$ and, choosing the numbers $a \geqslant 0$ from this interval, we obtain $0 \leqslant a \leqslant -1+\sqrt{2}$.

The final answer may be written thus:

for $0 \leqslant a < -1+\sqrt{2}$	$z_{1,2} = a + (-1 \pm \sqrt{1-a^2-2a})\, i$
for $a = -1+\sqrt{2}$	$z = -1+\sqrt{2} - i$,
for $a > -1+\sqrt{2}$	there are no solutions.

19. *Solve the following system of equations in terms of complex numbers:*

$$z^{13}w^{19} = 1$$
$$z^5w^7 = 1$$
$$z^2 + w^2 = -2$$

It need not come as a surprise that the system includes three equations for only two unknowns. There is nothing strange in the fact that it is required to find two numbers which satisfy three conditions.

Let us approach the solution of this system in the common way. We will derive various corollaries from it and after obtaining some solutions from them, we will use substitution to verify that they satisfy the original system or are extraneous.

Cubing both sides of the second equation and dividing the result by the first equation, we get $z^2w^2=1$. But then $z^6w^6=1$ and, dividing this equation by the second equation, we obtain $z=w$. Now, from the third equation we find that $w^2=-1$, whence $w_1=i$, $w_2=-i$, or $z_1=i$, $z_2=-i$.

A check has still to be made. It is done directly and we find that both pairs are solutions of the original system.

A natural question might arise as to how we guessed right in combining the equations in just that way and in getting the answer so quickly. Firstly, there is a still shorter solution (by raising the second equation to the eighth power and dividing the result by the first equation squared, we immediately get $z=w$). Secondly, brevity is not a necessary condition for a solution. It is also possible to solve this system in the ordinary way by eliminating one of the unknowns.

For example, a very natural approach is the following. Raising the first equation to the fifth power and dividing the result by the second equation raised to the thirteenth power, we get

$$w^4 = 1$$

One should not hurry now to extract the root or solve the equation $w^4-1=0$: we would then obtain four distinct values for w and, finding the values of z which correspond to each value of w, we would have a large number of distinct pairs w, z which have to be verified to obtain a solution. A much simpler approach is to raise the first equation of the system to the seventh power, the second to the nineteenth power and then divide the second by the first to get

$$z^4 = 1$$

Then the first equation may be rewritten as $zw^3=1$, whence $z=1/w^3$ or $z=w$ (noting that $w^4=11$). The solution is then concluded in the same way as above.

20. *Solve the following system of equations in terms of complex numbers*:

$$z^3+\bar{w}^7=0$$
$$z^5\cdot w^{11}=1$$

As in the preceding example, we derive various corollaries from this system. From the first equation we have $z^3=-\bar{w}^7$; from the second, $z^5=1/w^{11}$, whence we respectively obtain $z^{15}=-\bar{w}^{35}$ and $z^{15}=1/w^{33}$ and, hence, $-\bar{w}^{35}=1/w^{33}$, or $w^{33}\bar{w}^{35}=-1$.

From this equation it follows that $|w^{33}\bar{w}^{35}|=1$. Using the properties of moduli and conjugate numbers, we obtain $|w^{33}\bar{w}^{35}|=|w|^{33}\cdot|\bar{w}|^{35}=|w|^{68}=1$ so that $|w|=1$. Reverting to the equation $w^{33}\bar{w}^{35}=-1$, we rewrite its left-hand member: $(w^{33}\bar{w}^{33})\bar{w}^2=(w\bar{w})^{33}\bar{w}^2=(|w|^2)^{33}\bar{w}^2=\bar{w}^2$ (here we utilized yet another property of conjugate numbers). We have thus arrived at the equation $\bar{w}^2=-1$ or $\bar{w}_1=i$, $\bar{w}_2=-i$, whence $w_1=-i$, $w_2=i$.

Now compute the respective values of z. If $w=-i$, then from the first equation of the original system we have

$$z^3=-i^7=i$$

and from the second,

$$z^5=\frac{1}{(-i)^{11}}=\frac{1}{i}=-i$$

Dividing the second of these equations by the first, we have $z^2=-1$ and, since $z^3=i$, it follows that $z=-i$. Similarly, we find that if $w=i$, then $z=i$.

Since in this approach we considered consequences from the original system instead of the original system of equations, we have to check to see that the values found do indeed satisfy the original system. This is done by direct substitution, which convinces us that the given system has two solutions:

$$z_1=-i,\quad w_1=-i \quad\text{and}\quad z_2=i,\quad w_2=i$$

Exercises

1. Let $|z|=5$. Locate the points representing the complex numbers (a) $-4z$, (b) $2-z$, (c) $-1+3z$.

Locate the points representing the complex numbers z for which

2. $|z|<1$.
3. $|z|\geqslant 2$.
4. $1\leqslant|z|<2$.
5. $|z|<\left|\frac{z}{2}\right|+1$.
6. $|z+1|=3$.
7. $|i-z|<1$.
8. $|z+1-2i|=\sqrt{7}$.
9. $|i-1-2z|>9$.
10. $2\leqslant|z+i|\leqslant 3$.
11. $|z|=\left|z+\frac{1}{3i}\right|$.

12. $|z-1|=|z+1|=|z-i\sqrt{3}|$.
13. $|z|-4=|z-i|-|z+5i|=0$.
14. $|z-i\sqrt{2}|-|z+4|=|z|-1=0$.
15. $|z-1|^2+|z+1|^2=4$.
16. $|z-1|^2+|z+1|^2=5$.
17. Find the complex number z which simultaneously satisfies the equations

$$\left|\frac{z-12}{z-8i}\right|=\frac{5}{3} \quad \text{and} \quad \left|\frac{z-4}{z-8}\right|=1$$

18. Given two complex numbers z_1 and z_2. Find the complex number corresponding to the midpoint of the line segment between z_1 and z_2.

19. The complex numbers z_1, z_2, z_3 are the vertices of a triangle. Find all the complex numbers z which make the triangle into a parallelogram.

Represent the following numbers in trigonometric form:

20. $z=-\cos 30°+i\sin 30°$.
21. $z=1+\cos 40°+i\sin 40°$.
22. $z=-\cos\alpha+i\sin\alpha$.
23. $z=\sin\alpha-i\cos\alpha$.
24. $z=\tan\alpha-i$, $0\leqslant\alpha<\pi$, $\alpha\neq\pi/2$.

Locate the points representing the complex numbers z for which:

25. $\arg z=\pi/4$.
26. $\arg z=-5\pi/6$.
27. $\pi/3<\arg z\leqslant 3\pi/2$.
28. $\arg z=\pi$, $|z|<1$.
29. $|z-i|=1$, $\arg z=\pi/2$
30. $0<\arg z<\pi/4$, $|z-6i|=\sqrt{3}$

31. Among the complex numbers z which satisfy the condition $|z-25i|\leqslant 15$. find the number having the least positive argument.

32. Find the argument of the complex number $z_1=z^2+\bar{z}$ if $z=\cos\varphi+i\sin\varphi$, $0\leqslant\varphi<2\pi$.

33. If the complex numbers z_1 and z_2 are such that the product $z_1\cdot z_2$ is a real number, are they conjugate complexes?

34. If the complex numbers z_1 and z_2 are such that the sum z_1+z_2 is a real number, are they conjugate complexes?

35. Prove that if the complex numbers z_1 and z_2 with nonzero imaginary parts are such that the product $z_1\cdot z_2$ and the sum z_1+z_2 are real numbers, then z_1 and z_2 are conjugate complex numbers.

36. Demonstrate that the complex number $a+bi$ whose modulus is equal to unity, $b\neq 0$, can be represented as

$$a+bi=\frac{c+i}{c-i}$$

where c is a real number.

Solve the following equations.

37. $\bar{z}=z$.
38. $\bar{z}=-z$.
39. $\bar{z}=2-z$.
40. $\bar{z}=-4z$.
41. $z^2+\bar{z}=0$.
42. $z^2+|z|=0$.

43. For what real values of x and y are the numbers $-3+ix^2y$ and x^2+y+4i conjugate complex?

44. Locate the complex numbers $z=a+bi$ for which:

(a) $\log_{\frac{1}{2}} \frac{|z-1|+4}{3|z-1|-2} > 1$, (b) $\log_{\sqrt{3}} \frac{|z|^2-|z|+1}{2+|z|} < 2$.

45. For every real number $a \geqslant 1$, find all the complex numbers z that satisfy the equation

$$z+a|z+1|+i=0$$

46. For every real number $a \geqslant 0$, find all the complex numbers z that satisfy the equation

$$2|z|-4az+1+ia=0$$

47. Solve the following system of equations in terms of complex numbers:

$$z^3+w^5=0$$
$$z^2 \cdot \overline{w}^4=1$$

1.6 Logarithms

When studying logarithms it is important to note that all the properties of logarithms are consequences of the corresponding properties of powers, which means that the student should have a good working knowledge of powers as a foundation for tackling logarithms. This close relationship between logarithms and powers stems from the definition of a logarithm in terms of the concept of a power.

Here is a definition taken from a commonly used textbook: "The logarithm of a given number to a given base is the exponent of the power to which the base must be raised in order to obtain the given number." Thus, a number x is the logarithm of a number N to the base a if $a^x=N$.

There is one very essential detail in this definition: no restrictions are imposed on the phrase "to a given base", and so if we are to follow this definition literally (and a definition must always be followed literally), then we will have to concede that 3 is the logarithm of -8 to the base -2 (since $(-2)^3=-8$), 2 is the logarithm of 4 to the base -2 (since $(-2)^2=4$), and so forth. As for the base 1, the situation is stranger still: any number x is the logarithm of 1 to the base 1 because $1^x=1$ for every x.*

Any person acquainted with the school course of mathematics will say that these examples are meaningless since we have to consider only logarithms *to a positive base different from* 1. True enough, that is the convention, but it is much better to impose this restriction on the base directly in the definition. And so the definition should read:

Let there be a number $a>0$ and $a \neq 1$. A number x is called the logarithm of a number N to the base a if $a^x=N$.

The more attentive readers have perhaps noticed that we have not once written $x=\log_a N$ but have always stated: x is the logarithm of

* Besides, any positive number is the logarithm of 0 to the base 0 since $0^x=0$ for all values of $x>0$.

N to the base a. The explanation is very simple. Until we are sure that no number can have two distinct logarithms to a given base, we have no right to use the equals sign. Indeed, imagine for a moment that some number N has two distinct logarithms to the same base a; then, using the equals sign, we would be able to write $\alpha=\log_a N$ and $\beta=$ $=\log_a N$, whence $\alpha=\beta$.*

For this reason, we will introduce a notation for logarithms only when we are convinced that *no number can have two distinct logarithms to the same base*. Indeed, if two distinct numbers α and β were logarithms of the number N to the base a, then, by definition, the following equations would hold true:

$$a^\alpha=N \text{ and } a^\beta=N \qquad (*)$$

whence $a^\alpha=a^\beta$. But then, by the properties of powers with positive base different from unity, we would arrive at the equation $\alpha=\beta$. Thus, if the number N has a logarithm to a base a, then this logarithm is **unique**; *we denote it by the symbol* $\log_a N$.

Thus, by definition,

$$x=\log_a N \text{ if } a^x=N$$

Consequently, the equations $x=\log_a N$ and $a^x=N$ (provided the restrictions imposed earlier on hold true) express one and the same relationship between the numbers x, a, N: in logarithmic form in the former case and in equivalent exponential form in the latter.

It is easy to prove that *negative numbers and zero do not have logarithms to any base a (with the usual provisions that $a>0$ and $a\neq 1$).* Indeed, if $N\leqslant 0$ and $x=\log_a N$, then $a^x=N\leqslant 0$, which contradicts the property of powers having a positive base.

As for positive numbers, we assume **without proof** that *any positive number to any base has a logarithm*. This assertion is taken in school to be self-evident and is not even stated, although it is no easy job to establish its validity (this would require invoking a highly developed theory of real numbers and the theory of limits).

Quite naturally, the student must have a thorough knowledge of the definition and of the **properties of logarithms** and must be able to prove them.

First of all, note the so-called *fundamental logarithmic identity*

$$a^{\log_a N}=N$$

* Recall that a very much similar situation arose when we defined the square root of a number (Sec. 1.4). There too we introduced the definition without the equals sign, and only later found that the definition with equality would have been simply impossible, since, on introducing the notation for a square root, we would have been able, straightway, to prove something like $2=-2$ (both numbers being "equal" to the root of 4 and hence equal). It is precisely for this reason that we have no symbol for the square root as such, but only the symbol $\sqrt{\ }$ for the principal square root of a positive number.

which is valid for every N and a such that $a>0, a\neq 1, N>0$. This identity follows directly from the equations (*).

Here are some formulas that are frequently used in problem solving (we stress once again that, according to the definition of a logarithm, all bases are positive and different from unity).

I. $\log_a MN = \log_a M + \log_a N \quad (M>0,\ N>0)$

II. $\log_a \frac{M}{N} = \log_a M - \log_a N \quad (M>0,\ N>0)$

III. $\log_a N^\alpha = \alpha \log_a N \quad (N>0,\ \alpha$ any number)

IV. $\log_{a^\beta} N^\alpha = \frac{\alpha}{\beta} \log_a N \quad (N>0,\ \alpha\neq 0,\ \beta\neq 0)$

V. $\log_b N = \frac{\log_a N}{\log_a b} \ (N>0)$

VI. $\log_b a \cdot \log_a b = 1$

Let us prove Formula I. Raise a to the power of $\log_a M + \log_a N$. By the property of powers and by the fundamental logarithmic identity we have

$$a^{\log_a M + \log_a N} = a^{\log_a M} \cdot a^{\log_a N} = MN$$

The resulting equation

$$a^{\log_a M + \log_a N} = MN$$

may be rewritten in logarithmic form [see (*)] thus: $\log_a M + \log_a N = \log_a MN$, which signifies the validity of Formula I.

Formula II is proved similarly.

To prove Equation III, raise a to the power $\alpha \log_a N$ and utilize the properties of powers:

$$a^{\alpha \log_a N} = (a^{\log_a N})^\alpha = N^\alpha$$

From this, by the definition of a logarithm, we obtain the required equation.

Equation IV follows from the manipulations

$$(a^\beta)^{\frac{\alpha}{\beta} \log_a N} = a^{\alpha \log_a N} = (a^{\log_a N})^\alpha = N^\alpha$$

It will prove useful to memorize the following two special cases of Formula IV:

IVa. $\log_{a^\beta} N = \frac{1}{\beta} \log_a N \ (N>0,\ \beta\neq 0)$

IVb. $\log_{a^\alpha} N^\alpha = \log_a N \quad (N>0,\ \alpha\neq 0)$

To prove V, let us first write it in the form $\log_a N = \log_a b \cdot \log_b N$. The proof is similar to that of the preceding case:

$$a^{\log_a b \cdot \log_b N} = (a^{\log_a b})^{\log_b N} = b^{\log_b N} = N$$

We can reason differently. Writing the fundamental logarithmic identity

$$b^{\log_b N} = N$$

we derive from it the equation

$$\log_a (b^{\log_b N}) = \log_a N$$

(equal numbers have the same logarithms!). Now, using Property III, we convince ourselves of the validity of Formula V.

Formula VI is a special case of the preceding one obtained for $N=b$. Equation V is usually called the *rule for changing the base of a logarithm*. This rule makes different tables of logarithms to various bases unnecessary, it suffices to have, say, tables of common logarithms (base 10). For instance, suppose it is required to compute $\log_5 13$. On the basis of Property V, we can write $\log_5 13 = \frac{\log_{10} 13}{\log_{10} 5}$. Using logarithmic tables, we find $\log_{10} 13 \approx 1.1139$ and $\log_{10} 5 \approx 0.6990$, and thus $\log_5 13 \approx 1.5937$.

Some other properties of logarithms that are absolutely necessary in the solution of inequalities are:

VII. *If* $a>1$, *then from* $0<x_1<x_2$ *it follows that* $\log_a x_1 < \log_a x_2$ *and from* $\log_a x_1 < \log_a x_2$ *it follows that* $0<x_1<x_2$. In other words, for $a>1$ the inequalities $0<x_1<x_2$ and $\log_a x_1 < \log_a x_2$ are equivalent (see Sec. 1.10).

VIII. *If* $0<a<1$, *then from* $0<x_1<x_2$ *it follows that* $\log_a x_1 > \log_a x_2$, *and from* $\log_a x_1 > \log_a x_2$ *it follows that* $0<x_1<x_2$. In other words, when $a<1$ the inequalitios $0<x_1<x_2$ and $\log_a x_1 > \log_a x_2$ are equivalent.

These two properties are proved in exactly the same way, and so we confine ourselves to proving Property VIII.

Let a number a be positive and less than unity. If the inequality $0<x_1<x_2$ holds, then there exist numbers $\log_a x_1$ and $\log_a x_2$. Using the fundamental logarithmic identity, rewrite the inequality $x_1<x_2$ in the form

$$a^{\log_a x_1} < a^{\log_a x_2}$$

Whence, by the properties of a power to a base less than unity, we conclude that $\log_a x_1 > \log_a x_2$.

Conversely, if the inequality $\log_a x_1 > \log_a x_2$ is true, then, firstly, both numbers x_1 and x_2 are positive. Secondly, raising the number a, $0<a<1$, to the powers $\log_a x_1$ and $\log_a x_2$, we get (again by the properties of powers to a base less than 1) the inequality

$$a^{\log_a x_1} < a^{\log_a x_2}$$

or $x_1<x_2$. Now since, as we have already mentioned, the numbers x_1 and x_2 are positive, it follows that $0<x_1<x_2$, which completes the proof.

The following statements are consequences of the properties that have just been proved:

VIIa. If $a > 1$, then the inequalities $\log_a x < \alpha$ and $0 < x < a^\alpha$ are equivalent.

VIIb. If $a > 1$, then the inequalities $\log_a x > \alpha$ and $x > a^\alpha$ are equivalent.

VIIIa. If $0 < a < 1$, then the inequalities $\log_a x < \alpha$ and $x > a^\alpha$ are equivalent.

VIIIb. If $0 < a < 1$, then the inequalities $\log_a x > \alpha$ and $0 < x < a^\alpha$ are equivalent.

To prove this it suffices to note that $\alpha = \log_a a^\alpha$.

From these statements it is easy to derive that logarithms of numbers exceeding 1 to bases exceeding 1 are positive and logarithms of numbers less than 1 (but positive!) are negative; and, conversely, logarithms to bases less than 1 are negative for numbers exceeding 1 and positive for numbers less than 1.

Let us now solve some problems involving the basic properties of logarithms.

1. *Compute* $\log_{3\sqrt{3}} 27$.

By Formula IV we have

$$\log_{3\sqrt{3}} 27 = \log_{3^{3/2}} 3^3 = \frac{3}{3/2} \log_3 3 = 2$$

2. *Compute* $2^{\log_{2\sqrt{2}} 15}$

By Formula IVa we have

$$\log_{2\sqrt{2}} 15 = \log_{2^{3/2}} 15 = \frac{2}{3} \log_2 15$$

Applying the fundamental logarithmic identity, we get

$$2^{\log_{2\sqrt{2}} 15} = 2^{2/3 \log_2 15} = (2^{\log_2 15})^{2/3} = 15^{2/3} = \sqrt[3]{225}$$

3. *Compute* $\log_3 5 \cdot \log_{25} 27$.

By Formula IV, we have

$$\log_3 5 \cdot \log_{25} 27 = \log_3 5 \cdot \log_{5^2} 3^3 = \frac{3}{2} \log_3 5 \cdot \log_5 3$$

And since, by Formula VI, $\log_3 5 \cdot \log_5 3 = 1$, it follows that $\log_3 5 \cdot \log_{25} 27 = 3/2$.

4. *Compute* $(\sqrt[3]{9})^{\frac{1}{5 \log_5 3}}$.

By Formula VI we have

$$\frac{1}{5 \log_5 3} = \frac{1}{5} \log_3 5$$

It then only remains to take advantage of the fundamental logarithmic identity and the laws of exponents:

$$(\sqrt[3]{9})^{\frac{1}{5\log_5 3}}=(9^{1/3})^{\frac{1}{5}\log_3 5}=(3^{2/3})^{\frac{1}{5}\log_3 5}$$
$$=(3^{\log_3 5})^{2/3\cdot 1/5}=5^{2/15}=\sqrt[15]{25}$$

5. *Compute* $\sqrt{\left(\frac{1}{\sqrt{27}}\right)^{2-\frac{\log_5 13}{2\log_5 9}}}$

Using in succession the laws of logarithms and exponents we compute the radicand:

$$\left(\frac{1}{\sqrt{27}}\right)^{2-\frac{\log_5 13}{2\log_5 9}}=\frac{1}{27}\cdot(\sqrt{27})^{\frac{1}{2}\log_9 13}=\frac{1}{27}\cdot(3^{\log_3 13})^{3/8}=3^{-3}\cdot 13^{3/8}$$

whence it is clear that the given number is equal to $3^{-3/2}\cdot 13^{3/16}$.

6. *Which is greater,* $\log_4 5$ *or* $\log_{1/16}\frac{1}{25}$?

By Formula IVb, we have

$$\log_{1/16}\frac{1}{25}=\log_{4^{-2}}5^{-2}=\log_4 5$$

so that the two numbers are equal.

7. *Compute* $\log_3 2\cdot\log_4 3\ \ldots\ \log_{10}9\cdot\log_{11}10$.

By Formula V,

$$\log_3 2=\frac{\log_{11}2}{\log_{11}3};\quad \log_4 3=\frac{\log_{11}3}{\log_{11}4};\ \ldots;\ \log_{10}9=\frac{\log_{11}9}{\log_{11}10}$$

whence

$$\log_3 2\cdot\log_4 3\ \ldots\ \log_{11}10=\frac{\log_{11}2}{\log_{11}3}\cdot\frac{\log_{11}3}{\log_{11}4}\ \cdots\ \frac{\log_{11}9}{\log_{11}10}\cdot\log_{11}10=\log_{11}2$$

8. *Prove that the ratio of the logarithms of two numbers is not dependent on the base; that is,*

$$\frac{\log_a N_1}{\log_a N_2}=\frac{\log_b N_1}{\log_b N_2}\quad (N_1>0,\ N_2>0,\ N_2\neq 1)$$

By Formula V we have

$$\frac{\log_a N_1}{\log_a N_2}=\log_{N_2}N_1 \text{ and } \frac{\log_b N_1}{\log_b N_2}=\log_{N_2}N_1$$

whence it is clear that our equation holds true.

9. *Which is greater,* $\log_2 3$ *or* $\log_{1/4}5$?

Since $\log_2 3>0$ and $\log_{1/4}5<0$, it follows that $\log_2 3>\log_{1/4}5$.

10. *Which is greater,* $\log_5 7$ *or* $\log_8 3$?

Since $\log_5 7 > 1$ and $\log_8 3 < 1$, it follows that $\log_5 7 > \log_8 3$.

11. *Compute* $\log_{ab} \frac{\sqrt[3]{a}}{\sqrt{b}}$ *if* $\log_{ab} a = 4$.

By the laws of logarithms we have

$$\log_{ab} \frac{\sqrt[3]{a}}{\sqrt{b}} = \frac{1}{3}\log_{ab} a - \frac{1}{2}\log_{ab} b = \frac{4}{3} - \frac{1}{2}\log_{ab} b$$

It remains to find the quantity $\log_{ab} b$. Since

$$1 = \log_{ab} ab = \log_{ab} a + \log_{ab} b = 4 + \log_{ab} b$$

it follows that $\log_{ab} b = -3$ and so

$$\log_{ab} \frac{\sqrt[3]{a}}{\sqrt{b}} = \frac{4}{3} - \frac{1}{2}\cdot(-3) = \frac{17}{6}$$

12. *Compute* $\log_6 16$ *if* $\log_{12} 27 = a$.

The chain of transformations

$$\log_6 16 = 4\log_6 2 = \frac{4}{\log_2 6} = \frac{4}{1+\log_2 3}$$

shows us that we have to know $\log_2 3$ in order to find $\log_6 16$. We find it from the condition $\log_{12} 27 = a$:

$$a = \log_{12} 27 = 3\log_{12} 3 = \frac{3}{\log_3 12} = \frac{3}{1+2\log_3 2} = \frac{3}{1+\frac{2}{\log_2 3}} = \frac{3\log_2 3}{2+\log_2 3}$$

which means that $\log_2 3 = \frac{2a}{3-a}$ (note that, obviously, $a \neq 3$). We finally have $\log_6 16 = \frac{4(3-a)}{3+a}$.

13. *Compute* $\log_{25} 24$ *if* $\log_6 15 = \alpha$ *and* $\log_{12} 18 = \beta$.

We have the equation

$$\log_{25} 24 = \frac{1}{2}\left(\log_5 3 + 3\log_5 2\right) = \frac{3}{2}\log_5 2 + \frac{1}{2}\log_5 3$$

which shows us that we have to determine $\log_5 2$ and $\log_5 3$. The equation $\log_6 15 = \alpha$ yields

$$\alpha = \log_6 15 = \log_6 3 + \log_6 5 = \frac{1}{1+\log_3 2} + \frac{1}{\log_5 2 + \log_5 3}$$

and the equation $\log_{12} 18 = \beta$ yields

$$\beta = \log_{12} 18 = \log_{12} 2 + 2\log_{12} 3 = \frac{1}{2+\log_2 3} + \frac{2}{1+2\log_3 2}$$

Taking logs to base 5 in all cases, we find, by Formula V,

$$\alpha=\frac{1}{1+\log_3 2}+\frac{1}{\log_5 2+\log_5 3}=\frac{1}{1+\frac{\log_5 2}{\log_5 3}}+\frac{1}{\log_5 2+\log_5 3}=\frac{1+\log_5 3}{\log_5 2+\log_5 3},$$

$$\beta=\frac{1}{2+\log_2 3}+\frac{2}{1+2\log_3 2}=\frac{1}{2+\frac{\log_5 3}{\log_5 2}}+\frac{2}{1+2\frac{\log_5 2}{\log_5 3}}=\frac{\log_5 2+2\log_5 3}{\log_5 3+2\log_5 2}$$

The last two equations may be regarded as a system of equations for determining $\log_5 2$ and $\log_5 3$:

$$\alpha\log_5 2+(\alpha-1)\log_5 3=1$$
$$(2\beta-1)\log_5 2+(\beta-2)\log_5 3=0$$

If $\alpha(\beta-2)-(\alpha-1)(2\beta-1)=-\alpha-\alpha\beta+2\beta-1\neq 0$ (see Sec. 1.11), then this system has the solution

$$\log_5 2=\frac{2-\beta}{\alpha+\alpha\beta-2\beta+1},\quad \log_5 3=\frac{2\beta-1}{\alpha+\alpha\beta-2\beta+1}$$

We finally get

$$\log_{25} 24=\frac{5-\beta}{2\alpha+2\alpha\beta-4\beta+2}$$

Now let us verify that the expression $\alpha+\alpha\beta-2\beta+1$ is indeed different from zero. Thus, we have

$$\alpha+\alpha\beta-2\beta+1=\log_6 15+\log_6 15\cdot\log_{12} 18-2\log_{12} 18+1$$
$$=(\log_6 15-\log_{12} 18+1)+\log_{12} 18\cdot(\log_6 15-1)$$

The second summand here is positive since $\log_{12} 18>0$ and $\log_6 15>1$. As to the first summand, using the properties of logarithms, we can write $\log_6 15>1$, $\log_{12} 18<2$ and so $\log_6 15-\log_{12} 18+1>0$. Thus, the expression $\alpha+\alpha\beta-2\beta+1$ is positive.

The properties of logarithms, among them the properties I to VIII given above, are widely used in solving a broad range of problems, such as logarithmic equations and systems, logarithmic inequalities, and so on. We give here some of the simpler kinds, leaving the complicated ones to Secs. 1.9 and 1.10.

14. *Solve the equation* $x+\log_{10}(1+2^x)=x\log_{10} 5+\log_{10} 6$.

Transposing $x\log_{10} 5$ to the left member of the equation and utilizing the laws of logarithms, we get

$$x+\log_{10}(1+2^x)-x\log_{10} 5=x\log_{10} 10-x\log_{10} 5$$
$$+\log_{10}(1+2^x)=\log_{10} 2^x(1+2^x)$$

The equation can thus be rewritten as $\log_{10} 2^x(1+2^x)=\log_{10} 6$, whence

$$(2^x)^2+2^x-6=0$$

Denoting $z=2^x$, we arrive at the quadratic equation $z^2+z-6=0$ which has the roots $z_1=-3$, $z_2=2$. Since the equation $2^x=-3$ is impossible (because 2^x is positive for all values of x), it remains to solve the equation $2^x=2$. It has the root $x=1$, which is the sole root of the original equation.

15. *Solve the equation*

$$\log_a(ax)\cdot\log_x(ax)=\log_{a^2}\frac{1}{a},\ \textit{where}\ a>0,\ a\neq 1$$

Clearly, the roots must satisfy the conditions $x>0$, $x\neq 1$. Using the properties of logarithms, transform the expressions that enter into this equation:

$$\log_x(ax)=1+\log_x a=1+\frac{1}{\log_a x}=\frac{\log_a x+1}{\log_a x},$$

$$\log_{a^2}\frac{1}{a}=-\frac{1}{2}\log_a a=-\frac{1}{2},\ \log_a(ax)=1+\log_a x$$

Our equation can now be rewritten as

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

whence $(\log_a x)^2+\frac{5}{2}\log_a x+1=0$. Solving this equation we get

$$x_1=\frac{1}{a^2},\quad x_2=\frac{1}{\sqrt{a}}$$

16. *Solve the system of equations*

$$5(\log_y x+\log_x y)=26$$
$$xy=64$$

It is clear that it must be true that $x>0$, $y>0$, $x\neq 1$, $y\neq 1$. Denoting $z=\log_x y$ and using Formula VI, we find that the first equation of the system can be rewritten as $5(z+1/z)=26$, whence $z_1=5$, $z_2=1/5$. This means that the solutions of the original system must be sought among the solutions of the system

$$\log_x y=5$$
$$xy=64$$

and of the system

$$\log_x y=1/5$$
$$xy=64$$

Solving these systems and choosing those solutions which satisfy the conditions $x>0$, $x\neq 1$, $y>0$, $y\neq 1$, we obtain the answer. The original system has two solutions: $x_1=2$, $y_1=32$, $x_2=32$, $y_2=2$.

17. *What can be said about the number x if it is known that for every real $a \neq 0$*

$$\log_x (a^2+1) < 0?$$

For every $a \neq 0$ the number $1+a^2 > 1$. But since the logarithm of a number greater than unity is negative only to a base less than unity, it follows that $x<1$. Furthermore, since logarithms are only considered to a positive base, $x>0$. And so finally we see that the number x of our problem is taken in the interval $0<x<1$.

18. *Find all x such that* $\log_{1/2} x > \log_{1/3} x$.

From Formula V we have

$$\log_{1/3} x = \frac{\log_{1/2} x}{\log_{1/2} \frac{1}{3}} = \log_{1/3} \frac{1}{2} \cdot \log_{1/2} x$$

And so our inequality can be rewritten as

$$\log_{1/2} x \left(1 - \log_{1/3} \frac{1}{2}\right) > 0$$

Since $1 - \log_{1/3} \frac{1}{2} > 0$, from the latter inequality we obtain $\log_{1/2} x > 0$, whence $x<1$. But the original inequality is meaningful only when $x>0$. Therefore all x that satisfy the original inequality lie in the interval $0<x<1$.

19. *Solve the inequality* $\frac{1}{\log_a x} > 1$, $a>1$.

The fraction $1/p$ is greater than unity if its denominator p lies between zero and unity. Thus, our task is to find values of x such that their logarithms (to the base $a>1$) lie between zero and unity, that is to say, so that the following two conditions hold true simultaneously: $0<\log_a x$ and $\log_a x < 1$. The first states that the values of x must exceed unity, the second that they must be less than a. Hence, the solution of the original inequality is the interval $1<x<a$.

We can also reason differently. The left member of the proposed inequality is meaningful only for positive values of x different from unity, and so the inequality may be rewritten as $\log_x a > 1$. This inequality holds true only for values of x which are greater than unity (since for $0<x<1$ we have $\log_x a < 0$ when $a>1$) but less than a (since for $x>a>1$ we have, by the logarithmic laws, $\log_x a < 1$).

In the foregoing examples, Formulas I to VI were used successfully to transform a variety of expressions both with concrete numbers and literal data. Such manipulations are necessary primarily in the solution of equations and inequalities.

But in many such cases these formulas are not sufficient. First of all, this is due to the fact that the letters in the formulas have to satisfy very stringent restrictions. A still greater drawback of Formulas

I to IV is that the right and left members are meaningful for **different** restrictions on the values of the literal elements that enter into them.

For example, in Formula I, $\log_a MN$ has meaning when the numbers M and N are both positive as well as when they are both negative. By contrast, the right-hand member of this formula is meaningful only in the first instance. But this means that if we transform an equation and replace the logarithm of a product of two expressions M and N containing the unknown by the sum of the logarithms of these expressions, then, for values of the unknowns which make M and N negative numbers, we change the meaningful expression $\log_a MN$ into a meaningless expression $\log_a M + \log_a N$. As is explained in Sec. 1.9, we would thus lose certain roots of the equation at hand.

The very same goes for Formulas II and III.

For these reasons, formulas of a more general nature are used in solving problems containing unknowns:

I*. $\log_a MN = \log_a |M| + \log_a |N| \quad (MN > 0)$

II*. $\log_a \frac{M}{N} = \log_a |M| - \log_a |N| \quad (MN > 0)$

III*. $\log_a N^{2k} = 2k \log_a |N| \quad (N \neq 0,\ k \text{ an integer})$

IV*. $\log_{x^{2k}} N = \frac{1}{2} \log_{|x|^k} N \quad (N > 0,\ k \neq 0 \text{ an integer},\ x \neq 0,\ |x| \neq 1)$

It should be noted that Formulas I* and II* also have the drawbacks stated above: their left and right members are meaningful for **different** restrictions on the values of the letters that enter into them. Namely, the right-hand members have meaning for arbitrary M and N different from zero, while the left-hand members are only meaningful for M and N having the same sign, which means that they are subject to more stringent restrictions. For this reason, replacing $\log_a MN$ by $\log_a |M| + \log_a |N|$ when solving equations can lead to extraneous solutions but not to the loss of solutions, as can happen when using Formulas I-IV. Since acquiring extraneous solutions of an equation is preferable to losing solutions (superfluous solutions may be discarded by verification, but lost solutions cannot be found!), one should use formulas I* to IV* when manipulating literal expressions.

Here are some problems which illustrate the importance of utilizing these properties.

20. *Simplify the expression*

$$\log_4 \frac{x^2}{4} - 2 \log_4 4x^4$$

and then compute its value for $x = -2$.

It is quite evident here that computations by Formulas I and III, that is,

$$\log_4 \frac{x^2}{4} - 2\log_4 4x^4 = 2\log_4 x - \log_4 4 - 2\log_4 4 - 8\log_4 x$$
$$= -3 - 6\log_4 x$$

are erroneous because the latter expression for $x=-2$ is meaningless, whereas the original one is meaningful and is equal to -6.

This paradoxical result is due to the fact that Formulas I and III are only applicable to positive values of the letters. Now if we use Formulas I * and III * in which the values of the letters may be negative as well, we get

$$\log_4 \frac{x^2}{4} - 2\log_4 4x^4 = 2\log_4 |x| - 1 - 2 - 8\log_4 |x| = -3 - 6\log_4 |x|$$

It is clear that for $x=-2$ this expression is equal to -6.

21. *Solve the system of equations*

$$\log_2 xy = 5$$
$$\log_{1/2} \frac{x}{y} = 1$$

Using Formulas I * and II *, rewrite the system as

$$\log_2 |x| + \log_2 |y| = 5$$
$$\log_{1/2} |x| - \log_{1/2} |y| = 1$$

Denoting $z_1 = \log_2 |x|$, $z_2 = \log_2 |y|$, we get

$$z_1 + z_2 = 5$$
$$z_1 - z_2 = -1$$

whence $z_1=2$, $z_2=3$, and so $|x|=4$, $|y|=8$.

But this does not mean that the original system has four solutions:

$$x_1 = 4, \quad y_1 = 8, \qquad x_2 = -4, \quad y_2 = -8,$$
$$x_3 = 4, \quad y_3 = -8, \quad x_4 = -4, \quad y_4 = 8$$

because it is required that the expressions $\log_2 xy$ and $\log_{1/2} \frac{x}{y}$ be meaningful. They will clearly have meaning only for x and y having the same signs. And so our system will only have two solutions: $x_1=4$, $y_1=8$, and $x_2=-4$, $y_2=-8$.

Thus, using Formulas I * and II * we acquired extraneous solutions which were readily discarded in a verification; now if we had used formulas I and II and had rewritten the system as

$$\log_2 x + \log_2 y = 5$$
$$\log_{1/2} x - \log_{1/2} y = 1$$

we would have lost the solution $x_2=-4$, $y_2=-8$.

Note also that the original system may be solved in a different way by reducing it directly to the system

$$xy = 32, \quad \frac{x}{y} = \frac{1}{2}$$

whence the required answer is obtained.

Exercises

Compute the following.

1. $(a^{\alpha})^{-\beta \log_{a^s} N^{\gamma}}$
2. $-\log_8 \log_4 \log_2 16.$
3. $\log_{\pi} \tan (0.25\pi).$
4. $\log_9 \tan \frac{\pi}{6}.$
5. $\log_{0.75} \log_2 \sqrt{\sqrt[-2]{0.125}}.$
6. $\sqrt[3]{5^{\frac{1}{\log_7 5}} + \frac{1}{\sqrt{-\log_{10} 0.1}}}.$
7. $a^{\frac{\log_b \log_b N}{\log_b a}}.$
8. $\log_{10} \tan 1° + \log_{10} \tan 2° + \ldots + \log_{10} \tan 89°.$
9. $\log_{10} \tan 1° \cdot \log_{10} \tan 2° \ \ldots \ \log_{10} \tan 89°.$
10. $2^{\log_3 5} - 5^{\log_3 2}.$
11. $\left(\frac{1}{49}\right)^{1+\log_7 2} + 5^{-\log_{1/5} 7}.$
12. Find x if $1 - \log_{10} 5 = \frac{1}{3}\left(\log_{10} \frac{1}{2} + \log_{10} x + \frac{1}{3} \log_{10} 5\right)$

In the following examples which number is greater:

13. $\log_3 2$ or $\log_2 3$?
14. $\log_4 5$ or $\log_6 5$?
15. $\log_2 3$ or $\log_3 11$?
16. $\log_2 a$ or $\log_3 a$?
17. $\log_a 2$ or $\log_a 3$?
18. $\sqrt[3]{0.01}$ or $\sqrt[5]{0.001}$?
19. Prove that if $\alpha = \log_{12} 18$ and $\beta = \log_{24} 54$, then $\alpha\beta + 5(\alpha - \beta) = 1$.
20. Find $\log_{54} 168$ if $\log_7 12 = a$ and $\log_{12} 24 = b$.
21. Find $\log_{30} 8$ if $\log_{30} 3 = a$ and $\log_{30} 5 = b$.
22. Prove the formula $\frac{\log_a N}{\log_{ab} N} = 1 + \log_a b$ and indicate the permissible values of the letters.
23. Prove the identity

$$\log_a N \cdot \log_b N + \log_b N \cdot \log_c N + \log_c N \cdot \log_a N = \frac{\log_a N \cdot \log_b N \cdot \log_c N}{\log_{abc} N}.$$

24. Compute the sum of

$$\frac{1}{\log_2 N} + \frac{1}{\log_3 N} + \frac{1}{\log_4 N} + \ldots + \frac{1}{\log_{1967} N}$$

where $N = 1967!$.

25. Prove that if a and b are the lengths of the legs and c is the length of the hypotenuse of a right triangle, $c - b \neq 1$ and $c + b \neq 1$, then

$$\log_{c+b} a + \log_{c-b} a = 2 \log_{c+b} a \cdot \log_{c-b} a$$

26. Prove that $|\log_b a + \log_a b| \geqslant 2$ (a and b are positive numbers not unity).

27. Prove that $\log_{\log_3 2} \frac{1}{2} > 0$.

28. Prove that $\log_2 17 \cdot \log_{1/5} 2 \cdot \log_3 \frac{1}{5} > 2$.

29. For what values of a and b is the following inequality valid:

$$\log_a (a^2 b) > \log_b \left(\frac{1}{a^5}\right) ?$$

30. Prove that $\log_2 3$ is an irrational number.

31. Find all real values of x for which the expression

$$\sqrt{\log_{\frac{1}{2}} \frac{x}{x^3-1}}$$

is a real number.

Solve the following equations:

32. $\dfrac{\log_8 \left(\frac{8}{x^2}\right)}{(\log_8 x)^2} = 3$.

33. $\sqrt{\log_2 x^4} + 4 \log_4 \sqrt{\frac{2}{x}} = 2$.

34. $\log_{10}^2 x^3 - 20 \log_{10} \sqrt{x} + 1 = 0$.

35. $\log_2 x - 8 \log_{x^2} 2 = 3$.

36. $\log_{\sqrt{x}} 2 + 4 \log_4 x^2 + 9 = 0$.

37. $\log_x 3 \cdot \log_{\frac{x}{3}} 3 + \log_{\frac{x}{81}} 3 = 0$.

38. $\log_x (ax) \cdot \log_a x = 1 + \log_x \sqrt{a}$.

39. $\log_3 a - \log_x a = \log_{\frac{x}{3}} a$.

40. $x^{(\log_3 x)^3 - 3 \log_3 x} = 3^{-3 \log_{2\sqrt{2}} 4 + 8}$.

Solve the following systems of equations:

41. $2^{\log_{1/2} (x+y)} = 5^{\log_5 (x-y)}$

$\log_2 x + \log_2 y = \frac{1}{2}$.

42. $\log_{a^2} y = 2y^2$

$\log_a \sqrt{xa} + 2 \log_{a^4} \frac{y}{\sqrt{a}} = \frac{1}{2}$.

43. $\log_2 xy \cdot \log_2 \frac{x}{y} = -3$

$\log_2^2 x + \log_2^2 y = 5$.

44. $x^{2/3} y^{1/3} - x^{1/3} y^{2/3} = 2$

$\log_3 x + \log_3 y = 1$.

45. $\log_a x - \log_{a^2} y = m$

$\log_{a^2} x - \log_{a^3} y = n$.

46. $\log_2 x+\log_4 y+\log_4 z=2$
$\log_3 y+\log_9 z+\log_9 x=2$
$\log_4 z+\log_{16} x+\log_{16} y=2.$
47. $2x^y-x^{-y}=1$
$\log_2 y=\sqrt{x}.$
48. $x^2=1+6\log_4 y$
$y^2=2^x\cdot y+2^{2x+1}.$
49. Solve the inequality $\log_{1/3}x+\log_3 x>1.$

50. Solve the inequality $x^{\frac{1}{\log_{10}x}}\cdot\log_{10}x<1.$
51. For which values of a are the roots of the equation $x^2-4x-\log_2 a=0$ real?

1.7 Progressions

The study of arithmetic progressions does not usually cause any trouble, but occasionally there is some misunderstanding in the definition of a geometric progression. We shall therefore investigate this matter in some detail.

Some textbooks define a geometric progression as follows: "A geometric progression is a sequence of numbers each term of which, from the second on, is equal to the preceding one multiplied by some *nonzero* number which is constant for the given sequence." Other textbooks fail to give the restriction "nonzero" number. Thus, from the viewpoint of the first definition, the sequence

$$2,\ 0,\ 0,\ 0,\ \ldots,\ 0,\ \ldots \qquad (1)$$

is not a geometric progression, whereas the second definition permits considering it a geometric progression with "common ratio zero."

There can of course be a certain choice of definition. That freedom always exists. It is meaningless to argue about whether a definition is correct or not, since a definition is never proved. But when stating a definition, one should be guided by considerations of expediency.

Let us consider the second definition from this point of view: we will allow for zero common ratios.

The introduction of the general notion of a geometric progression was due to the desire to study sequences with positive terms in which each successive term is greater than the preceding one by a *definite* number of times. Such sequences occur frequently in a broad range of problems. But if we take the sequence (1), the question of how many times greater is the third term than the second is devoid of any meaning. It is extremely desirable that a geometric progression be reconstructed *uniquely* on the basis of any term and the common ratio. However, if it is known that the common ratio of a progression is equal to zero and its third term is also zero, then it is clearly impossible to determine the first term unambiguously. What is more, if (given a common ratio of zero) the third term is nonzero (say unity), then there is no geometric progression that can satisfy these conditions.*

From the foregoing it is evident that sequences like (1) possess anomalous properties; it is quite inexpedient to call such sequences geometric progressions. It is thus reasonable to demand that the definition of a geometric progression require the common ratio to be nonzero.

Even this definition has certain drawbacks. Within the framework of this definition, there is nothing to prevent us from regarding the sequence

$$0,\ 0,\ 0,\ 0,\ \ldots,\ 0,\ \ldots \qquad (2)$$

as a geometric progression with common ratio 2 or 1/3. It is quite undesirable to have such ambiguousness in the common ratio of a progression.**

To eliminate this possibility, it is best to define a geometric progression as follows: *A geometric progression is a sequence of numbers such that the first term is nonzero, and each of the succeeding terms is equal to the preceding one multiplied by a certain nonzero number that is constant for the given sequence.* It follows from this definition that no term of a geometric progression can be zero.

Thus, if in the solution of a problem in which the first term b_1 of a geometric progression is a function of an unknown quantity we obtain for the unknown a value that makes b_1 vanish, then we must—according to the definition that we have just given—discard this value of the unknown as not satisfying the condition of the problem.

The definition of progressions is also involved in a question of the following kind: *Is the sequence* 1, 1, 1, ..., 1, ... *an arithmetic or a geometric progression?* This sequence can actually be regarded as an arithmetic progression (with common difference 0) or as a geometric progression (with common ratio 1).

Note that all the definitions and formulas in the theory of progressions (both arithmetic and geometric) remain valid when the terms of the progressions are *complex* numbers, but in most problems involving progressions it is assumed (if not otherwise stated) that the terms of the progressions are real numbers.

Let us now examine some problems that present difficulties.

1. *For what values of x do the three numbers* $\log_{10} 2$, $\log_{10}(2^x-1)$ *and* $\log_{10}(2^x+3)$ *taken in that order constitute an arithmetic progression?*

---

* There are also other "undesirable" effects which stem from the second definition given above. Note that all these effects have no analogues in the theory of arithmetic progressions. Yet it is extremely desirable for the theories of arithmetic and geometric progressions to be analogous.

** Also note that from the definition of the sum S of an infinite sequence it is easy to find the magnitude of the sum of the terms of the sequence (2): $S=0$. Therefore, if (2) is regarded as a geometric progression with common ratio 2, we would have an instance of a geometric progression having a sum but not being an infinitely decreasing progression.

The solution of this problem requires a knowledge both of progressions and the properties of logarithms.

Taking advantage of the definition of an arithmetic progression, we can reduce this problem to the equation $2\log_{10}(2^x-1)=\log_{10}2+\log_{10}(2^x+3)$, which has to be solved. Let us rewrite it as $(2^x-1)^2=2(2^x+3)$ or, denoting $2^x-1=z$, we have $z^2-2z-8=0$, whence $z_1=4$, $z_2=-2$. The root z_2 is extraneous since the inequality $2^x-1>0$ must hold true (the inequality $2^x+3>0$ is automatically valid for arbitrary x). The root z_1 leads to the equation $2^x-1=4$ from which we find $x=\log_2 5$.

2. *Given an arithmetic progression and a geometric progression with positive terms. The first terms of these progressions coincide, the second terms likewise coincide. Prove that any other term of the arithmetic progression does not exceed the corresponding term of the geometric progression.*

This problem is rather instructive for it points up an interesting relationship between an arithmetic and a geometric progression.

Thus, we have two progressions, the arithmetic progression $a_1, a_2, a_3, \ldots, a_n, \ldots$ and the geometric progression $b_1, b_2, b_3, \ldots, b_n, \ldots$ with $a_1=b_1$, $a_2=b_2$. Since all the terms of the arithmetic progression are positive, $a_1>0$ and the common difference $d\geqslant 0$. Then the equation $a_2=b_2$ shows that the common ratio $q=\frac{d}{a_1}+1\geqslant 1$.

It is necessary to prove that $b_{n+1}\geqslant a_{n+1}$, $n=2, 3, \ldots$, that is, that $a_1q^n-a_1-nd\geqslant 0$. This inequality can easily be obtained if we take advantage of the binomial theorem

$$a_1q^n-a_1-nd=a_1\left(1+\frac{d}{a_1}\right)^n-a_1-nd$$

$$=a_1\left(1+n\frac{d}{a_1}+C_n^2\frac{d^2}{a_1^2}+\ldots+\frac{d^n}{a_1^n}\right)-a_1-nd=C_n^2\frac{d^2}{a_1}+\ldots\geqslant 0$$

where the sequence of dots indicates that some positive terms are not shown. It is evident from this that the equation $a_n=b_n$ is only possible for all values of n when $d=0$, that is, if all the terms of the arithmetic progression are equal.

Another solution might be suggested that does not make use of the binomial theorem. It is this. Above, we found that the common ratio $q\geqslant 1$ and is connected with the common difference d by the relation $d=a_1q-a_1$, and so

$$a_1q^n-a_1-nd=a_1q^n-a_1-a_1nq+a_1n=a_1(q^n-1)-a_1n(q-1)$$

$$=a_1(q-1)[q^{n-1}+q^{n-2}+\ldots+q+1-n]\geqslant 0$$

because $a_1>0$, $q\geqslant 1$, and the expression in square brackets is also nonnegative since from $q\geqslant 1$ follows

$$q^2\geqslant 1,\ q^3\geqslant 1,\ \ldots,\ q^{n-1}\geqslant 1, \text{ or } q^{n-1}+q^{n-2}+\ldots+q^2+q+1\geqslant n$$

For all values of n, the equation $a_n=b_n$ is possible only when $q=1$, that is to say when all the terms of the geometric progression are equal.

Before going over to problems involving geometric progressions, we shall take up a theoretical question. In problem solving it is rather often necessary to write down the condition under which given numbers constitute a geometric progression. Ordinarily, the condition is written this way: the numbers b_1, b_2, b_3 constitute a geometric progression if

$$b_2:b_1=b_3:b_2$$

or

$$\frac{b_2}{b_1}=\frac{b_3}{b_2} \tag{3}$$

It is easy to see that the latter equation is not equivalent to the definition of a geometric progression $b_2=b_1q$, $b_3=b_2q$. Indeed, the triples 2, 0, 0 and 0, 0, 0 constitute geometric progressions (in the former case the common ratio is 0, in the latter, it is an arbitrary number), but (3) is meaningless for either of them. It is therefore incorrect to use (3) as a criterion of a geometric progression, particularly in cases when b_1 and b_2 are certain functions of an unknown, since it is not known beforehand whether they are always nonzero or not.

The required condition is more properly written as $b_2^2=b_1b_3$, which is meaningful for all values of b_1, b_2, b_3 (including the value zero), and not in the form of (3). This condition can be stated still more generally. As we know, the following property holds true: *in a geometric progression, the square of any term (except of course the first and last) is equal to the product of the adjacent terms.* It can readily be verified that the converse is also true: *if n numbers are arranged in some order and are such that the square of each one (except the first and last) is equal to the product of the adjacent numbers, then these numbers constitute a geometric progression.* Using this assertion we could straightway write down the condition ($n-2$ equations) under which the given n numbers constitute a geometric progression.

This property of geometric progressions is sometimes paraphrased by students as follows: any term (except the extreme terms) of a geometric progression is a geometric mean of the adjacent terms. It is quite evident that in this form the assertion is only valid for progressions with real positive terms. If we write it down, say, for the geometric progression $1, -3, 9$, then we get the erroneous equation $-3=\sqrt{1\cdot 9}$.

3. *Prove that the three numbers* $\frac{\sin\alpha}{6}$, $\cos\alpha$, $\tan\alpha$, *taken in that order, constitute a geometric progression only when* $\alpha=\pm\pi/3+2\pi k$, $k=0, \pm 1, \pm 2, \ldots$.

Although the problem is stated as a trigonometric problem, its solution reduces to the study of the roots of a third-degree polynomial.

The given numbers constitute a geometric progression if the following equation holds true:

$$\cos^2\alpha = \frac{\sin\alpha\tan\alpha}{6} \qquad (4)$$

Thus, the indicated three numbers constitute a geometric progression only if α satisfies equation (4), which may be rewritten as

$$6\cos^3\alpha + \cos^2\alpha - 1 = 0 \qquad (5)$$

Let us check to see if the given numbers do form a geometric progression for $\alpha = \pm\pi/3 + 2\pi k$. For that value of α we have $\cos\alpha = 1/2$ and substitution convinces us that this is a root of (5).

The student who confines himself to this reasoning has not yet solved the problem. We have verified that for the indicated values of α the given numbers do indeed form a geometric progression. It remains to prove that no other values of α permit constructing a geometric progression. Denote $\cos\alpha = z$ and consider the polynomial $6z^3 + z^2 - 1$. We already know one root $z = 1/2$ and we can factor the polynomial:*

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

Since the discriminant of the resulting quadratic trinomial is negative, equation (5) does not have any real roots other than $\cos\alpha = 1/2$. Consequently, (4), which is equivalent to the definition of a geometric progression, is only satisfied for the values of α indicated in the hypothesis.

4. *Find all the complex numbers x and y such that the numbers x, $x+2y$, $2x+y$ form an arithmetic progression while the numbers $(y+1)^2$, $xy+5$, $(x+1)^2$ form a geometric progression.*

We now consider progressions with complex terms.

Since the numbers x, $x+2y$, $2x+y$ form an arithmetic progression, we have the equation $(x+2y) - x = (2x+y) - (x+2y)$, whence $x = 3y$. Since the numbers $(y+1)^2$, $xy+5$, $(x+1)^2$ must form a geometric progression, the following equation must be true:

$$(xy+5)^2 = (x+1)^2(y+1)^2$$

Make a substitution $x = 3y$ and use the formula for the difference of two squares. Then we have an equation for determining y:

$$(y-1)(3y^2 + 2y + 3) = 0$$

which has three roots: one real and two imaginary. The condition of the problem is thus satisfied by three pairs of numbers:

$$x_1 = 3, \ x_2 = -1 + 2\sqrt{2}i, \ x_3 = -1 - 2\sqrt{2}i,$$
$$y_1 = 1, \ y_2 = \frac{-1 + 2\sqrt{2}i}{3}, \quad y_3 = \frac{-1 - 2\sqrt{2}i}{3}$$

* For example, by dividing the polynomial $6z^3 + z^2 - 1$ by $z - 1/2$. Then, in the resulting factorization, we can make all the coefficients integral.

In the first case, the arithmetic progression has the form 3, 5, 7 and the geometric progression, 4, 8, 16; in the second case, the arithmetic progression has the form

$$-1+2i\sqrt{2},\ \frac{-5+10i\sqrt{2}}{3},\ \frac{-7+14i\sqrt{2}}{3}$$

$\left(\text{with common difference } \frac{2+4i\sqrt{2}}{3}\right)$ and the geometric progression,

$$\frac{-4+8\sqrt{2}}{9},\ \frac{8-4i\sqrt{2}}{3},\ -8$$

(with common ratio $-2-i\sqrt{2}$). The progressions can be computed in similar fashion for the third case as well.

5. *Find a geometric progression of real numbers if it is known that the sum of the first four terms is equal to* 15 *and the sum of their squares is equal to* 85.

This problem does not present any difficulties as far as progressions are concerned, but the resulting equation merits special consideration.

Denote the first term of the desired progression by a and the common ratio by q. Then the other three terms are aq, aq^2 and aq^3. By hypothesis, we have the following system of equations:

$$a(1+q+q^2+q^3)=15$$
$$a^2(1+q^2+q^4+q^6)=85$$

Squaring the first equation and dividing the result by the second equation, we obtain an equation in q:

$$\frac{(1+q+q^2+q^3)^2}{1+q^2+q^4+q^6}=\frac{45}{17}$$

Simplifying in the ordinary way, we obtain an equation of the sixth degree. But let us approach it differently. Transform the left-hand member as follows:

$$\frac{(1+q+q^2+q^3)^2}{1+q^2+q^4+q^6}=\frac{\left(\frac{q^4-1}{q-1}\right)^2}{\frac{q^8-1}{q^2-1}}=\frac{(q^4-1)^2}{(q-1)^2}\cdot\frac{q^2-1}{q^8-1}=\frac{q^4-1}{q-1}\cdot\frac{q+1}{q^4+1}$$
$$=\frac{(q^3+q^2+q+1)(q+1)}{q^4+1}=\frac{q^4+2q^3+2q^2+2q+1}{q^4+1}$$

This yields the equation

$$\frac{q^4+2q^3+2q^2+2q+1}{q^4+1}=\frac{45}{17}$$

which can be reduced to the quartic (fourth-degree) equation

$$14q^4-17q^3-17q^2-17q+14=0 \tag{6}$$

This is a special kind of equation known as a *reciprocal* equation of the fourth degree: its coefficients equidistant from the ends are equal. The general form of a reciprocal equation of the fourth degree is

$$Ax^4+Bx^3+Cx^2+Bx+A=0$$

and any equation of this kind can be solved by means of an artificial device. We will demonstrate how this technique works in our specific case* but its generality is apparent at once.

It is clear that $q=0$ is not a root of (6) and so both members of this equation can be divided by q^2; then the left-hand member can be transformed as follows:

$$14\left(q^2+\frac{1}{q^2}\right)-17\left(q+\frac{1}{q}\right)-17=14\left[\left(q+\frac{1}{q}\right)^2-2\right]-17\left(q+\frac{1}{q}\right)$$
$$-17=14\left(q+\frac{1}{q}\right)^2-17\left(q+\frac{1}{q}\right)-45$$

Denoting $q+1/q$ by t, we obtain the equation $14t^2-17t-45=0$ whose roots are $t_1=5/2$, $t_2=-9/7$. It remains to solve two equations:

$$q+\frac{1}{q}=\frac{5}{2},\quad q+\frac{1}{q}=-\frac{9}{7}$$

The roots of the first equation are found without any difficulty: $q_1=2$, $q_2=1/2$; the second equation does not have any real roots. Substituting in succession the values of q into the first equation of the original system, we get $a_1=1$, $a_2=8$.

Thus the hypothesis of the problem is satisfied by the following two geometric progressions:**

$$1,\ 2,\ 4,\ 8,\ 16,\ \ldots \text{ and } 8,\ 4,\ 2,\ 1,\ \frac{1}{2},\ \ldots$$

Occasionally, the concept of a progression is linked up with some geometric fact. The following two problems are an illustration.

6. *For which values of the common ratio can three successive terms of a geometric progression be the sides of a triangle?*

It is quite obvious that for a common ratio equal to unity, any three terms of a geometric progression (with positive first term) can be the sides of an equilateral triangle.

---

* In our case $B=C$; in the general case the coefficients B and C are distinct.

** It was no accident that the equation (6) obtained in the solution of this problem turned out to be reciprocal. Reciprocal equations have the following property: *if a number a is a root, then 1/a is also a root*. The proof of this fact is carried out by direct verification. Equation (6) defines the common ratio of the progression. However, in this problem the order of the desired numbers is immaterial and so if some number q is a common ratio that satisfies the condition of the progression, then $1/q$ is likewise a common ratio of a progression satisfying this condition (this progression consists of the same numbers as that corresponding to the common ratio q but written in reverse order). Hence it is natural to expect that the equation for q would be a reciprocal equation.

Suppose we are given three unequal positive numbers a, aq, aq^2. For which values of q can these numbers be the sides of a triangle?

As we know, three line segments can form a triangle if and only if each of them is less than the sum of the other two. Thus, to see if three given line segments will form a triangle, it suffices to verify that the largest one is less than the sum of the other two.

Using this property we can solve the problem. Our task thus is to figure out for which values of q the largest one of the three numbers a, aq, aq^2 is less than the sum of the other two. Clearly, for $q>1$ the number aq^2 is the largest number, and for $q<1$, a is. We will consider two cases:

(a) Let $q>1$. Then the inequality

$$aq^2 < aq + a$$

must hold true. Since a is a positive number, we find that the ratio q must satisfy the inequality $q^2-q-1<0$. All the q's in the interval

$$\frac{-\sqrt{5}+1}{2} < q < \frac{\sqrt{5}+1}{2}$$

are solutions of this inequality. Noting that in the case at hand $q>1$, we obtain that q can vary in the interval

$$1 < q < \frac{\sqrt{5}+1}{2}$$

Thus, the three numbers a, aq, aq^2 can be the lengths of sides of a triangle for all values of q.

(b) Let $0<q<1$. Similarly, we can find an interval over which the common ratio q must vary. But we will approach this differently.

Rewrite the progression a, aq, aq^2 in reverse order aq^2, aq, a and denote its common ratio, $1/q$, by q'. Since $q<1$, then $q'=1/q>1$. We thus arrive at the earlier case and find that q' must vary over the interval $1<q'<\frac{\sqrt{5}+1}{2}$, or $1/q$ lies in the interval $1<\frac{1}{q}<\frac{\sqrt{5}+1}{2}$. From this we find that q in the case at hand can vary over the interval

$$\frac{\sqrt{5}-1}{2} < q < 1$$

Combining all the foregoing cases, we find that the common ratio q of the progression can only vary within the interval

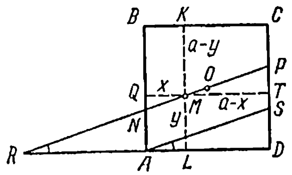
$$\frac{\sqrt{5}-1}{2} < q < \frac{\sqrt{5}+1}{2}$$

7. *A straight line is drawn through the centre of a square ABCD intersecting side AB at the point N so that AN : NB=1 : 2. On this line take an arbitrary point M lying inside the square. Prove that the distances from M to the sides of the square AB, AD, BC, CD, taken in that order, form an arithmetic progression.*

Referring to Fig. 14, denote the distance from M to AB by x and to AD by y; then the distances to the sides BC and CD will be $a-y$ and $a-x$, respectively, where a is a side of the square.

If we draw $AS \| NP$ through point A, it is easy to see that $\tan \angle SAD = 1/3$. Now extend NP to intersection with the prolongation of AD at point R and consider $\triangle RML$. Since $PD=2SD$, we

Fig. 14



conclude, by the property of parallel lines, that $RD=2AD=2a$, and so $RL=2a-(a-x)=a+x$. Furthermore we have $\frac{ML}{LR}=\tan \angle MRL$ whence $y=\frac{a+x}{3}$.

It now remains to verify that for arbitrary $0<x<a$ the four numbers x, $\frac{a+x}{3}$, $a-\frac{a+x}{3}=\frac{2a-x}{3}$, $a-x$ form an arithmetic progression. This verification is performed directly on the basis of the definition of an arithmetic progression:

$$\frac{a+x}{3}-x=\frac{2a-x}{3}-\frac{a+x}{3}=a-x-\frac{2a-x}{3}$$

At the examination, many students made the following serious mistake. It was asserted that the four numbers MQ, ML, MK, MT form an arithmetic progression simply because the sum of the extreme terms $MQ+MT=a$ is equal to the sum of the middle terms, $ML+MK=a$. This is quite illegitimate. If four numbers form an arithmetic progression, then the sum of the extreme numbers is equal to the sum of the middle numbers. However, this may be true, yet the four numbers need not necessarily constitute an arithmetic progression, as witness 1, 6, 5, 10.

In some problems the application of familiar formulas for arithmetic or geometric progressions is only possible after certain special transformations.

8. *Find the general term of the sequence* 2, 4, 7, 11, ... *having the property that the differences between successive terms constitute an arithmetic progression.*

We will indicate a technique for finding the general term of any sequence in which the differences between successive terms constitute an arithmetic progression.

Denote the given sequence by $a_1, a_2, a_3, \ldots$; then

$$\begin{aligned} a_2-a_1 &= r_1, \\ a_3-a_2 &= r_2, \\ a_4-a_3 &= r_3, \\ &\cdots\cdots \\ a_n-a_{n-1} &= r_{n-1} \end{aligned}$$

and, by hypothesis, the numbers $r_1, r_2, r_3, \ldots$ form an arithmetic progression. Adding all these equations and collecting terms, we get

$$a_n = a_1 + r_1 + r_2 + \ldots + r_{n-1}$$

In our case, the first term of the sequence is 2, and the sum of the arithmetic progression (the first term 2 and common difference 1) is readily found to be

$$r_1 + \ldots + r_{n-1} = \frac{2r_1 + d(n-2)}{2}(n-1) = \frac{(n-1)(n+2)}{2}$$

And so

$$a_n = 2 + \frac{(n-1)(n+2)}{2} = 1 + \frac{n(n+1)}{2}$$

9. *Find the sum of*

$$1 + 2\cdot 2 + 3\cdot 2^2 + 4\cdot 2^3 + 5\cdot 2^4 + \ldots + 100\cdot 2^{99}$$

The proposed sum is not the sum of a geometric progression. However, with a little ingenuity the solution is easily found on the basis of familiar formulas. Namely, write the proposed sum, call it S, as

$$\begin{aligned} S = 1 + 2 + 2^2 + 2^3 + \ldots &+ 2^{98} + 2^{99} \\ + 2 + 2^2 + 2^3 + \ldots &+ 2^{98} + 2^{99} \\ + 2^2 + 2^3 + \ldots &+ 2^{98} + 2^{99} \\ + 2^3 + \ldots &+ 2^{98} + 2^{99} \\ \cdots\cdots\cdots & \\ &+ 2^{98} + 2^{99} \\ &+ 2^{99} \end{aligned}$$

It is apparent that each of these lines is a sum of some geometric progression. Applying the appropriate formula, we get

$$\begin{aligned} S &= \frac{1(2^{100}-1)}{2-1} + \frac{2(2^{99}-1)}{2-1} + \frac{2^2(2^{98}-1)}{2-1} + \frac{2^3(2^{97}-1)}{2-1} + \ldots + \frac{2^{98}(2^2-1)}{2-1} \\ &+ \frac{2^{99}(2-1)}{2-1} = (2^{100}-1) + (2^{100}-2) \\ &+ (2^{100}-2^2) + (2^{100}-2^3) + \ldots + (2^{100}-2^{98}) + (2^{100}-2^{99}) \\ &= 100\cdot 2^{100} - (1 + 2 + 2^2 + \ldots + 2^{99}) \\ &= 100\cdot 2^{100} - \frac{1(2^{100}-1)}{2-1} = 99\cdot 2^{100} + 1 \end{aligned}$$

10. *Find the coefficient of* x^n *in the expansion of* $(1 + x + 2x^2 + 3x^3 + \ldots + nx^n)^2$.

Let us first try to figure out how terms in x^n can appear in the expansion. By the formula of the square of a polynomial, the desired expansion will contain, firstly, the squares of all the terms of the given sum and, secondly, all the doubled paired products of these terms. It is readily seen that if n is odd, then the terms in x^n in the expansion can only appear in the doubled products; but if n is even, the terms in x^n in the expansion will also stem from certain doubled products, but besides the same kind of term will result from squaring the term $\frac{n}{2}x^{\frac{n}{2}}$.

It is natural, therefore, when solving this problem, to consider two separate cases: n even and n odd.

We first consider the case when n is an even number. Then the proposed sum has $n+1$ terms, which is an odd number, and the given expression may be written as follows (isolating the "middle" term):

$$\left[1+x+2x^2+\ldots+\left(\frac{n}{2}-1\right)x^{\frac{n}{2}-1}+\frac{n}{2}x^{\frac{n}{2}}\right.$$
$$\left.+\left(\frac{n}{2}+1\right)x^{\frac{n}{2}+1}+\ldots+(n-1)x^{n-1}+nx^n\right]^2$$

Obviously, from among the squares of the terms of our sum, only the square of the "middle" term will yield a term in x^n; and only the doubled products of terms equidistant from the ends will yield the desired power. Thus, the desired coefficient is

$$\begin{aligned}S=\left(\frac{n}{2}\right)^2&+2\cdot1\cdot n+2\cdot1\cdot(n-1)+2\cdot2\cdot(n-2)+\ldots\\&+2\left(\frac{n}{2}-1\right)\left(\frac{n}{2}+1\right)=\left(\frac{n}{2}\right)^2+2n+2\cdot1\cdot n\\&+2\cdot2\cdot n+2\cdot3\cdot n+\ldots+2\left(\frac{n}{2}-1\right)n\\&-2\cdot1\cdot1-2\cdot2\cdot2-2\cdot3\cdot3-\ldots-2\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-1\right)\\&=\frac{n^2}{4}+2n+2n\left[1+2+3+\ldots+\left(\frac{n}{2}-1\right)\right]\\&\qquad-2\left[1^2+2^2+3^2+\ldots+\left(\frac{n}{2}-1\right)^2\right]\end{aligned}$$

The sum in the first square bracket is computed as the sum of an arithmetic progression, the sum in the second square bracket is determined from the formula for the sum of the squares of integers:

$$1^2+2^2+\ldots+k^2=\frac{k(k+1)(2k+1)}{6}$$

This formula is very useful and it is well to be able to prove it. A simple computation shows that for even n the coefficient of x^n is

$$S=\frac{n(n^2+11)}{6}$$

Now let n be odd. The proposed sum has $n+1$ terms, which is an even number, and the given expression has two "middle" terms

$$\left[1+x+2x^2+\ldots+\frac{n-1}{2}x^{\frac{n-1}{2}}+\frac{n+1}{2}x^{\frac{n+1}{2}}+\ldots\right.$$
$$\left.+(n-1)x^{n-1}+nx^n\right]^2$$

It is easy to see that the terms in x^n can now only appear from doubled products of the terms equidistant from the ends. Consequently, the desired coefficient is

$$S=2\cdot 1\cdot n+2\cdot 1\cdot(n-1)+2\cdot 2\cdot(n-2)+\ldots$$
$$+2\cdot\frac{n-1}{2}\cdot\frac{n+1}{2}=2n+2\cdot 1\cdot n$$
$$+2\cdot 2\cdot n+2\cdot 3\cdot n+\ldots+2\cdot\frac{n-1}{2}\cdot n$$
$$-2\cdot 1\cdot 1-2\cdot 2\cdot 2-2\cdot 3\cdot 3-\ldots-2\cdot\frac{n-1}{2}\cdot\frac{n-1}{2}$$
$$=2n+2n\left[1+2+3+\ldots+\frac{n-1}{2}\right]$$
$$-2\left[1^2+2^2+3^2+\ldots+\left(\frac{n-1}{2}\right)^2\right]=\frac{n(n^2+11)}{6}$$

It thus appears that the final formula does not depend on whether the number n is even or odd.

Many problems involving the setting up of equations are connected with progressions, but most of them do not use anything more than the definitions of progressions. We will consider here a so-called mixture problem. Problems of this nature in which a geometric progression is implicit occur quite frequently and cause trouble.

11. *A total of a litres of pure acid were drawn from a tank containing 729 litres of pure acid and were replaced by water. The result was thoroughly mixed to obtain a homogeneous solution and then another a litres of solution were drawn off, and again replaced by water, and again thoroughly mixed. This procedure was performed six times and the tank then contained 64 litres of pure acid. Determine a.*

The first time that a litres of pure acid were drawn off and replaced by water left the tank with $729-a$ litres of pure acid. It is also clear now that one litre of solution contains $\frac{729-a}{729}$ litre of pure acid. During the second procedure, $a\cdot\frac{729-a}{729}$ litres of pure acid were drawn off, and the tank then contained only

$$729-a-a\cdot\frac{729-a}{729}=\frac{(729-a)^2}{729}\ \text{litres}$$

of acid. Thus, after the second replacement with water there will be $\frac{729-a}{729}:729=\frac{(729-a)^2}{729^2}$ litre of acid in one litre of solution. And so the third reduction in acid amounts to $a\cdot\frac{(729-a)}{729^2}$ litres, with

$$\frac{(729-a)^2}{729}-a\cdot\frac{(729-a)^2}{729^2}=\frac{(729-a)^3}{729^2}$$

litres of acid remaining in the tank.

It is easy to see that the amount of acid in the tank after the sixth operation should be

$$\frac{(729-a)^6}{729^5}\ \text{litres}$$

True, this conjecture does not take the place of a proof. To obtain complete rigour, we would have to repeat this procedure six times and make sure that there is precisely that amount of pure acid left in the tank.* The equation

$$\frac{(729-a)^6}{729^5}=64$$

if we note that $2^6=64$ and $3^6=729$, yields $a=243$ immediately. Thus, 243 litres of liquid were drawn off in each operation.

In conclusion, a few words about infinitely decreasing geometric progressions. The important thing here is to realize the fundamental difference between the question of *defining the concept* of the sum of such a progression and the question of *computing this sum.*

If we have a certain finite sequence of numbers, then the *sum* u_1+ $+u_2+u_3+\ \dots\ +u_n$ of these numbers has a very definite meaning in accordance with the concepts of arithmetic. We have to find the number $S_2=u_1+u_2$, then add the third given number to obtain $S_3=S_2+$

---

* It is also easy to prove that the amounts of acid in the tank after each drawing-off operation are

$$729-a,\quad \frac{(729-a)^2}{729},\quad \frac{(729-a)^3}{729^2},\ \dots$$

and constitute a geometric progression with common ratio $\frac{729-a}{729}=1-\frac{a}{729}$. Indeed, if after the nth operation there remained k_n litres of pure acid in the tank, then there will be $\frac{k_n}{729}$ litre of acid in one litre of solution and thus, as a result of the $(n+1)$th operation the amount of acid will be reduced by $\frac{ak_n}{729}$ litres and will be equal to $k_{n+1}=k_n-\frac{ak_n}{729}=k_n\left(1-\frac{a}{729}\right)$ litres.

$+u_3=u_1+u_2+u_3$, then add the fourth given number to obtain $S_4=$ $=S_3+u_4$, and so on until all the given numbers have been exhausted.

Now when there are **infinitely many** numbers, this definition of a sum is no longer applicable because we can never exhaust all the numbers, and so, for the time being, the notation $u_1+u_2+u_3+\ldots+u_n+$ $+\ldots$ is devoid of any meaning, for we do not know how to add an infinity of numbers.

In this case we do as follows. We compute so-called *partial sums*, which are sums containing a finite number of summands:

$$\begin{aligned} &S_1=u_1, \\ &S_2=u_1+u_2, \\ &S_3=u_1+u_2+u_3, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &S_k=u_1+u_2+u_3+\ldots+u_k, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

They may be computed in the ordinary way for any value of k. Then the sequence of numbers

$$S_1, S_2, S_3, \ldots, S_k, \ldots$$

either has a limit or it does not. If the sequence of partial sums does not have a limit as $k\to\infty$, the sum of numbers $u_1, u_2, u_3, \ldots, u_n, \ldots$ is not defined. *If such a limit exists, then, by definition, it is called the sum of the given numbers.* If that limit is equal to S:

$$\lim_{k\to\infty} S_k=S$$

then we write

$$u_1+u_2+u_3+\ldots+u_n+\ldots=S$$

Thus, the sum of an infinite set of numbers is quite a new concept that differs essentially from the notion of a sum as given in arithmetic.

If we now consider the *infinitely decreasing geometric progression**

$$a,\ aq,\ aq^2,\ \ldots,\ aq^n,\ \ldots,\ |q|<1$$

then, as the textbooks demonstrate, the sequence of partial sums in this case does indeed have a limit, and it is equal to $S=\frac{a}{1-q}$. Thus, *an infinitely decreasing geometric progression has a sum, and this sum is* $S=\frac{a}{1-q}$.

---

* A geometric progression with an infinite number of terms and with common ratio less than unity in absolute value is for short termed an *infinitely decreasing progression*. However this name is not quite exact: an infinitely decreasing progression may not be decreasing, and may be truly decreasing only if its first term is positive and the common ratio satisfies the inequality $0<q<1$. For instance, the infinitely decreasing geometric progression $1, -1/2, 1/4, -1/8, \ldots$ is not a decreasing progression by generally accepted terminology.

The important thing to bear in mind is that one must be convinced of the **existence** of a sum of an infinite sequence of numbers before attempting to **compute** that sum. This is vividly illustrated in the following example. Let us attempt to find the sum of the following infinite geometric progression: 2, 4, 8, 16, ..., 2^n, ... If we fail to pose the question of the existence of a sum and denote it directly by S

$$S=2+4+\ldots+2^{n-1}+2^n+\ldots$$

then, multiplying both members of this equation by 2, we get

$$4+8+\ldots+2^n+2^{n+1}+\ldots=2S$$

whence $2+2S=S$ and, hence $S=-2$. This strange result—the sum of an infinity of positive numbers is negative—is not difficult to explain: the original sequence does not have any sum and so our computations are meaningless. Indeed, the sequence of partial sums

$$S_1=2,\ S_2=2+4=6,\ S_3=2+4+8=14,\ \ldots,$$
$$S_k=2^{k+1}-2,\ \ldots$$

as $k\to\infty$ has no limit.

12. *Find an infinitely decreasing geometric progression having the property that its sum is twice the sum of the first k terms.*

This problem is interesting in that we do not arrive at an **unambiguous** answer, which is a situation that occasionally arises in solving problems involving progressions.

Denote the first term of the progression by a and the common ratio by q; then, by hypothesis, we can write

$$\frac{a}{1-q}=\frac{2a(1-q^k)}{1-q}$$

where k is the number given in the problem. From this we have

$$1=2(1-q^k),\quad q^k=\frac{1}{2}$$

If k is odd, then the equation has a unique root (we confine ourselves to progressions with real terms)

$$q=\sqrt[k]{1/2}$$

which in this case is the common ratio of the progression; but if k is even, then we have two roots:

$$q=\pm\sqrt[k]{1/2}$$

and both roots can be ratios of the desired progression. Thus, for k even, the common ratio of a progression is not defined uniquely.

The first term of the progression cannot be found since we lack the conditions. This means that any infinitely decreasing geometric progression with one of the indicated values of the common ratio and with an arbitrary first term has the property stated in this problem.

Exercises

1. The first term of a geometric progression is equal to $x-2$, the third term is $x+6$, and the arithmetic mean of the first and third terms stands in the ratio to the second term as 5 : 3. Determine x.

2. Let $\frac{1}{a+b}$, $\frac{1}{b+c}$ and $\frac{1}{c+a}$ be three successive terms of an arithmetic progression. Prove that in that case b^2, a^2 and c^2 are also successive terms of an arithmetic progression.

3. Three nonzero real numbers form an arithmetic progression and the squares of these numbers taken in the same order constitute a geometric progression. Find all possible common ratios of the geometric progression.

4. The sides of right triangle form a geometric progression. Find the tangents of the acute angles.

5. Let x_1 and x_2 be the roots of the equation $x^2-3x+A=0$ and let x_3 and x_4 be the roots of the equation $x^2-12x+B=0$. It is known that the numbers x_1, x_2, x_3, x_4 (in that order) form an increasing geometric progression. Find A and B.

6. Along a road lie an odd number of stones spaced at intervals of 10 metres. These stones have to be assembled around the middle stone. A person can carry only one stone at a time. A man started the job with one of the end stones by carrying them in succession. In carrying all the stones, the man covered a total of 3 km. How many stones were there?

7. Three brothers whose ages form a geometric progression divide a certain sum of money in proportion to their ages. If they do that three years later, when the youngest is half the age of the oldest, then he will receive 105 roubles, and the middle brother will get 15 roubles more than he gets now. Give the ages of the brothers.

8. Let b_1, b_2, b_3 be three successive terms of a geometric progression with common ratio q. Find all the values of q for which the inequality $b_3 > 4b_2 - 3b_1$ holds true.

9. All the harvesting combines on a certain farm can take in the harvest in 24 hours. However, according to the work plan, only one combine was in operation during the first hour, two during the second hour, three during the third hour, and so on, until all combines were in operation. All machines worked together for only a few hours before the harvest was completely taken in. The operating time provided by the plan would have been cut by 6 hours if all combines (except five) had started work from the very beginning. How many harvesting combines were there on the farm?

10. A tank was filled with petrol during the course of an integral number of hours, and the ratio of the amount added each succeeding hour to the amount added in the preceding hour was a constant. An hour before completion of the job there were 372 litres in the tank. If from a full tank we draw off the amount of petrol that was put in during the first hour, there will be 186 litres left; then if that quantity which was added during the second, last and second to the last hours together is drawn off, then the tank will have 72 litres. How much petrol was added to the tank during the first hour?

11. To a vessel containing pure water are added 6 litres of a solution containing 64% (by volume) alcohol. The solution is thoroughly mixed and an equal quantity (6 litres) of the solution is drawn off. How much water was there originally in the vessel if after performing this operation three times the vessel contained a solution with a 37% (by volume) concentration of alcohol?

12. It is given that the numbers 3, $3\log_y x$, $3\log_z y$, $7\log_x z$ form an arithmetic progression. Prove that $x^{18}=y^{21}=z^{28}$.

13. Solve the system of equations

$$\begin{aligned} 2x^4 &= y^4+z^4 \\ xyz &= 8 \end{aligned}$$

knowing that the logarithms $\log_y x$, $\log_z y$, $\log_x z$ form a geometric progression.

14. For an arithmetic progression $a_1, a_2, \ldots, a_k, \ldots$ the equation $s_m/s_n=m^2/n^2$ (s_k is the sum of the first k terms of the progression) holds true. Prove that

$$\frac{a_m}{a_n}=\frac{2m-1}{2n-1}$$

15. Let the numbers $a_1, a_2, \ldots, a_n$ constitute a geometric progression. Knowing the sums

$$s=a_1+a_2+\ldots+a_n \quad \text{and} \quad T=\frac{1}{a_1}+\frac{1}{a_2}+\ldots+\frac{1}{a_n}$$

find the product $P=a_1a_2 \ldots a_n$.

16. For which values of x do the numbers $2x^2$, x^4, 24 taken in that order form an arithmetic progression?

17. For which values of x do the numbers 1, x^2, $6-x^2$ taken in that order form a geometric progression?

1.8 Proving inequalities

It is quite common, at examinations, for the student to be asked to solve certain numerical or literal inequalities. This section is devoted to an analysis of proofs of literal and numerical inequalities. It would be nice of course if there were some unified method for proving all inequalities. Unfortunately, no such method exists. But we will give below a number of techniques that are of use in proving a rather large number of inequalities.

First we take up some inequalities that are frequently used in problem solving, such as the inequality between an arithmetic mean and a geometric mean, a consequence of this inequality concerning the sum of reciprocal quantities, and also the following trigonometric inequality:

$$-\sqrt{a^2+b^2}\leqslant a\sin x+b\cos x\leqslant\sqrt{a^2+b^2} \qquad (1)$$

The relation between an arithmetic mean and a geometric mean of two numbers reads:

For any two nonnegative numbers a and b the inequality

$$\sqrt{ab}\leqslant\frac{a+b}{2} \qquad (2)$$

holds true; equality occurs only when $a=b$.

A special case of (2) is the inequality

$$x+\frac{1}{x}\geqslant 2$$

which is valid for all $x>0$. In this inequality, the equals sign holds for $x=1$ only. It is useful to remember the verbal statement of this inequality.

The sum of two positive reciprocals does not exceed two, and is equal to two only when both numbers are equal to unity.

Also note that for any $x \neq 0$ the inequality

$$\left|x+\frac{1}{x}\right| \geqslant 2$$

or

$$\left|\frac{1+x^2}{2x}\right| \geqslant 1 \qquad (3)$$

holds true.

1. *Prove the inequality* $\frac{1}{\log_2 \pi}+\frac{1}{\log_\pi 2}>2.$

By the properties of logarithms, $1/\log_\pi 2=\log_2 \pi>0$, which means that the left member of our inequality is the sum of two positive reciprocals different from unity ($\log_2 \pi \neq 1$). Such a sum is greater than two. Hence, the original inequality holds true.

2. *Prove that if* $a>0$, $b>0$, $c>0$, *then*

$$\frac{bc}{a}+\frac{ac}{b}+\frac{ab}{c} \geqslant a+b+c$$

We take advantage of the following inequalities:

$$\frac{1}{2}\left(\frac{bc}{a}+\frac{ac}{b}\right) \geqslant c, \quad \frac{1}{2}\left(\frac{ac}{b}+\frac{ab}{c}\right) \geqslant a, \quad \frac{1}{2}\left(\frac{bc}{a}+\frac{ab}{c}\right) \geqslant b$$

(these inequalities are valid because the left members are arithmetic means and the right members are geometric means of positive numbers). Combining them term by term, we get the inequality that we wish to prove.

3. *Prove that if* $a>0$, $b>0$, $c>0$, *then*

$$(a+b)(b+c)(a+c) \geqslant 8abc$$

Taking the following inequalities [see formula (2)]

$$a+b \geqslant 2\sqrt{ab}, \quad b+c \geqslant 2\sqrt{bc}, \quad a+c \geqslant 2\sqrt{ac}$$

and multiplying them termwise, we get the desired inequality.

This inequality may be proved in a different manner by using the inequality between the arithmetic mean and the geometric mean for 8 positive numbers [see formula (5)]. Indeed, removing brackets in the left member of our inequality, we find that it can be rewritten as follows:

$$\frac{a^2b+b^2c+c^2b+a^2c+b^2a+c^2a+abc+abc}{8} \geqslant abc$$

On the left side we have the arithmetic mean of 8 positive numbers; on the right, as can readily be verified, we have their geometric mean, which completes the proof of the original inequality.

Before going on to the next problem, let us dwell on a typical mistake that is rather often made in proving inequalities. It is this. The student writes the inequality to be proved, then performs certain (quite legitimate!) manipulations and finally arrives at an obviously valid inequality (say $1<2$ or $(a-b)^2 \geqslant 0$), and then concludes: "hence, the inequality is proved." This is a crude logical error: from the fact that a true inequality has been obtained, we can by no means conclude that the original inequality was true! To be more exact, we proved the following: if one assumes that the proposed inequality is true, then the inequality obtained via a chain of transformations is also true. But it is obvious that the final inequality is true as it stands; and we continue to know nothing about the inequality which we set out to prove.

It is logically correct to reason in the reverse order. It is necessary to take some obviously valid inequality and perform manipulations (which of course must be legitimate from the viewpoint of algebra and trigonometry) that will bring us to the inequality to be proved. This is justified reasoning: we started with a valid inequality and via a chain of legitimate transformations arrived at the new inequality, which, hence, must also be valid.

Of course there remains the most important question. From what inequality are we to proceed so as to transform it into the required inequality? To answer this question we can perform the transformation of the proposed inequality that leads us to an obviously valid inequality. However, this stage in the solution of the problem must be regarded as an *exploratory search* for the proof, as an attempt to get the proper approach, but not as proof in itself. If as a result of this exploration (manipulations) we have obtained an obviously true inequality, then we can begin the proof proper: take this obviously correct inequality and manipulate it as we did in the exploratory search, but in reverse order; inverse the manipulations, so to say. If this "work backward procedure" is everywhere legitimate, then the inequality being proved is indeed valid.

Incidentally, a somewhat different procedure is often followed. If, in the process of exploring the proof via a reduction of our inequality to an obvious inequality, we always replaced the given inequality by an equivalent one (see Sec. 1.10), then the last inequality will be equivalent to the original one, and therefore its validity implies the validity of the original inequality. Hence, if at each stage in the transformation we specifically verified and stressed the equivalence of the inequalities, then the "work backward procedure" is not necessary.

We shall follow this reasoning in carrying out the proof of the following inequalities.

4. *Prove the inequality*

$$\frac{a^3+b^3}{2} \geqslant \left(\frac{a+b}{2}\right)^3 \text{ where } a>0,\ b>0$$

Replace this inequality by the equivalent one

$$\frac{a^3+b^3}{2}-\left(\frac{a+b}{2}\right)^3 \geqslant 0$$

Removing brackets and regrouping, we can write it in the equivalent form

$$\frac{3}{8}(a+b)(a-b)^2 \geqslant 0$$

Since $a>0$ and $b>0$, this inequality is obvious and, thus, the validity of the equivalent original inequality is proved.

5. *Prove that if $a>0$, $b>0$, then for any x and y the following inequality holds true*:

$$a\cdot 2^x+b\cdot 3^y+1 \leqslant \sqrt{4^x+9^y+1}\cdot\sqrt{a^2+b^2+1}$$

By hypothesis, both sides of this inequality are positive and so it is equivalent (see Sec. 1.10) to the following:

$$(a\cdot 2^x+b\cdot 3^y+1)^2 \leqslant (4^x+9^y+1)(a^2+b^2+1)$$

or to

$$a^2\cdot 4^x+b^2\cdot 9^y+1+2ab\,2^x3^y+2a2^x+2b3^y$$
$$\leqslant 4^xa^2+4^xb^2+4^x+9^ya^2+9^yb^2+9^y+a^2+b^2+1$$

Transposing all terms of this inequality to the right side, and then collecting like terms and regrouping, we can write it in the equivalent form

$$(a^29^y-2ab2^x3^y+4^xb^2)+(4^x-2a2^x+a^2)+(9^y-2b\cdot 3^y+b^2) \geqslant 0$$

Since each parenthesis is a perfect square, the original inequality is equivalent to the following obvious inequality:

$$(a3^y-b2^x)^2+(2^x-a)^2+(3^y-b)^2 \geqslant 0$$

Hence the original inequality is true.

Note that this inequality is also true for any real values of a and b (the proof of this fact is left to the reader).

6. *Prove that the inequality*

$$-1 \leqslant \frac{\sqrt{3}\sin x}{2+\cos x} \leqslant 1$$

is valid for arbitrary x.

This inequality (see Sec. 1.4) is equivalent to the inequality $|(\sqrt{3}\times \times\sin x)/(2+\cos x)| \leqslant 1$. Since both members of this inequality are nonnegative, then after squaring and multiplying by the positive expression $(2+\cos x)^2$, we get an equivalent inequality: $3\sin^2 x \leqslant (2+\cos x)^2$. Replacing $\sin^2 x$ by $1-\cos^2 x$ and grouping, we finally get $(2\cos x+$

$+1)^2 \geqslant 0$. This inequality holds true for all x, and since it is equivalent to the original one, the original inequality is also true, which is what we set out to prove.

The original inequality may be proved differently by making use of inequality (1). Indeed, since $2+\cos x > 0$ for all x, then, after multiplying by $2+\cos x$, we get the following double inequality which is equivalent to the original one:

$$-2-\cos x \leqslant \sqrt{3}\sin x \leqslant 2+\cos x$$

The inequality on the left may be written as

$$-2 \leqslant \sqrt{3}\sin x + 1\cdot\cos x$$

It is now evident that this is a special case of inequality (1), which is a true inequality. The validity of the inequality on the right is proved similarly.

7. *Prove that for arbitrary* α *the inequality* $4\sin 3\alpha+5 \geqslant 4\cos 2\alpha+ +5\sin\alpha$ *is valid.*

One of the most crude errors made in proving this inequality is the "proof" by substitution of specific values.

At an examination, a number of students reasoned something like this: "For $\alpha=0°$ the inequality holds because $5>4$, for $\alpha=30°$ the inequality is true because $4+5>4\cdot\frac{1}{2}+5\cdot\frac{1}{2}$, for $\alpha=45°, 60°, 90°$ it is also obviously true, which means it holds true for all values of α."

Actually, of course, these students proved the inequality only for several separate values of α and offered no proof whatsoever for the remaining values of α. A correct proof is as follows.

We know that $\sin 3\alpha=3\sin\alpha-4\sin^3\alpha$, $\cos 2\alpha=1-2\sin^2\alpha$, and so the original inequality may be rewritten as

$$4(3\sin\alpha-4\sin^3\alpha)+5 \geqslant 4(1-2\sin^2\alpha)+5\sin\alpha$$

or

$$16\sin^3\alpha-8\sin^2\alpha-7\sin\alpha-1 \leqslant 0$$

The latter inequality should be valid for all values of α. Denoting $\sin\alpha$ by x, we rewrite it as

$$16x^3-8x^2-7x-1 \leqslant 0$$

We now have to prove that this inequality is valid for arbitrary values of x in the interval $-1 \leqslant x \leqslant 1$.

Let us apply the grouping method illustrated earlier. The latter inequality can then be rewritten as

$$8x^2(x-1)+7x(x^2-1)+(x^3-1) \leqslant 0$$

or $(x-1)(4x+1)^2 \leqslant 0$. This inequality is clearly valid, and so the original inequality is proved.

The proof of certain inequalities requires skill in utilizing the properties of functions that enter into the inequalities.

8. *Prove that the inequality*

$$\cos(\cos x) > 0$$

is true for all x.

For all x we have $-1 \leqslant \cos x \leqslant 1$. Put $\alpha = \cos x$ to get $-1 \leqslant \alpha \leqslant 1$. Since $-\pi/2 < -1$ and $1 < \pi/2$, it follows all the more so that α satisfies the condition $-\pi/2 < \alpha < \pi/2$. The properties of the function $y = \cos x$ imply that $\cos \alpha$ is positive for all these values of α, which actually is what we set out to prove.

9. *Prove the inequality* $\cos(\sin x) > \sin(\cos x)$.

It can be rewritten as

$$\cos(\sin x) - \cos\left(\frac{\pi}{2} - \cos x\right) > 0$$

or

$$2\sin\left(\frac{\pi}{4} + \frac{\sin x - \cos x}{2}\right)\sin\left(\frac{\pi}{4} - \frac{\sin x + \cos x}{2}\right) > 0$$

We will show that the factors in the left-hand member are positive.

Since

$$|\sin x - \cos x| = |\sqrt{2}\sin(x - \pi/4)| \leqslant \sqrt{2} < \pi/2$$

it follows that

$$-\frac{\pi}{4} < \frac{\sin x - \cos x}{2} < \frac{\pi}{4}$$

and therefore

$$0 < \frac{\pi}{4} + \frac{\sin x - \cos x}{2} < \frac{\pi}{2}$$

Consequently

$$\sin\left(\frac{\pi}{4} + \frac{\sin x - \cos x}{2}\right) > 0$$

for all x. A similar proof is given that

$$\sin\left(\frac{\pi}{4} - \frac{\sin x + \cos x}{2}\right) > 0$$

A decisive factor in the examples which follow is the use of properties of the exponential function $y = a^x$: *if* $a > 1$, *a larger value of the argument is associated with a larger value of the function and, hence, a greater value of the function is associated with a greater value of the argument; if* $a < 1$, *to a greater value of the argument corresponds a smaller value of the function and, hence, to a larger value of the function corresponds a smaller value of the argument.*

10. *Prove that for positive numbers c and d and arbitrary* $\alpha > 0$, *the inequalities* $c < d$ *and* $c^\alpha < d^\alpha$ *are equivalent.*

Let c and d be positive numbers and $\alpha>0$. Consider the function $y=(c/d)^x$.

If $c<d$, then $0<c/d<1$. By the property of an exponential function with base less than unity, we have

$$\left(\frac{c}{d}\right)^{\alpha}<\left(\frac{c}{d}\right)^{0}$$

whence follows $c^{\alpha}/d^{\alpha}<1$, or $c^{\alpha}<d^{\alpha}$.

Conversely, if $c^{\alpha}<d^{\alpha}$, then $c^{\alpha}/d^{\alpha}<1$, or

$$\left(\frac{c}{d}\right)^{\alpha}<\left(\frac{c}{d}\right)^{0}$$

This means that the larger value of the argument ($\alpha>0$) of our function is associated with a smaller value of the function. But this is true only when the base is less than unity, that is, $c/d<1$, whence $c<d$.

The statement we have just proved is ordinarily formulated as follows: an inequality between positive numbers may be raised to any positive power; in particular, a root of any degree may be extracted.

11. *Prove the inequality*

$$(a^{\alpha}+b^{\alpha})^{1/\alpha}\leqslant(a^{\beta}+b^{\beta})^{1/\beta}$$

for $a\geqslant 0,\ b\geqslant 0,\ \alpha>\beta>0.$

If $a=0$ or $b=0$, then the proposition is obvious. Now let $a>0$ and $b>0$. It is clear that one of these numbers does not exceed the other. Suppose, say, $0<a\leqslant b$. Then $0<a/b\leqslant 1$, and since $\alpha>\beta$, it follows that

$$0<(a/b)^{\alpha}\leqslant(a/b)^{\beta} \text{ and } 1+(a/b)^{\alpha}\leqslant 1+(a/b)^{\beta}$$

From the latter inequality we get (see Example 10)

$$[1+(a/b)^{\alpha}]^{1/\beta}\leqslant[1+(a/b)^{\beta}]^{1/\beta}$$

Furthermore, since

$$1+(a/b)^{\alpha}\geqslant 1 \text{ and } 0<1/\alpha<1/\beta$$

it follows that

$$[1+(a/b)^{\alpha}]^{1/\alpha}\leqslant[1+(a/b)^{\alpha}]^{1/\beta}$$

Now we can write

$$[1+(a/b)^{\alpha}]^{1/\alpha}\leqslant[1+(a/b)^{\alpha}]^{1/\beta}\leqslant[1+(a/b)^{\beta}]^{1/\beta}$$

whence

$$\left(\frac{a^{\alpha}+b^{\alpha}}{b^{\alpha}}\right)^{1/\alpha}\leqslant\left(\frac{a^{\beta}+b^{\beta}}{b^{\beta}}\right)^{1/\beta}$$

Since $b>0$, the inequality being proved follows from the last inequality,

12. *Prove the inequality* $0<\sin^8 x+\cos^{14} x\leqslant 1$.

It is quite obvious that $\sin^8 x+\cos^{14} x\geqslant 0$. But the equality $\sin^8 x+\cos^{14} x=0$ is valid only if we simultaneously have $\sin^8 x=0$ and $\cos^{14} x=0$, which, of course, is impossible. Therefore, the strict inequality $\sin^8 x+\cos^{14} x>0$ holds true.

The properties of trigonometric functions imply that $\sin^2 x\leqslant 1$ and $\cos^2 x\leqslant 1$ for arbitrary real x. But since $8>2$ and $14>2$, it follows therefrom that

$$\sin^8 x\leqslant \sin^2 x \text{ and } \cos^{14} x\leqslant \cos^2 x$$

Combining these inequalities termwise and noting that $\sin^2 x+\cos^2 x=1$, we obtain

$$\sin^8 x+\cos^{14} x\leqslant 1$$

It is obvious here that, say, for $x=\pi/2$ we have equality; in other words, the weak inequality cannot be replaced by the strict inequality $\sin^8 x+\cos^{14} x<1$.

One of the techniques used in proving inequalities consists in the following. For instance, let it be required to prove the inequality $A<B$, where A and B are certain expressions. If we succeed in finding an expression C such that $A<C$ and at the same time $C\leqslant B$, then the required inequality $A<B$ will have thus been proved.

13. *Prove that for every positive integer n the following inequality holds true*:

$$\frac{1}{9}+\frac{1}{25}+\dots+\frac{1}{(2n+1)^2}<\frac{1}{4}$$

Noting that

$$\frac{2}{(2k+1)^2}<\frac{1}{2k}-\frac{1}{2k+2}$$

we replace the sum in the left member of the inequality to be proved by the *greater* expression

$$\frac{1}{3^2}+\frac{1}{5^2}+\dots+\frac{1}{(2n+1)^2}$$
$$<\frac{1}{2}\left[\left(\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{4}-\frac{1}{6}\right)+\dots+\left(\frac{1}{2n}-\frac{1}{2n+2}\right)\right]$$

However, this latter expression is equal to

$$\frac{1}{2}\left[\frac{1}{2}-\frac{1}{2n+2}\right]=\frac{1}{4}-\frac{1}{4n+4}$$

and, obviously, is less than $1/4$. Hence, the sum

$$\frac{1}{9}+\frac{1}{25}+\dots+\frac{1}{(2n+1)^2}$$

is all the more so less than $1/4$.

14. *Prove that for any positive integer $n > 1$ the inequality*

$$1+\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}+\ldots+\frac{1}{\sqrt{n}} > 2(\sqrt{n+1}-1)$$

holds true.

To prove this, reduce each term of the sum in the left-hand member:

$$\frac{1}{\sqrt{k}} > \frac{2}{\sqrt{k}+\sqrt{k+1}} = 2(\sqrt{k+1}-\sqrt{k})$$

Therefore, the left side of the inequality we want to prove can be reduced:

$$1+\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}+\ldots+\frac{1}{\sqrt{n}} > 2(\sqrt{2}-\sqrt{1})$$
$$+2(\sqrt{3}-\sqrt{2})+\ldots+2(\sqrt{n}-\sqrt{n-1})+2(\sqrt{n+1}-\sqrt{n})$$

Since the right side of this latter inequality is exactly equal to $2\sqrt{n+1}-2$, the original inequality is valid.

In the next example, an apt combining of factors leads us to the result we need.

15. *Prove that* $n! < \left(\frac{n+1}{2}\right)^n$, *where n is an integer exceeding unity.*

The validity of this inequality will follow from the validity of the equivalent inequality

$$(n!)^2 < \left(\frac{n+1}{2}\right)^{2n}$$

Let us multiply the number $n! = 1\cdot 2\ldots k\ldots(n-1)\,n$ by the number $n! = n(n-1)\ldots(n-k+1)\ldots 2\cdot 1$, arranging them one above the other:

$$\begin{array}{ccccccc} 1 & 2 & \ldots & k & \ldots & (n-1) & n \\ n & (n-1) & \ldots & (n-k+1) & \ldots & 2 & 1 \end{array}$$

Multiplying the numbers in each column, we get

$$(1\cdot n)[2(n-1)]\ldots[k(n-k+1)]\ldots[(n-1)\cdot 2](n\cdot 1)$$

In order to obtain $(n!)^2$ we have to multiply the terms of this row. Applying inequality (2) to each term of this row, we get

$$\sqrt{k(n-k+1)} \leqslant \frac{k+n-k+1}{2} = \frac{n+1}{2}, \quad k = 1, 2, \ldots, n$$

equality being achieved only when $k = n-k+1$, that is to say, for $k = (n+1)/2$. In other words, only for n odd; and only then for one term of our row in this inequality is equality possible. Hence, for all brackets except possibly one, the inequalities

$$[k(n-k+1)] < \left(\frac{n+1}{2}\right)^2$$

hold true. Since there are n terms in the row, we get

$$(n!)^2 < \left[\left(\frac{n+1}{2}\right)^2\right]^n$$

A sufficiently large number of inequalities can be proved by the method of mathematical induction.

16. *Prove that for any real number $\alpha \geqslant -1$ and any positive integer n the inequality*

$$(1+\alpha)^n \geqslant 1+n\alpha \qquad (4)$$

holds true.

The inequality is clearly true for $n=1$. Suppose that the inequality $(1+\alpha)^k \geqslant 1+k\alpha$ holds true; we will prove that in that case the inequality $(1+\alpha)^{k+1} \geqslant 1+(k+1)\alpha$ is valid. Indeed: $(1+\alpha)^{k+1}=(1+\alpha)^k(1+\alpha) \geqslant (1+k\alpha)(1+\alpha)=1+(k+1)\alpha+k\alpha^2 \geqslant 1+(k+1)\alpha$. This means that the original inequality holds true.

17. *Prove that the inequality $|\sin nx| \leqslant n\,|\sin x|$ is valid for any positive integer n.*

For $n=1$ the inequality is obviously true. Assuming that $|\sin kx| \leqslant k\cdot|\sin x|$, we will prove that $|\sin(k+1)x| \leqslant (k+1)\cdot|\sin x|$. Indeed, taking advantage of the inequality $|\cos kx| \leqslant 1$, we have

$$\begin{aligned}
|\sin(k+1)x| &= |\sin kx\cdot\cos x+\sin x\cdot\cos kx| \\
&\leqslant |\sin kx|\cdot|\cos x|+|\sin x|\cdot|\cos kx| \\
&\leqslant |\sin kx|+|\sin x| \leqslant k|\sin x|+|\sin x| \\
&= (k+1)\cdot|\sin x|
\end{aligned}$$

Hence, the original inequality is true.

18. *Prove the following theorem: if the product of $n \geqslant 2$ positive numbers is equal to 1, then the sum of the numbers is greater than or equal to n, that is, if*

$$x_1x_2\ldots x_n=1,\quad x_1>0,\quad x_2>0,\ \ldots,\ x_n>0,\ \textit{then}$$
$$x_1+x_2+\ldots+x_n \geqslant n$$

If $n=2$ we have to prove the statement: if $x_1x_2=1$, then $x_1+x_2 \geqslant 2$. But this is obvious since the arithmetic mean $(x_1+x_2)/2$ of two positive numbers is greater than or equal to the geometric mean $\sqrt{x_1x_2}=1$, or $x_1+x_2 \geqslant 2$. Besides, equality (that is, $x_1+x_2=2$) is attained only when $x_1=x_2=1$.

Using induction, we take any positive numbers $x_1, \ldots, x_k, x_{k+1}$ which satisfy the condition $x_1 \ldots x_{k-1}x_kx_{k+1}=1$. If each of these numbers equals 1, then the sum $x_1+\ldots+x_k+x_{k+1}=k+1$ so that in this case the original inequality is valid.

If this is not so, there will be a number among them less than 1 and a number greater than 1. Suppose that $x_k>1$, $x_{k+1}<1$. We have the equality

$$x_1\ldots x_{k-1}(x_kx_{k+1})=1$$

This is a product of k numbers and so the induction hypothesis is applicable and we can assert that

$$x_1+\dots+x_{k-1}+x_kx_{k+1}\geqslant k$$

But then

$$x_1+\dots+x_{k-1}+x_k+x_{k+1}\geqslant k-x_kx_{k+1}+x_k+x_{k+1}$$
$$=k+1+(x_k-1)(1-x_{k+1})>k+1$$

since $x_k-1>0$ and $1-x_{k+1}>0$ which completes the proof.

Note that we also established the fact that equality in the relation at hand is only possible if all $x_i=1$; now if not all x_i are equal to unity, then the strict-inequality sign holds true in this relation.

From this theorem follows the *generalized inequality between the arithmetic mean and the geometric mean for $n\geqslant 2$ positive numbers*:

$$\frac{x_1+\dots+x_n}{n}\geqslant \sqrt[n]{x_1\dots x_n},\quad x_1>0,\ \dots,\ x_n>0 \tag{5}$$

Indeed, denote $\sqrt[n]{x_1\dots x_n}$ by c and x_i/c by y_i. Then $y_1\dots y_n=$ $=(x_1\dots x_n)/c^n=1$. By what has been proved, $y_1+\dots+y_n\geqslant n$, whence $(x_1+\dots+x_n)/n\geqslant c$, and this completes the proof.

This inequality is widely used in the proof of other inequalities. For example, if we apply it to the numbers $1, 2, \dots, n$, then we immediately get the inequality

$$\sqrt[n]{1\cdot 2\dots n}<\frac{1+2+\dots+n}{n}$$

or $\sqrt[n]{n!}<(n+1)/2$, whence $n!<[(n+1)/2]^n$. We proved this inequality in Problem 15 via a special technique. This new proof is clearly simpler.

The foregoing examples show that the method of mathematical induction can successfully be applied in the proof of a variety of inequalities. At the same time, one should not overestimate the power of the induction method: there are many problems that would seem particularly suited to this method, whereas attempts to employ it encounter insuperable difficulties.

To illustrate, let us try to use induction on the inequality

$$\frac{1}{9}+\frac{1}{25}+\dots+\frac{1}{(2n+1)^2}<\frac{1}{4}$$

For $n=1$ it has the form $1/9<1/4$, which is true. Suppose that this inequality is valid for $n=k$:

$$\frac{1}{9}+\frac{1}{25}+\dots+\frac{1}{(2k+1)^2}<\frac{1}{4}$$

For $n=k+1$, the left side is

$$\frac{1}{9}+\frac{1}{25}+\dots+\frac{1}{(2k+3)^2}=\left[\frac{1}{9}+\frac{1}{25}+\dots+\frac{1}{(2k+1)^2}\right]+\frac{1}{(2k+3)^2}$$

By the induction hypothesis, the sum in the square brackets is less than $1/4$ and therefore

$$\frac{1}{9}-\frac{1}{25}+\dots+\frac{1}{(2k+3)^2}<\frac{1}{4}+\frac{1}{(2k+3)^2}$$

Quite obviously this inequality does not in the least imply that the left-hand member is less than 1/4. Thus, proof by induction has come to an impasse, whereas the inequality is very simply proved by an entirely different method. This is done in Problem 13.

In conclusion we offer two inequalities in the proof of which the techniques suggested above and certain others are used; these inequalities can be solved by several methods involving algebra, trigonometry and even geometry.

19. *Prove that if* $x^2+y^2=1$, *then* $-\sqrt{2}\leqslant x+y\leqslant\sqrt{2}$.

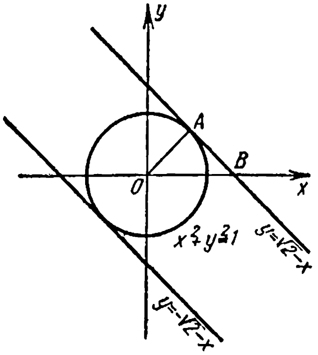
Algebraic solution. Let us write the obvious inequality $(x-y)^2\geqslant 0$, or $x^2+y^2\geqslant 2xy$, whence $2(x^2+y^2)\geqslant x^2+2xy+y^2$. Insofar as $x^2+y^2=1$, from the latter inequality we have $(x+y)^2\leqslant 2$, whence (see formula (3), Sec. 1.4) $|x+y|\leqslant\sqrt{2}$ or $-\sqrt{2}\leqslant x+y\leqslant\sqrt{2}$.

Trigonometric solution. If x and y satisfy the condition $x^2+y^2=1$, then we can find an angle α such that $x=\cos\alpha$, $y=\sin\alpha$. Then we have to prove that for any value of α

$$-\sqrt{2}\leqslant\cos\alpha+\sin\alpha\leqslant\sqrt{2}$$

Since $\cos\alpha+\sin\alpha=\sqrt{2}\sin(\alpha+\pi/4)$ and $-1\leqslant\sin(\alpha+\pi/4)\leqslant 1$, it follows that $-\sqrt{2}\leqslant\sqrt{2}\sin(\alpha+\pi/4)\leqslant\sqrt{2}$ for all values of α, which completes the proof of the required inequality.

Fig. 15



Geometric solution. We will consider x and y as coordinates of points in a plane in a given system of coordinates. Then the condition $x^2+y^2=1$ is satisfied by the points (x, y) lying on a circle of radius 1 centred at the origin (Fig. 15). The points which satisfy the inequality $x+y\leqslant\sqrt{2}$ lie on the straight line $y=\sqrt{2}-x$ and below that line (cf. Problem 27, Sec. 1.13).

Let B be the point of intersection of this straight line with the axis of abscissas and OA a perpendicular dropped on this line from the origin. Then $OB=\sqrt{2}$, $\angle ABO=45°$ and therefore $OA=1$. Hence, the point A lies on the circle and the straight line $y=\sqrt{2}-x$ is perpendicular to the radius OA at its endpoint, which is to say it is tangent to the circle.

Similarly, the inequality $-\sqrt{2}\leqslant x+y$ is satisfied by points lying on the straight line $y=-\sqrt{2}-x$ and above it; this line is also tangent to the circle $x^2+y^2=1$.

Thus, the double inequality to be proved is satisfied by points lying in the strip between the straight lines $y=\sqrt{2}-x$ and $y=-\sqrt{2}-x$ (these lines included). But the circle $x^2+y^2=1$ lies entirely inside this strip and so the coordinates of any point of it satisfy the inequality $-\sqrt{2}\leqslant x+y\leqslant\sqrt{2}$. The proof is complete.

20. *Let $a+b=2$, where a and b are real numbers. Prove that $a^4+b^4\geqslant 2$.*

Note that if one of the numbers, a or b, is negative, the inequality is almost obvious. Suppose, say, $b<0$. Then $a>2$ and the inequality $a^4+b^4\geqslant 2$ is true, since $b^4>0$ and $a^4>16$. We will therefore assume that $a\geqslant 0$ and $b\geqslant 0$.

First solution. Since $a+b=2$, then $(a+b)^2=4$. Using the inequality between the arithmetic mean and the geometric mean, $ab\leqslant(a^2+b^2)/2$, we have $4=(a+b)^2=a^2+b^2+2ab\leqslant 2(a^2+b^2)$, or $2\leqslant a^2+b^2$. Squaring this inequality (this is legitimate since the numbers on the right and left are positive), we get

$$4\leqslant(a^2+b^2)^2$$

On the basis of the inequality between the arithmetic mean and the geometric mean, $a^2b^2\leqslant(a^4+b^4)/2$. Therefore we have $4\leqslant(a^2+b^2)^2=a^4+b^4+2a^2b^2\leqslant 2(a^4+b^4)$ whence $2\leqslant a^4+b^4$, and the proof is complete.

Second solution. We again assume that $a\geqslant 0$ and $b\geqslant 0$. Since $a+b=2$, then $(a+b)^4=16$, or

$$\begin{aligned}(a+b)^4&=(a^2+2ab+b^2)(a^2+2ab+b^2)\\&=a^4+b^4+4ab(a^2+b^2)+6a^2b^2=16\end{aligned}$$

But since $a^2+b^2=4-2ab$, the last equality can be rewritten

$$a^4+b^4=16-16ab+2a^2b^2$$

If we are able to demonstrate that $16-16ab-2a^2b^2\geqslant 2$, then our inequality will have been proved.

By hypothesis, $ab\leqslant 1$. Indeed, $\sqrt{ab}\leqslant(a+b)/2$. Since $a+b=2$, it follows that $\sqrt{ab}\leqslant 1$, whence $ab\leqslant 1$. And so we have to prove the inequality $16-16ab+2a^2b^2\geqslant 2$, provided that $ab\leqslant 1$.

We set $x=ab$. Then we have to prove the inequality $x^2-8x+7\geqslant 0$ with the proviso that $x\leqslant 1$. The roots of the quadratic trinomial x^2-8x+7 are $x_1=1$, $x_2=7$. Therefore the last inequality may be written as $(x-1)(x-7)\geqslant 0$.

But for $x \leqslant 1$ this inequality is obvious. We have thus obtained $16-16ab+2a^2b^2 \geqslant 2$, which is what we set out to prove.

Third solution. Let $a=1+c$, $b=1-c$. Since we earlier assumed that $a \geqslant 0$ and $b \geqslant 0$, it follows that $-1 \leqslant c \leqslant 1$ and so we can take advantage of inequality (4) (see Problem 16 of this section):

$$(1+c)^4 \geqslant 1+4c, \quad (1-c)^4 \geqslant 1-4c$$

Thus

$$a^4+b^4=(1+c)^4+(1-c)^4 \geqslant (1+4c)+(1-4c)=2$$

In conclusion we note that a more general statement is valid: *if $a+b=2$, then $a^n+b^n \geqslant 2$ for any positive integer n.* This can easily be proved by, say, the third method given above.

Exercises

1. Prove that for all real numbers a, b, and c

$$a^2+b^2+c^2 \geqslant ab+bc+ca$$

2. Prove that if a, b, c are positive and unequal, then
(a) $(a+b+c)\ (a^{-1}+b^{-1}+c^{-1})>9$,
(b) $(a+b+c)\ (a^2+b^2+c^2)>9abc$.

3. Prove that $a^2+b^2+c^2+3 \geqslant 2(a+b+c)$.

4. Prove that for all x in the interval $0<x<\pi/2$ the inequality $\tan x+\cot x \geqslant 2$ holds true.

5. Prove that if a and b are positive numbers different from unity, then $|\log_a b+\log_b a| \geqslant 2$.

6. Prove that $(1/\log_2 \pi)+(1/\log_{4.5} \pi)<2$.

7. Prove that for all real x and y, the inequality $x^2+2xy+3y^2+2x+6y+3 \geqslant 0$ is valid.

8. Prove that $\sin^4 x-6\sin^2 x+5 \geqslant 0$ for all x.

9. Prove that the polynomial $x^8-x^5+x^2-x+1$ is positive for all real values of x.

10. Prove that if $a+b=c$, $a>0$, $b>0$, then $a^{2/3}+b^{2/3}>c^{2/3}$.

11. Let n be a positive integer. Prove the inequality

$$(1+1/n)^n<(1+1/2n)^{2n}$$

12. Prove that the following inequality holds true for every positive integer n:

$$\frac{1}{n+1}+\frac{1}{n+2}+\dots+\frac{1}{3n+1}>1$$

13. Prove that $(n!)^2>n^n$, where $n>2$ is a natural number.

14. Prove that $n!>2^{n-1}$, where $n>2$ is a natural number.

15. Prove that for arbitrary positive a and b and any positive integer n, the inequality $(a+b)^n<2^n(a^n+b^n)$ is valid.

16. Prove that the sum of the legs of a right triangle does not exceed the diagonal of a square constructed on the hypotenuse.

17. Prove that the sum of the cubes of the legs of a right triangle is less than the cube of the hypotenuse.

18. Prove that of all rectangles with a given perimeter the square has the largest area.

19. Prove that the area of an arbitrary triangle does not exceed one fourth the square of its semiperimeter.

Find the largest and smallest values of the following functions:

20. $y=5\cos 2x-4\sin 2x$. 21. $y=3^{x-1}+3^{-x-1}$

22. $y=\frac{x^2}{x^4+1}$. 23. $y=\frac{2x}{x^2+1}$

24. Prove that $2^{\sin x}+2^{\cos x}\geqslant 2^{1-\frac{1}{\sqrt{2}}}$ for all x. For which values of x is equality attained?

25. Prove that for all x in the interval $0<x<\pi/2$ the inequality $\cot(x/2)>1+\cot x$ holds true.

1.9 Solving equations

At examinations a rather strange situation would appear to develop around solving equations. These problems are ordinarily not considered difficult and most students handle them fairly well. Yet many serious mistakes are made.

This situation is strange only at first glance. The point is that very often there is a big gap between computational techniques and a conscious grasp of the logical foundations that underlie them.

Most students can of course simplify an equation by means of clear-cut manipulations, but by far not every student is capable of realizing that a solution has been lost or acquired, and many don't even give thought to such things. Still others may know certain parts of the theory pertaining to these matters but the knowledge is only formal and such students are often completely helpless in a slightly altered situation.

Let us say that the student is quite familiar with the fact that squaring both sides of an irrational equation can give rise to extraneous roots. Yet time and again students square trigonometric equations and then fail to discard extraneous roots! This mistake would not be made if the student realized why squaring leads to the introduction of extraneous roots.

Or take checking. There seem to be two opposing opinions among students here. Some regard checking as a whim of the teacher, something that simply has to be done in order to pass, while others regard checking as necessary in all cases without exception. They even go on to check the roots of quadratic equations. Both views are based on a total misconception of what checking really is and what place it occupies in problem solving.

In short, the student should have a firm grasp of the fundamentals of the theory that is needed in the solution of equations. Let us examine this minimum of theoretical knowledge.

First some definitions.

1. *The domain of the variable of an equation is the set of values of the variable (unknown) for which its left and right members are meaningful (defined)*; it is thus the set of all eligible replacements for a variable in an equation.

2. *A number a is a solution (root) of an equation if when substituted for the unknown (variable) makes the equation a true statement (converts it into a true numerical equation).*

According to this definition, the solution set (all the solutions) of an equation is a subset (a part) of the domain of the variable, otherwise substitution in place of the unknown would not yield a true statement and would be meaningless.

3. *To solve an equation means to find all the roots or prove that there are no roots.*

4. *If all the roots (solution set) of one equation are the roots of another equation, then the latter equation is a consequence of the former.*

5. *Two equations are termed equivalent if each is a consequence of the other.* From this definition it follows immediately that equivalent equations have the same solution sets.

6. *Two equations are equivalent on some set of values of the variable (unknown) if they have exactly the same solutions belonging to this set.*

Let us illustrate these concepts with two examples.

The domain of x in the equation $x-3=\sqrt{x}$ consists, according to the definition, of all x for which the left member $x-3$ and the right member $\sqrt{x}$ are meaningful. Clearly, the left member is defined for any x and the right member for $x\geqslant 0$. Therefore the domain of x in this equation consists of $x\geqslant 0$.

Yet many students erroneously state that the domain of the variable is $x\geqslant 3$, since "for $x<3$ the left member is negative and the right member cannot be negative." The quoted part of the statement is true and it is used in the solution of the given equation. It shows that the *roots* of the equation are not less than 3. But it does not follow therefrom that all *permissible values* are less than 3, because not all permissible (eligible) values are roots!

Consider the two equations

$$\log_6(x-2)+\log_6(x+3)=2 \text{ and } \log_6(x-2)(x+3)=2$$

Obviously every root of the former equation is a root of the latter one, so that the latter equation is a consequence of the former. The latter equation can readily be solved to yield the solution set (roots) $x_1=6$ and $x_2=-7$. The root x_2 does not satisfy the first equation, it is not even in the domain of the variable. Thus, the two given equations *are not equivalent*, but they *are equivalent in the domain of x of the first equation* (in this domain they have the one root $x=6$).

It is easy to see why this is so. The domain of x in the first equation consists of $x>2$, while the domain of x in the second equation is broader, including these x and also $x<-3$. It is therefore natural that in passing from the first equation to the second an extraneous root $x=-7$ appeared that does not belong in the domain of the variable of the first equation.

How do these newly introduced concepts operate in the solution of equations? The point is that in most cases a solution is obtained after a long chain of manipulations and transformations from one equation to the next. Thus, in the solution process, each equation is replaced by a new one, and quite naturally the new equation can have a new solution set (new roots). The prime task in a correct solution of any equation is to follow the variations in the solution set and not to allow for any loss of roots or any failure to discard extraneous roots.

The best method is clearly, each time, to replace the given equation with an equivalent equation. Then the roots (solution set) of the last equation will be the roots (solution set) of the original equation. In practice however this ideal version is rare. As a rule, an equation is replaced by a consequence that is not equivalent; then, by the definition of consequence, all the roots of the first equation are the roots of the second, that is to say *there is no loss of roots*, but *extraneous roots may appear* (on the other hand, they may not). And when in the process of manipulations the equation is replaced by a nonequivalent consequence, the roots have to be investigated. This is a check and it is necessary. Note here and now that, as we shall see later on, this investigation does not at all mean that we have to substitute the roots obtained into the original equation.

To summarize then, if a solution is carried through without an investigation of equivalence and sources of extraneous roots, then verification is a necessary part of the solution without which it cannot be regarded as complete even if no extraneous roots appeared in actual fact. Of course, if they did appear and were not discarded, the solution is simply incorrect. On the other hand, if each time the equation was replaced by an equivalent equation (which, incidentally, is an extremely rare case), and this fact is stipulated at each stage in the solution, then no verification is required. We thus see that the notion of checking plays a very definite and extremely essential role in the solution of equations and does not by any means merely reduce to a simple checking through of computations.

As for checking computations, that is up to the student. He may do it or he may not according to how carefully he feels the computations have been carried through. It is of course always best to check oneself at an examination, but this should be done on a separate sheet of paper and there is no need to include it in the solution.

It must be stressed that it is not permissible to replace a given equation by one which is not a consequence of the first, because then *there is a root of the first equation which is not a root of the second*, and so solving the second equation does not yield all the roots of the first. A *root will be lost* for good. That is the essential difference between loss of roots and the introduction of extraneous roots.

Such is the theory. In practical situations, one has to know the specific sources of introducing or losing roots. In the main, these sources are of two types: the so-called "identity transformations" and the performance of operations such as raising to a power, taking logarithms, antilogarithms, etc. in both members of an equation.

At first glance, "identity transformations" are quite harmless, but actually they often lead to nonequivalent equations since they change the domain of the variable (unknown). Say, if in the solution of an irrational equation we replace $(\sqrt{2x+1})^2$ by $2x+1$, we immediately extend the domain of the variable since $2x+1$ is meaningful for all x, while $(\sqrt{2x+1})^2$ is valid only for $x \geqslant -1/2$. The same goes for the example that we analyzed earlier: the use of the formula for the logarithm of a product led to an extension of the domain and, as a result, to the introduction of an extraneous root.

There is nothing strange in this: it is simply that most formulas used in transformations are such that their left and right members are meaningful for different values of the letters used. Such, for example, are the formulas

$$\sqrt{ab} = \sqrt{a}\sqrt{b}, \quad (\sqrt{x})^2 = x, \quad \log_a xy = \log_a x + \log_a y,$$

$$\log_a x^n = n \log_a x,$$

$$a^{\log_a b} = b, \quad \cot x = \frac{1}{\tan x}, \quad \sin 2x = \frac{2 \tan x}{1+\tan^2 x},$$

$$\tan(x+y) = \frac{\tan x + \tan y}{1 - \tan x \tan y}.$$

Therefore, replacing one part of a formula by another leads to an extension or narrowing of the domain of the variable. Extending the domain opens the way to acquiring extraneous roots, while narrowing (restricting) the domain makes possible the loss of roots, so that narrowing is never permissible. As for extraneous roots, if they are acquired through the extension of the domain, then it is not necessary to substitute them directly in the original equation in order to separate them from the roots of the original equation; it suffices to check to see whether they are in the domain of the variable. If they are not, discard them, if they are, leave them.

This fact is of exceptional importance in the solution of equations and so we will dwell on it in more detail.

A. *If in the process of transforming an equation, extraneous roots appear only due to extending the domain of the variable, then those roots (and only those) which appear in the domain will be the roots of the original equation.*

This rule relieves us of the necessity to substitute the roots found into the equation and the purely mechanical checking of numerical equations, which at times is exceedingly difficult or even sometimes completely impossible due to the fact that there are an infinity of numbers to be verified.

Thus, in place of *direct substitution* we can employ a *test for membership in the domain*, but only when the source of extraneous roots is extension of the domain of the variable, this case and no other. Therefore, when using the membership test the student must explicitly state in the solution process where and for what reasons extraneous roots can appear.

As to taking functions of both members of an equation, we consider only two of the more important cases: *squaring and the taking of antilogarithms.*

It often happens (especially in the solution of irrational equations) that one has to pass from a certain equation $f(x)=g(x)$ to the equation $[f(x)]^2=[g(x)]^2$. What happens to the roots in this transition? First of all it is clear that the second equation is a consequence of the first: if the number a is a root of the first equation, that is $f(a)=g(a)$, then $[f(a)]^2=[g(a)]^2$, and a is a root of the second equation. But, generally speaking, the converse is not true: the second equation is satisfied also by the roots of the "extraneous" equation $f(x)=-g(x)$. Thus, in squaring, roots are not lost but extraneous roots may appear.

A very important practical consequence follows from this statement.

B. *If both members of an equation are nonnegative on some set of values of the argument, then, upon squaring, we obtain an equation that is equivalent to the original one on that set.*

Indeed, in this case the "extraneous" equation clearly has no roots, except for those for which both sides vanish, but such are not extraneous for our equation. How this is done practically will be demonstrated in concrete examples below.

Similarly we consider taking antilogarithms in equations, that is, passing from the equation $\log_c f(x)=\log_c g(x)$ to the equation $f(x)=$ $=g(x)$. Let a be a root of the original equation, that is $\log_c f(a)=$ $=\log_c g(a)$. Then $c^{\log_c f(a)}=c^{\log_c g(a)}$, or $f(a)=g(a)$. Hence, any root of the original equation is a root of the second equation. On the other hand the domain of the variable of the second equation is greater than that of the first and so it is natural to expect extraneous roots, but they will be due precisely to the extension of the domain. Hence, it suffices to find the roots of the second equation and test them for membership in the domain of the first equation.

Such is the theoretical foundation that the student needs. It is also worth stressing that it is not always advisable to employ the whole theory; moderation is the best advice and one should strive for the simplest solution. If, say, in the solving process it becomes apparent that a simple verification of the resulting roots is not difficult, then there is no reason to seek out the sources of introducing roots or take an interest in the variation of the domain in the solution process, or even to find the domain at all. But if this verification is complicated, then theoretical reasoning can save the day: at the proper time (and in the final version) the student should investigate the transformation that might lead to extraneous roots.

At the same time, in any solution we must make sure that no loss of roots occurs. It is useful to state this explicitly, particularly if the transformation employed is sufficiently complicated.

Below we will illustrate the more typical cases and also the most insidious sources (though not all of them, naturally) that give rise to extraneous roots. They include the formulas for transforming radicals, the fundamental logarithmic identity, and the formulas for taking logarithms of a product and a power, clearing of fractions, the cancellation of similar terms, the replacement of an equation by a collection of simpler equations, and certain "verbal" arguments. We will then consider some sources of loss of roots. Some of the last and more involved problems will be analyzed with the aim of indicating certain difficulties of a different nature, not connected with the acquisition or loss of roots.

1. *Solve the equation* $\sqrt{2x-6}+\sqrt{x+4}=5$.

Squaring both sides and using the formulas for transforming radicals, we get the equation

$$2x-6+2\sqrt{(2x-6)(x+4)}+x+4=25$$

or

$$2\sqrt{2x^2+2x-24}=27-3x \tag{1}$$

Again squaring and getting rid of the radical, we arrive at the equation $x^2-170x+825=0$ whose roots are $x_1=5$ and $x_2=165$. Direct substitution of these values in the original equation shows that x_1 is a root and x_2 is not.

The above solution is almost exactly what any student would put down in his rough draft. With regard to this equation, there is no need to go into the theory and, say, investigate the source of the extraneous root; simply, when copying the solution onto the final clean sheet of paper, the student should indicate that in the process of transformations no roots could be lost and, at the end, he should check this by direct substitution. The solution will then be complete.

Still, it is well to note that the extraneous root appeared due to the squaring of equation (1), as a root of the "extraneous equation".

The following problem is just as simple as this one, but in the checking of a "good" root we suddenly come up against certain difficulties of a fundamental nature.

2. *Solve the equation*

$$\sqrt{5x+7}-\sqrt{3x+1}=\sqrt{x+3}$$

Squaring both members and manipulating, we get

$$2\sqrt{(5x+7)(3x+1)}=7x+5$$

whence, again squaring we obtain the quadratic equation $11x^2+$ $+34x+3=0$, whose roots are $x_1=-1/11$ and $x_2=-3$. Direct verification shows that $x=-1/11$ is the root of the original equation.

Checking the value $x=-3$, many students at the examination got the equation $\sqrt{-8}-\sqrt{-8}=0$. They considered this statement to be true on the grounds that the left member is a case of "equals subtracted from equals". Thus, the value $x=-3$ proved to be a root of the original equation. But this argument is baseless since the expression $\sqrt{-8}$ is devoid of meaning: as we know, irrational equations are only considered in the domain of real numbers, and the symbol $\sqrt{a}$ is used for real a only to denote the principal square root of a nonnegative number a. Therefore, the value $x=-3$ does not lie in the domain of the variable and, hence, is not a root of the original equation.

The situation is quite different in the problem which now follows. Here, checking the "bad" roots is very involved and the simplest approach to a solution is by applying the theory we have developed, when the sources of extraneous roots are taken into account in the very process of solution.

3. *Solve the equation* $\sqrt{x+3}+\sqrt{2x-1}=4$.

Both members of this equation are nonnegative in the domain of the variable, and so after squaring we obtain an equation which, according to Statement B, is equivalent to the original one in the domain of the variable:*

$$(\sqrt{x+3})^2+2\sqrt{x+3}\sqrt{2x-1}+(\sqrt{2x-1})^2=16$$

Using the formulas for transforming radicals, which clearly extend the domain, we get the equation

$$2\sqrt{2x^2+5x-3}=14-3x \qquad (2)$$

In these transformations, extraneous roots could appear only due to the extension of the domain of the variable.

Then we reason as follows. The left member of (2) is nonnegative for every (permissible) value of x; but the right member is negative for $x>14/3$. It is quite obvious that these values of x cannot be solutions of the equation, and so from now on we will consider equation (2) only in the domain $x\leqslant 14/3$. But in this domain, both members of (2) are nonnegative (for the permissible values of x with respect to (2), naturally) and, according to the Statement B, squaring yields an equation that is equivalent to (2) on the set $x\leqslant 14/3$:

$$(2\sqrt{2x^2+5x-3})^2=(14-3x)^2$$

* Actually, these equations are equivalent because the domains of the variables coincide, but this is not important since later on we will extend the domain so that we will not be able to dispense with a test for membership in the domain of the variable.

From this, once more extending the domain of the variable, we get the quadratic equation $x^2-104x+208=0$ whose roots are $x_{1,2}=52\pm$ $\pm 8\sqrt{39}$. Both these roots, as will readily be seen, lie in the domain of the variable of the original equation and for this reason we have only to check that they satisfy the condition $x\leqslant 14/3$. It is easy to compute that $x_1>14/3$ and $x_2<14/3$ so that x_2 is the only root of equation (2), and, consequently, of the original equation.

We once again stress that one should resort to this kind of detailed, "theoretical", solution only in case of necessity, only when working through the rough draft as the student sees that the roots are "bad", which is to say that a direct substitution into the equation leads to a rather complicated problem: the proof or disproof of the equations

$$\sqrt{55+8\sqrt{39}}+\sqrt{103+16\sqrt{39}}=4,$$
$$\sqrt{55-8\sqrt{39}}+\sqrt{103-16\sqrt{39}}=4$$

Incidentally, the first of these equations is clearly not true. The second one can easily be proved if one knows the formula for transforming expressions of the form $\sqrt{A\pm\sqrt{B}}$. The alternative approach of squaring involves considerable computational difficulties. It is quite clear that both these methods are more complicated than the one we gave, where all we had to do was to test the roots x_1 and x_2 for their validity under the condition $x\leqslant 14/3$. Nevertheless, in this problem it is still possible to overcome the difficulties of direct verification and avoid application of the theory.

However, in equations containing a parameter, direct verification is appreciably more difficult, and practically the only way out is to employ the theory.

4. *Solve the equation* $x-1=\sqrt{a-x^2}$.

The right side is nonnegative for all (permissible) x, and the left side is nonnegative for $x\geqslant 1$. Therefore, the given equation is, in the domain $x\geqslant 1$, equivalent to the equation $(x-1)^2=(\sqrt{a-x^2})^2$ which can be reduced to

$$2x^2-2x+1-a=0 \tag{3}$$

(in the process, the domain of the variable was extended and we will finally have to check the resulting roots to see if they lie in the domain). And so we have to solve equation (3) and choose the roots for which $x\geqslant 1$ and $a-x^2\geqslant 0$. The discriminant of this equation is equal to $2a-1$, so that for $a<1/2$ it does not have any real roots; all the more so, the original equation has no roots for these values of a.

Now we assume that $a\geqslant 1/2$; the roots of (3) are $x_{1,2}=(1\pm\sqrt{2a-1})/2$. The root x_2 clearly does not satisfy the condition $x\geqslant 1$ and so is not a root of the original equation. In order to find out about x_1 we have to solve the inequality $(1+\sqrt{2a-1})/2\geqslant 1$ or $\sqrt{2a-1}\geqslant 1$; it clearly holds true for $a \geqslant 1$. And so for $a < 1$ the original equation does not have any roots, but for $a \geqslant 1$ we still have to verify the validity of the inequality $a - x_1^2 \geqslant 0$, which is equivalent to the inequality $a \geqslant \sqrt{2a-1}$. Both members of this inequality are nonnegative (we are considering $a \geqslant 1$) (see Sec. 1.10) and they can be squared, yielding $a^2 \geqslant 2a - 1$ or $a^2 - 2a + 1 \geqslant 0$, which is valid for all values of a.

And so for $a < 1$ the original equation has no roots, but for $a \geqslant 1$ it has the root $x = (1 + \sqrt{2a-1})/2$.

Note that the verification of the last condition $a - x^2 \geqslant 0$, which logically speaking is obligatory, can be conducted without any computations at all. Indeed, x_1 and x_2 have been obtained as roots of the equation $(x-1)^2 = a - x^2$ and hence for $x = x_1$ and $x = x_2$ the right side is nonnegative.

We once again stress the fact that a direct substitution as a check of the roots would have reduced to equations in a:

$$\frac{a \pm \sqrt{2a-1}}{2} - \sqrt{\frac{a \mp \sqrt{2a-1}}{2}} = 1$$

the outward aspect alone of which is quite saddening. Thus, without a conscious mastering of the approach given here to the solution of equations, such problems can cause great difficulties.

One of the most common sources of extraneous roots is the use of various logarithmic formulas, in particular, the *formula for taking logarithms of a product*. Indeed, replacing $\log_a f(x) + \log_a g(x)$ by $\log_a f(x)\, g(x)$, we extend the domain of the variable, permitting values of the unknown x for which we simultaneously have $f(x) < 0$ and $g(x) < 0$. And so extraneous roots can appear, but *only due to extension* of the domain of the variable, so that to discard them, on the basis of Statement A, it suffices to verify their membership in the domain. Also note that the converse replacement—the logarithm of a product by the sum of the logarithms—can lead to a narrowing of the domain of the variable, and so is not permissible.

5. *Solve the equation* $\log_2(x+2) + \log_2(3x-4) = 4$.

Passing to the logarithm of a product, we get $\log_2(x+2)(3x-4) = 4$, whence $(x+2)(3x-4) = 16$. The roots of this equation are $x_{1,2} = (-1 \pm \sqrt{73})/3$. It is easy to see that only x_1 lies in the required domain of the original equation and, on the basis of Statement A, is its root.

A direct substitution of the "bad" root x_1 would not have required very cumbersome computations but there would be unpleasant enough "irrational-logarithmic" manipulations, whereas the employment of Statement A yielded the answer at once.

The appearance of extraneous roots as a result of applying the fundamental logarithmic identity ordinarily surprises the student, although there is actually nothing strange here. It is due to the extension of the domain when replacing the expression $a^{\log_a b}$ by b if a or b contains the unknown.

6. *Solve the equation* $x^{\log_{\sqrt{x}} 2x}=4$.

Replacing $\log_{\sqrt{x}} 2x$ by $\log_x (2x)^2$ (see Rule IV of Sec. 1.6), we get

$$x^{\log_x (2x)^2}=4$$

Now employing the fundamental identity, we get $(2x)^2=4$, which means $x_1=-1$, $x_2=1$. But neither x_1 nor x_2 lie in the domain of the variable of the original equation: $x_1<0$, and $\sqrt{x_2}=1$ cannot be a logarithmic base. Hence, the given equation does not have any roots.

The appearance of extraneous roots may not be so noticeable as in the problems given above. As a rule, this is due to the fact that the reasoning and computations employed lead to an extension of the domain of the variable. In the next problem, *extraneous roots appear in a mutual cancelling of like terms.* Again there is no cause for surprise: in cancelling, we remove the restriction that the eliminated terms must be meaningful and thus extend the domain of the variable.

7. *Solve the equation* $\log_{10}\sqrt{1+x}+3\log_{10}\sqrt{1-x}=\log_{10}\sqrt{1-x^2}+2$.
We transform $\log_{10}\sqrt{1-x^2}$:

$$\log_{10}\sqrt{1-x^2}=\log_{10}\sqrt{1+x}\sqrt{1-x}=\log_{10}\sqrt{1-x}+\log_{10}\sqrt{1+x}$$

It is easy to see that this manipulation does not change the domain of the variable, and the transformed equation

$$\log_{10}\sqrt{1+x}+3\log_{10}\sqrt{1-x}=\log_{10}\sqrt{1-x}+\log_{10}\sqrt{1+x}+2$$

is equivalent to the given one. Eliminating $\log_{10}\sqrt{1+x}$ in both members, we obtain the equation

$$2\log_{10}\sqrt{1-x}=2$$

whose domain consists of the numbers $x<1$, which, as is evident, is greater than that of the original equation. We should thus expect extraneous roots. Solving the last equation we get the root $x=-99$, which does not lie in the domain of the original equation and therefore is not its root. Thus, the given equation does not have any roots.

One of the sources of mistakes is the (explicit or implicit) *clearing of fractions.* But this causes an *extension of the domain of the variable*— those values of x are included for which the denominator is equal to 0.

8. *Solve the equation*

$$\frac{1}{\log_6 (3+x)}+\frac{2\log_{0.25}(4-x)}{\log_2 (3+x)}=1$$

Taking all logarithms to the base 2 and manipulating, we get the equivalent equation

$$\frac{\log_2 6-\log_2 (4-x)}{\log_2 (3+x)}=1 \qquad (4)$$

whence $\log_2 6 - \log_2 (4-x) = \log_2 (3+x)$ and then

$$\frac{6}{4-x} = 3 + x \qquad (5)$$

This last equation is reduced to a quadratic equation and its roots are found to be: $x_1 = 3$, $x_2 = -2$.

During the solution process, extraneous roots could have appeared only due to an extension of the domain of the variable because of clearing of fractions in equations (4) and (5). It is therefore sufficient to test the resulting roots for membership in the domain of the variable of the original equation. We thus find that x_2 does not lie in the domain, but x_1 does and, hence, it is a root of the original equation.

Mistakes that occur in solving equations in which the left member is a fraction and the right member is zero are due to this disregard for the domain of the variable. Frequently, the student simply discards the denominator in such cases and equates the numerator to zero. For a correct solution, one should equate the numerator to zero, find the roots of the resulting equation and discard those for which the denominator vanishes.

9. *Solve the equation* $\tan 3x = \tan 5x$.

Rewrite the equation as

$$\frac{\sin 3x}{\cos 3x} - \frac{\sin 5x}{\cos 5x} = 0$$

whence, after a few elementary manipulations, we get

$$\frac{\sin 2x}{\cos 3x \cos 5x} = 0$$

Now, solving the equation $\sin 2x = 0$, we get $x = k\pi/2$, $k = 0, \pm 1, \ldots$. Now discard extraneous solutions, which is to say, those for which the denominator $\cos 3x \cos 5x$ vanishes, which obviously happens when the values of k are odd. The solutions of the original equation will then be the angles $x = k\pi/2$, where k is even: $k = 2n$, $n = 0, \pm 1, \pm 2, \ldots$. That is,

$$x = n\pi, \quad n = 0, \pm 1, \pm 2, \ldots$$

Quite obviously it is a grave mistake to take the set of values $x = k\pi/2$ for the answer.

It is the same disregard for the domain of the variable that accounts for mistakes in equation solving in which the left-hand member has been factored and the right-hand member is zero. In solving such an equation, the student ordinarily equates each factor to zero in succession and combines the solutions obtained, completely disregarding the fact that for certain values of x which make one of the factors vanish the other factor may be meaningless, and in that case such values of x will not be roots of the proposed equation. Therefore a proper solution requires a check to see that all the values of x obtained do indeed lie within the domain of the variable. This can occasionally give rise to considerable difficulties.

10. *Solve the equation*

$$\sin 2x \cos^2 2x \sin^2 6x \tan x \cot 3x = 0$$

Equating each factor to zero in succession, we finally get five groups of roots:

$$x=\frac{k\pi}{2}, \quad x=(2k+1)\frac{\pi}{4}, \quad x=\frac{k\pi}{6}, \quad x=k\pi, \quad x=(2k+1)\frac{\pi}{6}$$

where, throughout, k is any integer.

But this is not yet the answer, the point being that $\tan x$ and $\cot 3x$ are not defined for all values of x and therefore many values of x in these groups may prove to be extraneous. Let us consider each group separately.

(1) $x=k\pi/2$. If k is even, $k=2l$, then $x=l\pi$ and $\cot 3x$ is meaningless; if k is odd, $k=2l+1$, then $x=l\pi+\pi/2$ and $\tan x$ is meaningless.

Thus, not a single angle x of the first group is actually a solution of the equation.

(2) $x=(2k+1)\pi/4$. Then, as is evident, $\tan x$ is meaningful. Besides, $3x=(6k+3)\pi/4$ so that $\cot 3x$ is likewise meaningful.

Thus, all angles x of the second group are solutions of the equation.

(3) $x=k\pi/6$. It is easy to see that, in the trigonometric circle, the terminal side of the angle x coincides with the vertical diameter for $k=6l+3$ and, hence, $\tan x$ is meaningful for $k\neq 6l+3$. Furthermore, $3x=k\pi/2$ and $\cot 3x$ has meaning only for odd values of k. Hence, the only suitable values are odd numbers k not equal to $6l+3$, or numbers k of the form $k=6l+1$, $k=6l+5$.

Thus, of the angles of the third group, the following angles are solutions:

$$x=l\pi+\pi/6, \quad x=l\pi+5\pi/6$$

where l is any integer.

(4) $x=k\pi$. In this case $\cot 3x$ is meaningless and so there are no solutions.

(5) $x=(2k+1)\pi/6$. By the trigonometric circle, it is evident that the terminal side of the angle x coincides with the vertical diameter for $k=3l+1$, and, hence, $\tan x$ will be meaningful for $k=3l, k=3l+2$.

Thus, in the fifth group we have the angles

$$x=l\pi+\pi/6, \quad x=l\pi+5\pi/6$$

where l is any integer; these are the same angles as in the third group.

The final answer can be written as follows:

$$x=(2n+1)\pi/4, \quad x=n\pi+\pi/6, \quad x=n\pi+5\pi/6$$

n an arbitrary integer, or, more compactly,

$$x=(2n+1)\pi/4, \quad x=n\pi\pm\pi/6$$

n an arbitrary integer.

Acquiring extraneous roots is not always so explicit as occurs in the last two examples. Sometimes the cause is what would appear at first glance to be quite harmless reasoning.

For example, the equation $\tan 3x=\tan 5x$ that was analyzed earlier is frequently solved as follows: "The tangents of two angles are equal if and only if the difference of the angles is equal to an integral multiple of π; hence, $2x=k\pi$, $x=k\pi/2$, $k=0, \pm 1, \pm 2, \ldots$." But we know that this answer is not correct.

Where does the mistake lie?

The explanation is rather simple: the assertion on which the solution is based is incorrect, although it is quite common among students. Indeed, if $\tan\alpha=\tan\beta$, then $\alpha-\beta=k\pi$, where k is an integer, but the converse is not true: if $\alpha-\beta=k\pi$, then the equation $\tan\alpha=\tan\beta$ may simply be meaningless (say if $\alpha=\pi/2$, $\beta=-\pi/2$). Therefore, in the replacement of equation $\tan 3x=\tan 5x$ by $2x=k\pi$, there was no loss of roots, but certain extraneous roots appeared.

Let us now consider some sources of the *loss of roots* and appropriate measures for avoiding such loss. Students most often lose roots when replacing a given equation by a new one having a more restricted domain of the variable. Such a restriction (narrowing) of the domain results, as we have already seen, from the use of logarithmic formulas, trigonometric formulas and certain "verbal" reasoning.

As we have already noted, replacing the logarithm of a product by a sum of the logarithms (Rule I of Sec. 1.6) leads to a narrowing of the domain, just as does Rule III which has to do with taking the logarithm of a power. To avoid such restrictions, one should employ Rules I * and III * instead of Rules I and III. The use of the former can at worst extend the domain, that is, lead to extraneous roots. And we already know what to do with extraneous roots.

That is how we will solve the following problem.

11. *Solve the equation*

$$\frac{3}{2}\log_{1/4}(x+2)^2-3=\log_{1/4}(4-x)^3+\log_{1/4}(x+6)^3$$

Since

$$\log_{1/4}(x+2)^2=2\log_{1/4}|x+2|,$$

$$\log_{1/4}(4-x)^3=3\log_{1/4}(4-x), \quad \log_{1/4}(x+6)^3=3\log_{1/4}(x+6),$$

the equation takes the form

$$\log_{1/4}|x+2|-1=\log_{1/4}(4-x)+\log_{1/4}(x+6)$$

whence

$$\log_{1/4}4|x+2|=\log_{1/4}(4-x)(x+6)$$

and consequently

$$4|x+2|=(4-x)(x+6)$$

(there is an extension of the domain in the last two transformations and so we can expect the appearance of extraneous roots!). The roots of this equation are $x_1=2$, $x_2=1-\sqrt{33}$.

Extraneous roots could appear during the process of solution only because of an extension of the domain of the variable and so, on the basis of Statement A, all we need to do is test the values x_1 and x_2 for membership in the domain. It is readily seen that all the expressions under the sign of the logarithm in the given equation for $x=x_1$ and for $x=x_2$ are positive so that both these numbers lie in the domain and are roots of the equation.

The restriction of the domain and, hence, the *loss of roots can also occur when passing to a new logarithmic base.*

12. *Solve the equation*

$$\log_{0.5x} x^2 - 14\log_{16x} x^3 + 40\log_{4x}\sqrt{x}=0$$

Here is a student's solution. Taking advantage of the change-of-base rule and taking x as the new logarithmic base, we have

$$\frac{\log_x x^2}{\log_x 0.5x}-\frac{14\log_x x^3}{\log_x 16x}+\frac{40\log_x \sqrt{x}}{\log_x 4x}=0$$

But it is quite evident that the new equation is devoid of meaning for $x=1$, whereas the original equation not only is meaningful for $x=1$ but has unity as its root. This is precisely where most students lose the root $x=1$.

We must therefore reason as follows: we want to pass to the base x; to do this we must be sure that $x>0$ and $x\neq1$. Since all the x of our domain are positive, the first condition $x>0$ is satisfied; on the other hand, unity lies in the domain and substitution shows that $x=1$ is a root. Thus, one root of the original equation has been found: $x=1$. Now let us seek roots that differ from unity. Then we can pass to the base x without losing roots.

From now on the solution is not difficult. Using the properties of logarithms and denoting $\log_x 2$ by y we have

$$\frac{2}{1-y}-\frac{42}{1+4y}+\frac{20}{1+2y}=0$$

This equation is reduced to the quadratic $2y^2+3y-2=0$, whose roots are $y_1=1/2$, $y_2=-2$. Then we get $\log_x 2=1/2$, whence $x=4$ and $\log_x 2=-2$, whence $x=1/\sqrt{2}$. Both of these values, 4 and $1/\sqrt{2}$, are roots of the original equation. Hence, the original equation has three roots.

A common and very grave mistake that results in a loss of roots is the *cancelling of a common factor from both sides of an equation.* It is clear that in the process, roots may be lost which make the common factor vanish.

In such cases it is best to transpose all terms to the left side, take out the common factor and consider two cases: (1) the common factor is equal to zero; (2) the common factor is not equal to zero; then of course the expression in the brackets is zero. It is also possible to consider first the case when the common factor is equal to zero and then cancel the common factor.

13. *Find all the solutions of the equation*

$$x^2 2^{x+1} + 2^{|x-3|+2} = x^2 2^{|x-3|+4} + 2^{x-1}$$

We consider two cases.

(a) Let $x \geqslant 3$. Here we have the equation

$$x^2 2^{x+1} + 2^{x-1} = x^2 2^{x+1} + 2^{x-1}$$

which is evidently satisfied for every x, and so in this case the solutions of the given equation will be all values of $x \geqslant 3$.

(b) Let $x < 3$. Then the equation takes the form

$$x^2 2^{x+1} + 2^{5-x} = x^2 2^{7-x} + 2^{x-1}$$

whence

$$2^{x-1}(4x^2-1) = 2^{5-x}(4x^2-1)$$

It was precisely at this point where many of the students at the examination were taken in by the exponential ("main") expressions and disregarded the "insignificant" power expressions and simply cancelled them obtaining the equation $2^{x-1}=2^{5-x}$. They then obtained the root $x=3$ and, what is more, failed to notice that it does not satisfy the condition $x<3$.

It is abundantly clear that before cancelling out $4x^2-1$ the students should have considered the case of $4x^2-1=0$. Then they would have found the roots $x_{1,2}=\pm 1/2$ which also satisfy the condition $x<3$.

Thus, the solutions of the given equation are: any $x \geqslant 3$, $x_1=1/2$, $x_2=-1/2$.

A common mistake made by students is the incorrect use of the following statement: "If two powers are equal and if their bases are equal and *different* from 0 and 1, then their exponents are equal as well." What is usually forgotten is the phrase "different from 0 and 1". The result is a loss of roots, namely those for which the base is equal to 0 or 1.

14. *Solve the equation* $x^{\sqrt{x}} = \sqrt{x^x}$.

This equation may be rewritten as

$$x^{\sqrt{x}} = x^{x/2}$$

Thus, the powers are equal and the bases are equal. So as not to lose any roots, let us see whether the base can be 0 or 1. Since the expression 0^0 is meaningless, the number 0 is not an element in the domain set and therefore $x=0$ is not a root of the equation. Contrariwise, $x=1$ is obviously a root. Now let us seek roots that are different from 0 and 1. Using the indicated rule, we obtain $\sqrt{x}=x/2$, whence we find the second root of the equation, $x=4$.

One sometimes hears the erroneous assertion: "If the power of a number is 1, then the exponent is equal to zero." This is only true if the base is different from 1, but if the base is 1, then for any exponent the power will be 1.

15. *Solve the equation* $|\cos x|^{\sin^2 x-\frac{3}{2}\sin x+\frac{1}{2}}=1$.

We reason as follows: if $|\cos x|=1$, then the power will be equal to 1 no matter what the exponent. But if $|\cos x|\neq 1$, then the exponent must necessarily be equal to zero. Thus, our equation separates into two:

$$|\cos x|=1 \quad \text{and} \quad \sin^2 x-\frac{3}{2}\sin x+\frac{1}{2}=0$$

The first equation yields $x_1=k\pi$, where k is any integer, and the second equation yields

$$\sin x=\frac{1}{2}, \text{ whence } x_2=(-1)^k\frac{\pi}{6}+k\pi,$$

$$\sin x=1, \text{ whence } x_3=\frac{\pi}{2}+2k\pi$$

A check shows that the angles of the second group do not lie in the domain of the variable (we get 0^0 in the left member, which is meaningless), the remaining roots satisfy the equation.

Finally, the solution set of the equation has the form

$$x_1=k\pi, \ x_2=(-1)^k\pi/6+k\pi \quad (k \text{ an integer})$$

Let us summarize: when applying the rule for passing from an equation of powers to an equation of exponents, it is necessary to consider three cases: the base is 0, the base is 1, and the exponents are equal. Reasoning in this manner we can avoid any loss of roots.

However, extraneous roots may appear. Indeed, in each case we have to solve an equation, and since all three equations are solved in isolation from one another it may happen that some of the solutions will not lie in the domain of the variable of the original equation. This is what occurred in the last example where some of the solutions of the second equation did not lie in the domain of the original equation and were therefore discarded.

For this reason, after applying the rule for passing from an equation of powers to an equation of exponents and after solving the respective equations, make a check. It will suffice to establish that the root belongs to the domain of the original equation, in which case it automatically satisfies the given equation.

A frequent source of loss of roots is the use of trigonometric formulas. It will be seen (see Sec. 2.2) that the left and right members of a trigonometric formula may have different domains of the variable. Such, for example, are the formulas of so-called "universal substitution" which express the sine and cosine in terms of the tangent of one-half an angle. In these formulas, the left member has a larger domain of the variable and therefore, when replacing the left member of the formula by the right, we restrict (narrow) the domain and run the risk of losing roots.

16. *Solve the equation* $\sin x - 2\cos x = 2$.

Passing to the tangent of one-half an angle, we get

$$\frac{2\tan\frac{x}{2}}{1+\tan^2\frac{x}{2}} - \frac{2\left(1-\tan^2\frac{x}{2}\right)}{1+\tan^2\frac{x}{2}} = 2$$

whence

$$\tan\frac{x}{2}=2 \text{ and } x=2\arctan 2+2n\pi,\ n=0,\ \pm 1,\ \pm 2,\ \ldots$$

However, this formula does not include all solutions: as can readily be verified, all the angles $x=(2n+1)\pi$, $n=0, \pm 1, \pm 2, \ldots$ are also solutions. These angles were lost when we introduced the tangent of half an angle. The original equation was meaningful for all values of x, while the second equation is meaningful only when $\tan(x/2)$ is meaningful, that is, for $x \neq (2n+1)\pi$.

For more details on trigonometric formulas see Sec. 2.2. Roots may be lost when solving equations by the *trial-and-error method*. The following are some illustrative examples.

17. *Solve the equation* $3^x+4^x=5^x$.

It is obvious that $x=2$ is a root of the equation. Is the equation thus solved? Of course not, for we may easily have not noticed some other root. Therefore, it is a grave mistake to stop at this point. Let us continue as follows: we divide both sides by 5^x to get

$$\left(\frac{3}{5}\right)^x+\left(\frac{4}{5}\right)^x=1$$

whence it is evident that if $x<2$, then, by the property of an exponential function with base less than unity, $(3/5)^x>(3/5)^2$, $(4/5)^x>$ $>(4/5)^2$ so that $(3/5)^x+(4/5)^x>(3/5)^2+(4/5)^2=1$, consequently, $x<2$ cannot be a root of the equation. Similarly, for $x>2$ we will always have the inequality $(3/5)^x+(4/5)^x<1$.

Thus, the chosen root, $x=2$, is the only one. Now the equation can be considered as solved. We found a root (it is immaterial just how this was done) and then proved that there are no other roots.

It is quite evident, as this example shows so clearly, that *trial-and-error solutions are legitimate if after guessing one or another of the roots* *we give rigorous proof that there can be no other roots.* Using this technique, it is very easy to solve the equation given in Problem 1:

$$\sqrt{2x-6}+\sqrt{x+4}=5$$

It is easy to choose a root $x=5$. But if $x>5$, then $\sqrt{2x-6}>\sqrt{10-6}=2$, $\sqrt{x+4}>3$, which means that for $x>5$ the left member exceeds 5. Similarly, for $x<5$ the left member is less than 5. Hence, $x=5$ is the only root of the equation.

However, if we confine ourselves to guessing and do not prove that there are no other roots, then roots will very often be lost. Such, for instance, is the danger in the problem that follows.

18. *Solve the equation* $3^x\cdot 8^{\frac{x}{x+2}}=6$.

Some students solved this equation thus: rewriting it as

$$3^x\cdot 2^{\frac{3x}{x+2}}=3^1\cdot 2^1$$

they chose a root x so that the exponents of the respective bases were the same:

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

whence the "answer" $x=1$.

But this "answer" is incorrect in the sense that only one root of the equation is found and nothing has been said about any other roots. Actually, if the exponents on the appropriate bases are equal, then the products of these powers are equal, however the converse is not in any way implied and is simply incorrect. For instance, the equation

$$3^1\cdot 2^1=3^2\cdot 2^{\log_2(2/3)}$$

is valid, but $1\neq 2$ and $1\neq\log_2(2/3)$. Therefore, the foregoing reasoning may lead to a loss of roots, and this is exactly what occurred in the equation at hand.

Taking logarithms of both members of the original equation to the base 10, we get

$$x\log_{10}3+\frac{3x}{x+2}\log_{10}2=\log_{10}6$$

or

$$x^2\log_{10}3+x(3\log_{10}2+2\log_{10}3-\log_{10}6)-2\log_{10}6=0$$

We now have to solve this quadratic equation. This can be done using a familiar formula, but we will try to simplify the solution by an ingenious device, since we have already seen, by trial and error, that $x_1=1$ is a root of the original equation and, consequently, satisfies the equivalent quadratic equation. For this reason, by Viète's theorem the second root of the quadratic equation is $x_2=(-2\log_{10}6)/\log_{10}3=$

$=-2\log_3 6$ and so the original equation has two roots: $x_1=1$, $x_2=$ $=2\log_3 6$.

Thus, it is useful to be able to guess a root, but never consider the guessing as the whole solution.

The main difficulty often consists not in the loss or the introduction of roots but in other things that are no less involved, as will be seen in the next few problems.

19. *Solve the equation*

$$\log_{1/(8\cos^2 x)} \sin x = \frac{1}{2}$$

By the definition of a logarithm, we get the following equation:

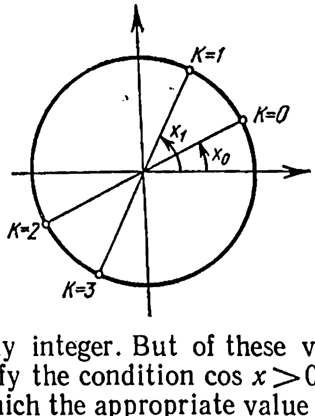
$$\sin x = \frac{1}{\sqrt{8\cos^2 x}}$$

This equation is a consequence of the original one, but clearly has a larger domain; indeed, the domain consists of all values of x for which $\cos x \neq 0$, whereas the original equation requires that two other conditions be satisfied as well: $1/\sqrt{8\cos^2 x} \neq 1$ and $\sin x > 0$. Nevertheless these equations are equivalent since every root of the second equation lies in the domain of the original one. Indeed, if $\sin x_0=$ $=1/\sqrt{8\cos^2 x_0}$, then, firstly, $\sin x_0 > 0$ and, secondly, $1/8\cos^2 x_0 \neq 1$, otherwise we would have $\cos^2 x_0 = 1/8$ and $\sin x_0 = 1$, which of course is impossible.

We now have $\sin x\,|\cos x| = 1/(2\sqrt{2})$ (see Sec. 1.4). To solve this equation it will be convenient to consider two cases.

(a) $\cos x > 0$. We then have the equation $\sin x \cos x = 1/(2\sqrt{2})$ or $\sin 2x = 1/\sqrt{2}$. Its solutions are given by the formula $x=(-1)^k \pi/8+$

Fig. 16



$+k\pi/2$, where k is any integer. But of these values we have to select only those that satisfy the condition $\cos x > 0$. To do this, determine the values of k for which the appropriate value of x lies in the first and fourth quadrants. This is easily done by depicting the solutions in the trigonometric circle. For values $k=0$, 1, 2, 3, the corresponding angles are shown in Fig. 16 (for other values of k, the angles repeat every four units). The only angles that suit us are $x_0=\pi/8$ and $x_1= =(-\pi/8)+(\pi/2)=3\pi/8$ and the groups which they generate (for $k= =4n$ and $k=1+4n$), that is, the angles

$$x=\frac{\pi}{8}+2n\pi, \quad x=\frac{3\pi}{8}+2n\pi$$

where n is any integer.

(b) $\cos x<0$. This case is examined in similar fashion. It yields four groups of solutions:

$$x=\frac{\pi}{8}+2n\pi, \quad x=\frac{3\pi}{8}+2n\pi, \quad x=\frac{5\pi}{8}+2n\pi,$$
$$x=\frac{7\pi}{8}+2n\pi$$

n any integer.

These groups can be combined into two:

$$x_1=(-1)^n\frac{\pi}{8}+n\pi, \quad x_2=(-1)^n\frac{3\pi}{8}+n\pi$$

20. *Solve the equation* $\tan(\pi\cos x)=\cot(\pi\cos 2x)$.

Transform the right member:

$$\cot(\pi\cos 2x)=\cot[\pi(2\cos^2 x-1)]=\cot(2\pi\cos^2 x)$$
$$=\tan(\pi/2-2\pi\cos^2 x)$$

Thus, we get the equation

$$\tan(\pi\cos x)=\tan(\pi/2-2\pi\cos^2 x)$$

whence

$$\pi\cos x-\left(\frac{\pi}{2}-2\pi\cos^2 x\right)=k\pi$$

where k is an arbitrary integer. Note right off that in passing to this equation we extended the domain of the variable; in the original equation the domain is defined by the condition

$$\pi\cos x\neq(\pi/2)+k\pi, \quad (\pi/2)-2\pi\cos^2 x\neq(\pi/2)+l\pi$$

That is, $\cos x\neq(1/2)+k$, $\cos^2 x\neq -l/2$ (where k, l are integers), while in the new equation the domain includes all values of x.

Furthermore, we have the equation $2\cos^2 x+\cos x-1/2=k$ or $4\cos^2 x+2\cos x-(2k+1)=0$, whence

$$\cos x=\frac{-1\pm\sqrt{8k+5}}{4}$$

We now have to determine for which values of k the equations

$$\cos x=\frac{-1-\sqrt{8k+5}}{4} \text{ and } \cos x=\frac{-1+\sqrt{8k+5}}{4}$$

have solutions. It is clear that $k\geqslant 0$ (otherwise $8k+5<0$).

The first equation has a solution for

$$-1 \leqslant \frac{-1-\sqrt{8k+5}}{4} \leqslant 1$$

The right inequality is automatically valid, and from the left we have $\sqrt{8k+5} \leqslant 3$ or $8k+5 \leqslant 9$, whence $k \leqslant 1/2 < 1$. Hence, the first equation has a solution only when $k=0$, and its solutions are

$$x = \pm \arccos\left(\frac{-1-\sqrt{5}}{4}\right) + 2n\pi, \ n \text{ any integer} \tag{6}$$

All these values of x clearly lie in the domain of the variable of the original equation because

$$\cos x = \frac{-1-\sqrt{5}}{4} \neq \frac{1}{2}+k, \qquad \cos^2 x = \frac{3+\sqrt{5}}{8} \neq -\frac{1}{2}$$

and consequently are its solutions.

The second equation has a solution for

$$-1 \leqslant \frac{-1+\sqrt{8k+5}}{4} \leqslant 1$$

or $-3 \leqslant \sqrt{8k+5} \leqslant 5$ whence $k \leqslant 5/2$. Hence, the second equation has a solution when $k=0, 1, 2$, and its solutions can be written as

$$x = \pm \arccos \frac{-1+\sqrt{8k+5}}{4} + 2n\pi \ (k=0, 1, 2; \ n \text{ any integer}) \tag{7}$$

All these values of x lie in the domain of the variable of the original equation and consequently are solutions of that equation.

Thus, the solution set of the original equation is given by the formulas (6) and (7).

Exercises

Solve the following equations:

1. $\sqrt{2x-3}+\sqrt{4x+1}=4$. 2. $\sqrt{4x-1}-\sqrt{x-2}=3$.
3. $\sqrt{x-1}+\sqrt{2x+2}=4$. 4. $\sqrt{x+1}+2\sqrt{2x-3}=-3$.
5. $\sqrt{x-1}+\sqrt{x+2}=\sqrt{34+x}-\sqrt{7+x}$.
6. $\sqrt{x+1}-1=\sqrt{x-\sqrt{x+8}}$.
7. $\sqrt{14-x}=\sqrt{x-4}+\sqrt{x-1}$.
8. $(2x+1)^{3/2}-(13x/2)=1$.
9. $\sqrt[3]{x-1}+\sqrt[3]{x-2}=\sqrt[3]{2x-3}$.
10. $6\sqrt[3]{x-3}+\sqrt[3]{x-2}=5\sqrt[6]{(x-2)(x-3)}$.
11. $a\sqrt[4]{1+x}+\frac{a}{x}\sqrt[4]{1+x}=\sqrt[4]{x}$.
12. $2^{2x+2}-6^x-2\cdot 3^{2x+2}=0$.

13. $8^x - 3\cdot 4^x - 3\cdot 2^{x+1} + 8 = 0.$

14. $4^x = 2\cdot 14^x + 3\cdot 49^x.$

15. $(2+\sqrt{3})^{x^2-2x+1} + (2-\sqrt{3})^{x^2-2x-1} = \dfrac{101}{10(2-\sqrt{3})}.$

16. $\log_3(4^x + 15\cdot 2^x + 27) - 2\log_3(4\cdot 2^x - 3) = 0.$

17. $(1 + x/2)\log_2 3 - \log_2(3^x - 13) = 3\log_{\sqrt{5/25}} 5 + 4.$

18. $\log_3(3^x - 1)\log_3(3^{x+1} - 3) = 6.$

19. $\log_5\left[(2+\sqrt{5})^x - (\sqrt{5}-2)^x\right] = \dfrac{1}{2} - 3\log_{1/5} 2.$

20. $\dfrac{\log_8(8/x^2)}{(\log_8 x)^2} = 3.$

21. $\dfrac{1}{2}\left(2^{\sqrt[3]{6x}} - 10\cdot 2^{-\sqrt[3]{6x}} + 1\right) = 3(\log_7\sqrt{49} - 1).$

22. $\log_{10}^2 x^3 - 20\log_{10}\sqrt{x} + 1 = 0.$

23. $\log_x 3\cdot\log_{x/3} 3 + \log_{x/81} 3 = 0.$

24. $1 + 2\log_x 2\cdot\log_4(10 - x) = 2/\log_4 x.$

25. $x^{x+1} = x.$

26. $x^{\log_x (x+3)^2} = 16.$

27. $\sqrt{x^{\log_{10}\sqrt{x}}} = 10.$

28. $x^{(\log_3 x)^3 - 3\log_3 x} = 3^{-3\log_{\sqrt{2}} 4} + 8.$

29. $x^{\log_3^2 x^2 - \log_3(2x) - 2} + (x+2)^{\log_{(x+2)^2} 4} = 3.$

30. $\dfrac{3}{x^{(\log_3 x^2)^3}} = (\sqrt{x})^{-\log_3 x + \frac{1}{\log_3\sqrt{x}}}.$

31. $3^{\log_a x} + 3\cdot x^{\log_a 3} = 2.$

32. $\sqrt{\log_{10} x} + \log_{10}\sqrt{x} = -1/2$

33. $\log_a(1 - \sqrt{1+x}) = \log_{a^2}(3 - \sqrt{1+x}).$

34. $\log_{\sqrt{2x-1}}(2x-3) = 2\log_8 4 + \log_2\left(1/\sqrt[3]{2}\right).$

35. $1/2\log_3(-x-16) - \log_3(\sqrt{-x} - 4) = 1.$

36. $\dfrac{1 + \log_2(x-4)}{\log_{\sqrt{2}}(\sqrt{x+3} - \sqrt{x-3})} = 1.$

37. $\sqrt{1 + \log_2 x} + \sqrt{4\log_4 x - 2} = 4.$

38. $\sqrt{1 + \log_x\sqrt{27}\log_3 x} + 1 = 0.$

39. $\log_{1/2}(x-1) + \log_{1/2}(x+1) - \log_{1/\sqrt{2}}(7-x) = 1.$

40. $\log_{1/\sqrt{1+x}} 10\cdot\log_{10}(x^2 - 3x + 2) = -2 + \log_{1/\sqrt{1+x}} 10\cdot\log_{10}(x-3).$

41. $\log_3(-x^2 - 8x - 14)\cdot\log_{x^2+4+4x} 9 = 1.$

42. $2\log_8(2x) + \log_8(x^2 + 1 - 2x) = 4/3.$

43. $\log_{x+1}(x^2 + x - 6)^2 = 4.$

44. $\log_3(\sqrt{x} + |\sqrt{x} - 1|) = \log_9(4\sqrt{x} - 3 + 4|\sqrt{x} - 1|).$

45. $2\log_2 x - \log_{1/2}(13-x) = \log_2 (x-10)^2 + 2\log_4 (8-x)$.

46. $2^{1+2\cos 5x} + 16^{\sin^2 (5x/2)} = 9$.

47. $3^{\sin 2x + 2\cos^2 x} + 3^{1-\sin 2x + 2\sin^2 x} = 28$.

48. $\frac{1}{2} + 16^{\sin x} = \frac{6}{16^{\cos^2\left(\frac{x}{2}+\frac{\pi}{4}\right)}}$.

49. $3(\log_2 \sin x)^2 + \log_2 (1-\cos 2x) = 2$.

50. $(\log_{\sin x} \cos x)^3 = 1$.

51. $\sqrt{\log_{\sin x} \cos x} = 1$.

52. $\log_{\sqrt{2}\sin x} (1+\cos x) = 2$.

53. $\log_{10} \sin 2x - \log_{10} \sin x = \log_{10} \cos 2x - \log_{10} \cos x + 2\log_{10} 2$.

54. $\log_2 \cos 2x - \log_2 \sin x - \log_2 \cos x = 1$.

55. $\log_{10} \sin (x/2) = \log_{10} (\cos x - \sin x) + \log_{10} (\cos x + \sin x)$.

56. $\log_2 \sin x - \log_2 \cos x - \log_2 (1-\tan x) - \log_2 (1+\tan x) = 1$.

57. $(\sin x)^{-\sin x} - 1 = \cot^2 x$.

58. $(\tan x)^{\cos^2 x} = (\cot x)^{\sin x}$.

59. $\left|\cos\frac{x}{2} + 2 - \frac{3}{4\cos\frac{x}{2}}\right|^{\sqrt{x^2+3x-10}} = 1$.

60. $(\cos 2x - \cos^4 x)\cot 3x + \frac{\sin 5x - \sin x}{8\sin 3x} = 0$.

61. $\cos^2 x \sin 6x + 2\sin^2 x \cdot \sin 3x \cdot \cos 3x + \left(\frac{\cos^2 x}{\cos 3x} - \sin^2 x\right)\sin 6x = 0$.

62. $\left(1+\frac{1}{\cos^2 x}\right)\sin 2x \cos 2x \cot 3x = 0$.

63. $\sqrt{\sin x} + \cos x = 0$.

64. $\sin 4x \sin x - \sin 3x \sin 2x = 1/2 \cos 3x + (1+\cos x)^{1/2}$.

65. $(\tan x + \sin x)^{1/2} + (\tan x - \sin x)^{1/2} = 2\sqrt{\tan x}\cos x$.

66. $5^x + 12^x = 13^x$.

67. $\left(\sqrt{2+\sqrt{3}}\right)^x + \left(\sqrt{2-\sqrt{3}}\right)^x = 2^x$.

1.10 Solving inequalities

A great deal of mistakes are made in the solution of inequalities. The point is that in most cases the solution of inequalities given at examinations does not require any particular ingenuity or artificial techniques, and so, as a rule, the student sees at a glance what steps must be taken. However, in carrying out the manipulations, the student makes serious mistakes due to a failure to recognize the fundamental theoretical propositions involving inequalities.

Actually, solving inequalities hardly requires anything more than the ability to reduce an inequality to the solution of elementary inequalities (without either losing a solution or introducing any extraneous ones), and then to solve these elementary inequalities. To carry out the latter part, the student has to know the fundamental properties of the functions studied at school (algebraic, exponential, logarithmic and trigonometric functions); to carry out the former part, the student must be able to handle the basic concepts involving the equivalence of inequalities, the sources of loss of solutions and of the introduction of extraneous solutions.

The basic definitions needed in the solution of inequalities repeat almost word for word those required for equations (Sec. 1.9). Note the following two differences in terminology however: the term "root" is not used when speaking of inequalities; one always uses the term "solution"; also, for the sake of brevity, one speaks of the solution being a certain set of values of x, for instance, the interval $a<x<b$, whereas in actuality every value of x of the set is a solution.

The similarity of equations and inequalities is quite naturally not confined to that of the basic definitions. It is obvious, for example, that everything that has been said about transforming equations which extend or restrict the domain of the variable is just as valid when applied to inequalities.

However, it must be stressed that solving inequalities has its peculiarities in that the same manipulations applied to equations and inequalities lead to different results. For instance, when multiplying both members of an equation by some nonzero factor (which is meaningful in the domain of the variable), an equation is replaced by an equivalent equation, whereas for inequalities we have to deal with the additional restriction that the factor be nonnegative in the domain of the variable. In the same way, squaring both sides of an equation does not lead to a loss of roots, while squaring an inequality can lead either to a loss of solutions or to the introduction of solutions. Students often lose sight of these peculiarities and make mistakes in the solution of inequalities that they never would make when solving equations.

It is a matter of wonder that so many mistakes are made by students when solving the simplest kind of inequality. Apparently this is due to a formally understood analogy between equations and inequalities. The reasoning goes roughly like this: "Since the solution of the equation $\log_{1/2} x=1$ is $x=1/2$, the solution of the inequality $\log_{1/2} x>1$ constitutes the values $x>1/2$." Similarly, solutions to the inequalities $(1/5)^x<2$ are written as $x<\log_{1/5} 2$, and so on. Yet the actual solutions to the two foregoing inequalities are different: in the first case, $0<x<1/2$, in the second, $x>\log_{1/5} 2$. A false analogy between equations and inequalities led to these mistakes.

Actually, when the student tackles an elementary inequality, he should consciously take advantage of the properties of the functions participating in the inequality.

Let us now consider examples in solving some elementary inequalities.

We wish to note first of all that the solution of linear (first-degree) and quadratic (second-degree) algebraic inequalities is usually quite thoroughly explained in textbooks and hardly ever causes any trouble. The solution of inequalities involving absolute values was discussed in Sec. 1.4.

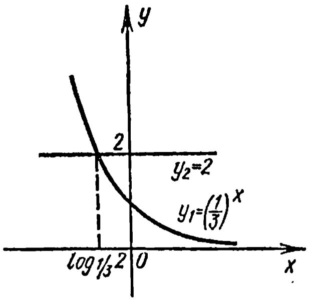
Here we wish to dwell on elementary exponential, logarithmic, and trigonometric inequalities.

An elementary *exponential inequality* is an inequality of the type $a^x > a^b$ ($a^x < a^b$). When handling such inequalities, it must be remembered that the properties of an exponential function differ for bases greater than unity and less than unity.

1. *Solve the inequality* $-1 \leqslant (1/3)^x < 2$.

To solve a double inequality means to find all the values of x which **simultaneously** satisfy the two inequalities: $(1/3)^x \geqslant -1$ and $(1/3)^x < 2$.

Fig. 17



Since an exponential function is always positive, the first of these inequalities is valid for all values of x.

Rewriting the second inequality as $(1/3)^x < (1/3)^{\log_{1/3} 2}$, we take advantage of the property of an exponential function: to a base less than unity, the greater value of the function is associated with the smaller value of the argument and conversely, to the smaller value of the function corresponds the greater value of the argument. This inequality is therefore equivalent to the inequality $x > \log_{1/3} 2$.

This solution is well illustrated by the graph shown in Fig. 17, namely, the solutions are those values of x for which the graph of the function $y = (1/3)^x$ lies **below** the horizontal straight line $y = 2$; that is, all x **to the right** of the abscissa of the point of intersection of these graphs (this abscissa is a solution of the equation $(1/3)^x = 2$). Thus, the solution of our inequality is the interval $x > \log_{1/3} 2$.

When solving inequalities containing the unknown under the sign of the logarithm, one must also bear in mind that the properties of a logarithmic function differ depending on whether the base is less than or greater than unity. However, another essential point in solving these inequalities is that the logarithmic function is not defined for all values of x. This is lost sight of by many students when solving an inequality like $\log_2 x < 1$. They reason this way: "We rewrite the inequality as $\log_2 x < \log_2 2$. The greater number to a base greater than 1 has the larger logarithm, and so the inequality is valid for $x < 2$."

Nothing would seem to be wrong in this argument, but still the answer is faulty because extraneous solutions were introduced. Indeed, any negative number is less than 2, but the original inequality is meaningless for negative values of x (because negative numbers do not have logarithms).

Why were extraneous solutions introduced? When "solving" the inequality, we passed from $\log_2 x < \log_2 2$ to $x < 2$. The latter inequality is meaningful for all values of x while the original inequality has meaning only for those values of x for which $\log_2 x$ is meaningful, that is to say, for $x > 0$. Hence, extraneous solutions were introduced simply because the fact was disregarded that a logarithmic function is defined only for positive values of x.

A correct answer is obtained if we choose from among the solutions of the latter inequality those whose values of $x > 0$; thus, the solution of our inequality is the interval $0 < x < 2$.

This simple example makes it abundantly clear that one should bear in mind, when solving logarithmic inequalities in this manner, that a logarithmic function is only defined for positive values of x. However, these inequalities may be solved in a different way: instead of using the domain of definition of the logarithmic function and its property of monotonicity we can immediately take advantage of Properties VII and VIII of logarithms (Sec. 1.6).

Thus, using Property VII in the above example, we can directly replace the inequality $\log_2 x < \log_2 2$ by the equivalent inequality $0 < x < 2$, which yields the answer.

Taking into account the simplicity of solving logarithmic inequalities by means of Properties VII and VIII, we will henceforth solve such inequalities by using these properties.

2. *Solve the inequality* $\log_{1/2} x > \log_{1/3} x$.

Taking the logarithm of the right member to the base 1/2 (Rule V, Sec. 1.6), we get an equivalent inequality:

$$\log_{1/2} x \left(1 - \log_{1/3} \frac{1}{2}\right) > 0$$

Since $1/2 > 1/3$, it follows that $\log_{1/3} 1/2 < \log_{1/3} 1/3$ or $1 - \log_{1/3} 1/2 > 0$.

Noting that $0 = \log_{1/2} 1$, we find that the original inequality is equivalent to $\log_{1/2} x > \log_{1/2} 1$.

Applying Property VIII to this inequality, we get the solution of the original inequality: $0 < x < 1$.

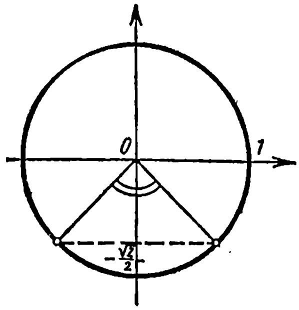
Now let us examine *trigonometric inequalities*. Despite the fact that the solutions of the more elementary trigonometric inequalities are thoroughly explained in the standard textbooks, students continue to make serious mistakes even when solving the simplest inequalities. We now examine a few typical mistakes of this nature.

(a) Knowing that the solutions of the equation $\sin x = a\,(|a| \leqslant 1)$ are given by the formula $x = (-1)^k \arcsin a + k\pi$, where $k = 0, \pm 1, \pm 2, \ldots$, many students write that "the solution of the inequality $\sin x < a$ consists of all values of $x < (-1)^k \arcsin a + k\pi$, $k = 0, \pm 1, \pm 2, \ldots$."

It is quite often difficult to convince the student of the absurdity of such an answer.

(b) Many mistakes are made that are connected with the formal use of the symbols $\arcsin a$, $\arccos a$, etc. These symbols are frequently employed when the student has not yet investigated whether they are meaningful or not. For instance, the solution to the inequality $\sin x \leqslant \log_4 5$ is written as $\arcsin(\log_4 5)$, which is meaningless since $\log_4 5 > 1$. Yet this inequality is valid for all values of x; this is evident from the very start because $\log_4 5 > 1$.

Fig 18.



(c) Mistakes occur due to improper use of the trigonometric circle. For example, when solving an inequality like $\sin x \leqslant -\sqrt{2}/2$, the students correctly indicate the angles that yield the solutions of the inequality (Fig. 18) but err when they give the analytic notation as

$$\frac{5\pi}{4} + 2k\pi \leqslant x \leqslant -\frac{\pi}{4} + 2k\pi, \quad k = 0, \pm 1, \pm 2, \ldots$$

It is clear that this notation is meaningless since the left member of the inequality is greater than the right member for all values of k.

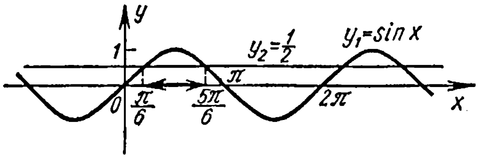
When solving elementary trigonometric inequalities, it is best to make use of the graphs of trigonometric functions. This is a practical guarantee against mistakes and makes for a pictorial representation of the regions in which the inequality is valid. When giving their analytic notation, it is convenient to take advantage of the following fact: if $f(x)$ is a periodic function, then to solve the inequality

$f(x) > a$ it suffices to find the solution in any interval that is equal to the length of the period of the function $f(x)$, then all values of x thus found and also all x that differ from these values by an integral number of the periods of the function $f(x)$ constitute a solution of our inequality.

3. *Solve the inequality* $\sin x > 1/2$.

We construct the graphs of the functions $y_1 = \sin x$ and $y_2 = 1/2$ (Fig. 19). This inequality is satisfied for all values of x for which the

Fig. 19



first graph lies above the second one. Since the period of the function $\sin x$ is 2π, it is sufficient for us to solve the proposed inequality on some interval of length 2π. It is easy to see that the most convenient interval is that from 0 to 2π: the solutions can most simply be written then as $\pi/6 < x < 5\pi/6$.

Thus, the complete solution of the inequality is

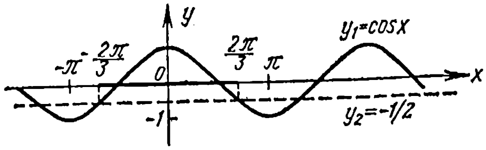
$$\frac{\pi}{6} + 2k\pi < x < \frac{5\pi}{3} + 2k\pi, \quad k = 0, \pm 1, \pm 2, \ldots$$

This notation is to be understood as follows: there is a certain interval for each integer k, and the set of all these intervals constitutes the solution of the inequality.

4. *Solve the inequality* $\cos x \geqslant -1/2$.

We construct the graphs of the functions $y_1 = \cos x$ and $y_2 = -1/2$ (Fig. 20). The period of the function $\cos x$ is also equal to 2π, but

Fig. 20



the drawing shows us that it is no longer convenient to take the interval from 0 to 2π for the basic interval because the solution of the inequality there will consist of two "pieces". It is therefore more convenient to seek the solution of this inequality on the interval from $-\pi$ to π. This is the interval $-2\pi/3 \leqslant x \leqslant 2\pi/3$. Consequently, the complete solution is

$$-\frac{2\pi}{3} + 2k\pi \leqslant x \leqslant \frac{2\pi}{3} + 2k\pi, \quad k = 0, \pm 1, \pm 2, \ldots$$

5. *Solve the inequality* $|\tan x| < 1/7$.

The period of the function $|\tan x|$ is equal to π. We consider the inequality on the interval from $-\pi/2$ to $\pi/2$ and construct the graphs of the functions $y_1=|\tan x|$ and $y_2=1/7$ (Fig. 21). It is evident that the solution will consist of all x lying in the interval $-x_0<x<x_0$, where x_0 is the abscissa of the intersection point of the graphs under consideration that lies between 0 and $\pi/2$, that is, the root of the equation $\tan x=1/7$ located in the interval $0<x<\pi/2$. Hence, $x_0=$

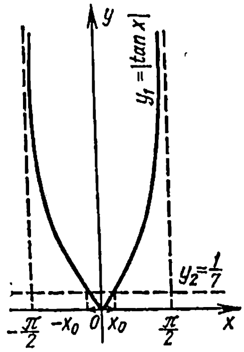


Fig. 21

$=\arctan(1/7)$. Taking into account the period of the function $y=|\tan x|$, we find that the solution of our inequality consists of all values of x located in the intervals

$$-\arctan\frac{1}{7}+k\pi<x<\arctan\frac{1}{7}+k\pi,\quad \text{where } k=0,\ \pm1,\ \pm2,\ \ldots$$

Note that the original inequality can be written as a double inequality $-1/7<\tan x<1/7$ and solved by using the graph of the function $y=\tan x$.

6. *Solve the inequality* $\sin x-\cos x>0$.

Using a consequence of the addition formula and $\pi/4$ as an auxiliary angle (we call this the auxiliary-angle formula), we get the inequality $\sqrt{2}\sin[x-(\pi/4)]>0$. Of course it can be solved by considering the graph of the function $y=\sin[x-(\pi/4)]$. However, it is best to do otherwise. Denoting $x-(\pi/4)$ by z, let us consider the inequality $\sin z>0$. Its solution $2\pi k<z<\pi+2\pi k$, $k=0, \pm1, \pm2, \ldots$ is directly obtained from the graph of the function $y=\sin z$. Now, substituting $x-(\pi/4)$ in place of z, we find the appropriate intervals of variation of x:

$$\frac{\pi}{4}+2k\pi<x<\frac{5\pi}{4}+2k\pi,\quad k=0,\ \pm1,\ \pm2,\ \ldots$$

This technique—replacing $x-(\pi/4)$ by z—enabled us to dispense with constructing the graph of the function $y=\sin[x-(\pi/4)]$. Its convenience is still more evident when solving elementary trigonometric inequalities with a complicated argument. For example, it allows us to get around constructing an extremely involved graph when solving inequalities like $\sin(\sqrt{2}x+7)>-1/2$. Here of course it is easier to denote $\sqrt{2x+7}$ by z and solve the inequality $\sin z > -1/2$ using the graph of the function $y = \sin z$, and then pass to x.

Higher-degree algebraic inequalities can also be classed as elementary inequalities. Students sometimes solve them by investigating various cases, which is to say, by passing to a solution of several systems of inequalities. Confusion often begins when the student is not able to find the common portion of the solutions and is undecided about whether or not to combine these solutions, yet there is a unified standard method for solving such inequalities. It is the so-called *method of intervals* that we now give.

Suppose, for example, we have to solve the inequality

$$(x-x_1)(x-x_2)\ldots(x-x_{n-1})(x-x_n) < 0$$

where $x_1, x_2, \ldots, x_n$ are *distinct real* numbers. We will assume that

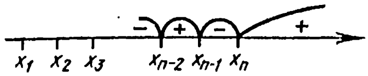
$$x_1 < x_2 < \ldots < x_{n-1} < x_n$$

Plot these points on the real number line (Fig. 22) and consider the polynomial

$$P(x) = (x-x_1)(x-x_2)\ldots(x-x_{n-1})(x-x_n) \qquad (1)$$

It is clear that for all $x > x_n$ all the parenthetic expressions in (1) are positive and, hence, for $x > x_n$ we have $P(x) > 0$. Since for $x_{n-1} < x < x_n$ the last parenthesis in the expression $P(x)$ is negative, and

Fig. 22



all the other parentheses are positive, it follows that for $x_{n-1} < x < x_n$ we have $P(x) < 0$. Similarly we obtain $P(x) > 0$ for $x_{n-2} < x < x_{n-1}$, and so on.

That is the underlying idea of the method of intervals. On the number line, the numbers $x_1, x_2, \ldots, x_n$ must be arranged in order of increasing magnitude. Then place the plus sign in the interval to the right of the largest number. In the next interval (from right to left) place the minus sign, then the plus sign, then the minus sign, etc. The solution of the inequality $P(x) < 0$ will then consist of intervals having the minus sign.

7. *Solve the inequality*

$$x(x+1)(-x+\sqrt{2})(x^2-x+1)(3x+1)^2(x+\sqrt{17})^3$$
$$\times(1-x)(2x-\pi^2)(-x+\pi)(x-\sin^2 1) < 0$$

It is quite obvious that if we reduce this inequality to systems of inequalities, then we will have a large number of cases to consider.

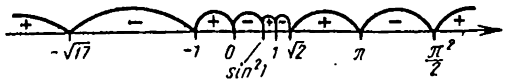
Let us solve it by the method of intervals. First, we have to reduce it to the proper form. Note that $x^2-x+1 > 0$ for any value of x and for this reason this factor can be cancelled from both members of the inequalities. Further note that $(3x+1)^2>0$ for $x \neq -1/3$ and therefore this factor can likewise be cancelled. Remember however that $x=-1/3$ is not a solution of the inequality. Besides, it is clear that the sign of $(x+\sqrt{17})^3$ coincides with that of $x+\sqrt{17}$ and therefore we can replace $(x+\sqrt{17})^3$ by $x+\sqrt{17}$ without impairing the inequality. Finally, represent each factor as $x-a$, where a is a number.

All these manipulations result in the inequality

$$(x-0)\,[x-(-1)]\,(x-\sqrt{2})\,[x-(-\sqrt{17})]$$
$$\times\,(x-1)\left(x-\frac{\pi^2}{2}\right)(x-\pi)\,(x-\sin^2 1)>0$$

which is equivalent to the original one for all $x \neq -1/3$ (since we multiplied three parentheses by -1, the sense of the inequality is reversed).

Fig. 23



Plot the numbers 0, -1, $-\sqrt{17}$, 1, $\pi^2/2$, π, $\sqrt{2}$ and $\sin^2 1$ on the real number line (Fig. 23). Then the last inequality is true for x located in the intervals

$$x<-\sqrt{17},\ -1<x<0,\quad \sin^2 1<x<1,$$
$$\sqrt{2}<x<\pi,\ \frac{\pi^2}{2}<x$$

The solutions of the original inequality are these values of x, with the exception of $x=-1/3$, that is,

$$x<-\sqrt{17},\ -1<x<-\frac{1}{3},\ -\frac{1}{3}<x<0,$$
$$\sin^2 1<x<1,\ \sqrt{2}<x<\pi,\ \frac{\pi^2}{2}<x$$

It is also to be noted that the weak inequality

$$(x-x_1)(x-x_2)\ \dots\ (x-x_n)\leqslant 0$$

can also be solved by the method of intervals, but the answer is written in the form of the intervals $x_i \leqslant x \leqslant x_{i+1}$ with the endpoints included.

Frequently, problems involving inequalities can be reduced to elementary inequalities by means of simple algebraic manipulations and the introduction of a new unknown.

8. *Solve the inequality*

$$9^x-10\cdot 3^x+9\leqslant 0$$

Denoting 3^x by y, rewrite the inequality thus: $y^2-10y+9\leqslant 0$. This quadratic inequality is true for all values of y in the interval $1\leqslant y\leqslant 9$.

Substituting 3^x in place of y, we obtain that the original inequality holds true for all x satisfying the double inequality $1 \leqslant 3^x \leqslant 9$.

Solving this elementary exponential inequality, we get the answer $0 \leqslant x \leqslant 2$.

9. *Solve the inequality*

$$\log_2^2 x + 3\log_2 x \geqslant \frac{5}{2}\log_{4\sqrt{2}} 16$$

Denoting $y = \log_2 x$ and noting that $5/2\ \log_{4\sqrt{2}} 16 = 4$, we rewrite our inequality thus: $y^2 + 3y - 4 \geqslant 0$. The solution set of this quadratic inequality is made up of all $y \geqslant 1$ and also all $y \leqslant -4$. Hence, the original inequality will hold true for all x for which $\log_2 x \geqslant 1$ and also for those x for which $\log_2 x \leqslant -4$. Solving these elementary logarithmic inequalities by means of Property VIII of logarithms, we get the answer: $x \geqslant 2$, $0 < x \leqslant 2^{-4}$.

10. *Solve the inequality*

$$\left(\frac{1}{2}\right)^{(x^6 - 2x^3 + 1)^{1/2}} < \left(\frac{1}{2}\right)^{1-x}$$

If we disregard the exponents, we can say that this is an elementary exponential inequality with base less than unity: $(1/2)^a < (1/2)^b$. Solving it, we find that the original inequality is equivalent to the inequality $(x^6 - 2x^3 + 1)^{1/2} > 1 - x$.

Since $(x^6 - 2x^3 + 1)^{1/2} = \sqrt{(x^3-1)^2} = |x^3 - 1|$ (see Sec. 1.4), it follows that we have yet to solve the inequality

$$|x^3 - 1| > 1 - x$$

Since the left member here is nonnegative, it is automatically satisfied for $1 - x < 0$, that is, when $x > 1$.

We now consider $x \leqslant 1$. In this case, $x^3 \leqslant 1$, and so $|x^3 - 1| = 1 - x^3$ and we have the inequality $1 - x^3 > 1 - x$ or

$$x(x-1)(x+1) < 0$$

Solving this inequality by the method of intervals, we find that it is true for $x < -1$ and for x located in the interval $0 < x < 1$. All these values of x lie in the domain $x \leqslant 1$ under consideration and so are solutions of the original inequality.

Thus, the original inequality is valid for $x < -1$, $0 < x < 1$, $x > 1$.

11. *Solve the inequality*

$$5 + 2\cos 2x \leqslant 3\,|2\sin x - 1|$$

Taking advantage of the formula for the cosine of a double angle and denoting $\sin x$ by y, we can rewrite our inequality as $7 - 4y^2 \leqslant 3\,|2y - 1|$. To get rid of the absolute-value sign, consider two cases: $y \geqslant 1/2$ and $y < 1/2$.

(a) Suppose $y \geqslant 1/2$, then our inequality is written $7-4y^2 \leqslant 3(2y-1)$ or $2y^2+3y-5 \geqslant 0$. The solution set of the latter inequality is $y \geqslant 1$ and $y \leqslant -5/2$. But taking into account that we only consider $y \geqslant 1/2$, we find that this condition is satisfied by $y \geqslant 1$ alone.

(b) Let $y < 1/2$. Then the original inequality is rewritten $7-4y^2 \leqslant$ $\leqslant -3(2y-1)$ or $2y^2-3y-2 \geqslant 0$. The solution set of this last inequality consists of $y \geqslant 2$ and $y \leqslant -1/2$. But Condition (b) is satisfied solely by $y \leqslant -1/2$.

Thus, the solutions of the inequality in y are $y \leqslant -1/2$ and $y \geqslant 1$. If in these inequalities we replace y by $\sin x$, we find the solutions of the original inequality to be all x that satisfy the elementary trigonometric inequality $\sin x \leqslant -1/2$ and all x satisfying the inequality $\sin x \geqslant 1$.

The solution set of the first inequality consists of all x lying in the intervals

$$-\frac{5\pi}{6}+2k\pi \leqslant x \leqslant -\frac{\pi}{6}+2k\pi, \quad k=0, \pm 1, \pm 2, \ldots$$

The second inequality will be true only for those values of x for which $\sin x = 1$; that is, for

$$x=\frac{\pi}{2}+2k\pi, \qquad k=0, \pm 1, \pm 2, \ldots$$

Thus, finally, the solution set of the original inequality consists of all $x=\pi/2+2k\pi$ and all x located in the intervals

$$-\frac{5\pi}{6}+2k\pi \leqslant x \leqslant -\frac{\pi}{6}+2k\pi, \quad k=0, \pm 1, \pm 2, \ldots$$

12. *Solve the inequality*

$$\log_5 \sin x > \log_{125}(3\sin x-2)$$

Noting that $\log_5 \sin x = \log_{125} \sin^3 x$, we rewrite our inequality as

$$\log_{125} \sin^3 x > \log_{125}(3\sin x-2)$$

Now applying Property VIII of logarithms (see Sec. 1.6), we see that our inequality is equivalent to the inequality $\sin^3 x > 3\sin x-2 > 0$.

Denoting $y=\sin x$, we arrive at the system of inequalities

$$y^3-3y+2>0$$
$$3y-2>0$$

Regrouping, represent the left member of the first inequality as

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

It then follows that this inequality is true for all $y>-2$ with the exception of $y=1$.

The second inequality of this system is valid for $y>2/3$. Hence, the solution of the system includes all $y>2/3$, except $y=1$.

Returning to x, we find that the original inequality is equivalent to the following double inequality: $2/3<\sin x<1$.

The solutions of this elementary trigonometric inequality are given by the intervals

$$\arcsin 2/3+2k\pi<x<\pi/2+2k\pi,$$
$$\pi/2+2k\pi<x<\pi-\arcsin 2/3+2k\pi,$$
$$k=0,\ \pm 1,\ \pm 2,\ \dots$$

13. *Solve the inequality*

$$\cos[\pi(x^2-10x)]-\sqrt{3}\sin[\pi(x^2-10x)]>1$$

Putting $y=\pi(x^2-10x)$, rewrite the inequality as

$$\frac{1}{2}\cos y-\frac{\sqrt{3}}{2}\sin y>\frac{1}{2}$$

Using the auxiliary-angle formula, we get $\cos[y+(\pi/3)]>1/2$. The solution of this elementary inequality consists of the intervals

$$-\frac{\pi}{3}+2k\pi<y+\frac{\pi}{3}<\frac{\pi}{3}+2k\pi,\quad k=0,\ \pm 1,\ \pm 2,\ \dots$$

Reverting to x, we find that for every integer k we have to solve the following system of quadratic inequalities:

$$x^2-10x-2k<0$$
$$x^2-10x-2k+2/3>0$$

The first inequality has solutions if and only if the discriminant of the quadratic expression $x^2-10x-2k$ is positive, that is, $25+2k>0$ or $k\geqslant -12$ (k an integer). And so the second inequality of the system will also be considered only for $k\geqslant -12$.

Note that for these k, the discriminant of the second inequality is also positive. For any fixed $k\geqslant -12$, the solution of the first quadratic inequality is the interval $5-\sqrt{25+2k}<x<5+\sqrt{25+2k}$ while the solutions of the second one consist of two infinite intervals: $x>5+$ $+\sqrt{25+2k-(^2/_3)}$ and $x<5-\sqrt{25+2k-(^2/_3)}$. The common portions of the solutions of these two inequalities (in the terminology of set theory we would say "the intersection of the solutions of these two inequalities") yield the solution of the system and, hence, of the original inequality. Clearly, $\sqrt{25+2k-(^2/_3)}<\sqrt{25+2k}$ for all $k\geqslant -12$.

Taking this remark into account, it is easy to write out the answer:

$$5-\sqrt{25+2k} < x < 5-\sqrt{25+2k-2/3},$$
$$5+\sqrt{25+2k-2/3} < x < 5+\sqrt{25+2k}$$

where k is an integer $\geqslant -12$.

Besides inequalities that are combinations of elementary inequalities, the student often has to deal with inequalities in the solution of which he has to apply various transformations and the associated concepts.

In the following simple examples we will show how the concept of the domain of the variable is used.

14. *Solve the inequality* $\sqrt{x} > -1$.

Since the left member is a nonnegative expression, the inequality is true for all values of x for which it is meaningful, that is to say, in the domain of the variable x. But the domain of this inequality consists of the set $x \geqslant 0$; this is the solution of the inequality.

15. *Solve the inequality* $\sqrt{\log_{10} x} > 0$.

This time again, the expression on the left-hand side is nonnegative and so the inequality holds true for all x in the domain of the variable, with the exception of those for which the left member vanishes. This domain is determined by the condition $\log_{10} x \geqslant 0$, which is to say it is the set $x \geqslant 1$. But when $x=1$, the left member vanishes and so this value of the unknown is not a solution of the inequality; the interval $x > 1$ constitutes the solution of the original inequality.

16. *Solve the inequality* $\log_{2-x}(x-3) \geqslant -5$.

The domain of the variable here is defined by the conditions $x-3 > 0$, $2-x > 0$, $2-x \neq 1$. But the inequalities $x-3 > 0$ and $2-x > 0$ do not have common solutions. Hence, the domain of our inequality does not contain a single number and so the inequality does not have a solution.

17. *Solve the inequality* $\sqrt{x+2}+\sqrt{x-5} \geqslant \sqrt{5-x}$.

The domain of the variable is defined by the inequalities $x+2 \geqslant 0$, $x-5 \geqslant 0, 5-x \geqslant 0$. But this system of inequalities has the sole solution $x=5$. Hence, the domain of the original inequality consists of the unique solution $x=5$. Therefore, no transformations are needed to solve this inequality since it is sufficient to verify that it is satisfied for $x=5$. A direct verification shows that $x=5$ is the solution.

18. *Solve the inequality* $\sqrt{2+x-x^2} > x-4$.

The domain of this inequality is the interval $-1 \leqslant x \leqslant 2$. Thus, the left member of the original inequality assumes real and nonnegative values for $-1 \leqslant x \leqslant 2$. It is meaningless for other values of x. But it is obvious that the right member of the inequality is negative for all $x < 4$ and, in particular, for all x in the interval $-1 \leqslant x \leqslant 2$; thus the proposed inequality is valid. Hence, the solution of the inequality is the interval $-1 \leqslant x \leqslant 2$.

19. *Solve the inequality* $\sqrt{\sin x + 2\cot x} < -1$.

The left member of this inequality is nonnegative for all permissible x and, consequently, it cannot be true for any value of x, which means there are no solutions.

The foregoing examples make it clear that we cannot give a general recipe of how to employ the notion of the domain of the variable of an inequality in various specific cases. In the first two examples we simply could not have found the solutions without computing the domain, in the third, fourth and fifth we first found the domain and this immediately gave us our answer. On the contrary, in the sixth example it would have been a complicated job to find the domain; what is more, it would have been senseless since there were no solutions anyway among the permissible values of x.

For this reason, when solving complicated problems, it is sometimes useful to find the domain at the start, but occasionally this is useless since later on it turns out to be superfluous for the given case. A general piece of advice may be given: if computing the domain is not complicated, then it is best to do so (since it will never do any harm), but if it is a complicated affair, then put off computing the domain until it is really needed.

At examinations one often encounters problems that require *transformations* which can result in a loss of solutions or the introduction of extraneous solutions. Here again, as in the case of equation solving, a principal role is played by the concept of *equivalence*. In Sec. 1.9 we examined the equivalence of equations and demonstrated why the student has to be sure that the newly derived equations and the original equations are equivalent. All this basically holds true for inequalities as well, in fact it is still more important than for equations.

Indeed, for equations it usually suffices to point out that for a certain transformation certain extraneous roots may be introduced and then to check the roots. In the case of inequalities, it is not possible to verify solutions by substitution since ordinarily there are an infinity of solutions. It is therefore necessary to pay special attention to the derived and original inequalities being equivalent.

It is to be noted that the transformations which lead to nonequivalence of equations (see Sec. 1.9) naturally lead to nonequivalent inequalities.

Certain manipulations only extend or restrict the domain of the variable of the inequalities. A general procedure can be suggested for such transformations: manipulations restricting the domain are forbidden since that might result in a loss of solutions; as for manipulations extending the domain, first carry them out and then choose from the solutions of the final inequality those values which enter into the domain of the original inequality. These will yield the answer.

The most common types of transformations that alter domains are the "identity transformations", which have already been mentioned in

Sec. 1.9. Besides these, the solution of inequalities involves other transformations as well: clearing of fractions, taking certain functions of both members. These include powering, taking logarithms, antilogarithms, and the like. We will now take these up in more detail.

We start with the most "harmless" one, that of *clearing fractions.* Recall equations. There is no loss of solutions when clearing fractions, and extraneous solutions are introduced only due to the extension of the domain of the variable, which is to say, via adding to the domain of the original equation those values of the unknown which make the denominator vanish.

Many think that the same holds true of inequalities, and so they "solve" the inequality $1/x<1$ this way: "clearing fractions we get $1<x$; all these values of x yield the solutions of the original inequality since the denominator of the original inequality does not vanish for any value."

But it is easy to see that the original inequality holds true for all negative values of x as well. All these solutions are thus lost by the student because clearing of fractions in equations is quite different from that operation in inequalities.

Actually, clearing of fractions in an equation (or inequality) consists in multiplying both members of the equation (or inequality) by the expression in the denominator. In this operation, equations remain equivalent if they are multiplied by a nonzero expression, but for inequalities this property is more involved: multiplication of both members of an inequality by a positive expression does not change the sense of the inequality, multiplication by a negative expression reverses the sense of the inequality.

Therefore, when multiplying both members of the inequality at hand by x, one should have taken into account that the x could have assumed negative values as well as positive values, and then he should have reversed the sense of the inequality in the latter case.

Thus, in every case when we wish to multiply both members of an inequality by an expression that is dependent on x and assumes both positive and negative values, the student should examine the two appropriate cases. This rule is often forgotten and is the cause of a lot of trouble.

20. *Solve the inequality*

$$(x-2)/(x+2) \geqslant (2x-3)/(4x-1)$$

The domain of the variable in this inequality consists of all values of x except $x=-2$ and $x=1/4$. From now on we will consider only those values of x which lie in the domain. At the examination, many students cleared fractions and wrote that it can be replaced by the following inequality:

$$(x-2)(4x-1) \geqslant (2x-3)(x+2) \qquad (2)$$

This is clearly wrong since the manipulation actually amounts to multiplying both members of the original inequality by the expression $(x+2)(4x-1)$, which may be negative as well as positive. The original inequality may be replaced by (2) if and only if the expression $(x+2)(4x-1)$ is positive, and so also we have to consider the case when it is negative. Thus, the solution of the original inequality reduces to solving systems of inequalities.

It is simpler however to do as follows. Transpose all terms of the original inequality to the left side and reduce it to a common denominator:

$$\frac{2(x^2-5x+4)}{(x+2)(4x-1)} \geqslant 0$$

The roots of the quadratic expression x^2-5x+4, i.e. $x_1=1$ and $x_2=4$, are the solutions of our inequality.

We will now assume that $x \neq 4$ and $x \neq 1$, and we will solve the inequality

$$\frac{(x-1)(x-4)}{(x+2)(x-1/4)} > 0 \tag{3}$$

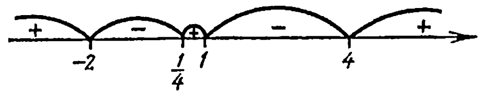
At this point, students often reduce the inequality to two systems of inequalities: the numerator and denominator are both greater than zero or are both less than zero. It is simpler however to solve it by the method of intervals.

Multiply both sides of the last inequality by the expression $(x+2)^2 \times (x-1/4)^2$, which is positive for the x under consideration. Then for all these values of x our inequality will be equivalent to the following one:

$$(x+2)(x-1/4)(x-1)(x-4) > 0 \tag{4}$$

This inequality is in a form convenient for application of the method of intervals. Fig. 24 shows us that the solutions of the last inequality

Fig. 24



consist of all x in the intervals $x<-2$, $1/4<x<1$, $4<x$. Since we have already found that $x=1$ and $x=4$ are solutions to the original inequality, we get the answer:

$$x<-2, \ \frac{1}{4}<x\leqslant 1, \ 4\leqslant x$$

In the foregoing solution, we replaced the inequality (3) by the inequality (4) by multiplying the first one by the square of the denominator. Similarly, we can assure ourselves that, generally, the inequalities

$$\frac{P(x)}{Q(x)} > 0 \quad \text{and} \quad P(x)\,Q(x) > 0$$

are equivalent. Therefore, to solve the inequality

$$\frac{P(x)}{Q(x)} > 0$$

where $P(x)$ and $Q(x)$ are polynomials, one applies the method of intervals to the inequality $P(x)\,Q(x) > 0$, which need not even be written out explicitly, it being sufficient to locate the roots of the polynomials $P(x)$ and $Q(x)$ on the number line and affix the appropriate sign to each of the resulting intervals.

21. *Solve the inequality*

$$\log_{1/2}\frac{5x+4}{x-2} > \tan\frac{5\pi}{4}$$

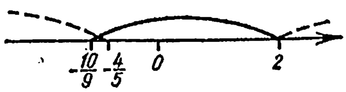
Noting that $\tan(5\pi/4)=1$ and applying Property VII of logarithms (Sec. 1.6), we see that our inequality is equivalent to the double inequality $0<(5x+4)/(x-2)<1/2$ or, what is the same thing, to the system of inequalities

$$\frac{5x+4}{x-2} > 0$$
$$\frac{5x+4}{x-2} < \frac{1}{2}$$

Transposing 1/2 to the left member of the second inequality and carrying out the obvious manipulations, we rewrite it in the form $[x+(10/9)]/(x-2)<0$.

We use the method of intervals to solve each of the inequalities of this system and find that the first inequality holds true for $x>2$ and for $x<-4/5$, while the second one is valid for x located in the interval $-10/9<x<2$. We now have to find the common part of these solutions (their intersection, in the terminology of set theory). This

Fig. 25



is conveniently done on the number line (Fig. 25). Plotting the points $-10/9$, $-4/5$ and 2, we denote the solutions of the first inequality by the broken line and the solutions of the second one by the solid line. The overlapping (common) portion of these two ranges is readily found to be $-10/9<x<-4/5$. This is the solution of the original inequality.

22. *Solve the inequality*

$$(\log_x 2)(\log_{2x} 2)(\log_2 4x) > 1$$

Using the properties of logarithms, this inequality may be rewritten

$$\frac{\log_2 4x}{\log_2 x\cdot\log_2 2x} > 1$$

Denoting $\log_2 x$ by y, we rewrite the last inequality as

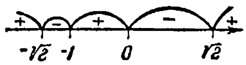
$$\frac{2+y}{y(1+y)} > 1$$

Transposing all terms to the right and reducing to a common denominator, we get

$$\frac{y^2-2}{y(1+y)} < 0$$

Factoring the numerator, we locate the roots of the numerator and denominator on the number line (Fig. 26) and then apply the method of intervals to get the solution of the inequality: all values of y in the intervals $-\sqrt{2} < y < -1$ and $0 < y < \sqrt{2}$.

Fig. 26



Recalling that $y = \log_2 x$, we see that the solution of the original inequality includes all values of x that satisfy either the inequality $-\sqrt{2} < \log_2 x < -1$ or the inequality $0 < \log_2 x < \sqrt{2}$. The solution of the first inequality consists of the x in the interval $2^{-\sqrt{2}} < x < 2^{-1}$; the solution of the second inequality consists of the values of x located in the interval $1 < x < 2^{\sqrt{2}}$.

Now let us consider *raising to a power*. In the sequel we will frequently make use of the following statement.

Theorem. *If $f(x) \geqslant 0$ and $\varphi(x) \geqslant 0$ on some set of values of x, then the inequalities $f(x) > \varphi(x)$ and $[f(x)]^2 > [\varphi(x)]^2$ are equivalent on that set.*

Proof. Let x_0 be an arbitrary solution of the first inequality taken from the set of values of x under consideration. If $\varphi(x_0) > 0$, then from the validity of the inequality $f(x_0) > \varphi(x_0)$ follows, on the basis of the theorem on raising numerical inequalities to a power, the validity of the inequality $[f(x_0)]^2 > [\varphi(x_0)]^2$. But if $\varphi(x_0) = 0$, then it is obvious that the validity of the inequality $f(x_0) > 0$ implies $[f(x_0)]^2 > 0$. This proves that every solution of the inequality $f(x) > \varphi(x)$ is a solution of the inequality $[f(x)]^2 > [\varphi(x)]^2$.

The converse is proved in similar fashion: that every solution of the inequality $[f(x)]^2 > [\varphi(x)]^2$ is a solution of the inequality $f(x) > \varphi(x)$.

The proof of the theorem is complete.

Note that in the statement of the theorem the strict inequalities $f(x) > \varphi(x)$ and $[f(x)]^2 > [\varphi(x)]^2$ may be replaced by the weak inequalities $f(x) \geqslant \varphi(x)$ and $[f(x)]^2 \geqslant [\varphi(x)]^2$. The proof of this fact is carried out in the same way as the proof of the theorem.

As shown in Sec. 1.9, when equations are raised to a power, it is only possible to introduce extraneous solutions, which may occur due to an extension of the domain of the variable or when the signs of the two sides of the equation are disregarded. Similarly, extraneous solutions can be introduced in the solution of inequalities; they too are introduced because of an extension of the domain of the variable and also when the signs of the two members of the inequality are disregarded. Below are some examples which illustrate how extraneous solutions are introduced in both cases.

However, unlike the case of equations, raising an inequality to a power can result in the *loss* of solutions as well. The reason why students make mistakes here is that they remember that raising an equation to a power cannot result in the loss of a solution but forget that raising an inequality to a power can result in the loss of solutions. We will show below how it is possible to lose a solution when raising an inequality to a power.

Let us begin with an example that illustrates how extraneous solutions are introduced due to extension of the domain of the variable when raising an inequality to a power.

23. *Solve the inequality* $\sqrt{(x-3)(2-x)} > \sqrt{4x^2+12x+11}$.

Some students gave this solution: "Since the right and left members of this inequality are nonnegative (this is because we have principal square roots on the right and left), the inequality may be squared to obtain the equivalent inequality $5x^2+7x+17>0$. The quadratic trinomial in the left-hand member of this equation does not have any real roots and therefore this inequality holds true for all real values of x. It then follows, because the inequalities are equivalent, that the original inequality too holds true for all values of x." This reasoning appears to be correct, but there is one serious defect. It is true only in the domain of the variable of the original inequality.

The proper solution is: in the domain of the variable, both members of the original inequality are nonnegative; for this reason it is equivalent, in the domain, to the inequality $5x^2+7x+17>0$ and hence is true for all values of x in the domain. It is now easy to find the domain of the original inequality and thus to obtain the answer: $2 \leqslant x \leqslant 3$.

In the problem that follows, extraneous solutions are introduced not because of an extension of the domain of the variable but because of raising to a power without investigating the signs of both members of the inequality.

24. *Solve the inequality* $x+1>\sqrt{x+3}$.

Here is an instance of reasoning that gives rise to extraneous solutions: "The domain of the variable of our inequality is $x \geqslant -3$. For any x in the domain we have a nonnegative number (principal square root) on the right; hence, the number on the left is a positive number. For this reason, squaring yields the equivalent inequality $x^2+x-2>0$, the solution of which is $x>1$ and also $x<-2$. Taking into account the domain of the original inequality, we get the answer: the solution of the original inequality consists of all values of $x>1$ and also of all values of x located in the interval $-3\leqslant x<-2$."

Actually, all values of x in the interval $-3\leqslant x<-2$ are not solutions to the original inequality. The point is that for x in the domain, the right member of the inequality is indeed nonnegative, whereas the left member is negative for certain values of x located in the domain and is nonnegative for others. It is clear that for those values of x in the domain for which the left member is negative, the inequality is invalid and so there are no solutions of our inequality among them. It is thus necessary to seek solutions of the original inequality among those values of x in the domain for which the left-hand member of the inequality is nonnegative, which is to say among $x\geqslant-1$.

For these x, both members of the inequality are indeed nonnegative, and it can be squared to obtain the inequality $x^2+x-2>0$, which is equivalent to the original inequality on the set $x\geqslant-1$. It is now necessary to choose from among the solutions of the inequality x^2+ $+x-2>0$ those which satisfy the condition $x\geqslant-1$. They will yield the solutions of the original inequality, which are $x>1$.

The mistake that was made in the earlier reasoning was due to the fact that the student did not notice the shift in concepts. It is true that for any value of x which is a *solution* of the original inequality there is a nonnegative number (principal square root) on the right and a positive number on the left. However, it is obvious that not all values of x located in the domain will be solutions of the original inequality, and so the number on the left will not be positive for all x of the domain. The student replaced the words "for any x which is a solution" by the phrase "for every value of x in the domain." This was his mistake.

25. *Solve the inequality*

$$\sqrt{4-\sqrt{1-x}}-\sqrt{2-x}>0$$

Difficulties here spring up when we begin to compute the domain of the variable. The domain of this inequality is defined from the conditions: $2-x\geqslant0$, $1-x\geqslant0$, $4\geqslant\sqrt{1-x}$. The first two of these inequalities are true for $x\leqslant1$. But both sides of the third inequality are nonnegative for these values of x, and so it can be squared to get an equivalent inequality: $x\geqslant-15$. Thus, the domain of the original inequality is $-15\leqslant x\leqslant1$. We rewrite our inequality thus: $\sqrt{4-\sqrt{1-x}}>\sqrt{2-x}$. Within the domain, both members of this inequality are nonnegative, therefore squaring yields an inequality, $2+x>\sqrt{1-x}$, that is equivalent in the domain. For values of $x<-2$ and such that enter into the domain, the left member of this inequality is negative, while the right member is nonnegative, which means that there are no solutions to the original inequality among these values of x. It remains to consider the values of x in the interval $-2\leqslant x\leqslant 1$. For these x, both members of the inequality $2+x>\sqrt{1-x}$ are nonnegative, and so squaring yields the quadratic inequality $x^2+5x+3>0$, which is equivalent to the original inequality on the set $-2\leqslant x\leqslant 1$. This latter inequality holds true for $x>(-5+\sqrt{13})/2$ and for $x<(-5-\sqrt{13})/2$. Now to get the answer we have to choose from among these solutions those which lie in the interval $-2\leqslant x\leqslant 1$. These consist of all values of x in the interval $(-5+\sqrt{13})/2<x\leqslant 1$. They are the ones which constitute the answer to this problem.

Note that if we had not taken the domain of the variable into account, we would have introduced extraneous solutions, for example, all $x>1$; and if we had not taken into consideration that the inequality $2+x>\sqrt{1-x}$ has solutions only for $-2\leqslant x\leqslant 1$, we would also have introduced extraneous solutions, for example, all the values of $x<(-5-\sqrt{13})/2$.

Let us now examine some problems in which one can lose solutions by raising the inequality to a power.

26. *Solve the inequality* $\sqrt{x+2}>x$.

If we square this inequality at once, we will lose solutions even if we take into account the domain of the variable. Indeed, the domain for this inequality is $x\geqslant -2$. Squaring, we get the inequality $x+2>x^2$, whose solution will consist of all the values of x in the interval $-1<x<2$. All these values of x enter into the domain, and so some students wrote that these values constitute the answer to the problem. Actually, in thus reasoning they lost the solutions $-2\leqslant x\leqslant -1$, because it is easy to see that for any number in this interval the left member of the inequality is nonnegative, while the right member is negative.

So as not to lose solutions, the student must keep careful watch of the signs of the left and right members. The proper solution of this inequality is as follows.

The domain of the variable in this inequality consists of all $x\geqslant -2$. The left member of the given inequality is nonnegative in the domain, while the right member may be positive or negative. Clearly, the original inequality will be true for all those values of x in the domain for which the right member is negative. Hence, all the values of x in the interval $0>x\geqslant -2$ are solutions of the original inequality.

Now let us consider the remaining values of x, that is, $x\geqslant 0$. Both members of the original inequality are nonnegative for all these x, and so the inequality can be squared to obtain $x+2>x^2$, which is an equivalent inequality for all $x\geqslant 0$.

The solution of the last inequality consists of all x in the interval $-1<x<2$. In this case, the solution of the original inequality consists of all values of x in the interval $0\leqslant x<2$.

Combining these two cases, we find that the solution to the original inequality will consist of all values of x lying in the interval $-2 \leqslant x < 2$.

In the next problem, it will be possible to lose solutions if one fails to take into account the signs of the right and left members of an intermediate inequality.

27. *Solve the inequality*

$$\sqrt{x^2+3x+2} < 1+\sqrt{x^2-x+1}$$

The domain of the variable here consists of two intervals: $x \leqslant -2$ and $x \geqslant -1$. In the domain, both members of our inequality are nonnegative and so squaring yields the equivalent (in the domain) inequality $2x < \sqrt{x^2-x+1}$.

(a) For $x \leqslant -2$ and $-1 \leqslant x < 0$, this inequality is true since for each of these values of x there is a negative number on the left and a positive number on the right. Thus, all these values of x are solutions to the original inequality.

(b) For $x \geqslant 0$, both members of the inequality $2x < \sqrt{x^2-x+1}$ are nonnegative and so squaring yields the equivalent (for these x) inequality $3x^2+x-1 < 0$. The solution of this inequality consists of values of x in the interval $(-1-\sqrt{13})/6 < x < (-1+\sqrt{13})/6$.

Taking Condition (b) into account, we find that in the latter case the solution of the original inequality will consist of all values of x in the interval $0 \leqslant x < (-1+\sqrt{13})/6$.

Combining both cases we get the answer: $x \leqslant -2$ and also $-1 \leqslant x < (1+\sqrt{13})/6$.

It will be noted that those students who did not consider the cases (a) and (b) and squared the inequality $2x < \sqrt{x^2-x+1}$ from the start naturally lost some of the solutions. Most likely what happened was that since at the beginning of the solution of the inequality the signs of the left and right members had already been investigated, there was a kind of loss of "vigilance" in the second squaring.

28. *Solve the inequality*

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

Denoting $\tan x$ by y, rewrite the inequality as

$$2\sqrt{3+2y-y^2} \geqslant 1+3y \tag{5}$$

The domain of the variable in (5) is the interval $-1 \leqslant y \leqslant 3$. Our inequality is obvious for those values of y in the domain for which $1+3y < 0$; that is, all values of y in the interval $-1 \leqslant y < -1/3$ are solutions of inequality (5).

It remains to consider Case (b): $-1/3 \leqslant y \leqslant 3$. Here both members of (5) are nonnegative and so squaring in the case at hand yields the equivalent inequality $13y^2-2y-11 \leqslant 0$.

The solution of the last inequality consists of all values of y in the interval $-11/13 \leqslant y \leqslant 1$. Taking into consideration Condition (b), we find that in this case the solution of inequality (5) consists of all values of y in the interval $-1/3 \leqslant y \leqslant 1$.

Combining both cases we find the solution to inequality (5) to be all values of y in the interval $-1 \leqslant y \leqslant 1$.

Returning to x, we find the solution of the original inequality to be all x satisfying the inequality $-1 \leqslant \tan x \leqslant 1$. We solve this elementary trigonometric inequality to get the answer:

$$-\frac{\pi}{4}+k\pi \leqslant x \leqslant \frac{\pi}{4}+k\pi, \text{ where } k=0, \pm 1, \pm 2, \ldots$$

Ordinarily, *taking antilogarithms* of inequalities is only employed in the solution of inequalities involving the unknown under the logarithmic sign. We have already considered the solution of elementary logarithmic inequalities and have seen that they are very simply solved by taking advantage of Properties VII and VIII of logarithms (Sec. 1.6). The more complicated logarithmic inequalities should therefore also be solved on the basis of these properties. This will help the student to avoid many mistakes.

One more remark is in order: despite the fact that taking antilogs is always involved in the solution of logarithmic inequalities (either with regard for the domain of definition of the logarithmic function or by Properties VII and VIII), that term is not always used and we find phrases like this: "on the basis of the properties of logarithms (or the logarithmic function) we have...".

Taking antilogarithms is investigated in the next few problems.

29. *Solve the inequality*

$$\log_{\frac{25-x^2}{16}}\left(\frac{24-2x-x^2}{14}\right) > 1$$

The natural thing to do is to take antilogarithms. Since the logarithmic base contains x and since the properties of a logarithmic function differ according as the base is greater than or less than unity, we cannot take antilogs straight off and will have to consider two cases.

(a) Let $\frac{25-x^2}{16} > 1$, that is, $x^2 < 9$. In this case the given inequality is equivalent to

$$\frac{24-2x-x^2}{14} > \frac{25-x^2}{16}$$

This inequality may be rewritten as $x^2+16x-17<0$. The solution of this inequality consists of all values of x in the interval $-17<x<1$, but the condition of this case (that is, $-3<x<3$) is only satisfied by those x located in the interval $-3<x<1$. All these values of x constitute the solution of the original inequality in the case at hand.

(b) Now let $0<(25-x^2)/16<1$. Here, the original inequality is equivalent to the double inequality

$$0<\frac{24-2x-x^2}{14}<\frac{25-x^2}{16}$$

Thus, in this case we have to solve the following system of double inequalities:

$$0<\frac{25-x^2}{16}<1$$
$$0<\frac{24-2x-x^2}{14}<\frac{25-x^2}{16}$$

The first one is readily reduced to $9<x^2<25$ and its solution consists of two intervals: $-5<x<-3$ and $3<x<5$. The second double inequality is equivalent to the system of inequalities

$$x^2+2x-24<0$$
$$x^2+16x-17>0$$

The first inequality of this system has the solution $-6<x<4$, the solution of the second one consists of two infinite intervals $x>1$ and $x<-17$, so the solution of the latter system is the interval $1<x<4$.

We now have to choose from these values of x those which satisfy the first double inequality; they are the values of x in the interval $3<x<4$.

Thus, combining the two cases, we have the solution of the original inequality, which consists of two intervals: $-3<x<1$ and $3<x<4$.

30. *Solve the inequality*

$$\log_{\frac{2\cos x}{\sqrt{3}}}\sqrt{1+2\cos 2x}<1$$

Denoting $\cos x$ by y and taking advantage of the formula for the cosine of a double angle, we rewrite the inequality as

$$\log_{2y/\sqrt{3}}\sqrt{4y^2-1}<1 \qquad (6)$$

The domain of this inequality is defined by the conditions $y^2>1/4$, $y>0$, $y\neq\sqrt{3}/2$, or $y>1/2$ and $y\neq\sqrt{3}/2$.

Since the logarithmic base may be greater than 1 or less than 1 for various values of y, we consider two cases:

(a) Let $1/2<y<\sqrt{3}/2$. Then the base is less than 1 and we obtain the equivalent inequality $\sqrt{4y^2-1}>2y/\sqrt{3}$ or, since both members are positive, $4y^2-1>4y^2/3$. This inequality is true for $y^2>3/8$, or $y>\sqrt{6}/4$. Taking into account the condition of the case at hand, we find the solution of inequality (6) to be the interval $\sqrt{6}/4<y<\sqrt{3}/2$.

(b) Let $y > \sqrt{3}/2$. Then we get the equivalent inequality $\sqrt{4y^2-1} < 2y/\sqrt{3}$ whence, after squaring, follows $y^2 < 3/8$. But under the condition of our case $y^2 > 3/4$, that is, we do not obtain any new solutions to inequality (6).

It thus remains to solve the elementary trigonometric inequality $\sqrt{6}/4 < \cos x < \sqrt{3}/2$, which is satisfied for all values of x in the intervals

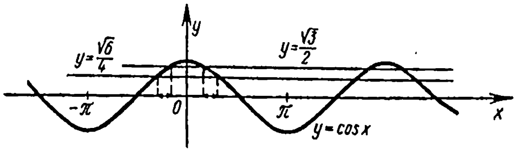
$$-\arccos\frac{\sqrt{6}}{4} + 2k\pi < x < -\frac{\pi}{6} + 2k\pi,$$

$$\frac{\pi}{6} + 2k\pi < x < \arccos\frac{\sqrt{6}}{4} + 2k\pi$$

where k is any integer (Fig. 27).

With respect to *taking logarithms* of inequalities, it is easy to see in which cases this operation leads to an equivalent inequality. However, it is well to bear in mind that unwise logarithm-taking of inequalities can result in a restriction of the domain of the variable and a loss

Fig. 27



of solutions. Therefore, prior to taking logarithms always check to see that both members of the inequality are positive. Only then (and naturally with regard for the base of the logarithm) are we able to generate an equivalent inequality.

Earlier (see Problem 1) we solved the elementary inequality $(1/3)^x < 2$, using the properties of the exponential function. Let us now solve this inequality by taking logarithms. Since both members of the inequality $(1/3)^x < 2$ are positive, we can take advantage of Property VIII of logarithms (Sec. 1.6) and take logarithms of the inequality to the base 1/3 to get $\log_{1/3}(1/3)^x > \log_{1/3} 2$ (note that the sense of the inequality has been reversed!), whence $x > \log_{1/3} 2$.

Thus, the solution set of the inequality is the set $x > \log_{1/3} 2$.

Note that all the elementary exponential inequalities discussed above could have been solved via logarithms.

Let us solve a few problems by taking logarithms.

31. *Solve the inequality*

$$x^4 \cdot 7^{\log_{7^{1/3}} 5} \leqslant 5^{-\log_{1/x} 5}$$

The domain of the variable here consists of all $x > 0$, except $x = 1$.

Since $7^{\log_{7^{1/3}} 5} = 5^3$ and $\log_{1/x} 5 = -\log_x 5$, our inequality may be rewritten in the form

$$x^4 \cdot 5^3 \leqslant 5^{\log_x 5}$$

Both members are positive within the domain of the variable and so we can take logarithms of both sides of the inequality to the base 5 (greater than unity) and obtain the equivalent (in the domain) inequality $4\log_5 x + 3 \leqslant \log_x 5$. Denoting $\log_5 x$ by y and transposing all terms to the left-hand side, we rewrite the inequality thus: $4y + 3 - 1/y \leqslant 0$ or, reducing to a common denominator, thus: $(y+1)(y - 1/4)/y \leqslant 0$.

Now we apply the method of intervals and find the solution to be $y \leqslant -1$ and y in the interval $0 < y \leqslant 1/4$.

Now reverting to x, we see that the original inequality is true for those values of x for which $\log_5 x \leqslant -1$ and also for those x for which $0 < \log_5 x \leqslant 1/4$. Solving these elementary logarithmic inequalities, we get the answer: $0 < x \leqslant 1/5$, $1 < x \leqslant \sqrt[4]{5}$.

32. *Solve the inequality*

$$(x^2 + x + 1)^x < 1$$

For arbitrary real x, the quadratic expression $x^2 + x + 1$ is positive and therefore the domain of the variable consists of all real values of x.

Since both sides of the original inequality are positive for all x, we take logs to the base 10 to get the equivalent inequality $x \log_{10}(x^2 + x + 1) < 0$. This inequality holds true in two cases: when x satisfies the system of inequalities

$$x > 0$$
$$\log_{10}(x^2 + x + 1) < 0$$

and when x satisfies the system of inequalities

$$x < 0$$
$$\log_{10}(x^2 + x + 1) > 0$$

Let us solve the first system of inequalities. From the properties of logarithms we find that it is equivalent to the system

$$x > 0$$
$$x^2 + x + 1 < 1$$

Since the solution of the second inequality of the system is $-1 < x < 0$ and the solution of the first is $x > 0$, this system is inconsistent, which means that in this case the original inequality does not have a solution.

The second system is equivalent to

$$x < 0$$
$$x^2 + x + 1 > 1$$

The solution set of this system consists of all $x < -1$, whence the solution of the original inequality is the set of all values of $x < -1$.

A different solution of this inequality may be suggested. Since the properties of a power depend on whether the base is greater or less than unity, it is natural to consider two cases.

(a) Suppose that $x^2+x+1<1$, or $-1<x<0$. For all these values of x, the quadratic x^2+x+1 is raised to a negative power x. And since for all these values of x the trinomial $x^2+x+1<1$, it follows that for them $(x^2+x+1)^x>1$, which contradicts the condition. Hence, these values of x cannot be solutions of our inequality.

(b) Suppose that $x^2+x+1>1$. This is clearly valid for $x>0$ and for $x<-1$. Therefore we have to consider two cases here.

Let $x>0$. Then $x^2+x+1>1$ and after raising the expression to a positive power x the sense of the inequality remains unchanged, which means that for these x we have $(x^2+x+1)^x>1$. Hence, neither can these values of x be solutions of our inequality.

Let $x<-1$. Then $x^2+x+1>1$. If the quadratic expression x^2+x+1 is now raised to a negative power x, the result will be less than unity, which means that for all $x<-1$ we have $(x^2+x+1)^x<1$.

Thus, the solution set of the original inequality consists of all values of $x<-1$.

We have been making considerable use of the concept of domain of the variable of our inequalities. However, with the exception of just a few very elementary cases, we did not stress whether this has been helpful or not in solving inequalities. We will therefore consider two examples involving inequalities to see whether it is necessary to compute the domain of the variable beforehand.

In the next problem, the solution will be appreciably simplified if the domain of the variable is found beforehand.

33. *Solve the inequality*

$$\log_{x^2}\left(\frac{4x-5}{|x-2|}\right)\geqslant\frac{1}{2}$$

The domain of x here is defined from the conditions $(4x-5)/|x-2|>0$, $x^2>0, x^2\neq 1$, whence $x>5/4$ and $x\neq 2$. But for all these values of x we have $x^2>1$ and so our inequality, by the property of logarithms to a base exceeding unity, is equivalent (within the domain of x) to

$$\frac{4x-5}{|x-2|}\geqslant x$$

Since $x\neq 2$, the expression $|x-2|$ is positive and therefore the original inequality is equivalent, within the domain of x, to

$$4x-5\geqslant x|x-2|$$

We now consider two cases.

(a) Let $x>2$. Then our inequality will be rewritten as $4x-5\geqslant x^2-2x$ or $x^2-6x+5\leqslant 0$. The solution set of the last inequality consists of all x in the interval $1\leqslant x\leqslant 5$, and the solution set of the original inequality in this case consists of all values of x in the interval $2<x\leqslant 5$.

(b) Now let $5/4\leqslant x<2$. Then our inequality takes the form $4x-5\geqslant -x^2+2x$ or $x^2+2x-5\geqslant 0$. Its solution set will consist of all values of x in the intervals $x\geqslant\sqrt{6}-1$ and $x\leqslant -\sqrt{6}-1$. The solution set of the original inequality will in this case consist of all values of x in the interval $\sqrt{6}-1\leqslant x<2$.

Combining both cases, we find the solution of the original inequality to consist of all values of x in the intervals $\sqrt{6}-1\leqslant x<2$ and $2<x\leqslant 5$.

In the next problem, it is not advisable to establish the domain of the variable beforehand since it does not simplify the solution and is a rather complicated matter.

34. *Solve the inequality*

$$\log_x 2x\leqslant\sqrt{\log_x (2x^3)}$$

Without finding the domain of x, we note only that in the domain, $x>0$ and $x\neq 1$.

Denote $\log_x 2$ by y and rewrite the inequality as

$$y+1\leqslant\sqrt{y+3} \tag{7}$$

Here the domain of y consists of all $y\geqslant -3$. But the inequality is obvious for y in the interval $-3\leqslant y<-1$, which means that all these values of y constitute the solution.

Now let $y\geqslant -1$. Then both members of (7) are nonnegative; this inequality may be squared to obtain the equivalent (for $y\geqslant -1$) inequality $(y+1)^2\leqslant y+3$, whose solution consists of all y in the interval $-2\leqslant y\leqslant 1$. In this case, the solution of (7) consists of all values of y in the interval $-1\leqslant y\leqslant 1$.

Combining both cases we see that inequality (7) is satisfied for $-3\leqslant y\leqslant 1$.

Now returning to x we find that the original inequality will have solutions for all x that satisfy the double inequality $-3\leqslant\log_x 2\leqslant 1$.

This inequality may be solved in two ways.

First solution. Since the properties of logarithms differ for bases greater than or less than unity, we consider two cases: $x>1$ and $0<x<1$.

(a) Let $x>1$. Then $\log_x 2>0$, and all the more so $\log_x 2\geqslant -3$. It remains to solve the inequality $\log_x 2\leqslant 1$, whence $2\leqslant x$. Thus, here the solution set of the original inequality consists of all values of $x\geqslant 2$.

(b) Let $0<x<1$. Then $\log_x 2<0$ and all the more so $\log_x 2\leqslant 1$. It remains to solve the inequality $-3\leqslant\log_x 2$, whence $x^{-3}\geqslant 2$, or $x\leqslant 2^{-1/3}$. Thus in this case the solution set of the original inequality consists of all x in the interval $0<x\leqslant\sqrt[3]{1/2}$.

Combining both cases we have the solution of the original inequality: $0<x<\sqrt[3]{1/2}$ and $x\geqslant 2$.

Second solution. Since $\log_x 2=1/\log_2 x$, then, by denoting $\log_2 x$ by z, we get the system of inequalities

$$\frac{1-z}{z}\leqslant 0$$

$$\frac{1+3z}{z}\geqslant 0$$

Solving each of these inequalities by the method of intervals, we find that the system will hold true for $z\geqslant 1$ and also for $z\leqslant -1/3$.

To get the answer we have to solve the elementary logarithmic inequalities $\log_2 x\geqslant 1$ and $\log_2 x\leqslant -1/3$, whence we obtain $x\geqslant 2$ and $0<x\leqslant\sqrt[3]{1/2}$.

We conclude this section with a problem containing two danger spots at once: the introduction and loss of solutions.

35. *Solve the inequality*

$$\sqrt{4\sin^2 x-1}\,\log_{\sin x}\frac{x-5}{2x-1}\geqslant 0$$

At the very start, many students made the mistake of discarding the first factor. Their reasoning probably went like this: since $\sqrt{4\sin^2 x-1}\geqslant 0$, it is necessary for the second factor to be nonnegative as well. This argument contains two mistakes at once: firstly, the domain of the variable is extended when the radical is discarded; secondly, when the first factor is zero, the inequality is valid even when the second factor is negative. The first mistake introduces extraneous solutions, the second mistake results in a loss of solutions.

A proper solution must take into account both of these items and can be carried out as follows. The given inequality is valid in two cases: when the second factor is greater than or equal to zero and when the first factor is zero. Then, we naturally have to take only those solutions of the resulting inequality and equation which enter into the domain of the variable of the original inequality.

This domain is defined by the system of inequalities

$$4\sin^2 x-1\geqslant 0,\quad 0<\sin x<1,\quad \frac{x-5}{2x-1}>0$$

From the first two inequalities it follows that $1/2\leqslant\sin x<1$; the third is satisfied for $x<1/2$ and $x>5$.

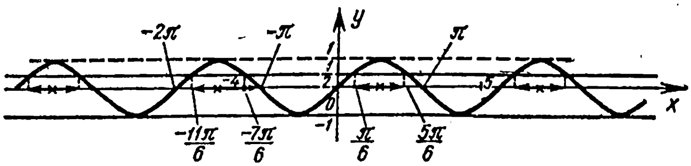
Let us begin with the inequality $\log_{\sin x}[(x-5)/(2x-1)]\geqslant 0$. Since $\sin x<1$, it follows that the inequality is equivalent, in the domain, to the inequality $(x-5)/(2x-1)\leqslant 1$, the solutions of which are $x\leqslant -4$ and $x\geqslant 1/2$. Of these values of x, only $x\leqslant -4$ and $x>5$ lie in the domain of x. It remains to take into account the inequality $1/2\leqslant\sin x<1$.

The graph shown in Fig. 28 makes it evident that the solution of this double inequality consists of the intervals

$$\frac{\pi}{6}+2n\pi \leqslant x \leqslant \frac{5\pi}{6}+2n\pi, \quad n=0, \ \pm 1, \ \pm 2, \ \ldots$$

with the points $x=\pi/2+2n\pi$ eliminated. Since we only need $x\leqslant-4$ and $x>5$, it follows (this too is found from the graph) that the values $n=0$ and $n=-1$ do not satisfy us, and there remains a portion of the

Fig. 28



interval corresponding to $n=-1: -(11\pi)/6\leqslant x\leqslant-4$ without $x=-3\pi/2$.

We now consider the equation $4\sin^2 x-1=0$, from which, taking into account that $\sin x>0$ in the domain, we get $\sin x=1/2$. However, we have just solved an inequality which is clearly satisfied by the solutions of the equation $\sin x=1/2$. Therefore all the solutions of the equation at hand have been obtained and there would be no reason to include them if (and this is yet another underwater reef of the given problem) the solutions of the double inequality $1/2\leqslant\sin x<1$ had not partially been discarded because of the conditions $x\leqslant-4$ and $x>5$. In this operation, the values $-7\pi/6$, $\pi/6$ and $5\pi/6$ were eliminated; the latter two do not enter into the domain of the original inequality, and the first one, that is, $x=-7\pi/6$, is to be adjoined to the intervals obtained above.

We thus get the answer:

$$\frac{\pi}{6}+2n\pi\leqslant x\leqslant\frac{5\pi}{6}+2n\pi, \quad x\neq\pi/2+2n\pi$$

where n is any integer except 0 and 1, $-11\pi/6\leqslant x\leqslant-4$ and also $x=-7\pi/6$.

Exercises

Solve the following inequalities.

1. $\log_x \dfrac{x+3}{x-1} > 1.$
2. $\log_{0.1}(x^2+1) < \log_{0.1}(2x-5).$
3. $\left(\tan\dfrac{\pi}{8}\right)^x-\left(\tan\dfrac{\pi}{8}\right)^{-x}<3.$
4. $\dfrac{1+\log_a^2 x}{1+\log_a x}>1, \ 0<a<1.$

5. $\dfrac{x^2+2}{x^2-1} < -2.$

6. $\cos^2 x\,(\tan x+1) > 1.$

7. $(1.25)^{1-(\log_2 x)^2} < (0.64)^{2+\log_{\sqrt{2}} x}.$

8. $(\log_2 x)^4-\left(\log_{1/2}\dfrac{x^5}{4}\right)^2-20\log_2 x+148 < 0.$

9. $\left(\dfrac{1}{2}\right)^{\log_3\log_{1/5}\left(x^2-\frac{4}{5}\right)} < 1.$

10. $\log_2(2^x-1)\cdot\log_{1/2}(2^{x+1}-2) > -2.$

11. $1-\sqrt{1-8(\log_{1/4}x)^2} < 3\log_{1/4}x.$

12. $4x^2+3^{\sqrt{x}+1}+x\cdot 3^{\sqrt{x}} < 2x^2\cdot 3^{\sqrt{x}}+2x+6.$

13. $\log_{x^2}(2+x) < 1.$

14. $4\log_{16}\cos 2x+2\log_4\sin x+\log_2\cos x+3 < 0 \quad (0 < x < \pi/4).$

15. $|x|^{x^2-x-2} < 1.$

16. $\log_{1/3}\sqrt{x+1} < \log_{1/3}\sqrt{4-x^2}+1.$

17. $\left[\dfrac{3(\sin x+\cos x)-\sqrt{2}}{2\sqrt{2}-\sin x-\cos x}\right]^{1/2} > 1.$

18. $x+4 < -\dfrac{2}{x+1}.$

19. $\dfrac{x^2-2x+3}{x^2-4x+3} > -3.$

20. $\tan\dfrac{1}{1+x^2} \geqslant 1.$

21. $\dfrac{5}{4}\sin^2 x+\dfrac{1}{4}\sin^2 2x > \cos 2x.$

22. $\cos^2 x < \dfrac{1}{2}.$

23. $\tan\dfrac{x}{2} > \dfrac{\sin x-2\cos x}{\sin x+2\cos x} \quad (0 < x < \pi).$

24. $4^x < 2^{x+1}+3.$

25. $\sqrt{1-x} \leqslant \sqrt[4]{5+x}.$

26. $\dfrac{1}{\log_2 x} \leqslant \dfrac{1}{\log_2\sqrt{x+2}}.$

27. $\log_x\dfrac{4x+5}{6-5x} < -1.$

28. $(\log_{|x+6|}2)\cdot\log_2(x^2-x-2) \geqslant 1.$

29. $\log_3\dfrac{|x^2-4x|+3}{x^2+|x-5|} \geqslant 0.$

30. $3\sin 2x > \sin x+\cos x+1.$

31. $\log_{1/2}(x+1) > \log_2(2-x).$

32. $\sqrt{3-x}-\sqrt{x+1} > \dfrac{1}{2}.$

33. $\sqrt{\log_a \frac{3-2x}{1-x}} < 1$.

34. $\log_{1/3} \sqrt{x^2-2x} > \sin \frac{11\pi}{6}$.

35. $\log_4 \frac{2x-1}{x+1} < \cos \frac{2\pi}{3}$.

36. $\log_{x+\frac{5}{2}} \left(\frac{x-5}{2x-3}\right)^2 > 0$.

37. $\log_a (\sqrt{25-x^2}-1) \geqslant \log_a (|x|+1)$.

38. $\sqrt{x-5}-\sqrt{9-x} \geqslant 1$.

39. $\log_{\tan x} \sqrt{\sin^2 x-\frac{5}{12}} < -1$.

40. $\log_{\sqrt{2x^2-7x+6}} \left(\frac{x}{3}\right) > 0$.

41. $2\cos 2x+\sin 2x > \tan x$.

42. $|3^{\tan \pi x}-3^{1-\tan \pi x}| \geqslant 2$.

43. $(\log_{\sin x} 2)^2 < \log_{\sin x} (4\sin^3 x)$.

44. $\log_4 (2x^2+x+1)-\log_2 (2x-1) \leqslant -\tan \frac{7\pi}{4}$.

45. $\log_{\frac{2x+1}{x^2-4}} 2 \leqslant \frac{1}{2} \log_{\sin (\pi/3)} \frac{4}{3}$.

46. $\sqrt{\tan x-1}\ [\log_{\tan x} (2+4\cos^2 x)-2] \geqslant 0$.

47. $\log_{(3x^2+1)} 2 < 1/2$.

48. $x^{\log_{10} \sin x} \geqslant 1\ (x > 0)$.

49. $\log_{5/8} (2x^2-x-3/8) \geqslant 1$.

50. $-9\sqrt[4]{x}+\sqrt{x}+18 \geqslant 0$.

51. $\log_{\sin x-\cos x} (\sin x-5\cos x) \geqslant 1$.

52. $\cot (5+3x)\cdot(\cot 5+\cot 3x) \geqslant \sqrt{\cot 3x} - 1$.

53. $\frac{\sqrt{24-2x-x^2}}{x} < 1$.

54. $x^2-|3x+2|+x \geqslant 0$.

55. $25^x-2^{2\log_2 6-1} < 10\cdot 5^{x-1}$.

56. $\sqrt{2-\sqrt{3+x}} < \sqrt{4+x}$.

57. $\log_{|\sin x|} (x^2-8x+23) > \frac{3}{\log_2 |\sin x|}$.

58. $\log_{x/2} 8+\log_{x/4} 8 < \frac{\log_2 x^4}{\log_2 x^2-4}$.

59. $2\tan 2x \leqslant 3\tan x$.

60. $\sqrt{5-2\sin x} \geqslant 6\sin x-1$.

1.11 Systems of equations

In the solution of systems of equations, as in the solution of single equations, equivalence and the associated notions discussed in detail in Sec. 1.9 play an important role. Sec. 1.9 however discusses only such questions as apply to single equations. The point is that despite the total logical similarity of theoretical reasoning pertaining to the equivalence of equations and the equivalence of systems of equations, practical application in the case of systems (simultaneous equations) involves much greater difficulties than in the case of single equations. For this reason, we will not, as a rule, make use of this concept.

The special difficulties that occur when solving systems of equations are of course connected solely with the transformations of the system which involve several equations (in everything that pertains to a single equation, we can of course make full use of the notions and recommendations given in Sec. 1.9). There are many such manipulations, as every student knows. It is easy to recall the great variety of ingenious devices that one often has to employ in the solving of systems of equations.

For this reason, in solving systems of equations the student ordinarily applies one of the two approaches indicated in Sec. 1.9: *the deriving of consequences that are not necessarily equivalent to the given system, and the subsequent discarding of extraneous roots.* In this procedure we do not even pass to systems which are consequences of the original system, but to separate equations, each one of which is a consequence of the original system, that is, such as is satisfied by any solution of the system. By combining these equations, we obtain systems that are consequences of the original system, and finally we get certain sets of unknowns. Then by means of a check (direct substitution, as a rule), we discard any extraneous solutions.

To do all this, it is of course necessary to be able to avoid transformations that would lead to a loss of solutions (such as, say, the division of both members of an equation by both members of another equation), for it is clear that in the process the solutions of the system for which both sides of the second equation vanish will be lost. On the other hand, most of the frequently employed transformations such as addition, subtraction, multiplication of the equations, substitution of the unknown of one equation into another one, and the like, cannot result in a loss of any solutions and are therefore permissible. Ordinarily one has to examine each case as to whether or not a more or less involved transformation can result in a loss of solutions.

Direct verification of the final solutions can undoubtedly present considerable difficulties in certain cases. To avoid them, one can, as a rule, take advantage of the suggestions given for the solution of single equations. The following problem will illustrate this point.

1. *Find all complex solutions of the system*

$$\frac{2x^2}{1+x^2}=y$$
$$\frac{2y^2}{1+y^2}=x \qquad (1)$$

It will be noticed that by inverting the fraction in the left members of the two equations and by replacing the right members by $1/y$ and $1/x$ respectively, we obtain a simple system in $u=1/x$ and $v=1/y$:

$$\frac{u^2}{2}+\frac{1}{2}=v$$
$$\frac{v^2}{2}+\frac{1}{2}=u \qquad (2)$$

But is such inverting of the equation permissible? How do the solutions fare in that case? It is easy to see that this manipulation does not introduce any extraneous solutions, but it can result in a loss of solutions. Solutions in which both members of the equation are zero will be lost. Thus, in order to avoid any loss of solutions, this possibility must be investigated prior to passing to system (2).

The left member of the first equation vanishes when $x=0$, and the right member vanishes when $y=0$. Hence, the only solution that can be lost (and indeed is lost) is $x=0$, $y=0$. Therefore, having found the solution $x=0$, $y=0$, we can find the remaining solutions from system (2).

Subtracting the second equation of (2) from the first, we get the equation $u^2-v^2=2(v-u)$, whence either $u-v=0$ or $u+v=-2$.

In the first case, substituting $v=u$ into the first equation of system (2), we get $u=1$. Direct substitution convinces us that the pair $u_1=1$, $v_1=1$ satisfies the second equation as well, and, hence, system (2).

In the second case, substituting $v=-u-2$ into the first equation, we get $u^2+2u+5=0$, whence $u_{2,3}=-1\pm 2i$ and $v_{2,3}=-1\mp 2i$. Direct substitution of the values $u_{2,3}$ and $v_{2,3}$ into the second equation results in two equations:

$$(-1-2i)^2+1=-2+4i,\ (-1+2i)^2+1=-2-4i$$

Their proof only requires removing the brackets.

Thus, the system of equations (2) has the solution set

$$u_1=1,\ v_1=1,\ u_2=-1+2i,\ v_2=-1-2i,$$
$$u_3=-1-2i,\ v_3=-1+2i$$

Recalling that $x=1/u$ and $y=1/v$, we obtain the solution set of the original system of equations (1):

$$x_1=1,\ y_1=1,\ x_2=\frac{1}{-1+2i},\ y_2=-\frac{1}{1+2i},$$
$$x_3=-\frac{1}{1+2i},\ y_3=\frac{1}{-1+2i}$$

Aside from these solutions we must not forget the earlier found solution, $x_4=0$, $y_4=0$.*

We now take up the question of investigating a *system of two linear equations in two unknowns*. Such systems confront the student with serious logical difficulties. And as usual the first stumbling block is the definitions. *What is a system of two linear equations in two unknowns? What is the solution set of such a system? In what case is such a system termed determinate (indeterminate, consistent, inconsistent)?* For many, these questions are insurmountable. Very often the student gives a correct general definition but gets entangled as soon as he is asked to state *whether, say, the following systems are systems of linear equations in two unknowns*:

$$\begin{array}{llll} x+y=1 & 2x-y=1 & x=0 & 0\cdot x+0\cdot y=0 \\ x+y=1 & 2x-y=2 & y=1 & 0\cdot x+0\cdot y=0 \end{array}$$

All these systems fit the general definition. The difficulty stems from the fact that they are "too simple". It is immediately evident that the first system is indeterminate, the second is inconsistent, the third clearly has the unique solution $x=0$, $y=1$, and the solution set of the fourth system of equations consists of any pair of numbers x, y.

In studying this topic, it is often necessary first of all to realize under what restrictions on the coefficients can a system of two linear equations in two unknowns be investigated. And most important, pay special attention to the fact that after the manipulations performed on the original system, say,

$$\begin{aligned} a_1x+b_1y&=c_1 \\ a_2x+b_2y&=c_2 \end{aligned}$$

we obtain a system like

$$\begin{aligned} (a_1b_2-a_2b_1)\,x&=c_1b_2-c_2b_1 \quad (=\Delta_1) \\ (a_1b_2-a_2b_1)\,y&=a_1c_2-a_2c_1 \quad (=\Delta_2) \end{aligned} \tag{3}$$

which is a consequence of the original system but in the general case is not necessarily an equivalent system, so that in passing to it we do not lose any solutions but we may **introduce** some extraneous solutions.

If $\Delta=a_1b_2-a_2b_1\neq0$, then this system is clearly determinate, that is, it has the unique solution $x=\Delta_1/\Delta$, $y=\Delta_2/\Delta$. **Direct substitution** (verification) makes it clear that this solution is the solution of the original system as well.

If $\Delta=0$ but $\Delta_1\neq0$ or $\Delta_2\neq0$, then this system is inconsistent, which means it has no solutions. Hence, the original system of equations does not have any solutions.

* It is advisable to reduce to real-imaginary form the complex numbers x_2, y_2, x_3 and y_3: $x_2=(1+2i)/3$, etc.

Finally, if $\Delta=0$ and $\Delta_1=0$, then in this case $\Delta_2=0$ as well so that system (1) is of the form

$$0\cdot x=0$$
$$0\cdot y=0$$

Consequently, the solution set of (1) is obviously any pair of numbers x, y. But this does not yet mean that the original system of equations is satisfied by any pair of numbers because we might introduce extraneous solutions. That actually is the case: when $\Delta=\Delta_1=\Delta_2=0$, it can readily be shown that the following equation holds true for the original system:

$$\frac{a_1}{a_2}=\frac{b_1}{b_2}=\frac{c_1}{c_2}$$

Consequently, the first and second equations of the original system differ solely in the coefficient; for this reason the original system is equivalent to a single equation. But then it is evident that it is indeterminate, that is, it has an infinity of solutions: by taking arbitrary values of one unknown, we can use the equation to obtain corresponding values of the other unknown.

An investigation of this kind is not always sufficient. We now give the necessary definitions and will consider a complete investigation of a system of two linear equations in two unknowns with arbitrary coefficients.

Definition 1. *A system of two linear equations in two unknowns is a system of two equations of the form*

$$\begin{aligned} a_1x+b_1y&=c_1 \\ a_2x+b_2y&=c_2 \end{aligned} \tag{4}$$

*where a_1, a_2, b_1, b_2, c_1, c_2 are arbitrary real numbers.**

Definition 2. *The solution set of the system is a pair of real numbers (or complex numbers, if the coefficients are complex) x_0, y_0 which satisfy each of the equations of the system.*

In other words, the solution is a collection of two numbers x_0, y_0 such that substitution into the original system (4) of x_0 in place of the unknown x and of y_0 in place of the unknown y results in two true numerical equations:

$$a_1x_0+b_1y_0=c_1$$
$$a_2x_0+b_2y_0=c_2$$

It is thus necessary to distinguish two meanings with which we invest the words *solution of a system (of equations)*: the solution of a system is to be understood as the **process** of seeking the values of the unknowns and also any **collection** of values of the unknowns which

* There is nothing to prevent considering linear systems with complex coefficients. All subsequent investigation (except the geometric interpretation) remains valid in this more general case as well.

transform the equations of the system into true numerical equations. *To solve a system of equations means to find all the solutions (the solution set).**

Only a misunderstanding of the term "solution of a system" can account for the nonsense that is sometimes heard at examinations that goes like this: "The system

$$\begin{aligned} 2x+3y&=13 \\ -x+y&=1 \end{aligned} \tag{5}$$

has two solutions: $x=2$ and $y=3$." It is precisely the collection (set) of these two numbers $x=2$ and $y=3$ that constitutes the solution of the system (the solution set). These values of the unknowns are sometimes called the "roots of system (5)", which is wrong too because the term "root" is only applicable to a single equation in one unknown and is not used in the study of systems of equations. Finally, it is not proper to say that the "numbers 2 and 3 constitute the solution of system (5)" since the solution set consists of a pair of values of the unknowns and it is essential to point out which unknown (x or y) is equal to 2 and which is equal to 3 (the solution set is in the form of an ordered pair). The correct answer to the problem of seeking the solutions to system (5) should read thus: "System (5) has the unique solution $x=2$, $y=3$" or "the solution set of system (5) is the ordered pair $x=2$, $y=3$."

Definition 3. *A system of equations is termed:*
consistent if it has at least one solution;
inconsistent if it does not have any solutions;
determinate if it has a unique solution;
indeterminate if it has more than one solution.

In other words, systems of equations may be consistent and inconsistent; consistent systems in turn break down into determinate systems and indeterminate systems. Illustrative examples of each type follow.

The system

$$\begin{aligned} 0\cdot x+0\cdot y&=1 \\ x+y&=2 \end{aligned}$$

is inconsistent because for no values of x and y can the first equation become a true numerical equation. The system

$$\begin{aligned} 2x-y&=1 \\ 2x-y&=2 \end{aligned}$$

is inconsistent; although there are ordered pairs of numbers (say $x=0$, $y=-1$, etc.) which convert the first equation into a true numerical equation and also number pairs (say $x=1$, $y=0$, etc.) which convert

---

* We stress the fact that Definition 2 holds for arbitrary systems of equations. The remarks made here with respect to solutions of systems of equations apply to the solution of any system of equations.

the second equation into a true numerical equation, there is not a single pair which simultaneously makes both equations true numerical equations. The system

$$\begin{matrix} x=0 \\ x=2 \end{matrix} \quad \text{or} \quad \begin{matrix} x+0\cdot y=0 \\ x+0\cdot y=2 \end{matrix}$$

is likewise inconsistent.

The system

$$\begin{matrix} x=0 \\ y=1 \end{matrix} \quad \text{or} \quad \begin{matrix} x+0\cdot y=0 \\ 0\cdot x+y=1 \end{matrix}$$

is consistent; more precisely, it is determinate, since there is only one pair, $x=0$, $y=1$, that makes each one of the equations a true numerical equation. The system

$$\begin{matrix} 2x+3y=13 \\ -x+y=1 \end{matrix}$$

is consistent since it has the solution set $x=2$, $y=3$; more precisely, it is determinate (this requires supplementary substantiation that is not obvious).

The system

$$\begin{matrix} x+y=1 \\ x+y=1 \end{matrix}$$

is consistent because it has, for example, the solution $x_1=0$, $y_1=1$, and, moreover, it is indeterminate, since, for example, it has $x_2=1$, $y_2=0$ as another solution.

The system

$$\begin{matrix} 0\cdot x+0\cdot y=0 \\ 0\cdot x+0\cdot y=0 \end{matrix}$$

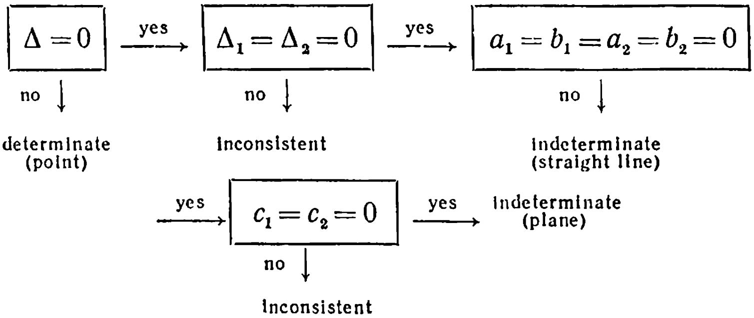
is also indeterminate since any number pair x, y converts both equations into true numerical equations.

Now let us investigate system (4). *To investigate a linear system means to decide whether it is inconsistent, determinate or indeterminate. In the two latter instances, it is necessary to find the solutions.* We now prove a theorem which amounts to an investigation of systems of two linear equations in two unknowns. We state the theorem rather unusually in the form of a diagram.

Theorem. *Given a system* (4). *We introduce the following notation:*

$$\Delta=a_1b_2-a_2b_1, \quad \Delta_1=c_1b_2-c_2b_1, \quad \Delta_2=a_1c_2-a_2c_1$$

Then



This diagram is to be understood as follows. We start with the first question: Does the condition $\Delta=0$ hold true? If it does not, the appropriate arrow indicates that the system is determinate. If it does, a different arrow states the next question: Does the condition $\Delta_1=\Delta_2=0$ hold? If it does not, the appropriate arrow shows that the system is inconsistent. If it does, a different arrow poses the next question; and so forth. The words "straight line" and "plane" in parentheses will be explained later on.

This diagrammatic statement of the theorem is more pictorial than the customary verbal formulation which follows.

If $\Delta \neq 0$, then the system is determinate and it has the solution $x=$ $=\Delta_1/\Delta$, $y=\Delta_2/\Delta$. If $\Delta=0$, but at least one of the numbers Δ_1, Δ_2 is nonzero, then the system is inconsistent. If $\Delta=0$ and $\Delta_1=\Delta_2=0$ but at least one of the numbers a_1, b_1, a_2, b_2 is nonzero, then the system is indeterminate. If $\Delta=\Delta_1=\Delta_2=0$ and all numbers a_1, b_1, a_2, b_2 are zero, but at least one of the numbers c_1, c_2 is nonzero, then the system is inconsistent. Finally, if all coefficients of the system are zero, the system is indeterminate.*

The awkwardness of this verbal formulation is obvious.

Proof. Consider system (4). Multiplying the first equation of the system by b_2 and subtracting b_1 times the second equation from the result yields

$$(a_1b_2-a_2b_1)\,x=c_1b_2-c_2b_1$$

Similarly, multiplying the second equation by a_1 and subtracting a_2 times the first equation from the result yields

$$(a_1b_2-a_2b_1)\,y=a_1c_2-a_2c_1$$

Thus, we find that system (3) must hold. This system can be rewritten as

$$\begin{aligned}\Delta\cdot x&=\Delta_1\\ \Delta\cdot y&=\Delta_2\end{aligned} \tag{6}$$

* It is easy to see that the condition $a_1=b_1=a_2=b_2$ implies $\Delta=\Delta_1=\Delta_2=0$.

System (6) is a consequence of the original system (4), which means that *every solution of* (4) *is a solution of* (6). We do not however assert that these systems are equivalent.

Suppose that $\Delta \neq 0$. Then (6) has the unique solution $x = \frac{\Delta_1}{\Delta}$, $y = \frac{\Delta_2}{\Delta}$. Direct substitution (**verification**) is still required to be assured that this number pair x, y is also the solution of the original system (4). Thus, in the case at hand, the original system of equations has a unique solution. This proves the first vertical arrow.

Taking a step rightwards, we consider the case $\Delta = 0$. Suppose that the condition $\Delta_1 = \Delta_2 = 0$ does not hold, that is, at least one of the numbers Δ_1, Δ_2 is nonzero. Then system (6) is inconsistent, but in that case so is system (4). This proves the legitimacy of the second vertical arrow.

Taking yet another step rightwards, we consider the case $\Delta = \Delta_1 = \Delta_2 = 0$. Suppose that the condition $a_1 = b_1 = a_2 = b_2 = 0$ is not valid, that is, one of the numbers a_1, b_1, a_2, b_2 is nonzero. For the sake of definiteness, let $a_1 \neq 0$. Then from the first equation of system (4) we have $x = \frac{c_1 - b_1 y}{a_1}$. Substituting this value into the second equation of the system, we get

$$a_2 x + b_2 y = \frac{a_2 c_1 - a_2 b_1 y}{a_1} + b_2 y = \frac{a_2 c_1 + (a_1 b_2 - a_2 b_1) y}{a_1} = \frac{a_2 c_1}{a_1} = c_2$$

(Here we made use of the equations $a_1 b_2 - a_2 b_1 = \Delta = 0$ and $a_2 c_1 = a_1 c_2$, the latter being a consequence of the condition $\Delta_2 = 0$.) Thus, every solution of the first equation is a solution of the second equation and, hence, system (4) reduces to the first equation. But the first equation is satisfied if we give the unknown y an arbitrary value and find the corresponding value of x from the formula $x = \frac{c_1 - b_1 y}{a_1}$. Thus, in the case at hand, system (4) has an infinity of solutions, all of them being given by the formula

$$x = \frac{c_1 - b_1 y}{a_1} \quad (y \text{ an arbitrary number}) .$$

Consequently, system (4) is indeterminate. This proves the legitimacy of the third vertical arrow.

Moving one step rightwards, we consider the case when all numbers a_1, b_1, a_2, b_2 are zero (then all the more so, $\Delta = \Delta_1 = \Delta_2 = 0$). In this case, system (4) has the altogether strange aspect

$$\begin{aligned} 0 \cdot x + 0 \cdot y &= c_1 \\ 0 \cdot x + 0 \cdot y &= c_2 \end{aligned}$$

Clearly, if $c_1 = c_2 = 0$, then it is satisfied by any pair of numbers x, y, which means the system is indeterminate and if at least one of the numbers c_1, c_2 is nonzero, then the system is inconsistent. This proves the legitimacy of the last two arrows. The proof of the theorem is complete.

Now let us consider a **geometric interpretation** of the systems examined above. We know that the set of points in a plane whose coordinates satisfy an equation of the form

$$ax+by=c$$

where either a or b (or both) are nonzero, constitute a straight line. And so to solve a system like (4) in which every equation has at least one unknown means to find the common point of two straight lines. Geometry tells us all the possible cases of the mutual positions of two straight lines in a plane. We then have the following correspondence:

In geometric terms:	*In algebraic terms:*
the lines intersect	the system is determinate (first vertical arrow)
the lines are parallel	the system is inconsistent (second vertical arrow)
the lines are coincident	the system is indeterminate (third vertical arrow)

The remaining two arrows cannot thus be interpreted since in those cases the system is devoid of any geometric meaning. However, in the last case, when any pair of numbers is a solution of the system, we can say that the whole plane constitutes the solution set of the system.

To illustrate, let us *investigate the system*

$$ax+y=a$$
$$x+ay=1$$

First, compute Δ, Δ_1, Δ_2:

$$\Delta=a^2-1, \quad \Delta_1=a^2-1, \quad \Delta_2=0$$

Now verify the starting condition: $\Delta=0$. If $\Delta=a^2-1\neq 0$, that is if $a\neq 1$ and $a\neq -1$, then the system is determinate (first vertical arrow). But if $\Delta=0$ (that is, $a=1$ or $a=-1$), then we come to the condition $\Delta_1=\Delta_2=0$ (via the horizontal arrow). This condition is valid here and so we follow the second horizontal arrow. Since not all the coefficients of the unknowns here are zero, the system is indeterminate for $a=1$ and $a=-1$ and reduces to a single equation.

Thus, if $a\neq 1$ and $a\neq -1$, the system is determinate and the solution is $x=1$, $y=0$; when $a=1$ and $a=-1$, the system is indeterminate. In these cases, the solutions are given by the formulas

$$y=1-x \quad \text{for } a=1 \quad (x \text{ any number})$$
$$y=x-1 \quad \text{for } a=-1 \quad (x \text{ any number})$$

Here is another example. Let us *investigate the system*

$$x \sin 2\alpha + y(1+\cos 2\alpha) = \sin 2\alpha$$
$$x(1+\cos 2\alpha) - y \sin 2\alpha = 0$$

We have

$$\Delta = -\sin^2 2\alpha - (1+\cos 2\alpha)^2 = -2(1+\cos 2\alpha),$$
$$\Delta_1 = -\sin^2 2\alpha, \quad \Delta_2 = -\sin 2\alpha(1+\cos 2\alpha)$$

If $\Delta \neq 0$, that is, $\cos 2\alpha \neq -1$, then the system is determinate. But if $\Delta = 0$, i.e. $\cos 2\alpha = -1$, then $\sin 2\alpha = 0$, and hence, $\Delta_1 = \Delta_2 = 0$. Besides, in this case all six coefficients of the system are equal to zero, and therefore any pair of numbers x and y can be a solution.

Thus, if $\cos 2\alpha \neq -1$, i.e., $\alpha \neq \frac{(2k+1)\pi}{2}$, then, no matter what the integer k, the given system is determinate and the solution in this case is

$$x = \frac{\Delta_1}{\Delta} = \sin^2 \alpha, \quad y = \frac{\Delta_2}{\Delta} = \sin \alpha \cos \alpha$$

When $\alpha = \frac{(2k+1)\pi}{2}$, where k is an integer, the system is indeterminate and its solution is any pair of numbers x and y.

2. *What complex numbers $a+bi$ can be represented in the form $(c+di)/(c-di)$, where c and d are real numbers?*

This problem may be rephrased as follows: *for what values of a and b does the equation $a+bi = (c+di)/(c-di)$, in which c and d are unknowns, have a solution*, or *for which values of a and b does the equation $(a+bi) \times (c-di) = c+di$ have a nonzero solution?*

Making the necessary transformations on the basis of the definition of the equality of complex numbers, we obtain a system of two linear equations:

$$\begin{aligned} (a-1)c + bd &= 0 \\ bc - (a+1)d &= 0 \end{aligned} \qquad (7)$$

It is a system in two unknowns c and d and involves two parameters a and b. To solve this system in the ordinary way is rather involved because it requires considering a large number of cases. For instance, in order to express one unknown in terms of the other, we have to consider separately the cases when the corresponding coefficient is zero or nonzero.

Now there is no necessity to solve system (7) since we are only interested in whether it has nonzero solutions or not. It is quite obvious however that it has the solution $c=0$, $d=0$, and so we want it to have more than one solution, so that it will be *indeterminate*.

According to the theorem, this occurs if and only if $\Delta = -(a^2-1) - b^2 = 0$. Then, in the diagram, we are on the third vertical arrow ($\Delta_1 = \Delta_2 = 0$, and the coefficients of the unknowns cannot all be zero because $a-1$ and $-(a+1)$ do not vanish simultaneously).

Thus, this representation is only possible for numbers $a+bi$ in which $a^2+b^2=1$ (that is, complex numbers with modulus unity and only such numbers).

Exercises

1. $\sqrt{x+y}+\sqrt{2x+4y}=\sqrt{2}+4$
   $\sqrt{x+2y}-\sqrt{2x+2y}=2\sqrt{2}-2.$
2. $\sqrt[x-y]{x+y}=2\sqrt{3}$
   $(x+y)\,2^{y-x}=3.$
3. $x^{\sqrt[4]{x}+\sqrt{y}}=y^{8/3}$
   $y^{\sqrt[4]{x}+\sqrt{y}}=x^{2/3}.$
4. $\log_2 x+\log_4 y+\log_4 z=2$
   $\log_3 y+\log_9 z+\log_9 x=2$
   $\log_4 z+\log_{16} x+\log_{16} y=2.$
5. $x^{y^2-15y+56}=1$
   $y-x=5.$
6. $x^2=1+6\log_4 y$
   $y^2=y\cdot 2^x+2^{2x+1}$
7. $x^2+4y^3=96$
   $\log_{y^2} 2=\log_{xy} 4.$
8. $(2^x+1)\,2^{y+1}=9$
   $\sqrt{x+y^2}=x+y.$
9. $\log_9(x^2+2)+\log_{81}(y^2+9)=2$
   $2\log_4(x+y)-\log_2(x-y)=0.$
10. $\log_9(x^2+1)-\log_3(y-2)=0$
    $\log_2(x^2-2y^2+10y-7)=2.$
11. $\dfrac{1+x^2+xy}{x+y}=2-y$
    $\log_{2^{1-y}} 2^{x^2}=1+y.$
12. $\sqrt[x-y]{x+y}=\dfrac{\sqrt{52-2x}}{\sqrt[4]{x-y}}$
    $\dfrac{3}{2}\log_8(x-y)-\log_{1/\sqrt{2}}(x-y)=5.$

13. Can a system of two linear equations in two unknowns have exactly two solutions?

14. For which values of k does a solution of the system

$$x+ky=3$$
$$kx+4y=6$$

exist and satisfy the inequalities $x>1$, $y>0$?

15. For which values of m does the system

$$x+my=1$$
$$mx-3my=2m+3$$

fail to have any solution?

16. For which values of α is any pair of numbers x, y, which is a solution of the system of equations

$$x\sin 2\alpha+y(1+\cos 2\alpha)=\sin 2\alpha$$
$$x(1+\cos 2\alpha)-y\sin 2\alpha=0,$$

simultaneously a solution of the system of equations

$$x\sin\alpha+y\cos\alpha=\sin\alpha$$
$$x\cos\alpha-y\sin\alpha=0?$$

1.12 Verbal problems

We use the term verbal, or story, problems for those which traditionally are called problems leading to equations. Examination problems nowadays frequently require the use not only of equations but of inequalities* as well, and sometimes of other conditions which are not written in the form of equations and inequalities. The most common feature of problems of this type is that the condition is given in verbal form without any formulas or even without any literal designations of the unknowns. The habit of most students to regard any verbal problem as a problem which involves the setting up of equations is sometimes a drawback in that they are psychologically unprepared when it turns out that the equations alone are not sufficient to obtain a solution.

The ordinary type of problem in which all the conditions can be written down in the form of equations do not, as a rule, cause any particular difficulty, though even in these problems there are certain items that occasionally are troublesome. Usually the real difficulty of more complicated problems lies in their unusual aspect, the necessity to reason and not merely to solve certain systems of equations or inequalities.

It frequently happens that simple arguments without any setting up of equations and inequalities, even if that is possible, proves to be faster and simpler. What is more, an occasional problem is more amenable to simple, everyday reasoning than to ordinary mathematical techniques. Incidentally, the "commonsense" solution is not always rigorous and so one has to supplement it with rigorous mathematical arguments.

* Note that, practically speaking, inequalities occur in nearly all problems of this kind. For instance, if, say, s is the distance, then $s>0$. However, they are not ordinarily written out explicitly, but are used in the solving of equations and in discarding extraneous solutions.

Let us begin with some *mixture problems*. The student frequently finds difficulty in setting up the equations in these problems.

1. *Given three mixtures consisting of three components A, B and C. The first mixture only contains components A and B in the weight ratio* 3 : 5, *the second mixture contains components B and C in the weight ratio of* 1 : 2, *and the third mixture contains only the components A and C in the weight ratio* 2 : 3. *In what ratio must we take these mixtures so that the resulting mixture of components A, B and C stand in the weight ratio* 3 : 5 : 2?

Some students have found the phrase "weight ratio" difficult to grasp, others fear the word "mixture". Actually, the problem is not at all difficult.

Since components A and B are in the ratio 3 : 5 by weight in the first mixture, each gram of the first mixture contains 3/8 gram of component A and 5/8 gram of component B. Similarly, 1 gram of the second mixture contains 1/3 gram of B and 2/3 gram of C, and 1 gram of the third mixture contains 2/5 gram of A and 3/5 gram of C.

If we take x grams of the first mixture, y grams of the second, and z grams of the third and mix them, we get $(x+y+z)$ grams of the new mixture, which will contain $(\frac{3}{8}x+\frac{2}{5}z)$ grams of A, $(\frac{5}{8}x+\frac{1}{3}y)$ grams of B and $(\frac{2}{3}y+\frac{3}{5}z)$ grams of C. We have to take the first, second and third mixtures in quantities such that in the new mixture the components A, B and C are in the weight ratio 3 : 5 : 2, that is, in one gram of the new mixture there must be 3/10 gram of A, 5/10 gram of B and 2/10 gram of C. But then $x+y+z$ grams of the new mixture will contain $\frac{3(x+y+z)}{10}$ grams of A, $\frac{5(x+y+z)}{10}$ grams of B and $\frac{2(x+y+z)}{10}$ grams of C. Equating the different expressions for one and the same quantity of components A, B, C, we get the following system of equations:

$$\begin{aligned}\frac{3}{8}x+\frac{2}{5}z&=\frac{3}{10}(x+y+z)\\ \frac{5}{8}x+\frac{1}{3}y&=\frac{5}{10}(x+y+z)\\ \frac{2}{3}y+\frac{3}{5}z&=\frac{2}{10}(x+y+z)\end{aligned}\tag{1}$$

Note from the start that although we have three equations in three variables, there are only two independent equations in this system. This can easily be shown, for example, by subtracting from the equation $x+y+z=x+y+z$ the sum of the first two equations; we get the third equation. For this reason, from system (1) we find only the ratio $x:y:z$, and not x, y and z themselves. For instance, eliminating x from the first two equations of (1), we find that $y=2z$. Substituting this value of y into any equation of the system, we get $x=(20/3)z$.

Consequently, $x:y:z=20:6:3$, which means the mixtures have to be taken in the weight ratio 20:6:3.

Percentage problems represent another difficult type of problem for students. Yet there is nothing hard about the notion of a percent. We can get rid of percentages simply by considering hundredths of a number. The following problem is one involving both mixtures and percentage.

2. *The percentages (by weight) of alcohol in three solutions form a geometric progression. If we mix the first, second and third solutions in the weight ratio of* 2:3:4, *we obtain a solution containing* 32% *alcohol. If we mix them in the weight ratio* 3:2:1, *we obtain a solution containing* 22 % *alcohol. What is the percentage of alcohol in each solution?*

Let there be x% alcohol in the first solution, y% in the second, and z% in the third. This means that 1 gram of the first solution contains $x/100$ gram of alcohol, 1 gram of the second solution, $y/100$ gram of alcohol, and 1 gram of the third solution, $z/100$ gram of alcohol. If we take 2 grams of the first solution, 3 grams of the second, and 4 grams of the third, we get 9 grams of a mixture containing $\left(2\cdot\frac{x}{100}+3\cdot\frac{y}{100}+4\cdot\frac{z}{100}\right)$ grams of alcohol. By the statement of the problem, the resulting mixture contains 32% alcohol, which means that 9 grams of the mixture contains $9\cdot\frac{32}{100}$ grams of alcohol. From this condition we get the equation

$$\frac{2x+3y+4z}{100}=\frac{9\cdot 32}{100}$$

In similar fashion, we get yet another equation:

$$\frac{3x+2y+z}{100}=\frac{6\cdot 22}{100}$$

Finally, by hypothesis, the numbers x, y, z form a geometric progression, and so $y^2=xz$.

It now remains to solve the system of equations

$$\begin{aligned}2x+3y+4z&=288\\3x+2y+z&=132\\y^2&=xz\end{aligned}$$

Solving the first two equations for y and z and substituting the resulting expressions into the third equation, we get the equation $x^2-76x+768=0$ with roots $x_1=64$ and $x_2=12$.

But the value $x_1=64$ does not satisfy the conditions of the problem since the corresponding value of $y=48-2x$ is negative. Hence, there remains only $x=12$, from which we easily find $y=24$ and $z=48$. Thus, the first solution contains 12% alcohol, the second 24%, and the third 48%.

In many cases difficulties arise in the solution of the systems, particularly when finding the unknown requires a certain amount of guesswork or an artificial device. Such techniques often simplify computations or even suggest the only way to solve the problem.

3. *A tributary flows into a river. On the tributary, at a certain distance from the mouth of the tributary, lies point A. On the river, point B lies at the same distance from the mouth of the tributary. The time required for a motorboat to cover the distance from A to the mouth of the tributary and back stands in the ratio of* 32 to 35 *to the time required to go from B to the mouth of the tributary and back. If the rate of the motorboat were* 2 km/hr *more, then this ratio would be* 15:16, *and if the rate of the motorboat were* 2 km/hr *less, then the ratio would be* 7:8. *Find the rate of flow of the river (distances are measured along the tributary and river, respectively).*

Let the rate of river flow be u km/hr, the rate of the motorboat in still water v km/hr and the rate of flow of the tributary, w km/hr. Furthermore, let the distance from A to the mouth of the tributary be equal to s km. Then the time the boat takes to cover the distance from A to the mouth of the tributary and back is $t_1 = s/(v+w) + s/(v-w) = 2sv/(v^2 - w^2)$ (hrs.).

Since the distance from B to the mouth of the tributary is also s km, the time required for the boat to go from B to the mouth of the tributary and back is $t_2 = s/(v+u) + s/(v-u) = 2sv/(v^2 - u^2)$ (hrs.). From the statement of the problem, $t_1 : t_2 = 32 : 35$ and we get the first equation

$$\frac{v^2 - u^2}{v^2 - w^2} = \frac{32}{35}$$

Two more equations are set up in similar fashion:

$$\frac{(v+2)^2 - u^2}{(v+2)^2 - w^2} = \frac{15}{16}$$

$$\frac{(v-2)^2 - u^2}{(v-2)^2 - w^2} = \frac{7}{8}$$

Simplifying we get the following system of equations:

$$3v^2 = 35u^2 - 32w^2$$
$$(v+2)^2 = 16u^2 - 15w^2$$
$$(v-2)^2 = 8u^2 - 7w^2$$

We have to find u. This is best done by first eliminating u, which is just the unknown we are seeking. Eliminating u, we obtain the system

$$2(v-2)^2 - (v+2)^2 = w^2$$
$$35(v-2)^2 - 24v^2 = 11w^2$$

Now eliminating w, we get the equation $13(v-2)^2 + 11(v+2)^2 - 24v^2 = 0$, whence $v = 12$. Now it is easy to find $w = 2$ and, finally,

$u = 4$. We thus have the answer: the rate of flow of the river is equal to 4 km/hr.

The next problem involves a system of three linear equations in three unknowns. This would appear to be easy to solve, but at the examinations many students failed because they got entangled in the manipulations with the *literal coefficients*. Problems involving literal data come up rather often at examinations and the student should be able to handle them.

4. *Two rivers flow into a lake. A steamship leaves port M on the first river, steams downstream to the lake then across the lake (still water) and up the second river (upstream) to port N, and then makes the return trip. The ship has a speed of v in still water, the rate of flow of the first river is v_1, that of the second river v_2, the transit time between M and N is t, the distance from M to N is S. The return-trip time from N to M along the same route is equal to t. What distance does the steamship cover across the lake in one direction?*

Denote by s_1 and s_2 respectively the distances from ports M and N to the lake, and by s the distance over the lake. By the statement of the problem, we have $s_1 + s + s_2 = S$. It is then easy to see that the time required for the steamship to cover the distance from M to N is

$$\frac{s_1}{v+v_1} + \frac{s}{v} + \frac{s_2}{v-v_2} = t$$

The time required for the return trip is figured similarly. We thus obtain a system of three equations in three unknowns s_1, s_2, s:

$$\begin{gathered} s_1 + s + s_2 = S \\ \frac{s_1}{v+v_1} + \frac{s}{v} + \frac{s_2}{v-v_2} = t \\ \frac{s_1}{v-v_1} + \frac{s}{v} + \frac{s_2}{v+v_2} = t \end{gathered} \tag{2}$$

We are interested in s.

This system looks rather imposing, though actually there is nothing really complicated in it, particularly if we recall that v, v_1, v_2, S, t are given constants; it is quite obvious that the system (2) is a system of three equations in three unknowns, which of course can always be solved, say, by successive elimination of the unknowns.

However, it often happens that what is simple in theory turns out to be extremely awkward in practice. In this problem, such an approach would be unwieldy in the extreme, involving cumbersome manipulations merely because the coefficients of system (2) are rather complicated.

We will solve (2) in a somewhat artificial manner, but one which is shorter. The second equation of the system can be rewritten as

$$v^2 s_1 - v v_2 s_1 + v^2 s + (v_1 - v_2) v s - v_1 v_2 s + v^2 s_2 + v v_1 s_2 \\ = tv(v^2 + v v_1 - v v_2 - v_1 v_2)$$

Replacing the sum $v^2s_1+v^2s+v^2s_2$ in the left member by v^2S on the basis of the first equation, and collecting terms, we get the equation

$$v^2S+v[v_1s_2-v_2s_1+(v_1-v_2)s]-v_1v_2s=tv(v^2+vv_1-vv_2-v_1v_2) \qquad (3)$$

In the same way we can transform the third equation of our system. But we can save on manipulations if we notice that the third equation is very much like the second: simply replace s_1 by s_2 and v_1 by v_2 and vice versa, and we have the second equation. In (3) replace s_1 by s_2 and v_1 by v_2 and vice versa to get the third transformed equation:

$$v^2S+v[v_2s_1-v_1s_2+(v_2-v_1)s]-v_2v_1s=tv(v^2+vv_2-vv_1-v_2v_1) \qquad (4)$$

Now adding (3) and (4) we have $2v^2S-2v_1v_2s=tv(2v^2-2v_1v_2)$, whence follows the distance over the lake:

$$s=v\frac{vS-v^2t+v_1v_2t}{v_1v_2}=vt+v^2\frac{S-vt}{v_1v_2} \qquad (5)$$

The solution of the problem is complete. Some students, however, considered it necessary, after obtaining the answer with literal data [formula (5), say], to determine under what relationships of the data the answer has "real meaning" (such requirements are imposed as positivity of rates, distances, etc.; conditions are introduced that guarantee nonvanishing denominators, and the like). Quite naturally, a properly conducted investigation does not detract from the solution of a problem, but this investigation is not logically necessary, since it is usually taken for granted that the events actually took place and, hence, the literal data already satisfy the necessary relations. Quite naturally, such an investigation must be carried out if it is explicitly required by the conditions of the problem.

It rather often happens that in problems demanding the setting up of equations, the resulting system has second-degree *homogeneous* equations in two variables (equations of the type $ax^2+bxy+cy^2=0$, where a, b and c are certain numbers). Unfortunately, however, few students realize that the presence of homogeneous equations helps to solve a system of equations. A homogeneous equation of second degree in two unknowns directly defines the relationship of the unknowns, and this naturally simplifies subsequent computations. Let us examine a problem whose solution makes essential use of this fact.

5. *An automobile leaves A for B and a motorcycle leaves B for A at the same time and with a smaller speed. After a time they meet, and at that instant another motorcycle leaves B for A and encounters the automobile at a point which is distant from the meeting point of the automobile and first motorcycle* 2/9 *of the distance from A to B. If the speed of the automobile were* 20 km/hr *less, the distance between the meeting points*

would be 72 km *and the first meeting would have occurred 3 hours after the automobile left A. Find the distance between A and B (the motorcycles have the same speeds).*

Let the speed of the automobile be u km/hr, that of the motorcycle v km/hr, the distance AB, s km, and let t hours elapse before the automobile and first motorcycle meet. We readily set up the following system of equations:

$$tu+tv=s$$
$$3(u-20)+3v=s$$
$$\frac{\frac{2}{9}s}{u}=\frac{vt-\frac{2}{9}s}{v}$$
$$\frac{72}{u-20}=\frac{3v-72}{v}$$

Eliminating the auxiliary unknown t and simplifying, we get the following system:

$$s=3(u+v-20)$$
$$9uv=2(u+v)^2$$
$$v(u-20)=24(u+v-20)$$

To find s we have to find u and v from the last two equations. Noting that the second equation is a homogeneous second-degree equation in two variables, we can easily find the ratio $u:v$.

Since we are interested in u and v different from zero, then dividing the second equation by v^2, we get a quadratic equation in the new variable $z=u/v$:

$$2z^2-5z+2=0$$

The roots of this equations $z_1=2$ and $z_2=1/2$, and so either $u=2v$ or $u=v/2$.

But from the statement of the problem $u>v$, and so we take $u=2v$.

Substituting this value of u into the third equation, we get either $v=40$ or $v=6$. But if $v=6$, then $u=12$, but it is given that $u>20$. The problem is satisfied only by $v=40$. But then $u=80$ and $s=300$. Therefore, the distance AB is found to be 300 km.

Almost insuperable difficulties stem from problems in which the student finds, after correctly setting up the system of equations, that the number of unknowns is greater than the number of equations. The following problem illustrates this point.

6. *Two boys with one bicycle between them set out from A in the direction of B, one by bicycle and the other on foot. At a certain distance from A the one riding the bicycle left it by the road and continued towards B on foot. The one who had started out on foot reached the bicycle and rode the rest of the distance. Both reached B at the same time. On the return trip from B to A, they did as before, but this time the cyclist rode one kilometre*

more than the first time and so his comrade arrived in A 21 minutes after he did. Find the rate of each of the boys on foot if they both did 20 km/hr cycling, and, on foot, the first takes 3 minutes less to cover each kilometre than the second.

Let us introduce the following notation:

s km for the distance between A and B;

v km/hr for the rate on foot of the first boy;

w km/hr for the rate on foot of the second boy;

a km for the distance that the first boy cycled from A to B (he thus left his bicycle a km from A and covered the rest of the distance to B on foot).

It is quite clear that the whole trip from A to B was made by the first boy in $a/20+(s-a)/v$ hours, by the second boy in $a/w+(s-a)/20$ hours. The fact that they set out at the same time and arrived at the same time yields the first equation:

$$\frac{a}{20}+\frac{s-a}{v}=\frac{a}{w}+\frac{s-a}{20}$$

The information on the B to A trip enables us to set up the second equation in similar fashion:

$$\frac{a+1}{20}+\frac{s-a-1}{v}=\frac{a+1}{w}+\frac{s-a-1}{20}-\frac{7}{20}$$

(21 minutes = 7/20 hr).*

Since the first boy spent $1/v$ hours per km and the second $1/w$ hours, we immediately get (from the statement of the problem) the third equation:

$$\frac{1}{w}-\frac{1}{v}=\frac{1}{20}$$

We now have a system of three equations in four unknowns. It is impossible to determine all the unknowns s, a, v, w from this system. In this sense, the system is indeterminate. But does this mean we are not able to solve the problem? Of course not. This is because we only have to find two unknown quantities, the rates v and w, and those can be found uniquely from the system of equations.

To do this, subtract the first equation from the second to get

$$\frac{1}{w}+\frac{1}{v}=\frac{9}{20}$$

and consider the result together with the third equation. Obvious computations yield $v=5$ km/hr and $w=4$ km/hr.

---

* We converted minutes to hours since all quantities must be given in the same units. For example, if the distance is in kilometres and the time in hours, then the speed is in km/hr. Only when the units are the same do the familiar physical equations (such as $s=vt$, etc.) used in the solution hold true.

In this problem we were able to find the unknowns we needed despite the fact that there were fewer equations than unknowns. In the next problem we will obtain a system of equations without being able to determine any of the unknowns, though we will be able to determine the greater one, which is what is required in the problem.

7. *A student spends a certain sum of money on a bookbag, a fountain pen and a book. If the cost of the bookbag were less by a factor of* 5, *the pen by a factor of* 2 *and the book by a factor of* 2.5, *the overall cost would be* 8 *roubles. Now if, compared to the original cost, the prices were reduced—twofold for the bookbag, fourfold for the pen, and threefold for the book—then the total outlay would be* 12 *roubles. How much money was spent and what item cost more, the bookbag or the fountain pen?*

Suppose the bookbag cost x roubles, the pen y roubles and the book z roubles. Together they cost $x+y+z$, and that is what we wish to find out.

The first equation is set up on the condition that under the original assumption the outlay was 8 roubles:

$$\frac{x}{5}+\frac{y}{2}+\frac{z}{2.5}=8$$

Similarly we set up the second equation:

$$\frac{x}{2}+\frac{y}{4}+\frac{z}{3}=12$$

We have a system of two equations in three unknowns and cannot of course determine all the unknowns, but we can find the total sum, which is what we need. To do this, rewrite the equations thus:

$$\begin{aligned} 2x+5y+4z&=80 \\ 6x+3y+4z&=144 \end{aligned} \tag{6}$$

Adding these two equations, we get the sum of the unknowns: $x+y+z=$ $=28$. The total outlay comes to 28 roubles, thus answering the first question.

Now let us try to determine which item, bookbag or fountain pen, is more expensive: in other words we want to know which of the inequalities $x>y$ or $y>x$ is valid.

If we subtract the first equation of (6) from the second, we obtain

$$2x-y=32 \tag{7}$$

From this it is clear that $x>y/2$, because otherwise we would have $32=2x-y<0$. But the inequality $x>y/2$ does not yet answer the question given in the problem. This is because we have not yet made full use of equation (7). Namely, we merely noticed that the difference $2x-y$ is positive. Now let us try to make use of the fact that it is equal to 32 and also take into account that $x+y+z=28$ and that all the unknowns, x, y, z, must, realistically, be positive numbers.

Rewrite (7) as $x+(x-y)=32$. Since the total outlay is 28 roubles, then certainly $x<28$ and from the latter equation follows $x-y>0$, which states that the bookbag is more expensive than the fountain pen.

Nearly all the foregoing problems implicitly involve inequalities. In Problem 5 for instance there were even two: $u>v$ and $u>20$. The inequalities that appear in such problems do not ordinarily upset the student; what does cause a lot of trouble is when the conditions of the problem have to be written out explicitly as inequalities. Many students get as far as writing down the system of equations and inequalities, but no farther. Apparently they are not ready psychologically to solve systems of this kind. To illustrate, take the following problem.

8. *A fast train leaves A for C at 9 A. M. At the same time two passenger trains leave B (located between A and C), the first with destination A, the second with destination C. The two passenger trains have the same speed. The fast train meets the first passenger train not later than 3 hours after departure, then passes B not earlier than 14:00 of the same day and, finally, arrives in C together with the second passenger train exactly 12 hours after meeting the first passenger train. Find the time of arrival at A of the first passenger train.*

Let the speed of the fast train be v_1 km/hr, that of the passenger train, v_2 km/hr, and let the distance AB be s km. From the statement that the fast train meets the first passenger train *not later* than three hours after departure, we get

$$\frac{s}{v_1+v_2}\leqslant 3$$

From the condition that the fast train arrived in B *not earlier* than 5 hours after departure, we obtain

$$\frac{s}{v_1}\geqslant 5$$

Since the time elapse prior to the first meeting is $s/(v_1+v_2)$ hours, it follows that the fast train will overtake the second passenger train in $12+[s/(v_1+v_2)]$ hours, and so

$$\left(12+\frac{s}{v_1+v_2}\right)(v_1-v_2)=s$$

We have to find $x=s/v_2$, whence $s=xv_2$. Substituting this expression for s in the preceding equations and inequalities and denoting v_1/v_2 by α, we get the system

$$x\leqslant 3(\alpha+1)$$
$$x\geqslant 5\alpha$$
$$x=6(\alpha^2-1)$$

It was precisely this system that stumped many students.

Actually it is not so complicated. It is necessary to eliminate either x or α and go over to a system of two inequalities in one unknown. Since it appears easier at first glance to eliminate x, let us do so. Putting $x=6(\alpha^2-1)$ into the first two inequalities, we obtain the system of inequalities

$$2\alpha^2-\alpha-3\leqslant 0$$
$$6\alpha^2-5\alpha-6\geqslant 0$$

The solutions of the first inequality are $-1\leqslant\alpha\leqslant 3/2$, the solutions of the second are $\alpha\geqslant 3/2$ and $\alpha\leqslant -2/3$. Hence, the solution of the system is $\alpha=3/2$, and also all α lying in the interval $-1\leqslant\alpha\leqslant -2/3$. Since we are only interested in positive values of α, the condition of the problem is satisfied by the sole value $\alpha=3/2$. From this it is easy to find $x=15/2$ and we get the answer: the first passenger train arrives at A at 16:30.

This problem allows for a solution in which all the conditions of the problem are written down in the form of equations. This is achieved by introducing supplementary unknowns and obtaining a system of equations in which the number of unknowns is greater than the number of equations. However, it is more difficult to solve that type of system of equations than it is to solve a system of inequalities.

Let us solve the problem the second way, retaining all the earlier notation. Let the fast train meet the first passenger train after an elapse of $(3-t_1)$ hours $(t_1\geqslant 0)$, let it pass B in $(5+t_2)$ hours $(t_2\geqslant 0)$ and let it catch up with the second passenger train in $[(3-t_1)+12]$ hours. Then we can easily set up the equations

$$(v_1+v_2)(3-t_1)=s$$
$$v_1(5+t_2)=s$$
$$(15-t_1)(v_1-v_2)=s$$
$$xv_2=s$$

Eliminating s and denoting v_1/v_2 by α, we get the system of equations

$$\begin{aligned}(\alpha+1)(3-t_1)&=x\\ \alpha(5+t_2)&=x\\ (\alpha-1)(15-t_1)&=x\end{aligned}\tag{8}$$

This is a system of three equations in four unknowns. We only want x. Proceeding as before, eliminate x to get

$$\begin{aligned}\alpha t_2+(\alpha+1)t_1&=3-2\alpha\\ (1-\alpha)t_1-\alpha t_2&=15-10\alpha\end{aligned}\tag{9}$$

Noting that the right member of the second equation is five times the right member of the first, multiply the first by 5 and subtract the second to get

$$6\alpha t_2+(6\alpha+4)\,t_1=0 \tag{10}$$

Since $\alpha>0$, $t_1\geqslant 0$, $t_2\geqslant 0$, it follows that this equation can only be valid for $t_1=0$ and $t_2=0$. But then from (9) it is easy to find $\alpha=3/2$ and from (8), $x=15/2$. We get the same answer. Many students failed to notice that (10) can be derived from (9) and therefore they could not conclude, from (9), that $t_1=t_2=0$ and so could not solve the problem.

The foregoing shows that the first method of solution is much easier than the second method.

9. *A cyclist starts out from city A at* 9 *A. M. and proceeds at a constant rate of* 12 km/hr. *Two hours later, a motorcyclist starts out with an initial speed of* 22 km/hr *and proceeds with uniformly decelerated motion so that in one hour the speed diminishes by* 2 km/hr. *A man in a car driving to A at* 50 km/hr *meets first the motorcyclist and then the cyclist. Will the car driver be able to reach A by* 19:00 *that same day?*

This problem can likewise be solved by setting up a system of equations and inequalities, but it would require extended reasoning. It is best to approach the problem by simple reasoning instead of attempting to solve it by formally setting up of a system of equations and inequalities.

If follows from the statement of the problem that first the motorcyclist catches up with the cyclist and then the cyclist catches up with the motorcyclist. Let the cyclist spend t hours prior to a meeting (first or second, no matter which). Then the motorcyclist will spend $(t-2)$ hours over the same distance. Since they cover the same distance before meeting, we can equate their paths to obtain

$$12t=22\,(t-2)-2\,\frac{(t-2)^2}{2}$$

Solving this equation, we see that up to the first meeting the cyclist rode 6 hours and thus covered 72 km; prior to the second meeting he rode 8 hours and hence covered 96 km. It is given that the car driver met the cyclist before the latter had covered 96 km, which means the motorist has less than 96 km left to drive to A. He will spend less than 96/50 hours. Since the cyclist will spend less than 8 hours before meeting the car, their meeting will occur earlier than 17:00. Thus, after meeting the cyclist there remain over two hours for the car driver to reach A before 19:00 hours. But he requires less than 96/50 hours, that is less than two hours, to cover this distance. And so the motorist will reach A before 19:00.

At examinations, problems appear in which the student is asked to find an optimal solution, say, to buy the largest possible quantity of goods for a given sum of money or to choose the best possible (cheapest) variant in transportation of goods, and the like.

Solutions of this type of problem may consist of setting up systems of equations and inequalities and their solution. However, the most necessary element in such cases is the reasoning that helps to choose the best possible variant.

10. *The task is to construct a number of identical dwelling houses with a total floor-space of* 40,000 *sq. metres. The cost of one house of N sq. metres* (m^2) *of dwelling floor-space consists of the cost of the above-ground portion (superstruction), which is proportional to* $N\sqrt{N}$ *and the cost of the foundation, which is proportional to* $\sqrt{N}$. *The cost of a house of* 1600 m^2 *is set at* 176,800 *roubles, in which case the cost of the above-ground portion is* 36% *of that of the foundation. Determine how many houses should be put up so that the overall cost is the smallest possible and find that sum.*

Suppose we decide to construct n identical structures each of which has y m^2 of dwelling floor-space. Then the equation $yn=40{,}000$ is valid. Let the cost of one structure of y square metres of floor-space be z thousand roubles; then the cost x of the construction job as a whole is computed from the equation $x=zn$.

The cost of a house consists of the cost, v, of the above-ground portion of the structure and the cost, w, of the foundation: thus, $z=v+w$. It is given that the cost of the above-ground portion of a house of y square metres is proportional to $y\sqrt{y}$, that is, $v=\alpha y\sqrt{y}$, where α is a certain coefficient. Similarly, $w=\beta\sqrt{y}$, where β is a certain coefficient.

In the particular case of the construction of a house of 1600 square metres, taking into account that the cost of the superstructure is 36 % of the cost of the foundation, we get

$$\alpha\cdot 1600\cdot\sqrt{1600}=\frac{36}{100}\cdot(\beta\cdot\sqrt{1600})$$

and taking into consideration that the construction of a house of 1600 square metres costs 176,800 roubles, we have

$$176{,}800=\alpha\cdot 1600\sqrt{1600}+\beta\sqrt{1600}$$

All the conditions of the problem have been written down, we now have to determine x as a function of n and then determine for which n, x will be a minimum.

It is easy to find α and β from the last two equations: $\alpha=$ $=117/160{,}000$, $\beta=13/4$. Substituting v and w into the expression for z, we get $z=(117/160{,}000)y\sqrt{y}+(13/4)\sqrt{y}$. Now putting this value of z and the value $y=40{,}000/n$ from the first equation into the second, we get

$$x=650\left(\frac{9}{\sqrt{n}}+\sqrt{n}\right)$$

We thus conclude that x, the cost of construction, is the above function of n, the number of houses. We now have to determine the smallest value of x. Applying to the right member of this equation the inequality between the arithmetic mean and the geometric mean, we find

$$x \geqslant 2\cdot 650\sqrt{9}=3900$$

equality being attained only when $8/\sqrt{n}=\sqrt{n}$, that is, for $n=9$. In other words, the cost of construction will always be at least 3.9 million roubles and will be exactly equal to this figure only when $n=9$.

Thus, in this housing construction project, the smallest expenditure will be in building 9 houses and the total cost in that case will be 3.9 million roubles.

11. *It is agreed to spend 100 roubles on christmas tree decorations. Such decorations are sold in sets consisting of 20 items per set costing 4 roubles a set, sets of 35 items at 6 roubles, and sets of 50 items at 9 roubles. How many of what kinds of sets should be bought in order to have the largest number of items?*

Let x, y, z be the number of sets, respectively, of the first, second and third kind which must be bought to ensure a maximum quantity of items (this solution is termed the optimal solution of the problem). Then

$$4x+6y+9z=100$$

This is the only equation that can be set up on the basis of the statement of the problem. But we also know that x, y and z are nonnegative integers and that the number of items of decoration in this purchase is greater than any other. These conditions, it turns out, are quite sufficient to ensure an unambiguous determination of all unknowns.

The first idea is to solve the equation by running through all possible values of the unknowns but this is rather hopeless due to the enormous number of cases.

However, this number may be appreciably reduced with the aid of arguments of an economic nature. Indeed, 12 roubles can buy 3 sets of the first type or 2 sets of the second; in the former case we get 60 items, in the latter 70. It is then clear that in an optimal solution the number of sets of the first type should not exceed 2. Comparing sets of the second and third type in similar fashion, we find that in the optimal solution there should not be more than one set of the third kind. We have thus obtained the inequalities $x \leqslant 2$, $z \leqslant 1$.

Now there are fewer cases and we can run through them. For $x=0$ we get the equation $6y+9z=100$ to determine y and z. It clearly does not have a solution since the left member is divisible by 3 but the right is not. Furthermore, when $x=1$ we get the equation $2y+3z=32$, which (taking into account the inequality $z\leqslant 1$) has the unique solution $y=16$, $z=0$. Finally, when $x=2$, there is no solution, as in the case of $x=0$.

Thus, to ensure the largest number of decorations, we have to buy 1 set of 20 items and 16 sets of 35 items.

It is possible in this solution to dispense with examining all cases if a more detailed use is made of divisibility, as witness: from the given equation it follows that the number x yields a remainder of 1 upon division by 3, and the number z is even. It therefore follows immediately from the inequalities $x\leqslant 2$ and $z\leqslant 1$ that $x=1$ and $z=0$; from the equation we get $y=16$.

Note in conclusion that in actuality the foregoing reasoning signifies that the optimality condition of the solution may be written in the form of the following system of equations and inequalities:

$$
\begin{gathered}
4x+6y+9z=100, \\
0\leqslant x\leqslant 2, \quad 0\leqslant y, \quad 0\leqslant z\leqslant 1
\end{gathered}
\tag{11}
$$

with the supplementary proviso that x, y and z are integers. Now the condition of x, y and z being integers means that $z=2n$ and $x=1+3k$, where n and k are also integers. Substituting these values of z and x into the appropriate equations, we get $n=k=0$, or $x=1$ and $z=0$. It is now easy to find $y=16$ from equation (11).

12. *A forestry has to deliver* 1590 *trees. The vehicles assigned to this job are* $1\frac{1}{2}$*-ton, three-ton and five-ton trucks. A* $1\frac{1}{2}$*-ton truck carries* 26 *trees at a time, a three-ton truck carries* 45 *trees, and a five-ton truck,* 75 *trees. The cost of one run of a* $1\frac{1}{2}$*-ton truck is* 9 *roubles, that of a three-ton truck,* 15 *roubles, and of a five-ton truck,* 24 *roubles. The forestry wishes to minimize the overall cost of the deliveries. How is this to be done (it is assumed all trucks are fully loaded)?*

Let x, y, z be the number of $1\frac{1}{2}$-ton, three-ton, and five-ton trucks, respectively, in the case of optimal distribution. Since all vehicles are fully loaded, the number of transported trees in this setup will be $26x+45y+75z$ and thus we get the equation $26x+45y+75z=1590$.

We are now in the same position as in Problem 11, but attempts to reduce the number of cases that succeeded then do not yield any substantial simplification. For example, x is divisible by 15. That's about all. We might add that 45 $1\frac{1}{2}$-ton trucks would cost 405 roubles a trip, 26 three-ton trucks carrying the same number of trees would cost 390 roubles so that in an optimal solution the number of $1\frac{1}{2}$-ton trucks should not exceed 44. And so for x we get three possibilities:

$x=0$, $x=15$, $x=30$. For each of these values we would have to solve the equation for y and z, which also have a good many solutions.

This is a very tedious approach, though in the absence of any other, it is acceptable.

Here's an attractive idea, which, unfortunately, doesn't do the job. From the statement of the problem it is easy to figure out that for 45 roubles using 5 $1\frac{1}{2}$-ton trucks we can deliver 130 trees and using 3 three-ton trucks, 135 trees. It would therefore appear that the number of $1\frac{1}{2}$-ton trucks should not exceed 4, otherwise those same trees could be delivered more cheaply. From this and from earlier considerations it follows that $x=0$ and the number of cases left to be examined is much less.

Actually, however, this argument only implies that for a given sum of money we can deliver a larger number of trees, but our aim is to deliver a given number of trees at minimum cost. Nevertheless, it is still possible to get around brute-force tactics by just a little common-sense reasoning.

Any reasonable person would first estimate which of the given types of trucks is the most efficient by determining the cost of delivering a single tree. We find the following: for a $1\frac{1}{2}$-ton truck, a three-tonner and a five-tonner the cost is 9/26, 1/3 and 8/25 rouble respectively. Since $9/26>1/3>8/25$, it is clear that it is more profitable to use five-ton trucks first, then, if necessary, three-tonners and, finally, $1\frac{1}{2}$-ton trucks.

It is easy to see that the greatest number of trees that can be delivered by five-ton trucks comes out to 1575. However, seeing that all vehicles must be fully loaded, we get 1500 trees for delivery by five-ton trucks. Then 90 trees can be delivered by three-tonners and so it is natural to suppose that the optimal distribution will be 20 five-ton and 2 three-ton trucks.

It is easy to demonstrate that this plan is indeed an optimal plan: if we reduce the number of five-tonners, then the "undelivered" trees apportioned to these vehicles would have to be delivered by $1\frac{1}{2}$-ton and three-ton trucks, but delivery costs per tree on these trucks are higher, and so the total cost of the assignment would increase.

We thus have the optimal distribution of 20 five-ton and 2 three-ton trucks, while all the unknowns and the single equation that was set up remain unused! To summarize, then, we set out in the ordinary way, but in the process of solution we found an approach which made all the earlier arguments unnecessary. Quite obviously the method given here would be quite sufficient at any examination.

Exercises

1. Three cyclists start out simultaneously from the same place in one direction around a circular track 1 km in length. The rates of the cyclists form, in a certain order, an arithmetic progression with common difference 5 km/hr. After some time, the second one catches up with the first, having made one extra circuit; 4 minutes later the third arrives at that point, having covered the same distance that the first did at the time he caught up with the second cyclist. Find their rates.

2. Three brothers, whose ages form a geometric progression, divide among themselves a certain sum of money in direct proportion to the age of each. If this were done in three years time, when the youngest becomes one-half the age of the oldest, then the youngest would receive 105 roubles, and the middle one, 15 roubles more than at the present time. Find the ages of the brothers.

3. Two groups of tourists start out from A in the direction of B at the same time. The first group goes by bus (at 20 km/hr) and reaches C, midway between A and B, and then continues on foot. The second group starts out on foot, but in one hour boards a car which proceeds at 30 km/hr and reaches B. The first group passes C 35 minutes before the second group, but arrives at B 1 hr and 25 minutes later than the second group. What is the distance from A to B if the rate of the first group (on foot) is 1 km/hr greater than that of the second group?

4. Two identical vessels are filled with alcohol. We draw off a litres of alcohol from the first vessel and add that amount of water. Then we draw off a litres of the resulting mixture of water and alcohol and add that amount of water. In the case of the second vessel, we draw off $2a$ litres of alcohol and add that amount of water, and then draw off $2a$ litres of the resulting mixture and add that amount of water. Determine what part of the volume of the vessel is taken up by a litres if the strength of the final mixture in the first vessel is 25/16 times the strength of the final mixture in the second vessel. (By strength is meant the ratio of volume of pure alcohol in the mixture to the total volume of the mixture. It is assumed that the volume of the mixture is equal to the sum of the volumes of its components.)

5. Two bodies are in uniform motion around a circle in the same direction, and one of them catches up with the other every 46 seconds. If these bodies were in motion in opposite directions, they would meet every 8 seconds. Determine the rates of the bodies if we know that the radius of the circle is equal to 184 cm.

6. Towns A and B are located on a river; B downstream from A. At 9 A.M. a raft starts floating downstream from A in the direction of B (the rate of the raft relative to the bank of the river is the same as the rate of the current). At that time, a boat starts out from B for A and meets the raft in 5 hours. Upon reaching A, the boat turns round and returns to B arriving at the same time as the raft. Did the boat and raft arrive at B by 9 P.M. (of the same day)?

7. Three workers receive an assignment, which each separately completes in a specified time, the third completing the job one hour faster than the first. Working together they can complete the job in one hour. But if the first worker does one hour and then the second 4 hours, together they can complete the job. How long does it take each worker separately to complete the full assignment?

8. We have two solutions of a salt in water. To obtain a mixture containing 10 grams of salt and 90 grams of water, one takes twice as much (by weight) of the first solution as the second. One week elapses and 200 grams of water has evaporated from each kilogram of the first and second solution. Now to obtain the same mixture as before, we require four times more (by weight) of the first solution than of the second. How many grams of salt did 100 grams of each solution originally contain?

9. A freight train going from A to B arrives at station C together with a passenger train going from B to A with a speed m times that of the freight train. The two trains stop at C for t hours and then continue on their ways, each train increasing its speed 25% over the original speed it had prior to arrival at C. Then the freight train arrives at B later by t_1 hours and the passenger train arrives at A later by t_2 hours than if they had gone nonstop at their original speeds. How much earlier did the freight train start out from A than the passenger train from B?

10. Three points A, B and C are connected by straight roads. Adjoining section AB of the road is a square field with a side of $1/2$ AB; adjoining BC is a square field with a side equal to BC, and adjoining CA is a rectangular section of the woods, length equal to AC and of width 4 km. The wooded area is 20 square kilometres greater than the sum of the areas of the square fields. Find the area of the woods.

11. Thirty students received marks of 2, 3, 4, 5 at an examination. The sum of the marks is 93, there are more 3's than 5's and fewer 3's than 4's. Besides, the number of 4's is divisible by 10 and the number of 5's is even. Determine the number of the various marks received by the thirty students.

12. A motorcycle and a car (Volga) leave A for B at the same time; and at the same time another car (Moskvich) leaves B for A. The Moskvich arrives in A in 5 hours and 50 minutes. The cars meet 2 hours 30 minutes later, and the motorcycle and Moskvich meet 140 km from A. If the rate of the motorcycle were twice what it was, it would have met the Moskvich car 200 km from A. Find the speeds of the motorcycle and the two cars.

13. An empty tank is being filled through two pipelines simultaneously with pure water and a constant concentration of an acid solution. When filled, the tank has a 5% solution of the acid. If when the tank were half full, the water supply were cut off, the full tank would have a 10% solution of the acid. Determine which pipeline delivers liquid faster and how much faster.

14. A car leaves point A for B. At the same time, a cyclist starts out in the opposite direction (from B). Three minutes after they meet, the car turns around and follows the cyclist; after catching up with the cyclist, it turns around and goes to B. If the car had turned around 1 minute after the meeting, and the cyclist (after the meeting) had increased his speed 15/7 times, the car would have spent the same amount of time for the entire trip. Find the ratio of the speeds of the bicycle and the car.

15. Two men start out at the same time from A in the direction of B, which is 100 km from A, one on a bicycle, the other on foot. Also at the same time, a car starts out from B and goes in the direction of A. An hour later the car meets the cyclist and, continuing another $14\frac{2}{17}$ km, it meets the man on foot, who boards the car, and the car overtakes the cyclist. Compute the speeds of the bicycle and the car if we know that the man on foot was doing 5 km/hr. The time required to get into the car and turn the car around is disregarded.

16. A laboratory has to order a certain number of identical spherical flasks with a total capacity of 100 litres. The cost of one flask consists of the cost of labour, which is proportional to the square of the surface area of the flask, and the cost of the material, which is proportional to the surface area. A flask of capacity 1 litre costs 1 rouble 25 kopeks, and in this case the labour cost is 20% of the cost of the flask (the wall thickness of the flask is negligible). Will 100 roubles (100 kopeks to a rouble) be enough to cover the cost?

17. Bus No. 1, which a student uses to get to his place of study (institute) without changing buses, covers the distance in 2 hours and 1 minute. He can also get to the institute by any one of the buses No. 2, 3, ... , No. K, but the only way to make a change to Bus No. P is from Bus No. $(P-1)$. The routes of these buses are such that if the student gets to the institute on one of them, he will spend en route (disregarding transfers) a time inversely proportional to the number of buses used. Moreover, he will have to spend 4 minutes at each transfer. Is there a route he can take such that the total time is less than 40.1 minutes?

18. Between town A and city E is a gasoline (petrol) station O and a water supply station B, which divide the distance between A and E into three equal parts ($AO=OB=BE$). A motorist and cyclist start out simultaneously from A in the direction of E and, at the same time, a truck starts out from E in the direction of A and at the water supply station passes the car and at the gasoline station passes the cyclist. At the gas station, the cyclist increases his speed 5 km/hr. The motorist reaches E and then sets out on the return trip at 8 km/hr slower than before. As a result, when the truck arrives in A, the cyclist still has 7.5 km to go to B, while the motorist is in between O and A, 14 km from O. Find the distance from the town to the city and also the speeds of the car, truck and bicycle.

19. A rectangular plot of area 900 square metres is to be fenced in; two adjoining sides to be brick and the two others to be wooden. One metre of the wooden fence costs 10 roubles and one metre of the brick fence costs 25 roubles. A total of 2000 roubles has been alotted for the job. Will this sum be sufficient?

20. The tank at a water supply station is filled with water by several pumps. At first, three pumps of the same capacity are turned on; 2.5 hours later, two more pumps (both the same) of a different capacity are set into operation. One hour after the additional pumps were set into operation the tank was almost filled to capacity (15 cubic metres were still lacking), in another hour the tank was full. One of the two additional pumps could have filled the tank in 40 hours. Find the volume of the tank.

21. At a 10,000-metre ski race, the first skier starts out and is followed shortly by a second one, the rate of the second skier being 1 m/sec more than that of the first one. When the second skier catches up with the first, the latter increases his rate by 2 m/sec, while the rate of the second skier remains unchanged. As a result, the second skier finishes 7 minutes and 8 seconds after the first one. If the distance had been 500 metres longer, then the second skier would have finished 7 minutes and 33 seconds after the first one. Find the time lapse between the start of the first and second participants.

22. Three skaters whose rates, taken in some order, form a geometric progression, start out at the same time on a skating circuit. After a time, the second one overtakes the first, covering 400 metres more. The third skater covers the same distance that the first did when the latter was overtaken by the second during a time that is 2/3 minutes more than the first. Find the rate of the first skater.

23. A farm has tractors of four models, A, B, C, D. Four tractors (2 of model B and one of models C and D) plough a field in two days. Two model A tractors and one model C tractor take three days to do this job, and three tractors of models A, B and C take four days. How long will it take to do the job if a team is made up of four tractors of different models?

24. Grass was mowed on three fields in the course of three days. On the first day, all the grass of the first field was mowed in 16 hours. On the second day, all the grass of the second field was mowed in 11 hours. On the third day, all the grass was cut on the third field in 5 hours, four hours of which the work was done with scythes and one hour by a mowing machine. During the second and third days, together, four times more grass was cut than on the first day. How many hours was the mowing machine in operation if in one hour it mows five times as much grass as is cut by hand. It is assumed that the hand and machine operations were separate in time (did not overlap) and there were no breaks in the work.

25. A factory has to deliver 1100 parts to a client. The parts are packed in boxes of three types. One box of type One holds 70 parts, one of type Two holds 40, and one of type Three holds 25 parts. The cost of delivery in a type One box is 20 roubles, in a type Two box, 10 roubles, in a type Three box, 7 roubles. What kind of boxes should be used in order to minimize the cost of delivery? All boxes must be used to full capacity.

26. A stamp collector decides to put all his stamps in a new album. If he puts 20 stamps on one sheet, there will be some left over; if he puts 23 stamps on a sheet, there will be at least one empty sheet left in the album. If the stamp collector is presented with an album of the same kind, each sheet holding 21 stamps, he will have a total of 500 stamps. How many sheets are there in the album?

27. Two pipelines operating together fill a pool 3/4 full of water. If one pipeline fills one-fourth of the pool first, and then the second (the first is then switched off) brings the volume of water up to 3/4 the capacity of the pool, then this will require 2.5 hours. Now if the first pipeline is in operation for one hour and the second for half an hour, they will bring the water level up to more than one-half the pool. How long will it take each pipeline separately to fill the pool?

28. Points A and B are located on a river so that a raft floating downstream from A to B with the rate of the current covers the distance in 24 hours. A motorboat goes from A to B and returns in less than 10 hours. If the rate of the motorboat in

still water were 40% greater, the same distance (from A to B and back) could be covered in not more than 7 hours. Find the time it takes the motorboat to go from A to B at the original rate (not increased).

29. At 8 A.M. a fast train leaves A for B. At the same time a passenger train and an express train leave B for A, the speed of the former being one half that of the latter. The fast train meets the express train not earlier than 10:30 A.M. and arrives at B at 13:50 the same day. Find the time of arrival of the passenger train at A if we know that not less than an hour elapsed between the meetings of the fast train and express train and the fast train and the passenger train.

30. At 9 A.M. a cyclist starts out from A in the direction of B. Two hours later a motorist sets out and overtakes the cyclist not later than 12:00 noon. The motorist continues on and reaches B, then immediately turns round and heads back for A. On the way, the motorist meets the cyclist and arrives in A at 17:00 hours that same day. When does the cyclist arrive in B if we know that no more than 3 hours elapsed between the two encounters of the cyclist and motorist.

1.13 Graphs of functions

The ability to represent geometrically functional relationships given by formulas is particularly important in higher mathematics. That is why examination papers contain problems involving the graphing of functions.

Graph construction is often a source of trouble to many students. This is largely due to the fact that questions of graphing functions are rather scattered in the school course of mathematics and general techniques for construction of graphs are hardly at all considered.

It is well to review the basic material in the standard textbooks, to recall the basic curves and to drill a good deal in the sketching of specific graphs.

It will be recalled that a *functional relation is a rule or law in accordance with which to each value of a quantity x (the independent variable or argument) from a certain set of numbers, called the domain of definition of the function, is associated one definite value of a quantity y (the dependent variable or function); the collection of values assumed by the dependent variable y is termed the range of the function.*

It is important to emphasize that the arguments of functional relationships studied at school *are always assumed to take on real values*, and *only real numbers are allowed* as values of the dependent variable.

If a functional relation (function) is given by a formula $y=f(x)$, then finding its domain of definition reduces to finding all the real values of the argument for which the expression $f(x)$ defining the function is meaningful, that is to say, assumes real values. Let us consider some examples.

1. *Find the domain of definition (or, simply, domain) of the function* $y=\log_x \cos x$.

It is obvious that the domain of this function includes only those values of x for which the following conditions are simultaneously valid: (a) $x>0$, $x\neq 1$ (since the logarithmic base must be positive and nonzero); (b) $\cos x > 0$ (since negative numbers and zero do not have logarithms).

Solving this system of inequalities, we find that the domain of the function at hand is the following set of numbers:

$$0 < x < 1, \quad 1 < x < \frac{\pi}{2}, \quad -\frac{\pi}{2} + 2k\pi < x < \frac{\pi}{2} + 2k\pi$$

where $k = 1, 2, 3, \ldots$ (represent it on the number line!).

2. *Find the domain of the function*

$$y = \frac{\cot x}{\sqrt{\sin x - \cos x}} \tag{1}$$

This function is not defined for those values of x for which $\sin x - \cos x = 0$ (the denominator of the fraction must be different from zero), and, besides, for those x for which $\sin x - \cos x < 0$ (because for these values of x the denominator assumes imaginary values). Thus, the domain of function (1) consists only of those values of x for which the inequality $\sin x - \cos x > 0$ is valid; solving this inequality (see Problem 6, Sec. 1.10), we find

$$\frac{\pi}{4} + 2k\pi < x < \frac{5\pi}{4} + 2k\pi, \quad k = 0, \pm 1, \pm 2, \ldots \tag{2}$$

However, it must be further noted that $\cot x$ is not defined for $x = n\pi$, where n is any integer. And so all the values of $x = n\pi$, $n = 0, \pm 1, \ldots$, likewise do not belong to the domain of the function and must be excluded from the resulting sequence of intervals (2). Thus, for the domain of function (1) we finally get the following set of real numbers:

$$\frac{\pi}{4} + 2k\pi < x < \pi + 2k\pi, \quad \pi + 2k\pi < x < \frac{5\pi}{4} + 2k\pi,$$
$$k = 0, \pm 1, \pm 2, \ldots$$

3. *Find the domain of the function*

$$y = \sqrt{\cos(\cos x)} + \arcsin \frac{1 + x^2}{2x} \tag{3}$$

We consider each summand separately. The domain of this function can only embrace those values of the argument for which the first term assumes real values, i.e. those values of x for which the radicand $\cos(\cos x)$ is nonnegative: $\cos(\cos x) \geqslant 0$. It is easy to see (see Problem 8 of Sec. 1.8) that this inequality holds true for all real values of x.

Now let us examine the second summand. By definition, the expression $\arcsin a$ is meaningful only for $|a| \leqslant 1$ (see Sec. 2.5); in other words, only those values of x belong to the domain of function (3) for which $|(1 + x^2)/2x| \leqslant 1$. However, it may be proved directly (see formula (3) of Sec. 1.8) that the inequality $|(1 + x^2)/2x| \geqslant 1$ holds for all nonzero real values of x, equality being achieved only when $x = 1$ and $x = -1$.

Consequently, the domain of (3) consists of two points only: $x=-1$ and $x=1$.

The foregoing examples show that in finding the domain of definition of a function one has to invoke various branches of algebra and trigonometry. Only when these sections are fully mastered can the student tackle such problems with ease.

The student should have a firm knowledge of the definitions and be able to investigate such general properties of functions as boundedness, monotonicity (intervals over which a function is increasing or decreasing), evenness and oddness, periodicity, and be able to find the range of a function, its zeros, extremal values, and the like.

The investigation of the properties of functions is carried out without invoking the concept of a derivative, which is an element of mathematical analysis and is outside the school curriculum.

The student should have a clear idea of a system of coordinates in the plane and be able to sketch, by memory, the graphs of the basic functions $y=kx+b$ (straight line); $y=ax^2+bx+c$ (parabola); $y=k/x$ (hyperbola); $y=|x-a|$; $y=x^3$; $y=\sqrt{x}$; $y=1/x^2$; $y=a^x$ $(a>0, a\neq 1)$; $y=\log_a x$ $(a>0, a\neq 1)$; $y=\sin x$ (sine curve); $y=\cos x$; $y=\tan x$; $y=\cot x$. The student should be able to sketch the graphs of these functions in each concrete case, giving a general picture and the characteristic peculiarities of behaviour of the curve and not be forced to compute each time a table of values and plot the curve.

The student should also be able to illustrate geometrically on the graph the properties of a function. When relating some property (say, the oddness of the sine), the student sometimes sketches the appropriate graph (sine curve) and then makes the mistake of saying "this property is evident from the drawing." Such reasoning is faulty because it is precisely by using the property of the function that one can more or less accurately sketch its graph. For this reason, all the properties of a function must be demonstrated by rigorous analysis as is done in the standard textbook.

Some examination problems call for constructing graphs of functions which are combinations of basic functions. Again, it is only the *approximate* behaviour of the curve that is required. Plotting can then be used as an auxiliary aid.

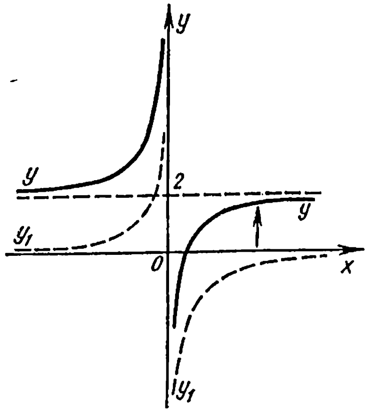
Let us examine some problems in which graphs are constructed by translation or a specific deformation of the graphs of basic functions.

4. *Construct the graph of the function* $y=2-1/x$.

The domain of this function consists of all real values of x except $x=0$. If we consider the function $y_1=-1/x$ (which is a hyperbola whose branches are located in the second and fourth quadrants), it is obvious that for each value $x=x_0$ the value of y is greater by 2 than the value of y_1 for the same value x_0 of the argument. It therefore suffices to translate the graph of the function y_1 2 units upwards along the axis of ordinates to get the desired graph of the function y (Fig. 29).

It is easy to see that this device enables us *to construct the graph of tne function $y=a+f(x)$, where a is a given number, if we have already constructed the graph of the function $y_1=f(x)$*: it is sufficient to translate

Fig. 29

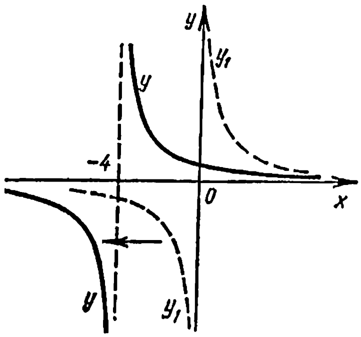


the graph of the function y_1 a units upwards if $a>0$ or $|a|$ units down if $a<0$.

5. *Construct the graph of the function* $y=\frac{3}{x+4}$.

Evidently, x can assume all values except -4. Compare this function with the function $y_1=3/x$. It is clear that the value of the function y corresponding to some value $x=x_0$ coincides with the value of y_1, which corresponds to the value of its argument equal to x_0+4. For

Fig. 30



example, the function $y=3/(x+4)$, when $x_0=1$, takes the value $y=$ $=3/5$, and the function $y_1=3/x$ assumes this very same value for the value of its argument equal to $5=x_0+4$. And so if we translate the graph of the function y_1 four units to the left along the x-axis, we get the graph of the function y that we want (Fig. 30).

It is not so hard to figure out that this same method enables one *to draw the graph of the function $y=f(x+b)$, where b is a given number, if we already have the graph of the function $y_1=f(x)$*: it suffices to translate the graph of y_1 b units leftwards if $b>0$ or $|b|$ units rightwards if $b<0$.

6. *Construct the graph of the function* $y=\frac{-x+5}{3x-2}$.

To draw the graph, first transform the fraction and represent the function as

$$y=-\frac{1}{3}+\frac{13/9}{x-(2/3)}$$

Arguing as in Problems 4 and 5, we see that the graph of the proposed function is an "ordinary" hyperbola $y=(13/9)/x$ translated 2/3 unit rightwards along the x-axis and 1/3 unit down along the y-axis (make the drawing!).

A similar device permits drawing the graph of any function

$$y=\frac{ax+b}{cx+d}$$

This is called a *linear fractional function*. Indeed, a simple transformation permits writing this function as *

$$y=\frac{a}{c}+\frac{\frac{ad-bc}{c^2}}{x+\frac{d}{c}}$$

It then remains to invoke the arguments given above in solving Problems 4 and 5.

Note that in the very same way, by combining remarks pertaining to Problems 4 and 5, we can readily *represent the graph of a function $y=a+f(x+b)$, where a and b are specified numbers, if the graph of the function $y_1=f(x)$ has already been constructed.*

7. *Draw the graph of the function* $y=\log_4(-x)$.

Sometimes a student gives an answer like this: "The graph of the function does not exist since negative numbers do not have logarithms." The mistake here is the failure to grasp the fact that $-x$ does not by any means always represent a negative number.

The domain of the function y under consideration is the set $x<0$. It is immediately clear that the value of this function, when $x=-x_0$, $x_0>0$, coincides with that of the function $y_1=\log_4 x$ for the value x_0 of its argument. Hence, to obtain the graph of the function y it is

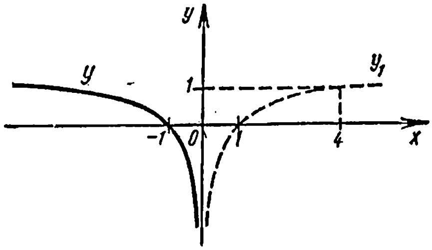
---

* It is assumed here that $c\neq 0$ (otherwise the function is simply linear) and that $ad-bc\neq 0$. If this latter condition is not fulfilled, then the original function, for all admissible x, is of the form $y=k$, where k is a constant. For instance, the function $y=(2x+2)/(x+1)$ can, in its domain of definition, i.e., when $x\neq -1$, be written as $y=2$. Hence, the graph of this function is the straight line $y=2$ with point $(-1, 2)$ deleted (but not the entire line $y=2$, as some students seem to think; cf. Problem 11).

sufficient to reflect the graph of the function y_1 about the y-axis (Fig. 31).

It will be noted that this same device permits *constructing the graph of the function $y=f(-x)$ if we have the graph of the function $y_1=f(x)$*: it suffices to reflect the graph of y_1 about the y-axis.

Fig. 31

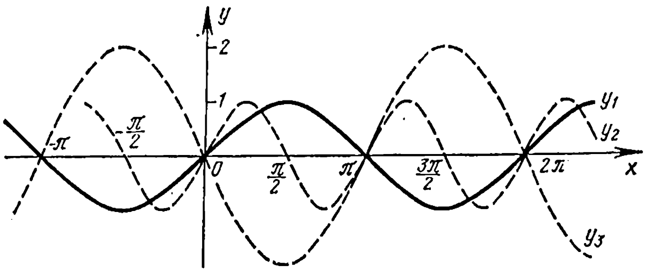


8. *Construct in a single drawing the graphs of the functions*

$$y_1=\sin x, \quad y_2=\sin 2x, \quad y_3=-2\sin x$$

The student is not always able to give a proper representation of all three curves in a single drawing and correctly to indicate their mutual positions (Fig. 32), to indicate the peculiarities of each of the sine curves and to explain how they are obtained one from another.

Fig. 32



For one thing, it is useful to remember that the smallest positive period of the function $y=A \sin \omega x$, where $\omega \neq 0$ and $A \neq 0$ are given numbers,* is equal to $2\pi/|\omega|$ (for instance the smallest positive period of the function $y=-3 \sin \pi x$ is the number $2\pi/\pi=2$, and for the function $y=1/4 \sin(-x/3)$ is the number $2\pi/|-1/3|=6\pi$), while its "amplitude" is equal to $|A|$ (thus, the "amplitude" of the function $y=-1/2 \sin 3x$ is $1/2$).

The foregoing of course refers also to all the other trigonometric functions.

It is important to stress the fact that *it is possible to construct the graph of the function $y=Af(\omega x)$, where $\omega \neq 0$ and $A \neq 0$ are given*

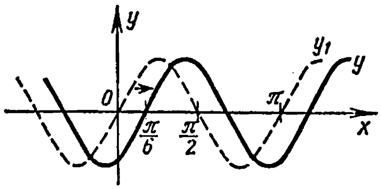
* Clearly, in the case of $\omega<0$, this function can be represented as $y=-A \sin|\omega| x$.

numbers, if we know the graph of the function $y_1=f(x)$. First compress the graph of y_1 along the x-axis ω times if $\omega>0$; but if $\omega<0$, then compress the graph of the function y_1 $|\omega|$ times along the x-axis and perform a reflection with respect to the y-axis (see the solution of Problem 7). Then take the resulting curve and stretch it along the y-axis A times if $A>0$; but if $A<0$, then perform an $|A|$-fold stretching along the y-axis and a reflection about the x-axis. Of course, if $|\omega|<1$, then the compression along the x-axis is actually a stretching; in the same way, the $|A|$-fold stretching along the y-axis for $|A|<1$ is actually a compression.

Note particularly an important special case: if the graph of the function $y_1=f(x)$ has been sketched, the graph of the function $y=-f(x)$ can be obtained from it by a reflection about the x-axis.

9. *Construct the graph of the function* $y=\sin[2x-(\pi/3)]$.

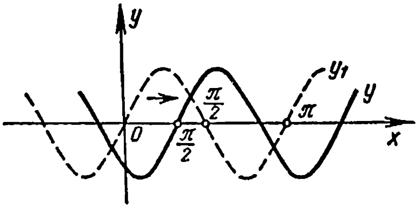
Fig. 33



Representing the given function in the form $y=\sin 2[x-(\pi/6)]$, we see immediately that for every value $x=x_0$ the value of y coincides with the value of $y_1=\sin 2x$, which corresponds to the value $x_0-(\pi/6)$ of its argument. And so to construct the graph of y, draw the graph of y_1 and then translate it $\pi/6$ units rightwards along the x-axis (Fig. 33).

A very common mistake in constructing the graph of the function y is as follows: the graph is drawn of the function y_1 and it is then translated rightwards by $\pi/3$ units along the x-axis (Fig. 34). It is

Fig. 34



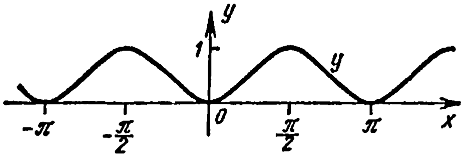
easy to see that this construction is incorrect, because the graph crosses the x-axis at the point $\pi/3$ (since the graph of the function y_1 cuts this axis at the origin and is then translated $\pi/3$ units rightwards!). Yet the value of the function y is clearly nonzero for the value of the argument $x=\pi/3$.

The technique used in this specific instance enables one to construct the graph of any function of the form $y=A\sin(\omega x+\varphi)$, $y=A\cos(\omega x+\varphi)$, $y=A\tan(\omega x+\varphi)$, etc., and also $y=a\sin\omega x+b\cos\omega x$.

This technique is of a general nature and permits *obtaining the graph of a function* $y=f(\omega x+\varphi)$, *where* $\omega\neq 0$ *and* φ *are specified numbers, if the graph of the function* $y_1=f(x)$ *has already been drawn*: it is sufficient to draw the graph of the function $y_2=f(\omega x)$ (it may be obtained by the method indicated in the solution of Problem 8) and then to translate it along the x-axis rightwards by an amount $|\varphi/\omega|$ if $\varphi/\omega<0$ or leftwards by φ/ω if $\varphi/\omega>0$ (see Problem 5).

It is sometimes useful to first transform the formula defining the functional relation; then the graph is readily drawn. In particular, it is always desirable to represent a complicated functional relationship as an easily surveyable combination of elementary functions, the graph of which combination is obtainable by familiar techniques (that was precisely how we constructed the graph in Problem 6).

Fig. 35



10. *Construct the graph of the function* $y=\sin^2 x$.

Since this function may be written as $y=1/2-1/2\cos 2x$, the graph of the function y is obtained by familiar techniques: the cosine curve $y_1=-1/2\cos 2x$, which is constructed by the technique described in the solution of Problem 8, must be translated 1/2 unit upwards (Fig. 35).

11. *Draw the graph of the function*

$$y=x^{\frac{1}{\log_{10} x}} \tag{4}$$

Employing familiar formulas involving logarithms, we see that $x^{1/\log_{10}x}=x^{\log_x 10}=10$, whence students often conclude immediately that the graph of the function (4) is the straight line $y=10$.

This conclusion is incorrect however. It is necessary to take into account the domain of definition of the function and the conditions under which the transformations that are carried out are legitimate.

The domain of the function (4) consists of the real numbers which satisfy the conditions: $x>0$, $x\neq 1$. Under these conditions, it is legitimate to carry out the transformation indicated above. And so the graph of the function (4) is the half-line $y=10$, $x>0$ with the point (1,10) deleted (Fig. 36, the arrowhead at any point indicates that that point does not belong to the graph).

12. *Sketch the graph of the function*

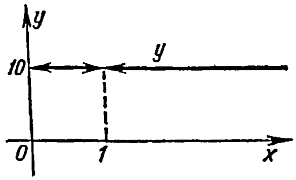
$$y=\log_{1/2}\left(x-\frac{1}{2}\right)+\log_2\sqrt{4x^2-4x+1} \tag{5}$$

First of all, perform an identity transformation of the second summand (see Sec. 1.4):

$$\log_2 \sqrt{4x^2-4x+1} = \log_2 \sqrt{(2x-1)^2} = \log_2 |2x-1| = 1 + \log_2 \left|x - \frac{1}{2}\right|$$

It is now clear that the domain of the function y is the set $x > 1/2$ (because the second term in the formula defining this function is

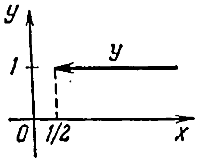
Fig. 36



meaningful for all $x \neq 1/2$, while the first is meaningful only for $x > 1/2$). However, the equation $\log_{1/2}(x-1/2) = -\log_2(x-1/2)$ is true for $x > 1/2$, and, hence, in its domain ($x > 1/2$) function (5) can be written $y=1$.

Thus the graph of the function y is the ray $y=1$, $x > 1/2$ [Fig. 37, the arrowhead at the point (1/2, 1) signifies that this point does not belong to the graph of the function (5)].

Fig. 37



Students often find it difficult to construct graphs of functions whose analytic expressions involve the absolute-value sign. The next few examples illustrate how the graphs of such functions are constructed.

13. *Construct the graph of the function* $y = |2 - 2^x|$.

First note that the proposed function can obviously be written in the form $y = |2^x - 2|$.

Consider the auxiliary function $y_1 = 2^x - 2$, the graph of which is readily drawn (by the technique described in the solution of Problem 4). How does the graph of the function y differ?

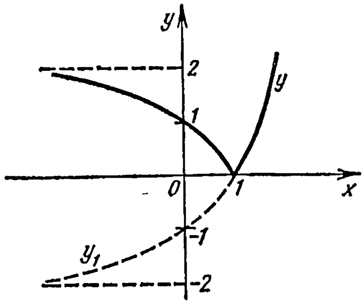
Recall the definition of absolute value; from this definition (see formula (2), Sec. 1.4) it follows that

$$y = \begin{cases} 2^x - 2 \text{ for values of } x \text{ for which } 2^x - 2 \geqslant 0, \\ \qquad\qquad \text{that is, for } x \geqslant 1, \\ -(2^x - 2) \text{ for values of } x \text{ for which } 2^x - 2 < 0, \\ \qquad\qquad \text{that is, for } x < 1 \end{cases}$$

It is then clear that the graph of the function y, for $x \geqslant 1$, coincides with the graph of the function y_1 and, for $x < 1$, is a curve symmetric to the graph of the function y_1 with respect to the x-axis (Fig. 38).

In precisely the same way we can *obtain the graph of the function* $y = |f(x)|$ *if the graph of the function* $y_1 = f(x)$ *has been drawn.* It suffices

Fig. 38

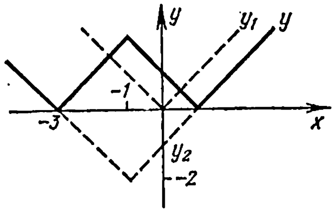


to replace the portions of the graph of y_1 lying below the x-axis by corresponding portions symmetric with respect to the x-axis [to find these portions we have to solve the inequality $f(x) < 0$].

14. *Construct the graph of the function* $y = ||x+1| - 2|$.

Here, without dropping the absolute-value signs, we can carry out the construction using the techniques indicated in the solution of Problems 6 and 13. Indeed, taking the graph of the function $y_1 = |x|$ (Fig. 39), translate it one unit leftwards along the x-axis and two units

Fig. 39



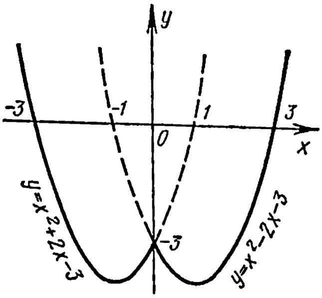
downwards along the y-axis. This yields the graph of the function $y_2 = |x+1| - 2$. Then replace the portion of the graph below the x-axis corresponding to $-3 \leqslant x \leqslant 1$, by the portion symmetric to it about the x-axis. The resulting polygonal line is the graph of the function y.

The general technique for constructing the graph of a function whose analytic expression contains an absolute-value sign consists in rewriting the expression of the functional relationship without using the absolute-value sign (see Sec. 1.4). In this case, the functional relationship on different portions of variation of the argument is, as a rule, described by different formulas. Quite naturally, on each of these portions, the graph must be constructed on the basis of the appropriate formula.

15. *Construct the graph of the function* $y=x^2-2|x|-3$.

To get rid of the absolute-value sign, consider separately two cases: $x\geqslant 0$ and $x<0$ (see Problem 1 of Sec. 1.4). If $x\geqslant 0$, then $y=x^2-2x-3$. It is easy to draw this parabola, then we take that portion which corresponds to nonnegative values of x. But if $x<0$, then $y=x^2+2x-3$. Draw this parabola and take that portion which corresponds to negative values of x. Taken together, the two pieces of the parabolas constitute the graph that interests us (Fig. 40).

Fig. 40



16. *Construct the graph of the function*

$$y=(|x+1|+1)(x-3) \tag{6}$$

By the definition of absolute value we can represent this function in the form

$$y=\begin{cases} [(x+1)+1](x-3)=(x+2)(x-3) & \text{if } x\geqslant -1, \\ [-(x+1)+1](x-3)=-x(x-3) & \text{if } x<-1 \end{cases}$$

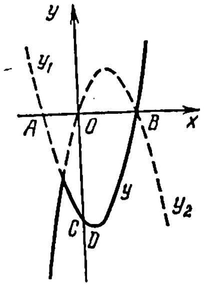
It now remains simply to sketch the curve, using the appropriate formula, for each of the indicated intervals ($x\geqslant -1$ and $x<-1$). Together, the two curves yield the graph of function (6).

Let us first consider the function $y_1=(x+2)(x-3)$. Ordinarily, students remove the brackets and perform a rather lengthy procedure of isolating a perfect square, whereas it is better not to remove the brackets because it is at once clear that this is a parabola, the graph of a quadratic trinomial; the parabola intersects the x-axis at the points $A(-2, 0)$ and $B(3, 0)$ (because -2 and 3 are the roots of the trinomial) and its branches are directed upwards (since the leading coefficient is positive). Substituting the value $x=0$ into the formula for the function y_1, we get the coordinates of the point C of intersection of this parabola with the axis of ordinates (y-axis): $C(0, -6)$. It is also easy to find the coordinates of the vertex D of this parabola. Since the parabola is symmetric about the vertical straight line passing through the vertex, its axis of symmetry bisects the line segment AB. It is therefore clear that the abscissa of the vertex is equal to $1/2$; the ordinate is computed directly and we get $D(1/2, -25/4)$.

Having constructed the parabola—the graph of the function y_1—we must isolate that portion which corresponds to the values $x \geqslant -1$ of the argument (Fig. 41).

The construction of the graph of the function $y_2 = -x(x-3)$ is similar. Take only that portion of the parabola which corresponds to the values $x < -1$ of the argument. The graph of function (6) is shown in Fig. 41 by the solid line.

Fig. 41



17. *Construct the graph of the function*

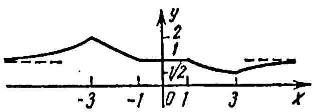
$$y = \frac{|x-3| + |x+1|}{|x+3| + |x-1|} \tag{7}$$

We first find the values of x for which each of the expressions under the absolute-value sign vanishes; they are $-3, -1, 1, 3$. By considering function (7) on each of the five intervals into which these values partition the number line, we obtain the following notation:

$$y = \begin{cases} 1 - \dfrac{2}{x+1} & \text{if } x < -3, \\ -\dfrac{x}{2} + \dfrac{1}{2} & \text{if } -3 \leqslant x < -1, \\ 1 & \text{if } -1 \leqslant x < 1, \\ \dfrac{2}{x+1} & \text{if } 1 \leqslant x < 3 \\ 1 - \dfrac{2}{x+1} & \text{if } 3 \leqslant x \end{cases}$$

The subsequent construction now follows familiar techniques (Fig. 42).

Fig. 42



It will be noted that if x increases without bound, the graph of the function (7) approaches without bound the straight line $y=1$, remaining all the time below it; but if x decreases without bound, then the graph approaches the same line without bound, remaining all the time above it.

18. *Construct the graph of the function* $y=|\sin x|+|\cos x|$.

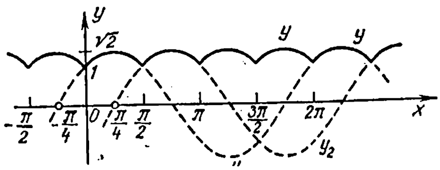
When drawing the graph of a periodic function it is often helpful to know that all values of such a function are repeated in every period. Thus, if a periodic function is given with period T, then it is sufficient to construct the graph on some segment of length T; for $0\leqslant x\leqslant T$, the portions of the graph on the intervals $T\leqslant x\leqslant 2T$, $2T\leqslant x\leqslant 3T$, $-T\leqslant x\leqslant 0$, etc., will have the very same shape.

It is clear that the number 2π is the period of the function y under consideration so that we can confine ourselves to the interval $0\leqslant x\leqslant 2\pi$. Partitioning this interval into four parts in each of which both $\sin x$ and $\cos x$ preserve sign, we get

$$y=\begin{cases} \sqrt{2}\sin\left(x+\frac{\pi}{4}\right) & \text{if } 0\leqslant x\leqslant \frac{\pi}{2}, \\ \sqrt{2}\sin\left(x-\frac{\pi}{4}\right) & \text{if } \frac{\pi}{2}\leqslant x\leqslant \pi, \\ -\sqrt{2}\sin\left(x+\frac{\pi}{4}\right) & \text{if } \pi\leqslant x\leqslant \frac{3\pi}{2}, \\ -\sqrt{2}\sin\left(x-\frac{\pi}{4}\right) & \text{if } \frac{3\pi}{2}\leqslant x\leqslant 2\pi \end{cases}$$

We now construct the graphs of $y_1=\sqrt{2}\sin[x+(\pi/4)]$ and $y_2=\sqrt{2}\times\sin[x-(\pi/4)]$ and then we take the portion of the curve y_1 on the interval from 0 to $\pi/2$, the portion of the curve y_2 on the interval from $\pi/2$ to π; and on the intervals from π to $3\pi/2$ and from $3\pi/2$ to 2π, we take the curves that are symmetric, about the x-axis, to the corresponding portions of the curves y_1 and y_2. Then, taking advantage of periodicity, we extend the resulting curve beyond the interval $0\leqslant x\leqslant 2\pi$ (the solid line in Fig. 43).

Fig. 43



It is clear from this graph that $\pi/2$ is the period of the given function so that we were overcautious in considering the interval from 0 to 2π. If we had realized from the very start that $\pi/2$ is the period of this function—and this is easy to demonstrate:

$$\left|\sin\left(x+\frac{\pi}{2}\right)\right|+\left|\cos\left(x+\frac{\pi}{2}\right)\right|=|\cos x|+|\sin x|$$

then the graph could be constructed much faster. This example shows that a careful preliminary analysis of the properties of a given function very often appreciably simplifies the construction of the graph.

19. *Graph the function*

$$y=\frac{\sin x}{\sqrt{1+\tan^2 x}}+\frac{\cos x}{\sqrt{1+\cot^2 x}} \tag{8}$$

At first glance this function might appear to be very complicated. However, by transforming the formula defining the given function we obtain a simpler notation for (8) that will permit drawing the required graph with comparative ease.

First of all, note that the domain of the function (8) is the entire number line with the exception of the points $x=n\pi/2$, where n is any integer (at each of these points, either $\tan x$ or $\cot x$ becomes meaningless).

Since for $x\neq n\pi/2$ the equations

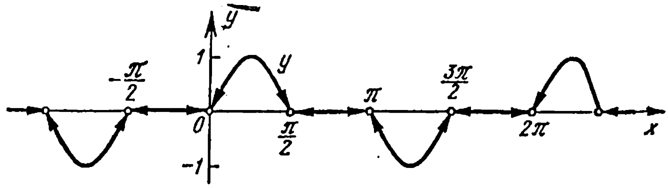
$$\sqrt{1+\tan^2 x}=\frac{1}{|\cos x|}, \quad \sqrt{1+\cot^2 x}=\frac{1}{|\sin x|}$$

are true, it is clear that the function (8) can, in its domain of definition, be written as

$$y=\sin x\cdot|\cos x|+\cos x\cdot|\sin x|$$

This is a periodic function with period 2π. The graph can be constructed as was done in Problem 18. It is shown in Fig. 44. Note once again, however, that the function (8) is not defined at the points $x=n\pi/2$, $n=0, \pm 1, \pm 2, \ldots$. In the figure this is indicated by arrowheads at the endpoints of the segments of the curve.

Fig. 44



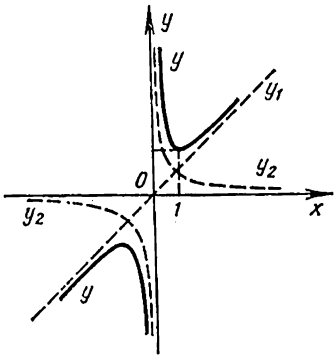
We will now consider some instances of the construction of complicated graphs in which the foregoing elementary devices do not suffice. Each of these examples has its peculiarities that must be taken into account when sketching the graph. In solving problems like those given below, one often has to reason in quite an unorthodox manner, so to speak. The approach should be to get onto an item that will give some clue to the construction.

20. *Graph the function* $y=\dfrac{x^2+1}{x}$.

Representing the given function as $y=x+(1/x)$, we apply a technique called *addition of graphs*, which means that the desired graph is constructed by "combining" two auxiliary graphs, $y_1=x$ and $y_2=1/x$. In other words, for each admissible value of the argument (that is, for every $x\neq 0$) the corresponding ordinate y is built up as an algebraic sum of the ordinates y_1 and y_2 corresponding to the same value of the argument (Fig. 45).

It is easy to figure out the shape of the graph of the function on the positive x-axis: for each value $x>0$, the corresponding ordinate of the straight line $y_1=x$ has to be increased by an ordinate of the hyperbola $y_2=1/x$ corresponding to the same value of x. It is quite obvious

Fig. 45



that for a positive x tending to zero, the expression $x+(1/x)$ tends to $+\infty$ (increases without bound), and for x tending to $+\infty$, the desired graph approaches the bisector $y_1=x$ without bound, since the summand $1/x$ becomes smaller and smaller. It is easy in this case to determine the smallest value of the function y (recall that so far we are only considering positive values of x): indeed, when $x>0$ the inequality $x+(1/x)\geqslant 2$ (see Sec. 1.8) holds true, which is to say the smallest value is equal to 2 and is reached when $x=1$.*

Construction of the graph is similar on the negative x-axis as well. Incidentally, we could take advantage of the fact that the function y is odd and, hence, its graph is symmetric about the origin.

21. *Construct the graph of the function* $y=x \sin x$.

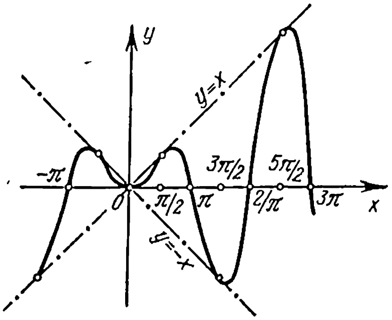
Take advantage of the fact that the formula defining this function is a product and we apply a technique called *multiplication of graphs*. The required graph will be constructed by "multiplying" two auxiliary graphs $y_1=x$ and $y_2=\sin x$. In other words, for each value of the argument, the corresponding ordinate y is constructed as a product of the ordinates y_1 and y_2 which correspond to the same value of the argument (Fig. 46).

We first construct the graph of the function y for nonnegative values of the argument. For each value of x we multiply the value of the corresponding ordinate of the straight line $y_1=x$ and the value of the ordinate of the sine curve $y_2=\sin x$, and are thus able to construct a

* It is somewhat more difficult but quite accessible to the student to prove that $x+(1/x)$ monotonically decreases when $0<x\leqslant 1$ and monotonically increases for $x\geqslant 1$

smooth curve that gives an approximate idea of the behaviour of th graph of the function y on the nonnegative x-axis. The aspect of th curve can be improved somewhat by plotting a few characteristi points. First of all, it is clear that $y=0$ for those values of x for whicl $\sin x=0$, and so the graph of the function y crosses the positive x-axi at the points $x=k\pi$, $k=0, 1, 2, \ldots$. Furthermore, for $x>0$ th obvious inequality $-x \leqslant x \sin x \leqslant x$ holds true; this means that fo

Fig. 46



positive values of the argument the graph of the function y does not extend above the straight line $y=x$ or below the straight line $y=-x$. In this case, the points of the graph of the function y that correspond to the values of $x>0$ for which $\sin x=1$, i.e., to the values $x=(\pi/2)+2k\pi$, $k=0, 1, 2, \ldots$, lie on the straight line $y=x$, and the points corresponding to the values of $x>0$ for which $\sin x=-1$, i.e., to the values $x=(3\pi/2)+2k\pi$, $k=0, 1, 2, \ldots$, lie on the straight line $y=-x$.

It is very easy to construct the graph of the function y on the negative x-axis: the function y is even and so its graph is symmetric about the y-axis.

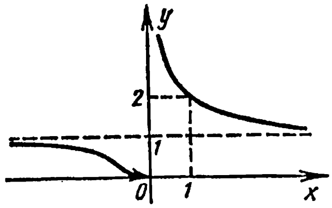
22. *Sketch the graph of the function* $y=2^{1/x}$.

Here we have to construct the graph of a function of a function. Such composite *functions* occur rather often at examinations. To graph such a function, the student must know the properties of the basic elementary functions and have a clear-cut idea of the consequent properties of combinations of these functions.

The domain of the function y consists of all real numbers except $x=0$. Since for $x>0$ the exponent $1/x>0$, it follows that $y>1$ for all positive values of the argument, by the property of an exponential function. Note that $y=2$ when $x=1$. If x increases without bound, then the expression $1/x$ decreases to zero monotonically, remaining positive (see the properties of a hyperbola), and so $2^{1/x}$ decreases to unity monotonically, remaining greater than unity (by the property of an exponential function). When x is positive and tends to zero, the exponent $1/x$ increases without bound and, hence, $2^{1/x}$ also increases without bound. This enables us to sketch a graph of the function y when $x>0$.

It is easy to demonstrate that the inequality $0<y<1$ is true on the negative x-axis. Using similar reasoning, we construct the graph of the function y for $x<0$ as well (Fig. 47; the arrowhead at the origin indicates that the origin does not belong to the graph).

Fig. 47



23. *Construct the graph of the function*

$$y=1-2^{1+\sin(x+1)}$$

If we represent this function as

$$y=1+(-2)\cdot 2^{\sin(x+1)} \tag{9}$$

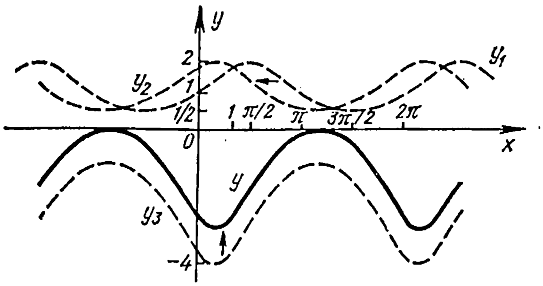
then it is clear that having the graph of the function $y_1=2^{\sin x}$ it is easy to obtain the graph of the function y by means of techniques discussed in the solution of Problems 4, 5, and 8.

And so let us first tackle the graph of the function y_1. This is a periodic function with period 2π; therefore, all we need to do is draw the graph on the interval $0\leqslant x\leqslant 2\pi$ (see Problem 18).

For $x=0$ the function y_1 assumes the value 1. If x is increased from 0 to $\pi/2$, then $\sin x$ monotonically increases from 0 to 1 and $2^{\sin x}$ monotonically increases from 1 to 2. If x then increases from $\pi/2$ to $3\pi/2$, then $\sin x$ monotonically decreases from 1 to -1 and $2^{\sin x}$ monotonically decreases from 2 to 1/2; in particular, for $x=\pi$ the function y_1 takes on the value 1. Finally, if x increases from $3\pi/2$ to 2π, then $\sin x$ monotonically increases from -1 to 0 and $2^{\sin x}$ monotonically increases from 1/2 to 1: the function y_1 has the value 1 for $x=2\pi$. All these statements about the behaviour of the function y_1 follow from the properties of a sine function and an exponential function (it is left to the reader to put a rigorous foundation under these statements). They permit determining the approximate behaviour of the graph of the function y_1 when $0\leqslant x\leqslant 2\pi$; it then remains to extend the resulting curve periodically over the entire x-axis (in Fig. 48, the graph of y_1 is shown as a dashed line).

Everything is now ready for the construction, using formula (9), of the graph of the function y. First translate the graph of the function y_1 one unit leftwards along the x-axis; this yield a curve which is the graph of the function $y_2=2^{\sin(x+1)}$ (see Problem 5). It is also a periodic function (with period 2π). It has a maximal value of 2 which it assumes at the points $x^* = (\pi/2) - 1 + 2k\pi$, $k = 0, \pm 1, \pm 2, \ldots$, it has a minimal value equal to 1/2 which it takes on at the points $x^{**} = -(\pi/2) - 1 + 2k\pi$, $k = 0, \pm 1, \pm 2, \ldots$ (Fig. 48). Stretching the curve y_2 by a factor of 2 along the y-axis and then reflecting it about the x-axis, we construct the graph of the function $y_3 = (-2) \cdot 2^{\sin(x+1)}$

Fig. 48



(see Problem 8). Note that this periodic function has a maximal value of -1 and a minimal value of -4 (Fig. 48). Finally, the graph of the function y is obtained by translating the curve of y_3 up one unit on the y-axis (see Problem 4).

The graph (solid line in Fig. 48) conveys the basic features of the behaviour of the function y. This is a periodic function (with period 2π) which vanishes at the points $x^{**} = -(\pi/2) - 1 + 2k\pi$, $k = 0, \pm 1, \pm 2, \ldots$ (with 0 the maximal value) and assumes a minimal value of -3 at the points $x^* = (\pi/2) - 1 + 2k\pi$, $k = 0, \pm 1, \pm 2, \ldots$. The function y varies monotonically in between the extremal values. When $x = 0$ the function y is equal to $1 - 2^{1+\sin 1}$ (note that $\sin 1$ is the sine of an angle of one radian!).

Of course, Fig. 48 gives only a rough idea of the graph of the function y, but that is as much as is ordinarily required at an examination.

24. *Construct the graph of the function* $y = \log_2 (1 - x^2)$.

First draw the graph of the auxiliary function $y_1 = 1 - x^2$. This parabola is shown in Fig. 49 by the dashed line. It is then necessary to construct the graph of the logarithm of this function.

For $x = 0$ we have $y = \log_2 1 = 0$. If x is increased from 0 to 1, then, as may be seen from the graph of the auxiliary function, $1 - x^2$ decreases from 1 to 0 and so $\log_2 (1 - x^2)$ decreases from 0 to $-\infty$. Similarly, if x decreases from 0 to -1, then $1 - x^2$ decreases from 1 to 0 and $\log_2 (1 - x^2)$ decreases from 0 to $-\infty$. For the remaining values of x, that is, for $x \leqslant -1$ and $x \geqslant 1$, we have $1 - x^2 \leqslant 0$, so that $\log_2 (1 - x^2)$ is meaningless. The graph of the function y is shown in Fig. 49 as a solid line.

Note that in the construction of this graph we did not start out by finding the domain of the function, which was obtained almost automatically. A preliminary determination of the domain of a function is frequently very useful, however.

25. *Construct the graph of the function* $y=\log_{\sin x} 1/2$.

The domain of this function is the collection of all values of x for which, simultaneously, $\sin x > 0$ and $\sin x \neq 1$, that is, the set

$$2k\pi < x < \frac{\pi}{2} + 2k\pi, \ \frac{\pi}{2} + 2k\pi < x < (2k+1)\pi,$$
$$k=0, \pm 1, \pm 2, \ldots$$

The function y is clearly periodic with period 2π. And so we can confine ourselves to an interval of length 2π, say, the interval $0 \leqslant x \leqslant$

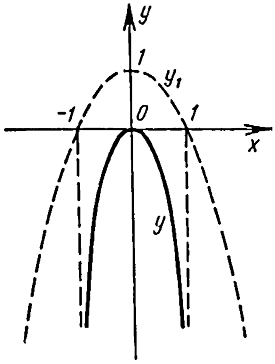


Fig. 49

$\leqslant 2\pi$. But not the whole of this interval lies in the domain of the function: the function is meaningful (over this interval) only for $0 < x < \pi/2$, $\pi/2 < x < \pi$. It is precisely on these intervals that we first of all have to construct the graph (then we can simply extend it over the entire domain because of its periodicity).

It will be seen that the function y can, in its domain, be rewritten as

$$y = \frac{1}{\log_{1/2} \sin x} \tag{10}$$

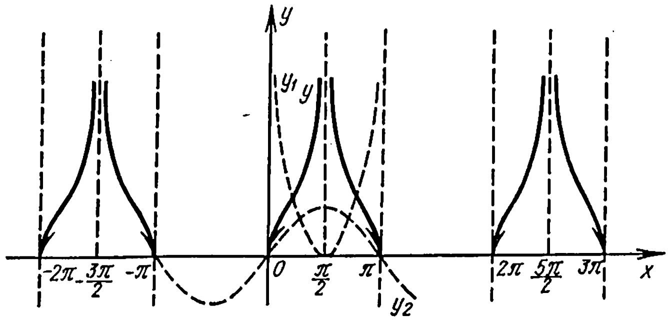
We first of all construct the graph of the auxiliary function $y_1 = \log_{1/2} \sin x$. It will only interest us over the interval $0 < x < \pi$. Taking the piece of the sine curve $y_2 = \sin x$ corresponding to this interval, we can use the same method as in the preceding problem (don't forget that the base of the logarithm $1/2 < 1$) to obtain the graph of the composite function y_1 (Fig. 50, the auxiliary graphs y_1 and y_2 are depicted by dashed lines).

We now consider the interval $0 < x < \pi/2$. Since for any value of x in this interval, the corresponding value of the function y is the *reciprocal* of the value of y_1 corresponding to the same value of the argument [see (10)], it is easy to obtain a rough sketch of the graph of y for

$0<x<\pi/2$ (the solid line in Fig. 50; the arrowhead on the curve at the origin indicates that this point does not belong to the graph).

It is easy to prove that by using familiar properties of elementary functions the function y monotonically increases when x varies from 0 to $\pi/2$; if x increases from 0 to $\pi/2$, then $\sin x$ increases monotonically

Fig. 50



from 0 to 1, and then $\log_{1/2}\sin x$ decreases monotonically from $+\infty$ to 0; and hence, [see (10)], the value of y increases monotonically from 0 to $+\infty$. Let us stress that if x approaches $\pi/2$, remaining all the time less than this value, then the value of the function y_1 tends to zero, remaining all the time positive, and therefore the value of y increases without bound. But if x approaches zero and remains positive, then the value of y_1 increases without bound and so the value of y tends to zero (although it does not take on the value 0).

The construction is similar for the graph of the function at hand when $\pi/2<x<\pi$.

It is useful to note that just about the same reasoning as is used in the construction of the graph of the function y on the basis of the graph of the auxiliary function y_1 [through the use of formula (10)] enables us *to construct the graph of* $y=1/f(x)$ *if the graph of the function* $y_1=f(x)$ *is known.*

Considering Fig. 50 in more detail, we note that we did not obtain a complete description of the behaviour of the graph of the given function in the foregoing solution (for instance, the fact that this graph is bent in the specific way as it is shown in the figure was not even discussed). But this is not required since it goes beyond the scope of the elementary means at the disposal of the student. And a *rough* sketch of the graph can, as we have seen, be made with relative ease.

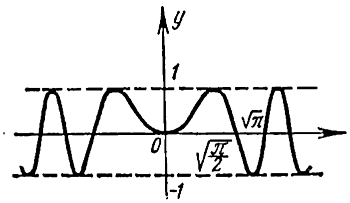
True, the shape of the curve could be improved a bit by computing a table of values of the function for "convenient" values of the argument and taking these into account when drawing the curve. As a rule, examination questions do not require such improvement. The important thing is to be able to sketch a curve that conveys the general aspect and characteristic features of the graph.

26. *Construct the graph of the function* $y=\sin x^2$.

First of all, do not confuse the notation $\sin x^2$ with $\sin^2 x$: the former means $\sin(x^2)$, the latter means $(\sin x)^2$.

Let us first consider the nonnegative values of the argument and partition the semi-axis $x \geqslant 0$ into intervals over which the function y increases or decreases. If x^2 increases from 0 to $\pi/2$ (which is to say that x increases from 0 to $\sqrt{\pi/2}$), then $\sin x^2$ increases from 0 to 1; if x^2 increases from $\pi/2$ to $3\pi/2$ (i.e., x increases from $\sqrt{\pi/2}$ to $\sqrt{3\pi/2}$), then $\sin x^2$ decreases from 1 to -1; if x^2 increases from $3\pi/2$ to $5\pi/2$ (i.e., x increases from $\sqrt{3\pi/2}$ to $\sqrt{5\pi/2}$), then $\sin x^2$ increases from -1 to 1, and so on. The graph of the function y therefore is of a wavelike nature with an amplitude of 1 (Fig. 51).* It is easy to

Fig. 51



obtain the x-intercepts of this graph: all you have to do is solve the equation $\sin x^2 = 0$. It is clear that the nonnegative roots of this equation are the numbers $x = \sqrt{k\pi}$, $k = 0, 1, 2, \ldots$.

On the negative x-axis, the graph is drawn at once since the function y is even.

We conclude this section with some problems of a different nature, but also connected with graphical constructions in the plane with a specified coordinate system.

27. *Find a set of points, in the plane, whose coordinates x and y satisfy the system of inequalities*

$$\begin{aligned} 5x + 3y &\geqslant 0 \\ y - 2x &< 2 \end{aligned} \qquad (11)$$

From the first inequality we have $y \geqslant -5x/3$. We first of all graph the function $y = -5x/3$ (Fig. 52). Then the points whose coordinates satisfy the *equation* $y = -5x/3$ lie on the constructed straight line and the points whose coordinate y *exceeds* $-5x/3$ will lie *above* this line. Thus, the set of points whose coordinates satisfy the first inequality of (11) will constitute the half-plane lying *above* the straight line $y = -5x/3$ (the straight line included; in Fig. 52 this region is denoted by vertical hatching).

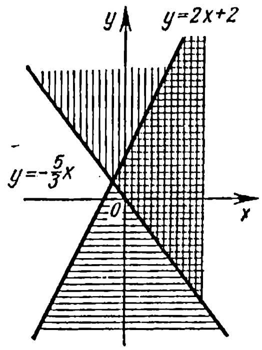
Similarly, from the second inequality of (11) we have $y < 2x + 2$ so that the set of points whose coordinates satisfy the second inequality

* It will be noted that $y = \sin x^2$ is not a periodic function.

of (11) will constitute the half-plane lying below the straight line $y=2x+2$ (the line itself is not included; in Fig. 52 this region is indicated by horizontal hatching).

Hence, the points of the plane whose coordinates x and y satisfy the system of inequalities (11) lie in the common portion (intersection)

Fig. 52



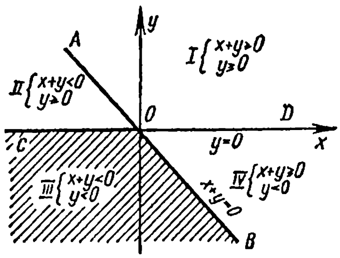
of the two resulting half-planes; this is an angular region (in Fig. 52 the desired set is indicated by double cross-hatching). Here, one of the bounding rays of the region—a piece of the straight line $y=-5x/3$—is included in the sought-for set, while the other—a piece of the straight line $y=2x+2$—is not included (the vertex A of the angular region, the intersection point of the straight lines $y=-5x/3$ and $y=2x+2$, does not belong to this set either).

28. *Determine the set of points, in a plane, whose coordinates x and y satisfy the relation*

$$|x+y|=|y|-x \tag{12}$$

As in the solution of other problems involving absolute values, it is useful first of all to attempt to get rid of the absolute-value sign.

Fig. 53



To do this (see Fig. 53), construct straight lines in the plane: $x+y=0$ (the bisector AOB of the second and fourth quadrants) and $y=0$ (the axis COD of abscissas). Clearly, the coordinates x and y of any point above the straight line $x+y=0$ satisfy the inequality $x+y>0$, and for any point below this line the inequality $x+y<0$ holds. Similarly, any point in the upper (relative to the x-axis) half-plane has a positive ordinate, and any point in the lower half-plane has a negative ordinate.

These straight lines partition the plane into four regions (Fig. 53), and it is clear that in each one of these regions the expressions $x+y$ and y preserve sign for all points (x, y). It is therefore advisable, in each of these regions, to seek separately the points whose coordinates x and y satisfy the relation (12).

For any point (x, y) of Region I (the angular region DOA including the bounding rays as well) we have the inequalities $x+y \geqslant 0$, $y \geqslant 0$. Hence, in Region I relation (12) takes the form $x+y=y-x$, or $x=0$. But this last equation is satisfied by the coordinates of the point of the positive y-axis (by no means all the points of the y-axis, since we are only interested in those points which lie in Region I, and the negative y-axis does not belong to this region).

For any point of Region II (the angular region AOC; of the boundary rays only the ray CO is included) we have the inequalities $x+y<0$, $y \geqslant 0$, and for this reason relation (12) takes the form $-(x+y)=y-x$, or $y=0$, in Region II. This latter equation is satisfied by the points of the negative x-axis (the other points of the x-axis do not lie in Region II).

The inequalities $x+y<0$, $y<0$ hold for any point of Region III (the angular region COB excluding the bounding rays), and so relation (12) in Region III assumes the form $-(x+y)=-y-x$, or $0=0$. This means that the coordinates of any point of Region III satisfy relation (12).

Finally, for any point of Region IV (the angular region BOD including only the ray BO) we have $x+y \geqslant 0$, $y<0$, and so relation (12) in Region IV assumes the form $x+y=-y-x$, or $x+y=0$. This latter equation is clearly satisfied by those points of Region IV which lie on the bisector of the fourth quadrant.

Thus, in the plane, the set of points whose coordinates x and y satisfy relation (12) is the angular region between the negative x-axis and the bisector of the fourth quadrant (including the bounding rays) and the positive y-axis (Fig. 53).

29. *A system of Cartesian coordinates is given in a plane. Represent the region of this plane filled with all the points whose coordinates satisfy the inequality*

$$\log_x \log_y x > 0 \tag{13}$$

Note right off that x and y which satisfy condition (13) are such that $x>0$, $y>0$, $x \neq 1$ and $y \neq 1$. Since the properties of logarithms are different for bases that exceed unity or are less than unity, it is natural to consider two cases.

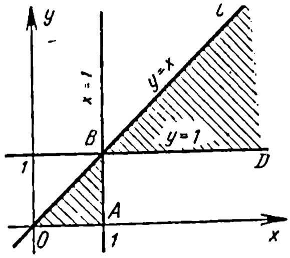
(a) Let $x>1$. Then by the properties of logarithms, the inequality (13) will hold true if the inequality $\log_y x > 1$ is valid. It will be recalled that the logarithms of numbers greater than unity to a base less than unity are negative. And so the inequality $\log_y x > 1$ cannot be valid for y in the interval $0 < y < 1$.

Thus, the inequality $\log_y x > 1$ can hold only when $y > 1$. But if $y > 1$, then all $x > y$ will be solutions to the inequality $\log_y x > 1$.

Thus, if $x > 1$, then for inequality (13) to hold, y must be greater than unity: $y > 1$, and the original inequality will be satisfied by those points for whose coordinates the condition $x > y$ is also valid.

If we depict the set of these points in a drawing, it will be seen that the set is the interior of the angular region CBD (Fig. 54).

Fig. 54



(b) Now let $0 < x < 1$. Reasoning in similar fashion, we find that the condition of the problem is satisfied by points for whose coordinates the conditions $0 < y < 1$ and $y < x$ are fulfilled. The set of these points is the interior of the triangle AOB (Fig. 54).

Consequently, points whose coordinates satisfy inequality (13) form the region cross-lined in Fig. 54 [the coordinates of the boundary points of this region do not satisfy relation (13)].

30. *Find all the points, in the plane, whose coordinates x and y satisfy the inequality*

$$\cos x - \cos y > 0$$

Using a familiar trigonometric identity, rewrite the given inequality as

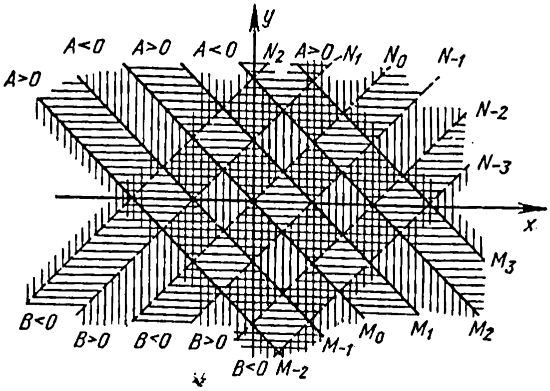
$$\sin \frac{x+y}{2} \sin \frac{y-x}{2} > 0$$

This inequality holds true for all points whose coordinates x and y are such that the expressions $A = \sin[(x+y)/2]$ and $B = \sin[(y-x)/2]$ have the same signs.

Let us first investigate the expression A. Solving the equation $\sin[(x+y)/2] = 0$, we find $x+y = 2k\pi$, $k = 0, \pm 1, \pm 2, \ldots$. Geometrically, this signifies that the expression A reduces to zero only the coordinates x and y of points in the plane which lie on one of the straight lines $y = -x + 2k\pi$, $k = 0, \pm 1, \pm 2, \ldots$ (in Fig. 55 these straight lines are indicated by solid lines). For the sake of brevity, we denote by M_k the straight line $y = -x + 2k\pi$ for integral k (thus, the straight line M_0 is the bisector of the second and fourth quadrants, M_{-1} has the equation $y=-x-2\pi$, etc.).

All the straight lines M_k are parallel and partition the plane into strips. Let us agree to call the strip between two adjacent lines M_k and M_{k+1} the strip $\{M_k, M_{k+1}\}$, the lines M_k and M_{k+1} themselves

Fig. 55



not being included in this strip. For example, $\{M_0, M_1\}$ is the strip between the straight lines $y=-x$ and $y=-x+2\pi$, that is, the set of points whose coordinates x and y satisfy the inequality $0<x+y<2\pi$. Analogously, in the general case, the strip $\{M_k, M_{k+1}\}$ is the set of points whose coordinates x and y satisfy the inequality $2k\pi<x+y<$ $<2(k+1)\pi$.

Now let us determine the set of points whose coordinates x and y satisfy the inequality $\sin[(x+y)/2]>0$. This inequality can easily be solved; it is valid for

$$2\cdot 2n\pi < x+y < 2(2n+1)\pi, \qquad n=0, \pm 1, \pm 2, \ldots$$

Geometrically, this signifies that the expression A is positive for the coordinates x and y of all points lying in each of the strips $\{M_{2n}, M_{2n+1}\}$, $n=0, \pm 1, \pm 2, \ldots$, i.e., in each strip bounded from below by the straight line M_{2n} with even index and from above by M_{2n+1}.

In the same way, by solving the inequality $\sin[(x+y)/2]<0$ we convince ourselves that expression A is negative for the coordinates x and y of all points lying in each of the strips $\{M_{2n-1}, M_{2n}\}$, $n=0$, $\pm 1, \pm 2, \ldots$. In each of the strips $\{M_k, M_{k+1}\}$ of Fig. 55 is indicated the sign of expression A: the strips where $A>0$ are marked with horizontal lines, the strips where $A<0$ are marked with vertical lines.

Let us now examine expression B. Similar reasoning shows that expression B is made to vanish by the coordinates x and y of points lying on the straight lines $y=x+2m\pi$, $m=0, \pm 1, \pm 2, \ldots$ (these lines are shown dashed in Fig. 55). For integral m we denote the straight line $y=x+2m\pi$ by N_m, and we will agree to call the strip between adjacent lines N_m and N_{m+1} the strip $\{N_m, N_{m+1}\}$ (the straight lines N_m and N_{m+1} themselves do not belong to this strip). It is easy to verify that the strip $\{N_m, N_{m+1}\}$ is a set of points whose coordinates x and y satisfy the inequality $2m\pi < y - x < 2(m+1)\pi$.

Solving the inequalities $B > 0$ and $B < 0$, we see that B is positive for the coordinates x and y of all points lying in each of the strips $\{N_{2p}, N_{2p+1}\}$, $p = 0, \pm 1, \pm 2, \ldots$, which is to say in each strip bounded from below by the line N_{2p} with even index and from above by N_{2p+1}. Furthermore, the expression B is negative for the coordinates x and y of all points lying in each of the strips $\{N_{2p-1}, N_{2p}\}$, $p = 0$, $\pm 1, \pm 2, \ldots$. In Fig. 55, in each strip $\{N_m, N_{m+1}\}$ is indicated the sign of the expression B: strips with $B > 0$ are marked with vertical lines, those with $B < 0$ are marked with horizontal lines.

It is now easy to describe the set of points, in the plane, whose coordinates x and y satisfy the inequality $A \cdot B > 0$: it includes all rectangles (excluding their contours) which are double cross-hatched in Fig. 55).

Exercises

Find the domains of the following functions.

1. $y = 5\sqrt{1 - 4x^2}$.
2. $y = \log_{x-1}(2 - x - x^2)$.
3. $y = \arccos 2^{|x|}$.
4. $y = \arcsin(\tan x)$.
5. $y = \sqrt[4]{x^{-2}(x-2)(x-3)}$.
6. $y = \sqrt{\log_{10}\cos(2\pi x)}$.

Construct the graphs of the following functions.

7. $y = -5$.
8. $y = \pi(x+1)$.
9. $y = x(1-x)$.
10. $y = x^2 + 5|x-1| + 1$.
11. $y = |-3x+2| - |2x-3|$.
12. $y = |x^2 - 3x + 2| + |5 - x|$.
13. $y = (x+1)(|x| - 2)$.
14. $y = |x+1| \cdot (|x| - 2)$.
15. $y = \dfrac{2x+1}{2-x}$.
16. $y = \dfrac{2x-6}{3-x}$.
17. $y = 1 - 1/|x|$.
18. $y = (1/3)^{-2x+1}$.
19. $y = 2 \cdot 3^{x+1} - 1$.
20. $y = 10^{-|x|}$.
21. $y = -\log_{10}(2x+1)$.
22. $y = \log_{|x|} 2$.
23. $y = |\log_{1/\pi} x^2|$.
24. $y = \sqrt{\log_{10}\sin x}$.
25. $y = \sin^2 x + \cos^2 x$.
26. $y = \sin^2 x - \cos^2 x$.
27. $y = \sqrt{3}\sin 2x + \cos 2x$.
28. $y = 2\sin|2x|$.
29. $y = 2\tan[-2x + (\pi/4)]$.
30. $y = -\cos^2[x - (\pi/6)]$.
31. $y = 1 + \sqrt{x}$.
32. $y = x + \sin x$.
33. $y = \log_{1/2}\dfrac{1}{1-x^2}$.
34. $y = \left(\dfrac{1}{2}\right)^{\tan x}$.
35. $y = \log \quad \tan x$.
36. $y = \sin(\arccos x)$.
37. $y = \cos 2x - \sqrt{1 - \sin 2x}\,(\sin x + \cos x)$.

38. $y=\frac{|x-2|+1}{|x+3|}$.

39. $y=3+2^{3\cos\frac{x}{3}}$.

How do the graphs of the following functions differ?

40. $y_1=\log_3 x^2$ and $y_2=2\log_3 x$.

41. $y_1=2^{\log_2 x}$ and $y_2=x$.

42. $y_1=\tan x\cot x$ and $y_2=1$.

Sketch, in a plane wit' a given coordinate system, the set of points whose coordinates x and y satisfy the following relations.

43. $|y-1|=x^2-4x+3$.

44. $|x|+x=|y|+y$.

45. $|x-2|+|y+1|\leqslant 1$

46. $|x-y|>2$.

47. $|2x+y|+|2x-y|<4$.

48. $|y|=\sin x$.

49. $\cos 2x+\cos y=0$.

50. $x\geqslant\sin|y|$.

*51. $\log_{|\sin x|} y>0$.

52. $x>\log_2|y|$, $y<x$.

53. A rectangular parallelepiped has altitude 1/2 and sides of the base x and y. Find the relationship between y and x and depict it graphically if we know that the lateral surface area of the parallelepiped is equal to the area of the base.

54. In a regular quadrangular pyramid of height y and side x of base, the area of the base is 1 unit more than the area of the cross section drawn through the vertex of the pyramid and the diagonal of the base. Find the relationship between y and x and represent it graphically.

55. The base of a first right cylinder is a circle of radius y, the base of a second right cylinder is a circle of radius x, the radius of the first circle being greater than that of the second. The altitude of the first cylinder is 1/8, the altitude of the second, 1/2. Find the relationship between y and x and represent it graphically if the difference between the area of the lateral surface and the area of the base of the first cylinder is equal to the area of the lateral surface of the second cylinder.

Chapter 2 TRIGONOMETRY

2.1 General remarks on trigonometry

The student usually knows the definitions of the trigonometric functions of an *angle*. However, like all elementary functions studied in algebra, the trigonometric functions are ultimately viewed as functions of a *numerical* argument. Yet difficulties are caused by phrases like this: the *sine of a given number*.

When asked *what* $\sin\pi$ *means*, the student ordinarily is fast with the following answer: $\sin\pi=0$. But it is not *what* $\sin\pi$ *is equal to* but *what the symbol stands for*, how one is to understand the notation $\sin\pi$. The sine of the number π, i.e. $\sin\pi$, is the sine of an angle of π radians, which is to say, an angle of 180°.

The definition of trigonometric functions of a numerical argument is approached gradually. These functions are first defined as the functions of an arbitrary (positive or negative) angle. Then the introduction of the radian measure of angles enables us to associate with every real number a a definite angle of a radians, and, conversely, every angle is uniquely associated with a real number, which is the size of the angle in radians. Then, finally, we can define the trigonometric functions of a numerical argument: *the trigonometric function of a number a is that same trigonometric function of an angle of a radians.* Thus, from a given number we find the corresponding angle, and the trigonometric functions have already been determined for every angle.

Thus, for example, $\sin 10$ is the sine of an angle of 10 radians. In other words, we have to take an xy-system of coordinates, the unit circle with centre at the origin and find the point M on the circumference of the circle such that the vector $\overrightarrow{OM}$ forms an angle of 10 radians ($\approx 570°$), measured counterclockwise, with the positive x-axis.* Then

* It is well to recall here the common illusion that in measuring angles in degrees we get a concrete number while radian measurement only yields an abstract number. Actually, any measurement will always yield a concrete number, whether 5 km, 28° or 10 radians. It is a purely conventional matter that when measuring angles in radians we leave out the name of the unit and often say "an angle equal to 3π" instead of "an angle equal to 3π radians".

the ordinate of the point M (which is a number!) will be the sine of an angle of 10 radians, in other words, it will equal $\sin 10$.

We see that the final definition of trigonometric functions does not involve any angles at all but establishes a relationship between numbers. The introduction of angles is only an *auxiliary*, intermediate, step, whose necessity is dictated solely by considerations of the teaching process.

The student often makes use of the symbol ∞, which is not generally used in elementary mathematics. For instance, a common but meaningless formula is $\tan 90^\circ=\infty$ or its verbal statement: "The tangent of a right angle is equal to infinity." On occasion the student will even "justify" it thus: $\tan 90^\circ=\sin 90^\circ/\cos 90^\circ=1/0=\infty$. Remember that all such manipulations are senseless.

Of course, trigonometric formulas must be memorized, but the student should be able to derive every one of them because the ability to *derive* a formula is far more important than a simple knowledge of the formula without an understanding of its derivation.

An analysis of the proofs of trigonometric formulas encountered in secondary school shows that any one of them can be obtained rather quickly if the student has memorized the definitions and the basic properties of the functions $\sin x$, $\cos x$, $\tan x$, $\cot x$,* the relation $\sin^2 x+\cos^2 x=1$ and the addition formulas. Using this as a basis, it is easy to derive the reduction formulas, the formulas for transforming a product of trigonometric functions into a sum and conversely, and so forth.

Suppose we need the formula of the sine of half an angle.** By the addition formula and the relation $\sin^2 x+\cos^2 x=1$, we can write down immediately

$$\cos\alpha=\cos\left(\frac{\alpha}{2}+\frac{\alpha}{2}\right)=\cos\frac{\alpha}{2}\cos\frac{\alpha}{2}-\sin\frac{\alpha}{2}\sin\frac{\alpha}{2}$$

$$=\cos^2\frac{\alpha}{2}-\sin^2\frac{\alpha}{2}=\left(1-\sin^2\frac{\alpha}{2}\right)-\sin^2\frac{\alpha}{2}=1-2\sin^2\frac{\alpha}{2}$$

whence

$$\sin^2\frac{\alpha}{2}=\frac{1-\cos\alpha}{2},\quad \text{i.e.}\quad \sin\frac{\alpha}{2}=\pm\sqrt{\frac{1-\cos\alpha}{2}} \tag{1}$$

On the other hand, it is not wise to overestimate the value of deriving any formula and thus not to memorize them, otherwise too much time at an examination will be taken up in deriving the needed formulas. The range of formulas that the student has in his so-called active memory should be rather extensive.

* The notations $1/\sin x=\operatorname{cosec} x$ (cosecant), $1/\cos x=\sec x$ (secant) are also encountered.

** It would be better to say the formula for the sine of half a number, the formula for the cosine of twice a number, etc., but we will not here depart from tradition and will use the word "angle" instead. Bear in mind, however, that "angle" and "number" may be used interchangeably throughout the sequel.

Many formulas can be derived in a variety of ways. Any one is as good as another. The student should pick the one he likes best. The only requirement is that the derivation be carried out *correctly*. The proofs of formulas given in standard textbooks should be viewed simply as possible versions of their derivation. If the student can offer a different derivation, all well and good. Preference should be given to simple derivations involving straightforward reasoning.

Take care that in deriving a formula you do not rely on another formula which itself is obtained from the one being proved. For example, the evenness of the function $\cos x$ is often demonstrated by the student as follows:

$$\cos(-x) = \cos(0-x) = \cos 0 \cos x + \sin 0 \sin x = \cos x \qquad (2)$$

This chain of equations utilizes the addition formula $\cos(\alpha-\beta)$ for the case when $\alpha < \beta$. For this reason, the proof given in (2) of the evenness of the cosine function may be regarded as proper only if the student is able to justify the addition formula $\cos(\alpha-\beta)$, $\alpha<\beta$, without recourse to this property of the cosine.

A word on the rather common confusion with regard to understanding the sign $\pm$ in formulas like (1). Some students believe that "the sine of half an angle can assume two values", others think that only one of these values can be chosen (that is, corresponding either to the plus sign or to the minus sign), but they are unable to explain properly when a given value is to be taken.

Actually, for any *fixed* value α we must take, in formula (1), *either* the value corresponding to the plus sign *or* the value corresponding to the minus sign (but never both values at the same time!).* The question of which value to choose depends on the *quadrant in which the angle $\alpha/2$ lies*: if it is in the first or second quadrant, then one should take the value with the plus sign, if it is in the third or fourth quadrant, then take the minus sign.**

Thus, the sign $\pm$ in formula (1) does not point to any kind of "two-signed nature" of the sine of half an angle. We have to put that sign because $\sin(\alpha/2)$ can (for different values of α) assume either positive or negative values, whereas the expression $\sqrt{1/2(1-\cos\alpha)}$ is nonnegative for all values of α.

Actually, formula (1) signifies that a knowledge of the value of $\cos\alpha$ does not *uniquely* define the value of $\sin(\alpha/2)$, but only defines the *absolute value* of $\sin(\alpha/2)$. In order to determine the value of $\sin(\alpha/2)$ one must know the $\cos\alpha$ and the quadrant in which the angle $\alpha/2$ lies.

* Provided $\cos\alpha \neq 1$, because when $\cos\alpha = 1$ both values are zero.

** It is easy to see that the plus sign is taken in (1) when $4n\pi \leqslant \alpha \leqslant 4n\pi + 2\pi$ and the minus sign when $4n\pi + 2\pi \leqslant \alpha \leqslant 4n\pi + 4\pi$ (where n is any integer). For $\alpha = 2n\pi$, it is immaterial which sign is chosen since for this value of the angle, $\sin(\alpha/2) = 0$.

It is therefore preferable to avoid notations like (1) and employ the more exact notation

$$\left|\sin\frac{\alpha}{2}\right|=\sqrt{\frac{1-\cos\alpha}{2}}$$

A similar situation is encountered in certain problems when it is required to compute the value of one trigonometric expression in terms of another one. Bear in mind that, generally speaking, on the basis of the value of one trigonometric function of an angle only the absolute values of the other functions of this angle can be defined unambiguously. To determine the values themselves of these functions, we have to know, for instance, in what quadrant the angle at hand is located.

Let us examine a problem.

It is known that $\sin\beta=4/5$ *and* $0<\beta<\pi$. *What is the ratio*

$$\frac{\sqrt{3}\sin(\alpha+\beta)-\frac{2}{\cos(\pi/6)}\cos(\alpha+\beta)}{\sin\alpha}$$

equal to if (a) *the angle* β *is acute*, (b) *the angle* β *is obtuse?*

This ratio, for brevity we denote it by M, can readily be reduced to the form

$$M=\frac{\sin\alpha\,(3\cos\beta+4\sin\beta)+\cos\alpha\,(3\sin\beta-4\cos\beta)}{\sqrt{3}\sin\alpha}$$

To compute the value of this expression we have to know $\sin\beta$ and also the value of $\cos\beta$. Since $\sin\beta=4/5$, it follows that we can immediately find the absolute value of the cosine of this angle: $|\cos\beta|=\sqrt{1-\sin^2\beta}=3/5$. The sign of the cosine must be determined depending on the quadrant under consideration.

When the angle β is *acute* we have $\cos\beta=3/5$ and, therefore, as is readily computed, $M=5/\sqrt{3}$. But if angle β is *obtuse*, then $\cos\beta=-3/5$ and, hence, $M=\sqrt{3}(7+24\cot\alpha)/15$.

Unfortunately, not all students are able to find the values of the argument for which a given formula is valid. One often hears things like this: "All the trigonometric formulas derived in the textbook are identities, which means they hold true for all values of the arguments." This is not so. The trigonometric formulas are valid only for *admissible* values of the arguments.

In particular, the formula $\sin 2\alpha=2\sin\alpha\cos\alpha$ is indeed valid for an arbitrary value of α, whereas the formula $\tan\alpha\cot\alpha=1$ is meaningful only for values of α different from $k\pi/2$, where k is an arbitrary integer (because for $\alpha=k\pi/2$ one of the functions that enter into the formula is not defined).

Therefore, when writing down a trigonometric formula always bear in mind the values of the letters for which it is valid.

The problem of finding these values is solved by determining the values of the arguments for which *each* of the component functions is meaningful. If for some value of the argument at least one of the component functions becomes meaningless, then that value of the argument must be discarded.

Consider, say, the formula

$$\tan\alpha-\tan\beta=\frac{\sin(\alpha-\beta)}{\cos\alpha\cos\beta} \quad (3)$$

The left member is defined if $\alpha\neq(\pi/2)+k\pi$ and $\beta\neq(\pi/2)+n\pi$, where k and n are arbitrary integers, since for each of these values either $\tan\alpha$ or $\tan\beta$ becomes meaningless. The right member of (3) is meaningful for those values of α and β for which $\cos\alpha\cos\beta\neq0$; it is easy to verify that we come to the same restrictions on α and β. We thus see that the left and right members of (3) exist under the same conditions:

$$\alpha\neq\frac{\pi}{2}+k\pi,\ \beta\neq\frac{\pi}{2}+n\pi\ (k,\ n \text{ integers})$$

These two inequalities yield the conditions under which the formula for the difference of two tangents is valid.

A somewhat more complicated analysis is required for the formula

$$\tan(\alpha-\beta)=\frac{\tan\alpha-\tan\beta}{1+\tan\alpha\tan\beta} \quad (4)$$

Its left member is defined if $\alpha-\beta\neq\pi/2+k\pi$; only for these values of α and β does the expression $\tan(\alpha-\beta)$ become meaningless. For the right member of (4) to be meaningful, $\tan\alpha$ and $\tan\beta$ must be defined, that is, it must be true that $\alpha\neq\pi/2+n\pi$, $\beta\neq\pi/2+m\pi$, where n and m are integers. But this is not all: the denominator $1+\tan\alpha\tan\beta$ must be nonzero. Since $\tan\alpha$ and $\tan\beta$ are defined (we have already assumed that), the condition $1+\tan\alpha\tan\beta\neq0$ may be rewritten as $\cos(\alpha-\beta)\neq0$. Hence, the denominator in the right member of (4) is nonzero if $\alpha-\beta\neq\pi/2+k\pi$, where k is an integer.* Hence, the formula for the tangent of a difference holds true provided that

$$\alpha\neq\frac{\pi}{2}+n\pi,\ \beta\neq\frac{\pi}{2}+m\pi,\ \alpha-\beta\neq\frac{\pi}{2}+k\pi\ (n,\ m,\ k \text{ integers})^{**} \quad (5)$$

---

* Note particularly that in order for a pair of values α, β to be excluded from the domain of admissible values of formula (4) the only thing required is that *either* $\alpha=\pi/2+n\pi$ for some integer n (β *arbitrary*) *or* $\beta=\pi/2+m\pi$ for some integer m (α arbitrary), *or* $\alpha-\beta=\pi/2+k\pi$ for an integer k. Hence, if *even one* of these equations is valid, then formula (4) is not true for the appropriate values of the arguments.

** The conditions (5) are sometimes written thus:

$$\alpha\neq\frac{\pi}{2}+k\pi,\quad \beta\neq\frac{\pi}{2}+k\pi,\ \alpha-\beta\neq\frac{\pi}{2}+k\pi,\ k=0,\ \pm1,\ \pm2,\ \dots$$

It is assumed here that in *each* of the inequalities the number k runs through all integers *independently* of the other two inequalities.

A comparison of the foregoing analysis of the conditions under which formulas (3) and (4) are meaningful points to one essential difference between these formulas. Whereas the left and right members of the formula of the difference of two tangents exist or do not exist *for one and the same* set of values of the arguments, the domains of existence of the left and right members of the formula for the tangent of a difference *are different*. For instance, the left member of (4) is meaningful for $\alpha=\pi/2$, $\beta=\pi/4$, whereas the right member is not defined for these values of α and β.

It is possible to give other instances of trigonometric formulas *in which the left and right members have different domains* (*domains of admissible values of the variable*). Such for example are the formulas expressing the sine and cosine in terms of the tangent of half an angle, the formula of the tangent of a sum, the formula of the tangent of a double angle (the reader is advised to make a careful analysis of these formulas).

This fact is paramount in the solution of equations (it is discussed in more detail in Sec. 1.9).

Exercises

1. Define (a) a negative angle, (b) the radian measure of an angle, (c) the tangent of a given angle, (d) cos 1, (e) arcsin a.
2. State whether each of the following assertions is a definition or a theorem: (a) $\sin^2\alpha+\cos^2\alpha=1$ for arbitrary α, (b) the sine of an angle φ is equal to the ordinate of the unit vector issuing from the origin and forming with the axis of abscissas the angle φ, (c) the graph of the function $y=\sin x$ passes through the origin, (d) $\cot 90°=0$.
3. Consider the theorem: "If the terminal side of the angle φ lies in the second quadrant, then $\cos\varphi\leqslant 0$." State the following theorems: the converse, the inverse and the contrapositive. Which of these theorems is valid?
4. Prove that if real numbers x and y satisfy the condition $x^2+y^2=1$, then there is an angle φ such that $x=\sin\varphi$ and $y=\cos\varphi$.
5. Which is larger: (a) sin 1° or sin 1, (b) tan 1 or tan 2?
6. How are the angles α and β related if it is known that (a) $\sin\alpha=\sin\beta$, (b) $\cos\alpha=\cos\beta$, (c) $\tan\alpha=\tan\beta$, (d) $\sin\alpha=\sin\beta$?
7. Express $\cos(\alpha/2)$ and $\sin(\alpha/2)$ in terms of $\sin\alpha$ if $270°\leqslant\alpha\leqslant 450°$.
8. Given $\sin x=1/4\,(\sqrt{5}-1)$. Compute $\sin 5x$.
9. Compute without the aid of tables
$$\cos\frac{\pi}{7}\cos\frac{2\pi}{7}\cos\frac{4\pi}{7}.$$
10. Verify the validity of the equation
$$\frac{1-\tan^2 15°}{1+\tan^2 15°}=\frac{\sqrt{3}}{2}.$$

2.2 Trigonometric transformations

At examinations the student encounters a diversity of problems involving the transformation of trigonometric expressions or the proof of trigonometric relations. All such problems are amenable to solution through the use of familiar formulas, though some of them appear quite often in "terrifying" form. Nevertheless, combining suitable formulas ordinarily does the trick and no fundamental difficulties should arise.

But for this the student should have a firm grasp of all formulas and be able to read them from left to right and from right to left, and "see" them in a variety of notations. For instance, in the notation $\sin x \times \sin 30° - \cos 30° \cos x$ the student should recognize $-\cos(x+30°)$ and not $\sin(x-30°)$, as some do.

This ability and the necessary habits are acquired by working with the basic trigonometric formulas and solving a sufficient number of problems.

When performing trigonometric transformations, take care *to observe all the rules of algebraic operations* and carry through the manipulations neatly and legibly because a simple thing like writing a plus sign instead of a minus sign or fumbling in the computation of a coefficient can nullify the final result. The student frequently has to make use of various algebraic identities which he should be able to apply to trigonometric expressions.

To take an instance, many mistakes are made in trigonometric transformations due to an improper understanding of the symbol $\sqrt{\ }$. In trigonometry, as in algebra, this symbol always denotes the *principal* square root (see Sec. 1.4) and so, for instance,

$$\sqrt{1-\sin 2x}=\sqrt{(\sin x-\cos x)^2}=|\sin x-\cos x|$$

but not $\sin x-\cos x$. In place of the expression $\sqrt{(1/2)(1-\cos 2\alpha)}$ one should write $|\sin\alpha|$ and not $\sin\alpha$, etc.

1. *Simplify the expression*

$$\frac{1}{\sqrt{b-a}}\cdot\frac{\sqrt{\frac{b-a}{a}}\sin x}{\sqrt{1+\left(\sqrt{\frac{b-a}{a}}\sin x\right)^2}}\sqrt{a+b\tan^2 x}$$

where $b>a>0$.

After a few simple manipulations, this expression (for brevity denote it by P) can be rewritten

$$P=\frac{\sin x\sqrt{a+b\tan^2 x}}{\sqrt{a+(b-a)\sin^2 x}}=\frac{\sin x\sqrt{a+b\tan^2 x}}{\sqrt{a\cos^2 x+b\sin^2 x}}$$

Some students handle this as follows:

$$\sqrt{a+b\tan^2 x}=\sqrt{a+b\frac{\sin^2 x}{\cos^2 x}}=\frac{\sqrt{a\cos^2 x+b\sin^2 x}}{\cos x}$$

and get a wrong answer: $P=\tan x$. The mistake here is this. In this transformation what we actually have to simplify is the expression

$\sqrt{\cos^2 x}$ which is equal to $|\cos x|$. And so the final result is $P = \sin x/|\cos x|$.

In the problem that follows the chief difficulty lies precisely in the algebraic aspect of the matter, in the necessity to indicate exactly for which values of the parameters a given transformation is legitimate and what the procedure should be applied in the case of other values of the parameters.

2. *Eliminate* θ *and* φ *from the relations*

$$m^2 \tan^2 \theta + n^2 \tan^2 \varphi = 1, \quad m^2 \cos^2 \theta + n^2 \sin^2 \varphi = 1,$$
$$m \sin \theta = n \cos \varphi$$

and find the relationship between m and n.

Here, it is necessary, assuming that the three relations are valid, to eliminate θ and φ. This can be done in a variety of ways. We indicate one.

In order to be able to take advantage of the third relation, rewrite the second so that it embodies the products $m \sin \theta$ and $n \cos \varphi$:

$$m^2 \sin^2 \theta + n^2 \cos^2 \varphi = m^2 + n^2 - 1$$

Then, taking into account the third given equation, we get

$$2n^2 \cos^2 \varphi = m^2 + n^2 - 1$$

Assuming $n \neq 0$ (the possibility of $n=0$ will be considered separately), we have

$$\cos^2 \varphi = \frac{m^2+n^2-1}{2n^2}$$

Furthermore, from the third relation, *on the assumption that* $m \neq 0$, we have

$$\cos^2 \theta = 1 - \sin^2 \theta = 1 - \frac{n^2}{m^2} \cos^2 \varphi = \frac{m^2 - n^2 + 1}{2m^2}$$

We can now already write down some relationships between m and n: if $m \neq 0$ and $n \neq 0$, then

$$0 < \frac{m^2+n^2-1}{2n^2} \leqslant 1, \quad 0 < \frac{m^2-n^2+1}{2m^2} \leqslant 1$$

Now rewrite the first of the given relations in the form

$$m^2 \left(\frac{1}{\cos^2 \theta} - 1 \right) + n^2 \left(\frac{1}{\cos^2 \varphi} - 1 \right) = 1$$

and substitute the expressions found for $\cos^2 \theta$ and $\cos^2 \varphi$ to obtain a relation between m and n:

$$\frac{2m^4}{m^2-n^2+1} + \frac{2n^4}{m^2+n^2-1} = m^2 + n^2 + 1$$

Now let us see what happens when $n=0$. The given relations become

$$m^2\tan^2\theta=1,\quad m^2\cos^2\theta=1,\quad m\sin\theta=0$$

The first equation shows that $m\neq 0$; then the third yields $\sin\theta=0$, which conflicts with the first one. Hence, from the given relations it follows that $n\neq 0$. It is analogously shown that $m\neq 0$ as well.

Thus, after eliminating θ and φ from the three original equations, we can make the following statements about m and n:

$$m\neq 0,\ n\neq 0,\ 0<\frac{m^2+n^2-1}{2n^2}\leqslant 1,\ 0<\frac{m^2-n^2+1}{2m^2}\leqslant 1,$$
$$\frac{2m^4}{m^2-n^2+1}+\frac{2n^4}{m^2+n^2-1}=m^2+n^2+1$$

The solution is complete.

Let us now dwell in more detail on the transformation of the expression $a\sin x+b\cos x$ by the *introduction of an auxiliary angle*. It will be recalled that in defining the angle φ from the conditions

$$\sin\varphi=\frac{b}{\sqrt{a^2+b^2}},\quad \cos\varphi=\frac{a}{\sqrt{a^2+b^2}} \tag{1}$$

(it is assumed that a and b are not zero simultaneously), we can reduce the expression $a\sin x+b\cos x$ to the form $\sqrt{a^2+b^2}\sin(x+\varphi)$.* It is easy to see that it is precisely the *signs* of the numbers a and b which determine the quadrant in which the angle φ lies.

This general method of introducing an auxiliary angle always leads to the desired end. In the actual solution of concrete problems, however, (say, in trigonometric equations) it is often more advantageous to introduce the auxiliary angle differently.

It is frequently convenient to define the auxiliary angle so that it should lie *between* 0 *and* $\pi/2$. If in place of formulas (1) we use

$$\sin\alpha=\frac{|b|}{\sqrt{a^2+b^2}},\quad \cos\alpha=\frac{|a|}{\sqrt{a^2+b^2}} \tag{2}$$

to define the auxiliary angle α, it will be clear that α may be chosen in the first quadrant. What is more, it is sufficient to use *only one of the formulas* of (2) because an angle in the first quadrant is uniquely defined by the value of the sine (or cosine).**

In the process, of course, the expression $a\sin x+b\cos x$ will not necessarily be reduced to $\sqrt{a^2+b^2}\sin(x+\alpha)$ but possibly to one of

---

* In standard textbooks it is shown that these conditions define a *unique* angle φ within the range $0\leqslant\varphi<2\pi$. Quite naturally, for the auxiliary angle we could choose any angle $\varphi+2k\pi$, $k=\pm 1, \pm 2, \ldots$.

** If so desired, when defining α (in the range from 0 to $\pi/2$) we can also use the relation $\tan\alpha=|b|/|a|$ (if $a\neq 0$) which follows from (2). It is easy to convince oneself that all these relations define *one and the same* angle α in the first quadrant.

the following expressions:

$$\sqrt{a^2+b^2}\sin(x-\alpha),\quad \sqrt{a^2+b^2}\sin(-x+\alpha),$$
$$\sqrt{a^2+b^2}\sin(-x-\alpha)$$

depending on the *signs* of the numbers a and b.

However, in transforming the concrete expression $a\sin x+b\cos x$ there is no need to recall formulas (2), it is much simpler to carry out the necessary manipulations directly.

Consider for example the *expression* $Q=-2\sin x+\sqrt{3}\cos x$. Since the square root of the sum of the squares of the coefficients of the sine and cosine is equal to $\sqrt{7}$, the expression Q may be rewritten as

$$Q=\sqrt{7}\left(-\frac{2}{\sqrt{7}}\sin x+\frac{\sqrt{3}}{\sqrt{7}}\cos x\right)$$

If we now take $2/\sqrt{7}$ for the cosine of the auxiliary angle α and $\sqrt{3}/\sqrt{7}$ for the sine of that angle, then

$$Q=\sqrt{7}(-\cos\alpha\sin x+\sin\alpha\cos x)$$
$$=\sqrt{7}\sin(\alpha-x)=-\sqrt{7}\sin(x-\alpha)$$

The angle α itself (in the range from 0 to $\pi/2$) may be determined from the equation

$$\sin\alpha=\frac{\sqrt{3}}{\sqrt{7}},\quad\text{or}\quad \alpha=\arcsin\frac{\sqrt{21}}{7}$$

and so, finally,

$$Q=-\sqrt{7}\sin\left(x-\arcsin\frac{\sqrt{21}}{7}\right)$$

If in determining the auxiliary angle α lying between 0 and $\pi/2$ we use the condition

$$\cos\alpha=\frac{2}{\sqrt{7}},\quad\text{or}\quad \alpha=\arccos\frac{2\sqrt{7}}{7}$$

then expression Q assumes a somewhat different form:

$$Q=-\sqrt{7}\sin\left(x-\arccos\frac{2\sqrt{7}}{7}\right)$$

It is sometimes more advantageous to reduce the expression $a\sin x+b\cos x$ to the *cosine* of a sum (or a difference) of the angle x and an auxiliary angle lying in the range from 0 to $\pi/2$. This transformation is also conveniently performed directly without recourse to general formulas.

Thus, the expression $R=\sin 2x-\sqrt{3}\cos 2x$ may be given in the form

$$R=2\left(\frac{1}{2}\sin 2x-\frac{\sqrt{3}}{2}\cos 2x\right)$$

It can be reduced to the cosine of a certain combination of angles if the expression in the brackets on the right is represented as an expanded cosine of the sum of the angle $2x$ and of a certain angle β (in the range from 0 to $\pi/2$). For this it suffices to take $1/2$ as the sine of the auxiliary angle β and $\sqrt{3}/2$ as its cosine. Since in this case $\beta=\pi/6$, it follows that

$$R=2\left(\sin\frac{\pi}{6}\sin 2x-\cos\frac{\pi}{6}\cos 2x\right)=-2\cos\left(2x+\frac{\pi}{6}\right)$$

As we know, a trigonometric expression is, generally speaking, *not* meaningful *for all* values of its arguments. For this reason, in problems that involve the transformation of a trigonometric expression it is always *assumed* (though not always explicitly stated in the problem) that the transformation of the proposed expression is to be carried out *in its domain of definition*, which is to say only for those values of the arguments for which the proposed expression is meaningful.

3. *Given that the angles* α, β, γ *are connected by the relation*

$$2\tan^2\alpha\tan^2\beta\tan^2\gamma+\tan^2\alpha\tan^2\beta+\tan^2\beta\tan^2\gamma+\tan^2\gamma\tan^2\alpha=1 \qquad (3)$$

Find $\sin^2\alpha+\sin^2\beta+\sin^2\gamma$.

According to what we have just mentioned, it is necessary to find the sum $\sin^2\alpha+\sin^2\beta+\sin^2\gamma$ (denote it by N) solely for angles α, β, γ such that for them all the trigonometric functions in (3) exist. In other words, we must assume that in the given problem the angles α, β, γ are such that $\tan\alpha$, $\tan\beta$, $\tan\gamma$ are meaningful.* It is precisely under this additional and implicitly stated condition that we will seek the quantity N.

Since it is assumed that we will seek N starting out from the value of a certain trigonometric expression involving tangents, it is natural to write down the desired quantity N in a different form, in terms of tangents:

$$N=\frac{\tan^2\alpha}{1+\tan^2\alpha}+\frac{\tan^2\beta}{1+\tan^2\beta}+\frac{\tan^2\gamma}{1+\tan^2\gamma}$$

---

* Note that the desired sum $\sin^2\alpha+\sin^2\beta+\sin^2\gamma$ has a definite meaning and a definite value for *all* angles α, β, γ. However, for those angles α, β, γ for which at least one of the numbers $\tan\alpha$, $\tan\beta$, $\tan\gamma$ does not exist, condition (3) of the problem becomes meaningless and so such angles cannot be considered "as being connected by the relation (3)".

Note that here we use the assumption that $\tan\alpha$, $\tan\beta$, $\tan\gamma$ exist. It is only then possible to make use of the formula connecting the sine and tangent of the same angle.

Subsequent manipulations do not cause any trouble. Reducing the expression obtained for N to a common denominator and using (3), we find that $N=1$.

In this problem, we did not actually have to find the admissible values of the arguments; the only important thing was that the transformations carried out be valid for all admissible values and that they should not alter the domains of definition.

However, if one has to use transformations that do change the domain, then great care must be taken. This is especially essential when it is necessary not only to transform the given expression but also to determine the values of the variable which make this expression vanish (which actually amounts to solving an *equation*). In this case, pay attention to the changes that the domain of definition undergoes. Do not permit a restriction of the domain, and make a check if it is extended.

4. *Simplify the expression*

$$\frac{1}{3\sin x}\left[(2\cos x-\sin x)\cot x+2\sin x+\cos x\right]\times\left[1+\left(\frac{\sqrt{3}\sin x}{2\cos x-\sin x}\right)^{-2}\right]^{-1}-\frac{\cos 2x\left[2(1-\sin x\cos x)+(\sin x+\cos x)^2\right]}{6(\sin x+\cos x)^2(1-\sin x\cos x)}$$

Find all the values of x for which this expression vanishes.

We will not begin by determining the admissible values of x and will perform only the formal transformations. Denoting the expression by A, the minuend by B and the subtrahend by C (so that $A=B-C$), we transform B and C to a simpler form:

$$B=\frac{1}{3\sin x}\cdot\frac{2}{\sin x}\cdot\frac{3\sin^2 x}{4(1-\sin x\cos x)}=\frac{1}{2(1-\sin x\cos x)}$$

$$C=\frac{3(\cos^2 x-\sin^2 x)}{6(\sin x+\cos x)^2(1-\sin x\cos x)}=\frac{\cos x-\sin x}{2(\sin x+\cos x)(1-\sin x\cos x)}$$

Consequently,

$$A=\frac{\sin x}{(\sin x+\cos x)(1-\sin x\cos x)}$$

Noting that $1-\sin x\cos x=\sin^2 x-\sin x\cos x+\cos^2 x$ and using the algebraic identity

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

we finally get

$$A=\frac{\sin x}{\sin^3 x+\cos^3 x}$$

Let us no examine the second part of the problem. From the final form of the given expression it is evident that it will vanish only for those values of x for which $\sin x=0$. However the original expression *is meaningless* for these values of x: the presence of $\sin x$ in the denominator and also of $\cot x$ presuppose that the original expression is considered for $x\neq k\pi$, k an integer. It therefore follows that the *given expression cannot vanish for any values of* x.

How are we to account for the fact that the final expression vanishes for certain values of x while the original expression is meaningless for these values of x? This is because in the process of transforming the original expression we *extended* its domain of definition. This occurred when we cancelled $\sin^2 x$ out of the numerator and denominator of the minuend. It was then that the values $x=k\pi$, $k=0, \pm1, \pm2, \ldots$, appeared in the domain of permissible values.

Note that the domain of the original expression fails to include not only $x=k\pi$, $k=0, \pm1, \pm2, \ldots$, but also the values of x for which the expressions $\sin x+\cos x$ and $2\cos x-\sin x$ vanish (the expression $1-\sin x\cos x$ never vanishes). As for the final expression, it is meaningless for those values of x for which $\sin x+\cos x=0$ but is meaningful for those values of x for which $2\cos x-\sin x=0$ (find the transformation following which the indicated values of x appeared in the domain!).

We have already discussed in detail the fact that trigonometric formulas are only valid for *admissible* values of the arguments [see, say, (3) and (4) of Sec. 2.1]. This applies in full measure to the trigonometric relations (identities) that are given to be proved at examinations.

In problems that require the student to substantiate a trigonometric relation, bear in mind that each relation must be considered *together* with a description of the set of values of the arguments for which it is valid. If the set on which the identity to be proved is valid is not indicated in the statement of the problem, then this means that the identity must be regarded in *the domain of its variable*. It is then necessary to find that domain and ensure the validity of the proof for *all* admissible values of the arguments.

5. *Prove the identity*

$$1=-\cos 2\alpha\left[1+\tan\left(\frac{\pi}{2}-\alpha\right)\tan(\pi-2\alpha)\right]$$

It is quite clear that if this identity is valid at all, then not for all values of α. Indeed, the left member is meaningful for all α (it is simply independent of α), while the right member is not defined for $\alpha=n\pi$, $\alpha=(\pi/4)+(m\pi/2)$, where n and m are integers (represent these values on the trigonometric circle!). All values of α, except those just mentioned, are admissible.

It is precisely for these admissible values of α that we have to prove the proposed identity. The proof is extremely simple if we take advantage of the basic formulas of trigonometry. The only thing to bear in mind is to check to see that the transformations are legitimate for all admissible values of α.

In order to prove a trigonometric relation, one ordinarily takes one of its parts and with the aid of various trigonometric and algebraic operations (and the data supplied in the statement of the problem) transforms it so as to obtain the expression in the other member of the relation being proved. This coincidence of the left and right members of the equation can also be attained by transforming them separately.

However, in more complicated cases, particularly if it is required to obtain the sought-for equation from the given equation, it is often rather difficult to see at a glance what manipulations will achieve the desired end. In such cases, it is common *to assume the relation to be valid* and then by means of various manipulations to reduce it to an obvious (or to the given) equation, thus probing for a path leading to the solution.

6. *Prove that if* $\cos x=\cos a\cos b$ *and if* $x\pm a\neq(2k+1)\pi$, $b\neq(2n+1)\pi$, k, $n=0, \pm1, \pm2, \ldots,$ *then*

$$1+\tan\frac{x+a}{2}\tan\frac{x-a}{2}=\sec^2\frac{b}{2} \tag{4}$$

It hardly seems possible in this problem to guess at once precisely which transformations of the equation $\cos x=\cos a\cos b$ will lead (taking into account all restrictions) to the desired equation (4).*

We will therefore *assume that equation* (4) *is true* and will transform it:

$$\left.\begin{aligned} \frac{\sin\frac{x+a}{2}\sin\frac{x-a}{2}}{\cos\frac{x+a}{2}\cos\frac{x-a}{2}}&=\frac{1}{\cos^2\frac{b}{2}}-1,\\ \frac{\cos a-\cos x}{\cos a+\cos x}&=\frac{1-\cos b}{1+\cos b},\\ \cos x&=\cos a\cos b \end{aligned}\right\} \tag{5}$$

These manipulations brought us to an equation which, *by the statement of the problem, is known to be true.*

---

* Note the essential nature of the restrictions imposed in the problem on x, a, b. Without them the assertion of the problem would not be true since the equation (4) we are interested in has a more restricted domain than the given relation $\cos x=\cos a\cos b$. For instance, for $x=a+\pi$, $b=\pi$, the relation $\cos x=\cos a\cos b$ is valid, whereas (4) becomes meaningless.

Here, many students made the grave mistake of concluding that "hence, the equation (4) to be proved is also true". There are of course no grounds for such a conclusion. The manipulations *do not prove* the validity of the required equation. Strictly speaking, we proved that if equation (4) is true, then $\cos x = \cos a \cos b$, which is to say that we proved the *converse* of what was required in the problem.

We continue to reason as follows. The manipulations of (5) will simply be regarded as an *exploratory search* of a solution in the rough. To achieve a real solution, let us start with the given equation $\cos x = \cos a \cos b$ and carry out all these transformations *in reverse order*.

Namely, we take the *true* (by the statement of the problem) equation $\cos x = \cos a \cos b$ and multiply both sides by 2, writing down the result like this:

$$-\cos x + \cos a \cos b = \cos x - \cos a \cos b$$

Adding the expression $\cos a - \cos b \cos x$ to both members, we get

$$(\cos a - \cos x)(1 + \cos b) = (\cos a + \cos x)(1 - \cos b)$$

Since $x \pm a \neq (2k+1)\pi$, it follows that $\cos a + \cos x \neq 0$. Since $b \neq (2n+1)\pi$, then $1 + \cos b \neq 0$, and so* both sides of this last equation can be divided by the product $(\cos a + \cos x)(1 + \cos b)$ to yield the *true* equation

$$\frac{\cos a - \cos x}{\cos a + \cos x} = \frac{1 - \cos b}{1 + \cos b}$$

In the left member, we transform the sum and difference of cosines into a product and apply the half-angle formula to the right member:

$$\frac{2 \sin \frac{x+a}{2} \sin \frac{x-a}{2}}{2 \cos \frac{x+a}{2} \cos \frac{x-a}{2}} = \frac{\sin^2 \frac{b}{2}}{\cos^2 \frac{b}{2}}$$

It now remains on the left to pass to tangents and on the right to express the sine of the angle $b/2$ in terms of its cosine in order to get the required equation (4).

But it is also possible to dispense with this reversal of manipulations. All we need to do is prove that all the transformations used in (5) to go from (4) to the relation $\cos x = \cos a \cos b$ are *reversible* in the domain of (4) (that is to say, not only does each equation obtained in the process of manipulations imply the subsequent equation, but that it itself follows from the subsequent equation) We have already discussed this method of reasoning (see Sec. 1.8) as applied to the proof of inequalities; it is fully applicable to the substantiation of equations, including trigonometric equations.

* We see that the restrictions given in the statement of the problem are utilized in an essential manner: it is only under these restrictions that we can carry through in reverse order the manipulations done in the rough draft.

Thus, we again have to analyze the transformations (5). It is easy to see that they are all reversible [in particular, the transition from the second equation in (5) to the third is reversible precisely because of the restrictions imposed in the statement of the problem on x, a, and b]. Now that we are supplied with the proof of the reversibility of each transformation, we can regard the manipulations of (5) as a complete solution: they permit us to conclude that the original equation (4) is valid (in the domain of x).

In conclusion, let us tackle another problem involving the proof of a trigonometric relation that is solved by means of an artificial device.

7. *Prove that for an arbitrary positive integer n the equation*

$$\sin\frac{\pi}{3}+\sin\frac{2\pi}{3}+\ldots+\sin\frac{n\pi}{3}=2\sin\frac{n\pi}{6}\sin\frac{(n+1)\pi}{6}$$

is true.

Denoting the left member by S for brevity, multiply it by $\sin(\pi/6)$ and then expand all the resulting products of the sines into differences of cosines:

$$S\cdot\sin\frac{\pi}{6}=\frac{1}{2}\left[\left(\cos\frac{\pi}{6}-\cos\frac{3\pi}{6}\right)+\left(\cos\frac{3\pi}{6}-\cos\frac{5\pi}{6}\right)+\ldots\right.$$
$$\left.+\left(\cos\frac{2n-1}{6}\pi-\cos\frac{2n+1}{6}\pi\right)\right]$$

Noting that in the resulting sum all intermediate summands cancel out, we get

$$S\cdot\sin\frac{\pi}{6}=\frac{1}{2}\left[\cos\frac{\pi}{6}-\cos\frac{2n+1}{6}\pi\right]=\sin\frac{n\pi}{6}\sin\frac{n+1}{6}\pi,$$
$$S=2\sin\frac{n\pi}{6}\sin\frac{(n+1)\pi}{6}$$

Thus, the left member of the original equation coincides with the right member; this confirms its validity.*

Exercises

1. Given that $\sin x=1/4(\sqrt{5}-1)$. Prove that in this case $\cos 4x=\sin x$ and find (in degrees) the angle x lying between 0° and 90°.
2. Prove that the expression $y=\cos^2 x+\cos^2(x+\alpha)-2\cos\alpha\cos x\cos(x+\alpha)$ is independent of x.
3. Simplify the expression

$$\frac{2\sin 2x\left(\sin x\cos x+\dfrac{\cos^3 x\sin x}{\sqrt{1+\cos^4 x}}\right)}{\cos^2 x+\sqrt{1+\cos^4 x}}+\frac{\cos^2 2x}{\sqrt{1+\cos^4 x}}$$

* The reader will see that this technique is applicable in a number of other problems as well. For instance, it permits *computing* sums of the form $s_1=\sin x+\sin(x+h)+\ldots+\sin(x+nh)$, $s_2=\cos x+\cos 2x+\ldots+\cos nx$, etc. [to do this, it suffices to consider the expressions $s_1\cdot\sin(h/2)$ and $s_2\cdot\sin(x/2)$, respectively].

4. Simplify the expression

$$\frac{1}{2\sqrt{-a^2-ab}}\left(\frac{\dfrac{\sqrt{-a^2-ab}}{\cos^2 x}}{a+\sqrt{-a^2-ab}\tan x}-\frac{\dfrac{\sqrt{-a^2-ab}}{\cos^2 x}}{-a+\sqrt{-a^2-ab}\tan x}\right)$$

where $a^2+ab<0$.

5. Simplify the expression

$$\frac{(\tan x)^{-1/2}}{\cos^2 x}-$$

$$-\frac{[1+(2\tan x)^{-1/2}][\tan x-(2\tan x)^{1/2}+1]-[1-(2\tan x)^{-1/2}][\tan x+(2\tan x)^{1/2}+1]}{2\sqrt{2}\cos^2 x\,[\tan x+(2\tan x)^{1/2}+1]\qquad[\tan x-(2\tan x)^{1/2}+1]}$$

$$+\frac{(\tan x-1)(2\tan x)^{-1/2}-(2\tan x)^{1/2}}{\sqrt{2}\cos^2 x(\tan x-1)^2}[1+2\tan x(\tan x-1)^{-2}]^{-1}.$$

Find all values of x for which this expression is equal to $3^{3/4}$.

6. Simplify the expression

$$\frac{2\sin a\cos x(1-\cos a\cos x)-\sin 2a\sin^2 x}{2(1-\cos a\cos x)^2}\left[1-\frac{\sin^2 a\sin^2 x}{(1-\cos a\cos x)^2}\right]^{-1/2}$$

if $0<a<\pi/4$, $\pi/4<x<\pi/2$.

7. Prove the identity

$$\cos^2\left(\frac{\pi}{2}-\alpha\right)-\sin\left(\frac{2\pi}{3}-\alpha\right)\sin\left(\alpha-\frac{\pi}{3}\right)=\frac{3}{4}.$$

8. Prove that if $0<x<\pi/2$, then

$$\sqrt{\tan x+\sin x}+\sqrt{\tan x-\sin x}=2\sqrt{\tan x}\cos\left(\frac{\pi}{4}-\frac{x}{2}\right).$$

9. Represent the expression $-\tan(x/2)+\cos x+\sin x$ as a product.

10. Simplify the expression

$$\sin(a-b)+\sin(b-c)+\sin(c-a)+4\sin\frac{a-b}{2}\sin\frac{b-c}{2}\sin\frac{c-a}{2}.$$

11. Eliminate θ and φ from the relations

$$a\sin^2\theta+b\cos^2\theta=1,\quad a\cos^2\varphi+b\sin^2\varphi=1,\quad a\tan\theta=b\tan\varphi$$

and find the relationship between a and b if $0<b<1$, $a>1$.

12. Eliminate θ and φ from the relations

$$p\cot^2\theta+q\cot^2\varphi=1,\quad p\cos^2\theta+q\cos^2\varphi=1,\quad p\sin\theta=q\sin\varphi$$

and find the relationship between p and q.

13. Given $\tan x=2b/(a-c)$, $a\neq c$. Compute the following expressions: $y=a\cos^2 x+2b\sin x\cos x+c\sin^2 x$, $z=a\sin^2 x-2b\sin x\cos x+c\cos^2 x$.

14. Prove that

$$\cos\frac{2\pi}{7}+\cos\frac{4\pi}{7}+\cos\frac{6\pi}{7}=-\frac{1}{2}.$$

15. Prove that $\tan 142°30'=2+\sqrt{2}-\sqrt{3}-\sqrt{6}$.

16. For which values of α and β does the equation $\tan(\alpha+\beta)-\tan\alpha=2\tan(\beta/2)$ follow from $\cos(2\alpha+\beta)=1$?

17. Prove that $\sin(\alpha+2\beta)=\sin\alpha$ if $\cot(\alpha+\beta)=0$.

18. Find $\tan(\alpha/2)$ if $\sin\alpha+\cos\alpha=\sqrt{7}/2$ and the angle α lies between $0°$ and $45°$.

19. Prove that if α and β are angles in the first quadrant, $\tan\alpha=1/7$, $\sin\beta=1/\sqrt{10}$, then $\alpha+2\beta=45°$.

20. Prove that if $\sin(x-\alpha)/\sin(x-\beta)=a/b$, $\cos(x-\alpha)/\cos(x-\beta)=A/B$ and $aB+bA\neq 0$, then $\cos(\alpha-\beta)=(aA+bB)/(aB+bA)$.

21. Knowing that $\tan\alpha$ and $\tan\beta$ are roots of the quadratic equation $x^2+px+q=0$, compute the expression $\sin^2(\alpha+\beta)+p\sin(\alpha+\beta)\cos(\alpha+\beta)+$ $+q\cos^2(\alpha+\beta)$.

22. Transform the expression $\cos\alpha\cos\beta\cos\gamma$ into a sum of sines provided that $\alpha+\beta+\gamma=\pi/2$.

23. Simplify the expression $\sin^2\alpha+\sin^2\beta+\sin^2\gamma-2\cos\alpha\cos\beta\cos\gamma$ if $\alpha+\beta+\gamma=\pi$.

24. The sum of three positive numbers α, β, γ is equal to $\pi/2$. Compute the product $\cot\alpha\cot\gamma$ if we know that $\cot\alpha$, $\cot\beta$, $\cot\gamma$ form an arithmetic progression.

25. Prove that if α, β, γ are angles of a triangle, then

$$\cot\frac{\alpha}{2}+\cot\frac{\beta}{2}+\cot\frac{\gamma}{2}=\cot\frac{\alpha}{2}\cot\frac{\beta}{2}\cot\frac{\gamma}{2}$$

26. Prove that if $0<\alpha<\pi/2$, $0<\beta<\pi/2$, $0<\gamma<\pi/2$ and $\cos\alpha+\cos\beta+$ $+\cos\gamma=1+4\sin\frac{\alpha}{2}\sin\frac{\beta}{2}\sin\frac{\gamma}{2}$, then $\alpha+\beta+\gamma=\pi$.

2.3 Trigonometric equations

The student should be able to write down correctly formulas that yield the roots of *elementary* trigonometric equations: $\sin px=a$, $\cos qx=b$, and so forth. Unfortunately, a good many mistakes are made in solving such equations because of a lack of skill (or understanding) in handling the symbols $\arcsin a$, $\arccos a$, etc.

To illustrate, consider the equation $\cos x=-1/2$. One often hears something like this: "the solution of this equation is $x=\pm(-\pi/3)+$ $+2k\pi$, where k is any integer." This clearly shows that the student does not understand the symbol $\arccos(-1/2)$ (see Sec. 2.5). The value is frequently computed correctly: $\arccos(-1/2)=\pi-(\pi/3)$, but for some reason the notation appears as $x=\pm\pi-(\pi/3)+2k\pi$ instead of $x=\pm(2\pi/3)+2k\pi$.

Serious mistakes are made when formulas are manipulated mechanically. Thus, in solving the equation $\sin x=(\sqrt{5}+1)/2$ a student gave the answer as

$$x=(-1)^k\arcsin\frac{\sqrt{5}+1}{2}+k\pi,\quad k=0,\ \pm1,\ \pm2,\ \ldots$$

whereas this equation does not have any roots at all because $(\sqrt{5}+1)/2>1$.

Sometimes the student writes the solution of a trigonometric equation in degrees. This is of course quite permissible, but the answer is preferably given in radians with x regarded as a *number* and not an angle. But it is not at all acceptable to write a combination of degrees and radians, such as, say, $x=(-\pi/8)+90^\circ\cdot k$.

An equation very close to the elementary trigonometric equations is one of the form

$$a\sin x+b\cos x=c \tag{1}$$

to which many other equations reduce. Equations like (1) are best worked by the *method of introducing an auxiliary angle* (see Sec. 2.2).

1. *Solve the equation*

$$\cos x+\sqrt{3}\sin x=2\cos 2x$$

Here, the left member is most conveniently transformed to the *cosine* of a difference (defining the auxiliary angle in the range from 0 to $\pi/2$):

$$\cos x+\sqrt{3}\sin x=2\cos\left(x-\frac{\pi}{3}\right)$$

The equation then assumes the form $\cos[x-(\pi/3)]=\cos 2x$. If we had transformed the left member to the sine of a sum: $\cos x+\sqrt{3}\sin x=2\sin[x+(\pi/6)]$, then after transposing $2\cos 2x$ to the left-hand side, we would have to take advantage of the reduction formula in order to pass to a product of trigonometric functions.* But now we can do this directly:

$$\cos\left(x-\frac{\pi}{3}\right)-\cos 2x=0,\quad \sin\frac{9x-\pi}{6}\sin\frac{3x+\pi}{6}=0$$

whence we get two groups (sets) of solutions:

$$x_1=\frac{\pi}{9}+\frac{2k\pi}{3},\ k=0,\ \pm 1,\ \pm 2,\ \ldots,$$

$$x_2=-\frac{\pi}{3}+2n\pi,\ n=0,\ \pm 1,\ \pm 2,\ \ldots$$

Trigonometric equations are nearly always found to have several groups of solutions. The important thing to realize is that the numbers k and n appearing in different groups *are in no way connected.* The result may also be written differently:

$$x_1=\frac{\pi}{9}+\frac{2k\pi}{3},\ x_2=-\frac{\pi}{3}+2k\pi,\ k=0,\ \pm 1,\ \pm 2,\ \ldots$$

Here however it is understood that in *each* of these equations the number k runs through all integers *independently* of the other equation.

---

* Rather often, equations of the form $\cos[x-(\pi/3)]=\cos 2x$ or $\sin[x+(\pi/6)]=\cos 2x$ are solved differently, by using the relationships between α and β that follow from the equations $\cos\alpha=\cos\beta$ or $\sin\alpha=\cos\beta$ (see Exercise 1). Examinations however show that the student makes many mistakes due to the incorrect use of these relations, which are outside the scope of the school curriculum and so need not be memorized, all the more so since their use does not substantially reduce the amount of computation.

When the auxiliary angle turns out to be "good" (say $\pi/3$, $\pi/4$, etc.), the equations of type (1) are always solved by the foregoing method. But if the auxiliary angle is not equal to a single one of the familiar values, then few students use this method. The usual procedure is universal substitution or squaring.

This is apparently due to an aversion for the symbols $\arcsin a$ and the like. In this "bad" case, the auxiliary angle can only be written down as an arc sine or arc cosine of some number (expression) which cannot be "computed". However, it is well to stress that such inconvenience is but slight when compared with the difficulties that await us in the use of other methods.*

2. *Solve the equation*

$$2\sin 4x + 16\sin^3 x\cos x + 3\cos 2x - 5 = 0$$

The natural way would seem to be to express $\sin 4x$ in terms of the trigonometric functions of the angle x:

$$2\sin 4x = 8\sin x\cos x - 16\sin^3 x\cos x$$

It is now apparent that the proposed equation can be reduced to the form (1):

$$4\sin 2x + 3\cos 2x = 5 \qquad (2)$$

Let α be an angle between 0 and $\pi/2$ that satisfies the relations $\sin\alpha = 3/5$, $\cos\alpha = 4/5$, that is, $\alpha = \arcsin 3/5$. Then (2) is equivalent to the equation $\sin(2x+\alpha) = 1$, whence

$$x = -\frac{1}{2}\arcsin\frac{3}{5} + \frac{\pi}{4} + k\pi, \qquad k = 0, \pm 1, \pm 2, \ldots$$

Relations defining the auxiliary angle α permit choosing $\alpha = \arccos 4/5$, which yields a different form of the answer:

$$x = -\frac{1}{2}\arccos\frac{4}{5} + \frac{\pi}{4} + k\pi, \qquad k = 0, \pm 1, \pm 2, \ldots$$

or, choosing $\alpha = \arctan 3/4$,

$$x = -\frac{1}{2}\arctan\frac{3}{4} + \frac{\pi}{4} + k\pi, \qquad k = 0, \pm 1, \pm 2, \ldots$$

Finally, we could introduce the auxiliary angle β, $0 < \beta < \pi/2$ so that equation (2) could be reduced to the form $\cos(2x - \beta) = 1$. For this purpose, it is sufficient to choose $\beta = \arcsin 4/5$ or $\beta = \arccos 3/5$,

* Actually, universal substitution (or the replacement of the sine and cosine by the tangent of half an angle) generally speaking restricts the domain of the variable and can result in a *loss* of roots (see Problem 15, Sec. 1.9). The operation of squaring both sides of the equation can *introduce* roots (see Sec. 1.9). Consequently, both these methods require supplementary investigation, whereas the method of introducing an auxiliary angle immediately leads to an equivalent elementary equation. That is precisely why this method is to be recommended when working equations of type (1).

which are written, respectively, as

$$x=\frac{1}{2}\arcsin\frac{4}{5}+k\pi, \qquad k=0,\ \pm 1,\ \pm 2,\ \ldots$$

and

$$x=\frac{1}{2}\arccos\frac{3}{5}+k\pi, \qquad k=0,\ \pm 1,\ \pm 2,\ \ldots$$

It is readily seen that these formulas are merely *different forms of notation* of the roots of equation (2), that is, they can be transformed one into the other (see Sec. 2.5).

3. *Solve the equation*

$$\tan\left(\frac{\pi}{2}\cos x\right)-\cot(\pi\sin x)=0$$

This equation has an unusual aspect due to the exotic form of the arguments of the tangent and cotangent. Actually, of course, there is nothing special here. The numbers $\sin x$ and $\cos x$ are quite definite for any x, and so the following expressions are quite meaningful: the tangent of the number $(\pi/2)\cos x$, provided $\cos x\neq\pm 1$, and the cotangent of the number $\pi\sin x$, provided $\sin x\neq 0$, $\sin x\neq\pm 1$. These conditions describe the domain of the variable of the original equation.

Transforming the difference of a tangent and cotangent and clearing fractions, we get the equation

$$\cos\left(\frac{\pi}{2}\cos x+\pi\sin x\right)=0 \tag{3}$$

The problem is to solve this equation and select the roots which lie in the domain of the original equation.

From (3) we immediately get

$$\frac{\pi}{2}\cos x+\pi\sin x=\frac{\pi}{2}+k\pi, \qquad k=0,\ \pm 1,\ \pm 2,\ \ldots$$

It is now necessary, for *each* integer k, to find *all* the roots of

$$2\sin x+\cos x=2k+1 \tag{4}$$

of type (1). Consequently, equation (3) breaks up into an *infinite* number of equations, and to find all its roots we have to find all the roots of equation (4) for $k=0$, all the roots of (4) for $k=1$, all the roots of equation (4) for $k=-1$, and so on.

This peculiarity of the problem under consideration caused much trouble at an examination. Some students obtained from (3) the relation $(\pi/2)\cos x+\pi\sin x=\pi/2$ and thus lost some of the roots; others thought that the number k should assume one very definite value and attempted in vain to find that value of k from equation (3).

It is not at all difficult to find all the roots of an infinite number of equations (4). Introducing an auxiliary angle, say, using the formula

$\beta = \arccos(1/\sqrt{5})$, we can reduce (4) to

$$\cos(x-\beta) = \frac{2k+1}{\sqrt{5}}, \quad k = 0, \pm 1, \pm 2, \ldots \tag{5}$$

Solving the inequality $-1 \leqslant (2k+1)/\sqrt{5} \leqslant 1$, we are readily convinced that it only has two *integral* solutions: $k=0$ and $k=-1$. Hence, out of the infinity of equations (5) we need the roots of only *two* equations that correspond to the values $k=0$ and $k=-1$:

$$\cos(x-\beta) = \frac{1}{\sqrt{5}} \text{ and } \cos(x-\beta) = -\frac{1}{\sqrt{5}}$$

The equations (5) corresponding to all the other values of k do not have any roots.

We write down the roots of the two equations in the form of four groups (x_1 and x_2 are the groups of roots of the first equation, and x_3 and x_4 are the groups of roots of the second equation):*

$$x_1 = \beta + \arccos\frac{1}{\sqrt{5}} + 2n\pi = 2\arccos\frac{1}{\sqrt{5}} + 2n\pi, \; n = 0, \pm 1, \pm 2, \ldots,$$

$$x_2 = \beta - \arccos\frac{1}{\sqrt{5}} + 2m\pi = 2m\pi, \quad m = 0, \pm 1, \pm 2, \ldots,$$

$$x_3 = \beta + \arccos\left(-\frac{1}{\sqrt{5}}\right) + 2p\pi = (2p+1)\pi, \quad p = 0, \pm 1, \pm 2, \ldots,$$

$$x_4 = \beta - \arccos\left(-\frac{1}{\sqrt{5}}\right) + 2q\pi = 2\arccos\frac{1}{\sqrt{5}} + (2q-1)\pi, \quad q = 0, \pm 1, \pm 2, \ldots$$

However, not all these groups lie in the domain of the original equation: for any root of the second and third groups it is obvious that $\sin x = 0$ and so $\cot(\pi \sin x)$ does not exist. Thus, these groups are not solutions of the original equation. Now the roots of the first and fourth groups lie in the domain of the original equation: the necessary computations (which may be performed, say, by analogy with the solution of Problems 2 to 4 of Sec. 2.5) show that for these roots $\cos x \neq \pm 1$, $\sin x \neq 0$, $\sin x \neq \pm 1$. Thus the solution of the original equation includes the two groups x_1 and x_4 given above.

It is sometimes possible to combine in one formula several groups of roots of a trigonometric equation. For instance, in the present problem, groups x_1 and x_4 can readily be represented as

$$x = 2\arccos\frac{1}{\sqrt{5}} + l\pi, \quad l = 0, \pm 1, \pm 2, \ldots$$

(this yields the first group for even $l=2n$ and the fourth group for odd $l=2q-1$). This however is not obligatory and neither is it always convenient. The final answer may be left in the form of several groups.

* In the transformation of the third group, use is made of the relation obtained in Problem 10, Sec. 2.5.

Trigonometric equations involving more or less complicated trigonometric expressions are a traditional part of many written examinations in mathematics. As we know, there is no unified method or procedure that can be used to solve every equation with trigonometric functions. But the general aim is to transform the trigonometric expressions in the equation in such a way that the equation under consideration is reduced to an elementary form or breaks up into several elementary equations.

For each specific example, one has to find the suitable mode of transformation. Sometimes a variety of different transformations and ideas are tried before the right approach is found. The surest way is via a good knowledge of the trigonometric formulas and the ability to carry through trigonometric transformations (see Sec. 2.2), and this can only be attained by practice.

Of course many trigonometric equations admit of several modes of solution, depending on the underlying idea of the solution and on how the trigonometric expressions involved are manipulated. We wish to stress that the form of notation of the roots is frequently dependent on the chosen procedure of solution and if we want to prove the equivalence of two different notations, we will have to resort to supplementary transformations.

This is an important point to bear in mind because sometimes the student solves a trigonometric equation and then begins to work it (as a check) by a different method and arrives at a different notation in the answer. Then thinking that this different form of the answer is an indication that the first solution is wrong, he tries to find nonexistent mistakes and spends a lot of time and energy instead of manipulating the answers and becoming convinced of their identity.

Incidentally, it is advisable at an examination to solve a given trigonometric equation in some one (preferably the simplest and shortest) way and dispense with transformations of the answer into other forms.

Keep track of equivalence in the process of manipulating the equation so as to avoid any loss of roots (say in cancelling a common factor from the right and left members of the equation) or the introduction of extraneous roots (say in squaring both sides of the equation). Besides, it is necessary to check to see whether all the resulting roots lie in the domain of the variable of the given equation. In all necessary cases (that is, when nonequivalent transformations are permitted) be sure to make a check.

All these questions connected with *solving equations* (and this goes for trigonometric equations as well) and also certain problems are discussed in great detail in Sec. 1.9. We shall not dwell further on them here.

The problems considered below illustrate several rather general recommendations that may be put to good use in the solving of *tri-*

gonometric equations. One should not of course presume that these suggestions are of a universal nature and applicable to all cases.

Many trigonometric equations involving a sine, cosine and tangent are frequently more easily solved if first reduced to a single function. For instance, it is sometimes possible to simplify the equation by means of universal substitution, that is, the replacement of all trigonometric functions in terms of the tangent of half an angle.

The student should bear in mind however that this transformation, generally speaking, restricts the domain of the variable and can therefore lead to a *loss of roots* (see Problem 15 of Sec. 1.9). Universal substitution must therefore be accompanied by an additional investigation.

4. *Solve the equation*

$$(1-\tan x)(1+\sin 2x)=1+\tan x$$

This equation may be solved in a variety of ways, but the fastest is by universal substitution: expressing $\sin 2x$ in terms of $\tan x$, we get

$$(1-\tan x)\left(1+\frac{2\tan x}{1+\tan^2 x}\right)=1+\tan x$$

which is *equivalent* to the original equation (since $\tan x$ exists in the domain of the original equation). Reducing the second factor of the left member to a common denominator and clearing fractions (this is possible since $1+\tan^2 x \neq 0$), we get the equation $\tan^2 x(1+\tan x)=0$ which has two groups of solutions:

$$x_1=k\pi, \quad x_2=-\frac{\pi}{4}+n\pi, \qquad k, n=0, \pm 1, \pm 2, \ldots$$

In solving trigonometric equations containing trigonometric functions of multiple arguments, students strive invariably to pass to functions of the argument itself; the result is an algebraic equation of high degree in $\sin x$ (or $\cos x$). In many cases, however, it is more convenient to pass from the *squares of trigonometric functions to the functions of a double argument* by the formulas

$$\cos^2 x=\frac{1}{2}(1+\cos 2x), \quad \sin^2 x=\frac{1}{2}(1-\cos 2x) \tag{6}$$

thus reducing the degree of the resulting equation. This frequently cuts manipulations and, besides, makes it possible to write down the answer automatically in a more compact form.

5. *Solve the equation*

$$\cos^4 x+\sin^4 x=2\cos\left(2x+\frac{\pi}{6}\right)\cos\left(2x-\frac{\pi}{6}\right)$$

The right member of this equation can readily be reduced to the form $1/2+\cos 4x$. If we furthermore express $\cos 4x$ in terms of $\cos x$, and in the left member of the original equation replace $\sin^2 x$ by $1 - \cos^2 x$, we get a biquadratic (in $\cos x$) equation: $12\cos^4 x - 12\cos^2 x + 1 = 0$.

Denoting $\cos^2 x$ by y, we obtain the quadratic equation $12y^2 - 12y + 1 = 0$, whose roots are

$$y_1 = \frac{3-\sqrt{6}}{6}, \quad y_2 = \frac{3+\sqrt{6}}{6}$$

The first root y_1 leads to the equation $\cos^2 x = (3-\sqrt{6})/6$. Since $0 < y_1 < 1$, it in turn separates into two equations:

$$\cos x = \sqrt{\frac{3-\sqrt{6}}{6}} \text{ and } \cos x = -\sqrt{\frac{3-\sqrt{6}}{6}}$$

whence we obtain two groups of solutions of the original equation. Similarly, the second root y_2 leads to two more groups of solutions of the original equation (it is left to the reader to write down all four groups).

Essentially, there are of course no objections to this mode of solution since it is quite correct. But the solution can be shortened and speeded up by a procedure that reduces to a consideration of a *single* elementary trigonometric equation and permits obtaining the answer as a single formula.

Namely, in the equation $\cos^4 x + \sin^4 x = 1/2 + \cos 4x$ we will not transform $\cos 4x$ but, on the contrary, we will express $\cos^4 x$ and $\sin^4 x$ in terms of the cosine of a quadruple argument, for which purpose we make use of (6). Very simple manipulations lead directly to the equation $\cos 4x = 1/3$, whence

$$x = \pm\frac{1}{4}\arccos\frac{1}{3} + \frac{k\pi}{2}, \quad k = 0, \pm 1, \pm 2, \ldots$$

It may be noted, in particular, that the use of formulas (6) always permits solving a *quadratic* equation in $\cos 2x$ instead of a *biquadratic* equation in $\cos x$ (or $\sin x$), thus cutting the amount of computation, and enables one to write down the answer in simpler form.

6. *Solve the equation*

$$4\cos^2(2-6x) + 16\cos^2(1-3x) = 13$$

Denoting $1 - 3x$ by y for brevity, we write the equation in the form $4\cos^2 2y + 16\cos^2 y = 13$. One can of course transform the first summand in the left member by the double-argument formula to obtain a biquadratic equation in $\cos y$. But if we transform the second summand in the left member by (6), we get the quadratic equation $4\cos^2 2y + 8\cos 2y - 5 = 0$.

The quadratic equation $4z^2 + 8z - 5 = 0$ in $z = \cos 2y$ has two roots: $z_1 = -5/2$, $z_2 = 1/2$. And so we have to solve two equations:

$\cos 2y = -5/2$ and $\cos 2y = 1/2$. Since $|\cos 2y| \leqslant 1$, the first equation has no roots, and from the second one we can readily find the roots of the original equation:

$$x = \frac{1}{3} \pm \frac{\pi}{18} + \frac{k\pi}{3}, \quad k = 0, \pm 1, \pm 2, \ldots$$

It is left to the reader to follow through the procedure and decide why it is possible to write down the answer in precisely this form.

The apt employment of formulas for *transforming a product of trigonometric functions into the sum (or difference) of such functions (or vice versa)* frequently produces the result in the shortest possible way. The student's ability to make a guess of the result of applying these formulas and to foresee the possibilities therein is very useful when solving trigonometric equations.

7. *Solve the equation*

$$\sin 4x + 3 \sin 2x = \tan x$$

Many students tackled this equation in the following manner. Multiplying it by $\cos x$ and using the double-argument formulas, they obtained the equation

$$\sin x \,(8 \cos^4 x + 2 \cos^2 x - 1) = 0$$

which is then solved in familiar ways.

A different approach makes things easier however. If after multiplying the original equation by $\cos x$ we expand the resulting products $\sin 4x \cos x$ and $3 \sin 2x \cos x$ into sums of trigonometric functions, we get the equation $\sin 5x + 4 \sin 3x + \sin x = 0$. Then we again transform the first and third terms on the left of this equation into a product of trigonometric functions to get $\sin 3x (\cos 2x + 2) = 0$. This equation separates into two: one ($\sin 3x = 0$) yields the group of solutions $x = k\pi/3$, $k = 0, \pm 1, \pm 2, \ldots$, and the second one ($\cos 2x + 2 = 0$) has no roots at all.

A check (which is necessary since multiplication by $\cos x$ extended the domain) shows that the group thus found is the solution of the original equation.

In many cases trigonometric equations are successfully solved by the following technique: *a new unknown is introduced to denote some combination of trigonometric functions and the equation is then solved for this unknown.* Of course, some experience is required to be able to see a suitable combination.

8. *Find all the solutions of the equation*

$$1 + \sin^3 x + \cos^3 x = \frac{3}{2} \sin 2x$$

It is easy to see that the left and right members of this equation can be expressed in terms of the sum and product of $\sin x$ and $\cos x$. Using the identity $a^3 + b^3 = (a + b)(a^2 - ab + b^2)$, we rewrite the equation as

$$1 + (\sin x + \cos x)(1 - \sin x \cos x) = 3 \sin x \cos x$$

In turn, the product $\sin x \cos x$ is expressed in terms of the sum $\sin x + \cos x$ by means of the trigonometric identity*

$$2 \sin x \cos x = (\sin x + \cos x)^2 - 1$$

It is therefore natural to denote the sum $\sin x + \cos x$ by y and to write the original equation as

$$1 + y - y\frac{y^2-1}{2} = 3\frac{y^2-1}{2}$$

We thus arrive at an algebraic equation in one unknown, y. Transposing all terms to the left and factoring out $y + 1$, we get the equation $(y + 1)(y^2 + 2y - 5) = 0$, which has three roots:

$$y_1 = -1, \quad y_2 = -1 + \sqrt{6}, \quad y_3 = -1 - \sqrt{6}$$

The first root leads to the equation

$$\sin x + \cos x = -1 \text{ or } \sin\left(x + \frac{\pi}{4}\right) = -\frac{\sqrt{2}}{2}$$

whence we get the group of solutions

$$x = -\frac{\pi}{4} + (-1)^{n+1}\frac{\pi}{4} + n\pi, \qquad n = 0, \pm 1, \pm 2, \ldots$$

As to the second and third roots, both exceed $\sqrt{2}$ in absolute value, whereas

$$|\sin x + \cos x| = \left|\sqrt{2}\sin\left(x + \frac{\pi}{4}\right)\right| \leqslant \sqrt{2}$$

and consequently these roots do not yield solutions of the original equation.

9. *Solve the equation*

$$\tan^2 2x + \cot^2 2x + 2 \tan 2x + 2 \cot 2x = 6$$

Note that $2 \tan 2x \cot 2x = 2$ and so the given equation can be rewritten in the form

$$(\tan 2x + \cot 2x)^2 + 2(\tan 2x + \cot 2x) - 8 = 0$$

Denoting the sum $\tan 2x + \cot 2x$ by z, we get the quadratic equation $z^2 + 2z - 8 = 0$, whose roots are $z_1 = -4$ and $z_2 = 2$. We now have to consider two possibilities corresponding to each of these roots:

$$\tan 2x + \cot 2x = -4 \text{ and } \tan 2x + \cot 2x = 2$$

* A similar formula, $2 \sin x \cos x = 1 - (\sin x - \cos x)^2$, is useful in a number of cases.

These can be reduced to quadratic equations in $\tan 2x$ by replacing $\cot 2x$ by $1/\tan 2x$. To each root of the quadratic equation corresponds a group of solutions.

Fewer manipulations are required if we first transform the sum of the tangent and cotangent:

$$\tan 2x + \cot 2x = \frac{2}{\sin 4x}$$

Then the root z_1 yields the equation $\sin 4x = -1/2$, whence we get the first group of solutions

$$x_1 = (-1)^{k+1}\frac{\pi}{24} + \frac{k\pi}{4}, \qquad k = 0, \pm 1, \pm 2, \ldots$$

and the root z_2 yields the equation $\sin 4x = 1$, which gives us a second group:

$$x_2 = \frac{\pi}{8} + \frac{n\pi}{2}, \qquad n = 0, \pm 1, \pm 2, \ldots$$

It is easy to see that all the values of both groups belong to the domain of the variable of the original equation.

Apt *grouping* of terms can also facilitate solutions of trigonometric equations. This is not always easy to do, however, and often a variety of possibilities have to be examined.

10. *Solve the equation*

$$4 \sin x + 2 \cos x = 2 + 3 \tan x$$

Though this seems to be a simple equation, it is rather involved. What might appear to be a natural solution — by means of a universal substitution—actually results in a fourth-degree equation in $\tan(x/2)$.

Let us attempt to group the terms of the equation so as to arrive at a decomposable equation. Multiplying all terms of the original equation by $\cos x$ (in this way we of course extend the domain so that a check will have to be made as to the possible introduction of extraneous roots) and transposing them to the left side, we get

$$4 \sin x \cos x + 2 \cos^2 x - 2 \cos x - 3 \sin x = 0$$

Can the left member of this equation be factored? At any rate it is not so obvious how this can be done and so we try out some variants.

It is rather easy to see that grouping the first term with the second and the third with the fourth and grouping the first with the fourth and the second with the third does not give us anything. Let us try grouping the first and third terms, and then the second and fourth:

$$2 \cos x (2 \sin x - 1) + (2 \cos^2 x - 3 \sin x) = 0 \qquad (7)$$

Now, the second term in (7) may be written as a quadratic (in $\sin x$) trinomial $2 \cos^2 x - 3 \sin x = 2 - 3 \sin x - 2 \sin^2 x$. But the trinomial $2y^2+3y-2$ can readily be factored: $2y^2+3y-2=(2y-1)(y+2)$. Therefore the second term in (7) can be represented as a product: $2\cos^2 x-3\sin x=-(2\sin x-1)(\sin x+2)$ and, hence, the equation (7) itself can then be written as

$$(2\sin x-1)(2\cos x-\sin x-2)=0$$

This equation separates into the elementary equation $\sin x=1/2$, whence

$$x_1=(-1)^k\frac{\pi}{6}+k\pi,\qquad k=0,\ \pm1,\ \pm2,\ \ldots$$

and the equation $\sin x-2\cos x=-2$ of type (1), whence

$$x_2=2n\pi,\ x_3=-2\arccos\frac{2}{\sqrt{5}}+2m\pi,\quad n,\ m=0,\ \pm1,\ \pm2,\ \ldots$$

All the roots of the three groups lie in the domain of the original equation and thus are its solutions.

11. *Solve the equation*

$$\cos(\pi 3^x)-2\cos^2(\pi 3^x)+2\cos(4\pi 3^x)-\cos(7\pi 3^x)$$
$$=\sin(\pi 3^x)+2\sin^2(\pi 3^x)-2\sin(4\pi 3^x)$$
$$+2\sin(\pi 3^{x+1})-\sin(7\pi 3^x)$$

To make this equation more compact, denote $\pi 3^x$ by y to get

$$\cos y-2\cos^2 y+2\cos 4y-\cos 7y$$
$$=\sin y+2\sin^2 y-2\sin 4y+2\sin 3y-\sin 7y$$

By transposing all terms to the left side and trying out a variety of grouping possibilities, we find the most acceptable one:

$$(\cos y-\cos 7y)+(\sin 7y-\sin y)+2(\cos 4y+\sin 4y)$$
$$-2(\cos^2 y+\sin^2 y)-2\sin 3y=0$$

whence

$$2\sin 4y\sin 3y+2\sin 3y\cos 4y+2(\cos 4y+\sin 4y)$$
$$-2(\sin 3y+1)=0\quad(8)$$

If we now factor $2\sin 3y$ out of the first two terms and compare the resulting expression with the third term, it will be clear that these three terms can be represented in the form of a product of two factors, one of which coincides with the last term of equation (8); and so (8) can be written in the decomposable form

$$(\sin 3y+1)(\sin 4y+\cos 4y-1)=0$$

This enables us to write down three groups of solutions:

$$y_1=-\frac{\pi}{6}+\frac{2k\pi}{3},\qquad y_2=\frac{n\pi}{2},\qquad y_3=\frac{\pi}{8}+\frac{m\pi}{2}$$

where k, n and m are arbitrary integers.

Recalling that $y = \pi 3^x$, we obtain an infinity of equations for determining the roots of the original equation:

$$\left.\begin{array}{ll} 3^x=-\frac{1}{6}+\frac{2k}{3}, & k=0,\ \pm1,\ \pm2,\ \ldots, \\ 3^x=\frac{n}{2}, & n=0,\ \pm1,\ \pm2,\ \ldots, \\ 3^x=\frac{1}{8}+\frac{m}{2}, & m=0,\ \pm1,\ \pm2,\ \ldots \end{array}\right\} \qquad (9)$$

In other words, *any* value of x that satisfies the first equation of (9) for *some integer* k is a solution of the original equation. We thus have to find all the roots of the first equation of (9) for each integral value of k. The same goes for the other two equations of (9).

At the examination, some students did not fully understand the set of relations (9), regarding them as a system of equations. That is, they sought only those values of x which (for certain integers k, n, m) satisfy the three equations (9) simultaneously. There were also mistakes in determining the roots of the equations (9). In some cases the roots of these equations were written down formally (for example, it was asserted that the roots of the first equation of (9) "are the numbers $\log_3 [(-1/6) + (2k/3)]$, where k is any integer") without the necessary and proper analysis of those (integral) values of k, n, m for which the equations of (9) have solutions.

Yet before solving the equations (9), one must recall that the equation $3^x=a$ has a (unique) root only for *positive* a and it is given by the formula $x = \log_3 a$. Therefore the equations (9) have solutions only for those (integral) values of k, n, m for which the corresponding right members of the relations (9) are positive.

It is easy to see that the right side of the first equation of (9) is positive for *integral* $k>0$; the right side of the second equation of (9) is positive for *integral* $n > 0$; and the right side of the third equation of (9) is positive for *integral* $m \geqslant 0$. Thus, we have to solve (9) only for the indicated values of k, m, n. The resulting values of x are then the roots of the original equation:

$$\begin{array}{ll} x=\log_3\left(-\frac{1}{6}+\frac{2k}{3}\right), & k=1,\ 2,\ \ldots, \\ x=\log_3\frac{n}{2}, & n=1,\ 2,\ \ldots, \\ x=\log_3\left(\frac{1}{8}+\frac{m}{2}\right), & m=0,\ 1,\ 2,\ \ldots \end{array}$$

Some examination problems call for finding not the whole set of roots of a trigonometric equation but only those which satisfy certain supplementary conditions indicated in the statement of the problem (say, roots lying in a certain interval of values).

Such problems may be worked in the following manner. Write out all roots of the equation at hand, then choose those for which the supplementary conditions are valid. Incidentally, it is sometimes simpler to seek only the solutions required and not write down all the solutions of the equation.

12. *Find all the solutions of the equation*

$$\sqrt{1+\sin 2x}-\sqrt{2}\cos 3x=0 \tag{10}$$

lying between π *and* $3\pi/2$.

This equation can be solved by squaring, but then at the end of the solution we will have to discard all extraneous roots and after that choose from the remaining ones those which satisfy the inequality $\pi < x < 3\pi/2$. We will approach the problem differently.

Since $\sqrt{1+\sin 2x} = |\sin x + \cos x|$, the equation (10) can be rewritten as

$$|\sin x+\cos x|-\sqrt{2}\cos 3x=0$$

First let us get rid of the absolute-value sign. It is not necessary however to consider all possible cases since we only need to find the roots of this equation that satisfy the inequality $\pi < x < 3\pi/2$. But in the third quadrant both sine and cosine are negative, and so (over the interval we are interested in!) the original equation can be reduced to

$$(\sin x+\cos x)+\sqrt{2}\cos 3x=0$$

or, after obvious manipulations,

$$\cos\left(2x-\frac{\pi}{8}\right)\cos\left(x+\frac{\pi}{8}\right)=0$$

The groups of solutions of this decomposable equation are then ordinarily written out and those are selected that lie between π and $3\pi/2$. But it is possible to get the answer at once without going through all these computations.

Indeed, consider first the equation $\cos[2x-(\pi/8)]=0$ or, denoting $2x-(\pi/8)$ by t, the equation $\cos t=0$. We are only interested in the values of x which satisfy the inequality $\pi < x < 3\pi/2$; from this it follows that

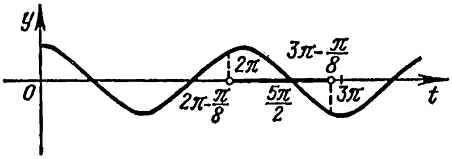
$$2\pi-\frac{\pi}{8}<2x-\frac{\pi}{8}<3\pi-\frac{\pi}{8}$$

Thus, we need only the roots of the equation $\cos t = 0$ that lie in the interval between $2\pi - (\pi/8)$ and $3\pi - (\pi/8)$. Referring to Fig. 56, the graph of the function $y = \cos t$, we see that on this interval the cosine vanishes only once, at the point $t = 5\pi/2$. Therefore,

$$2x - \frac{\pi}{8} = \frac{5\pi}{2}, \quad \text{whence } x = \frac{21\pi}{16}$$

In analogous fashion, we see that between π and $3\pi/2$ there is only one value of x, namely $x = 11\pi/8$, which satisfies the equation $\cos[x + (\pi/8)] = 0$.

Fig. 56



Choosing the solutions of a trigonometric equation is a frequent procedure when the equation is obtained from some original equation involving more than trigonometric functions (say from an equation involving logarithmic and trigonometric functions). In such cases, the role of "supplementary" conditions is often played by inequalities defining the domain of the original equation.

13. *Solve the equation*

$$\log_{\frac{-x^2-6x}{10}}(\sin 3x + \sin x) = \log_{\frac{-x^2-6x}{10}} \sin 2x \tag{11}$$

The logarithms involved in the statement of the problem are immediately eliminated, but it would be a grave mistake to assert that the original equation is equivalent to

$$\sin 3x + \sin x = \sin 2x \tag{12}$$

because the transition from (11) to (12) *extends* the domain of the variable, which means there may be extraneous solutions among the solutions of (12).

Hence, by Statement B of Sec. 1.9, to solve the original equation (11) it suffices to solve (12) and choose those roots that lie in the domain of (11), which is to say, that satisfy the inequalities

$$\sin 3x + \sin x > 0, \ \sin 2x > 0, \ -6 < x < 0 \tag{13}$$

(it is left to the reader to obtain these inequalities).

Equation (12) can then at once be rewritten as $2 \sin 2x \cos x = \sin 2x$. Since $\sin 2x > 0$ in the domain of the original equation [see (13)], we get $\cos x = 1/2$ after cancelling out $\sin 2x$, whence

$$x = \pm\frac{\pi}{3} + 2k\pi, \quad k = 0, \pm 1, \pm 2, \ldots$$

From among these roots we choose those that satisfy the conditions (13).

To do this, it is more convenient to consider, in place of the resulting single formula for the roots, two groups of solutions of equation (12):

$$x_1 = \frac{\pi}{3} + 2n\pi, \ x_2 = -\frac{\pi}{3} + 2m\pi, \qquad n, \ m = 0, \ \pm 1, \ \pm 2, \ \ldots$$

and for each of them separately to find those (integral) values of n and m for which the appropriate roots satisfy all three inequalities of (13).

Let us first consider the first group x_1. Apparently, the strongest of the restrictions (13) is the third one, and so we start with it since it will enable us to sweep out the greatest number of roots.

For the first group of solutions of (12) the third inequality of (13) is of the form

$$-6 < \frac{\pi}{3} + 2n\pi < 0 \tag{14}$$

Do not forget that we are only interested in the *integral* solutions of this inequality. Here the easiest way to find such solutions is by running through all cases. It is quite clear that for any integer $n \geqslant 0$ the middle portion of inequality (14) is positive, and so the inequality is not satisfied by a single nonnegative integral value of n. Furthermore, for an arbitrary integer $n \leqslant -2$ we have (using the fact that $\pi > 3$)

$$\frac{\pi}{3} + 2n\pi \leqslant \frac{\pi}{3} - 4\pi = -\frac{11\pi}{3} < -\frac{11 \cdot 3}{3} = -11 < -6$$

which is to say that inequality (14) is not satisfied by a single integer $n \leqslant -2$. Consequently, we have yet to check and see whether the value $n = -1$ satisfies (14). Since for this value of n the middle part of (14) is clearly negative and since (by virtue of the fact that $\pi < 3.2$)

$$\frac{\pi}{3} - 2\pi = -\frac{5\pi}{3} > -\frac{5 \cdot 3.2}{3} = -5\frac{1}{3} > -6$$

it is clear that the value $n = -1$ does indeed satisfy inequality (14).

Thus, of the entire group of x_1 solutions of (12) only one value, $x^* = -5\pi/3$, satisfies the third inequality of (13). Direct verification* shows that this value also satisfies two other inequalities of (13), that is, $x^* = -5\pi/3$ is a root of the original equation (11).

The second group x_2 of solutions of (12) could be investigated in similar fashion, but we can save time if we think to first check the second of the inequalities of (13). This check shows that not one of

* Incidentally, this verification need only be carried out in order to establish the inequality $\sin 2x^* > 0$. If, furthermore, we recall that x^* is a root of (12), that is, $\sin 3x^* + \sin x^* = \sin 2x^*$, then it will be clear that the inequality $\sin 3x^* + \sin x^* > 0$ follows from the inequality $\sin 2x^* > 0$.

the roots of the second group x_2 satisfies the condition $\sin 2x > 0$ and, hence, is not a root of equation (11).

There are also problems which involve choosing the roots of a trigonometric equation, but for a different reason, the aim being to find only those roots which are *common*, say, to *two* trigonometric equations.

14. *Solve the equation* $\sin 7x + \cos 2x = -2$.

At first glance, this problem does not seem to contain any peculiarities. But as we proceed with the manipulations it becomes clear that the equation is somewhat odd. It does not separate into several elementary equations but reduces to a *system of two* (elementary) *trigonometric equations in one unknown*.

Rewriting the original equation as

$$\left[1+\cos\left(\frac{\pi}{2}-7x\right)\right]+[1+\cos 2x]=0$$

and transforming each of the expressions in square brackets, we arrive at the relation

$$\cos^2\left(\frac{7x}{2}-\frac{\pi}{4}\right)+\cos^2 x=0 \qquad (15)$$

As we know, the sum of the squares of two quantities is zero if and only if both the quantities are equal to zero. Hence, the original equation *is equivalent to a system* of two equations in one unknown:

$$\left.\begin{array}{c}\cos\left(\frac{7x}{2}-\frac{\pi}{4}\right)=0\\ \cos x=0\end{array}\right\} \qquad (16)$$

We thus have to get all the solutions of system (16), that is, all the values of x which satisfy both equations of the system.

The first equation of system (16) has the following group of roots:

$$x=\frac{3\pi}{14}+\frac{2k\pi}{7}, \quad k=0, \pm 1, \pm 2, \ldots$$

the second equation, the roots

$$x=\frac{\pi}{2}+n\pi, \quad n=0, \pm 1, \pm 2, \ldots$$

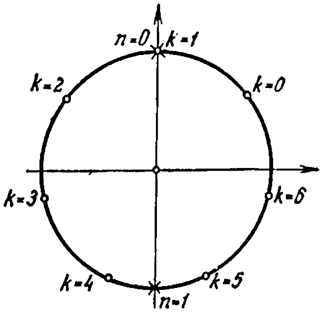
We have to choose all values of x which *simultaneously appear in both* groups (that is, to find all values of x which belong, for some integer k, to the first group and, for some integer n, to the second group).

To do this, we take advantage of the trigonometric circle,* on which we mark with points the values of x which lie in the first group for $k =$

* Selection of the values of x belonging to both the indicated groups can also be done in a purely analytic fashion (without resorting to the trigonometric circle), by the method used in Problem 4 of Sec. 1.2.

$= 0, 1, 2, \ldots, 6$ (Fig. 57). We note that the points representing the other values of x of this group (for the remaining values of k) are repeated every 7 units (for instance, a point associated with a value of x when $k = 9$ coincides with the point associated with the value of x for $k = 2$, and so on). The values of x of the second group are marked with crosses for $n = 0, 1$; the points associated with the other values of n are repeated every two units.

Fig. 57



From Fig. 57 it is clear that the groups under consideration have in common those values of x which are associated with the upper endpoint of the vertical diameter; these values are obtained from the second group for $n = 2p$, $p = 0, \pm 1, \pm 2, \ldots$ and from the first group for $k = 7q + 1$, $q = 0, \pm 1, \pm 2, \ldots$.

Thus the solution of system (16) and, hence, of the original equation, is

$$x = \frac{\pi}{2} + 2p\pi, \quad p = 0, \pm 1, \pm 2, \ldots$$

In this solution we made use of the transformation of the original equation to the form (15). A little ingenuity however will suggest how to dispense with that.

Take a closer look at the original equation. Its left member is the sum of a sine and a cosine, the right member is the number -2. But by the property of a sine and cosine, the inequalities $\sin 7x \geqslant -1$, $\cos 2x \geqslant -1$ hold true for arbitrary x, whence by combining these inequalities we get $\sin 7x + \cos 2x \geqslant -2$. Hence, the original equation is satisfied if and only if *both* terms of the left member are equal to -1, which is to say, when x *simultaneously* satisfies the two equations

$$\sin 7x = -1, \quad \cos 2x = -1$$

We again arrive at a system of equations in one unknown; its solution can be carried out in a manner similar to that used to solve system (16).

15. *Solve the equation*

$$2 \sin\left(x + \frac{\pi}{4}\right) = \tan x + \cot x \tag{17}$$

Elementary manipulations yield

$$\sin\left(x+\frac{\pi}{4}\right)\sin 2x=1 \qquad (18)$$

or

$$\cos\left(x-\frac{\pi}{4}\right)-\cos\left(3x+\frac{\pi}{4}\right)=2$$

The result is an equation of the same type as considered in the preceding problem. The reader will find no difficulty in performing the rest of the solution.

However, a careful examination of equation (18) suggests that an apt employment of the properties of trigonometric functions will save us extra manipulations. The key to the solution here is that the sine of any argument (number) does not exceed unity in absolute value.

Therefore the product in the left member of (18) can equal 1 in only two cases: when each of the factors is equal to 1 or when each of the factors is equal to —1. Thus, the number x will be the root of (18) if and only if the number satisfies *one of the following two* systems of equations:

$$\begin{aligned} \sin\left(x+\frac{\pi}{4}\right)&=1 \\ \sin 2x&=1 \end{aligned} \quad \text{or} \quad \begin{aligned} \sin\left(x+\frac{\pi}{4}\right)&=-1 \\ \sin 2x&=-1 \end{aligned}$$

We consider the first system. From the second equation we have $x = (\pi/4) + k\pi$, $k = 0, \pm 1, \pm 2, \ldots$. Substituting these values of x into the first equation, we get $\sin[(\pi/2) + k\pi] = 1$, which is only valid for *even* values of k, that is, when $k = 2n$, $n = 0, \pm 1, \pm 2, \ldots$. Hence the solutions of the first system are $x = (\pi/4) + 2n\pi$, where n is an arbitrary integer.

Solving the second system similarly, we see that it does not have any solution, and so the solutions of the first system are the roots of the original equation.

But the shortest solution of (17) is that which makes good use of inequalities. The right member has absolute value greater than or equal to 2 for any (admissible) value of x, whereas the absolute value of the left member does not exceed 2. And so equation (17) can only be satisfied by those values of x for which *both sides* of (17) are equal either to 2 or to —2. Let us consider these possibilities.

The right member of (17) assumes the value 2 when $\tan x = \cot x = 1$, that is, when $x = (\pi/4) + k\pi$, $k = 0, \pm 1, \pm 2, \ldots$. The left member of (17) assumes the same value when $x = (\pi/4) + 2n\pi$, $n = 0, \pm 1, \pm 2, \ldots$. The common part of these two groups, namely, the values $x = (\pi/4) + 2n\pi$, n any integer, yields the roots of equation (17).

Now, the right member of (17) is equal to —2 when $x = (3\pi/4) + n\pi$, $n = 0, \pm 1, \pm 2, \ldots$. But for these values of x the left member of (17) is 0 and, hence, the equation is not satisfied.

Trigonometric equations involving *parameters* besides unknowns constitute a special group. In these problems the first thing to decide is the values of the parameters for which solutions *exist*. And of course the solutions themselves must be found (depending on the parameters).

Although solving problems involving parameters does not presuppose any additional knowledge, the necessary investigation is at times rather difficult logically and technically.

16. *For every real number find all the real solutions to the equation*

$$\sin x + \cos(a+x) + \cos(a-x) = 2$$

Combining the second and third terms of the left side, we at once get an equation of type (1):

$$\sin x + 2\cos a \cos x = 2$$

It is natural to solve it by the auxiliary-angle method. But it will then be necessary to take into account the various possibilities which arise for various values of the parameter a.

The conditions defining the auxiliary angle may be written down as follows:

$$\sin\beta = \frac{1}{\sqrt{1+4\cos^2 a}}, \quad \cos\beta = \frac{2\cos a}{\sqrt{1+4\cos^2 a}} \tag{19}$$

Then the last equation can be reduced to the form

$$\cos(x-\beta) = \frac{2}{\sqrt{1+4\cos^2 a}} \tag{20}$$

This equation is known to have roots only if the right member does not exceed 1 in absolute value. But since it is positive (the principal square root!) for arbitrary a, equation (20) has solutions only for those values of the parameter a which satisfy the inequality

$$\frac{2}{\sqrt{1+4\cos^2 a}} \leqslant 1$$

(for the other values of a, equation (20) has no roots).

It is not difficult to solve this inequality (see Sec. 1.10). It reduces to the form $\cos^2 a \geqslant 3/4$, whence

$$-\frac{\pi}{6} + k\pi \leqslant a \leqslant \frac{\pi}{6} + k\pi, \quad k = 0, \pm 1, \pm 2, \ldots \tag{21}$$

To summarize: equation (20) (and, hence, the original equation) *has solutions only for those values of a that satisfy Condition* (21).

It is now easy to find the solutions of (20), (and, hence, of the original equation) that correspond to any value of a satisfying Condition (21): substitute the expression for the auxiliary angle β into the following equation:

$$x = \beta \pm \arccos\frac{2}{\sqrt{1+4\cos^2 a}} + 2n\pi, \quad n = 0, \pm 1, \pm 2, \ldots$$

Many students went straight to (19) without thinking very much and wrote down for β the expression obtained from the first formula of (19):

$$\beta = \arcsin \frac{1}{\sqrt{1+4\cos^2 a}}$$

But the angle β defined by this equation does not satisfy the second condition of (19) for all values of the parameter a. This angle lies in the first quadrant (because the expression under the arcsine symbol is positive for all values of a), and so its cosine will also be positive for all values of the parameter a. Yet the second formula of (19) indicates that for $\cos a < 0$ the cosine of the auxiliary angle must be negative.

The following hint may be an aid in choosing the expression for the auxiliary angle β. Since $\sin \beta$ is always positive [this is evident from the first formula of (19)], it follows that the angle β itself may be chosen either in the first or in the second quadrant. But this is precisely where the arc cosine lies. And so for the auxiliary angle we can take

$$\beta = \arccos \frac{2\cos a}{\sqrt{1+4\cos^2 a}}$$

The foregoing example shows that in problems involving a parameter it is not always possible to choose the auxiliary angle so that it lies in the irst quadrant and is expressed by a single formula suitable for *all* values of the parameter.

17. *Find all the solutions of the equation*

$$(\sin x+\cos x)\sin 2x = a(\sin^3 x+\cos^3 x)$$

located between $\pi/2$ and π. For which values of a does this equation have at most one solution satisfying the condition $\pi/2 \leqslant x \leqslant \pi$?

The given equation can at once be written in the decomposable form

$$(\sin x+\cos x)(\sin 2x - a + a\sin x\cos x) = 0$$

and so there is always (i.e., for any value of the parameter a) *at least one* root located in the interval at hand $\pi/2 \leqslant x \leqslant \pi$, namely the root $x = 3\pi/4$ of the equation $\sin x+\cos x = 0$.

Now let us seek the remaining solutions of the original equation, that is, the solutions of the equation $\sin 2x - a + a\sin x\cos x = 0$ or $(2+a)\sin 2x = 2a$.

For any value of the parameter a different from -2, this equation can be represented in the form

$$\sin 2x = \frac{2a}{2+a} \tag{22}$$

We are only interested in the roots x which are located between $\pi/2$ and π. But in this case, $\pi < 2x < 2\pi$ and so $\sin 2x$ must be nonpositive but not less than -1. Hence, equation (22) has roots located between $\pi/2$ and π only for those values of the parameter $a \neq -2$ for which

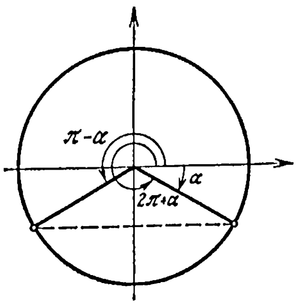
$$-1 \leqslant \frac{2a}{2+a} \leqslant 0 \tag{23}$$

This inequality can readily be solved: $-2/3 \leqslant a \leqslant 0$.

Thus, for every value of a satisfying the condition $-2/3 \leqslant a \leqslant 0$, equation (22) has roots lying between $\pi/2$ and π. There are no such roots for the remaining values of a different from -2.

Now let us find the roots x of (22) that lie between $\pi/2$ and π, assuming that $-2/3 \leqslant a \leqslant 0$. To do this, denote $2x$ by y and rewrite (22) as $\sin y = 2a/(2+a)$. We determine those roots y of this equation

Fig. 58



(for $-2/3 \leqslant a \leqslant 0$) which satisfy the condition $\pi \leqslant y \leqslant 2\pi$. Since, by virtue of inequality (23), the angle

$$\alpha = \arcsin \frac{2a}{2+a}$$

lies between $-\pi/2$ and 0 (see Sec. 2.5), it is easy to see, using the trigonometric circle in Fig. 58, that the equation $\sin y = 2a/(2+a)$ has only two solutions satisfying the condition $\pi \leqslant y \leqslant 2\pi$, namely,

$$y_1 = \pi - \arcsin \frac{2a}{2+a} \quad \text{and} \quad y_2 = 2\pi + \arcsin \frac{2a}{2+a}$$

And so the roots of equation (22) (for $-2/3 \leqslant a \leqslant 0$) lying between $\pi/2$ and π are

$$x_1 = \frac{\pi}{2} - \frac{1}{2} \arcsin \frac{2a}{2+a}, \quad x_2 = \pi + \frac{1}{2} \arcsin \frac{2a}{2+a}$$

Thus, to summarize, when $-2/3 \leqslant a \leqslant 0$, the original equation is satisfied by three numbers: $3\pi/4$, x_1, x_2. This is not however grounds for asserting that the original equation has *more* than one root for each value of a in the indicated interval, because it may turn out that $x_1 = x_2 = 3\pi/4$ for some of these values of a. That is exactly what happens when $a = -2/3$ and only for that value. This can easily be seen.

Consequently, for $-2/3 < a \leqslant 0$ the original equation has three roots in the interval $\pi/2 \leqslant x \leqslant \pi$, namely $3\pi/4$, x_1, x_2; for the remaining values of a, except $a = -2$, there is only one root, $3\pi/4$.

So far the case $a = -2$ has not been analyzed. Given this value of the parameter, we cannot consider equation (22) and have to make a direct study of the equation $(2 + a) \sin 2x = 2a$, which (for $a = -2$) assumes the form $0 = -4$. Since this equation does not have any roots, the original equation for $a = -2$ as well has only one root, $3\pi/4$, in the interval between $\pi/2$ and π.

Exercises

1. How are the angles α and β related if it is known that (a) $\sin\alpha = \sin\beta$, (b) $\cos\alpha = \cos\beta$, (c) $\tan\alpha = \tan\beta$, (d) $\sin\alpha = \cos\beta$?

2. Determine whether the equation $a \sin x + b \cos x = c$ has any solution.

Solve the following equations.

3. $\sin 2x - \tan(\pi/6) \cos 2x = 1$.

4. $\sin[2x - (\pi/2)] + \cos[2x - (\pi/12)] = \sqrt{2} \cos[3x + (\pi/6)]$.

5. $\sin 8x - \cos 6x = \sqrt{3}(\sin 6x + \cos 8x)$.

6. $\sin 3x + 4 \sin^3 x + 4 \cos x = 5$.

7. $2 \sin 4x - 3 \sin^2 2x = 1$.

8. $-2 + 4 \cos^2 z = \cos z + \sqrt{3} \sin z$.

9. $2 \sin^2 \left(\frac{\pi}{2} \cos^2 x\right) = 1 - \cos(\pi \sin 2x)$.

10. $\dfrac{1 + \tan x}{1 + \cot x} = 2 \sin x$.

11. $2\left[1 - \sin\left(\frac{3\pi}{2} - x\right)\right] = \sqrt{3} \tan \frac{\pi - x}{2}$.

12. $3 \cos^2 x - \sin^2 x - \sin 2x = 0$.

13. $\cot[(\pi/4) - x] = 5 \tan 2x + 7$.

14. $2 \cos 2x + \sin 3x - 2 = 0$.

15. $\frac{7}{4} \cos \frac{x}{4} = \cos^3 \frac{x}{4} + \sin \frac{x}{2}$.

16. $\tan x = (2 + \sqrt{3}) \tan(x/3)$.

17. $\cos(10x + 12) + 4\sqrt{2} \sin(5x + 6) = 4$.

18. $\cot^3\left(\frac{\pi - \pi x}{1 + x}\right) - \sqrt{3} \cot^2\left(\frac{\pi - \pi x}{1 + x}\right) = 6 \cot\left(\frac{\pi - \pi x}{1 + x}\right)$.

19. $4 \sin^4 x + \cos 4x = 1 + 12 \cos^4 x$.

20. $\sin 3t \cos t = \frac{3}{2} \tan t$

21. $\left(2 \sin^4 \frac{x}{2} - 1\right) \dfrac{1}{\cos^4 \frac{x}{2}} = 2$.

22. $\sin^8 x + \cos^8 x = \frac{17}{16} \cos^2 2x$.

23. $\sin^6 2x+\cos^6 2x=\frac{7}{16}$.

24. $\cos^2 x+\cos^2\left(\frac{3x}{4}\right)+\cos^2\left(\frac{x}{2}\right)+\cos^2\left(\frac{x}{4}\right)=2$.

25. $\sin y+\cos 3y=1-2\sin^2 y+\sin 2y$.

26. $\sin 2x\sin 4x\sin 6x=\frac{1}{4}\sin 4x$.

27. $\cos^2(x-\gamma)+\cos^2(0.5x+\beta-\gamma)$
$$-2\cos(0.5x-\beta)\cos(x-\gamma)\cos(0.5x+\beta-\gamma)=\frac{3}{4}.$$

28. $\sin^2 x+\sin 2x\sin 4x+\ldots+\sin nx\sin n^2x=1$.

29. $\sin x+\sin 2x+\sin 3x+\sin 4x=0$.

30. $\sin x\cos 2x+\sin 2x\cos 5x=\sin 3x\cos 5x$.

31. $\frac{\sin x+\sin 3x+\sin 5x}{\cos x+\cos 3x+\cos 5x}+2\tan x=0$.

32. $\cos x+\sin x=\frac{\cos 2x}{1-\sin 2x}$.

33. $\sin 3x+\sin x+2\cos x=\sin 2x+2\cos^2 x$.

34. $\frac{2}{\sqrt{3}}(\tan x-\cot x)=\tan^2 x+\cot^2 x-2$.

35. $\sqrt{17\sec^2 x+16\left(\frac{1}{2}\tan x\sec x-1\right)}=2\tan x(1+4\sin x)$.

36. $\sqrt{\frac{1}{16}+\cos^4 x-\frac{1}{2}\cos^2 x}+\sqrt{\frac{9}{16}+\cos^4 x-\frac{3}{2}\cos^2 x}=\frac{1}{2}$.

37. $\sin(5\pi 2^x)+\sin(\pi 2^x)-2\sin(3\pi 2^x)=8\sin^2(\pi 2^x)+2\cos(3\pi 2^x)$
$$-\cos(\pi 2^x)-\cos(5\pi 2^x).$$

38. $2\cos^2(\pi 4^x)-\sin(\pi 4^{x+1})+\sin(\pi 4^{x+1/4})-2\cos(\pi 4^{x+1/4})=0$.

39. $6\tan x+5\cot 3x=\tan 2x$.

40. $5(\sin x+\cos x)+\sin 3x-\cos 3x=2\sqrt{2}(2+\sin 2x)$.

41. Find all the solutions of the equation
$$\sin x+\sin\frac{\pi}{8}\sqrt{(1-\cos x)^2+\sin^2 x}=0$$
lying between $5\pi/2$ and $7\pi/2$.

42. Find all the solutions of the equation
$$\frac{\sqrt{1-\cos x}+\sqrt{1+\cos x}}{\cos x}=4\sin x$$
between 0 and 2π.

Solve the following equations.

43. $\tan 2x\tan 7x=1$.

44. $\log_{\frac{6x-x^2}{11}}(-\cos x-\cos 3x)=\log_{\frac{6x-x^2}{11}}(-\cos 2x)$.

45. $\log_{\frac{9x-x^2-14}{7}}(\sin 3x-\sin x)=\log_{\frac{9x-x^2-14}{7}}\cos 2x$.

46. $\sin x+\sin 9x=2$.

47. $\cos x-\sin 3x=-2$

48. $\sin\left(\frac{5x}{2}\right)-\sin\left(\frac{x}{2}\right)=2$.

49. $\cos x\cos 6x=-1$.

50. $(\sin x-\sqrt{3}\cos x)\sin 3x=2.$

51. $\cos(\pi\sqrt{x})\cos(\pi\sqrt{x-4})=1.$

52. $\sin x\sin 7x=1.$

53. $\tan x-\tan 2x=\sin x$

54. Prove that the equation

$$\sin 2x+\sin 3x+\ldots+\sin nx=n-1$$

has no solutions for an arbitrary integer $n>2$.

55. For which values of a does the equation

$$4\sin\left(x+\frac{\pi}{3}\right)\cos\left(x-\frac{\pi}{6}\right)=a^2+\sqrt{3}\sin 2x-\cos 2x$$

have solutions? Find these solutions.

56. Determine for which values of a the equation $a^2-2a+\sec^2\pi(a+x)=0$ has solutions and find them.

57. For which values of a does the equation $\sin^2 x+\sin x-a=0$ have solutions? Find the solutions in the interval $0\leqslant x<2\pi$.

58. Solve the equation $\sin 4x=m\tan x$, $m>0$.

59. For which values of b does the equation

$$\frac{b\cos x}{2\cos 2x-1}=\frac{b+\sin x}{(\cos^2 x-3\sin^2 x)\tan x}$$

have solutions? Find these solutions.

60. For which values of a does the equation

$$\frac{a^2}{1-\tan^2 x}=\frac{\sin^2 x+a^2-2}{\cos 2x}$$

have solutions? Find these solutions.

2.4 Systems of trigonometric equations

To solve a system of trigonometric equations means (in accord with the general definition of a solution to a system of equations given in algebra) *to find all sets of values of the unknowns that make all equations of the system true statements simultaneously.*

Ordinarily, when solving trigonometric systems the attempt is made either to eliminate one of the unknowns by expressing it in terms of the others in one of the equations of the system, or to reduce the trigonometric system to a system of algebraic equations by aptly introducing new unknowns or via a transformation of the equations of the system.

Naturally, when solving systems of trigonometric equations one must be careful not to lose any solutions or to discard any extraneous solutions if they have been introduced.

The solution of trigonometric systems does not require any special techniques or knowledge extending beyond the ordinary course of trigonometry. Nevertheless they involve certain difficulties, one of which is due to the fact that these systems, as a rule, have *infinitely many solutions*. For this reason, the proper notation of the set of values of the unknowns which constitute a solution, the choice of needed solutions, etc. may be complicated by the necessity to regard different cases or to solve auxiliary inequalities.

1. *Solve the system of equations*

$$\frac{1-\tan x}{1+\tan x}=\tan y$$
$$x-y=\frac{\pi}{6}$$

The second equation makes it easy to express one of the unknowns in terms of the other. This suggests that the system is best solved by direct elimination of one of the unknowns; it then reduces to an ordinary trigonometric equation.

Which unknown is eliminated is immaterial, we choose y. Since $y = x - (\pi/6)$, substitution of this expression into the first equation of the system gives a trigonometric equation in x:

$$\frac{1-\tan x}{1+\tan x}=\tan\left(x-\frac{\pi}{6}\right)$$

The left member of this equation can easily be reduced to $\tan[(\pi/4)-x]$. Then with the aid of the formula for the difference of tangents we get

$$\sin\left(2x-\frac{5\pi}{12}\right)=0 \quad \text{whence} \quad 2x-\frac{5\pi}{12}=k\pi, \quad k=0, \pm 1, \pm 2, \ldots$$

Hence, the solutions of the original system are:

$$x=\frac{5\pi}{24}-\frac{k\pi}{2}, \quad y=\frac{\pi}{24}-\frac{k\pi}{2}, \quad k=0, \pm 1, \pm 2, \ldots$$

A check, which is obligatory, convinces us that all the pairs x, y of values that we obtained satisfy the original system. We emphasize the fact that to each integer k there corresponds a *pair* of values x, y computed from these formulas; such a pair constitutes a *solution* of the original system. The original system has an infinity of solutions

2. *Solve the system of equations*

$$\tan\frac{x}{2}+\tan\frac{y}{2}-\cot\frac{z}{2}=0$$
$$\cos(x-y-z)=\frac{1}{2}$$
$$x+y+z=\pi$$

The last equation permits eliminating z at once. Substituting $z = \pi - (x+y)$ into the first two equations, we get a system of two equations in two unknowns:

$$\left.\begin{aligned}\tan\frac{x}{2}+\tan\frac{y}{2}&=\tan\frac{x+y}{2}\\ \cos 2x&=-\frac{1}{2}\end{aligned}\right\} \qquad (1)$$

The second equation can be solved immediately:

$$x = \pm\frac{\pi}{3} + k\pi, \qquad k = 0, \pm 1, \pm 2, \ldots \tag{2}$$

It would now seem natural to substitute this expression for x into the first equation of (1) and thus reduce the system (1) to a single equation. But this method leads to a rather unwieldy trigonometric equation in y (though it can be solved of course).

Let us try a different approach. We will transform the first equation of system (1). Applying the formula for a sum of tangents to the left member and performing obvious manipulations, we get the relation

$$\sin\frac{x+y}{2}\left(\cos\frac{x+y}{2} - \cos\frac{x}{2}\cos\frac{y}{2}\right) = 0 \tag{3}$$

Equating to zero the first factor, we get an algebraic relation involving x and y:

$$x + y = 2n\pi, \qquad n = 0, \pm 1, \pm 2, \ldots \tag{4}$$

If we now recall the expressions for x and z, it will be easy to get the first group of values of the unknowns:

$$\begin{aligned} &x_1 = \pm\frac{\pi}{3} + k\pi, && k = 0, \pm 1, \pm 2, \ldots, \\ &y_1 = \mp\frac{\pi}{3} + (2n-k)\pi, && n = 0, \pm 1, \pm 2, \ldots, \\ &z_1 = \pi - 2n\pi \end{aligned} \tag{5}$$

Equate the second factor of the left member of (3) to zero. Using the formula for the cosine of a sum, we at once get the relation $\sin(x/2) \times \sin(y/2) = 0$. But by (2), $\sin(x/2) \neq 0$ and so $\sin(y/2) = 0$, whence $y = 2m\pi$, m an integer. Recalling the expressions for x and z, we find the second group of values of the unknowns:

$$\begin{aligned} &x_2 = \pm\frac{\pi}{3} + k\pi, && k = 0, \pm 1, \pm 2, \ldots, \\ &y_2 = 2m\pi, && m = 0, \pm 1, \pm 2, \ldots, \\ &z_2 = \pi \mp \frac{\pi}{3} - (2m+k)\pi \end{aligned}$$

A check shows that both of the groups just found are indeed solutions of the original system.

A few words are in order concerning the notation of solutions to trigonometric systems.

In solving this problem, many students reasoned as follows: "Since it follows from (4) that $y = 2n\pi - x$, then, taking into account expression (2) for x, we find

$$y = 2n\pi \pm \frac{\pi}{3} - k\pi = \pm\frac{\pi}{3} + (2n-k)\pi\text{"} \tag{6}$$

Some even tried to substantiate the legitimacy of this formula: "Since in (2) we can take both signs (plus and minus), so in the expression for y we have to take both signs, and this is indicated in (6) by the $\pm$ sign there."

Actually, the expression (6) for y is not true: because if we take a certain value of x corresponding to the choice of a plus sign in (2), then the appropriate value of y corresponds to the choice of a minus sign in the second formula of (5). Thus, the signs in the formulas of (5) do not indicate an arbitrary choice of signs (plus or minus) in each of the formulas but a very definite choice: in all these formulas we take *either both upper signs simultaneously or both lower signs.*

It is important to understand properly the notation of (5). In particular, it means that to each choice of values k and n there correspond two solutions, or two number triples x, y, z, of the original system.

Another typical mistake of students is that they frequently denote arbitrary integers by one and the same letter. For example, in place of (4) many students wrote $x + y = 2k\pi$, $k = 0, \pm 1, \pm 2, \dots$ and, taking into account (2) and the expression for z, they obtained [in place of (5)] the group

$$x = \pm\frac{\pi}{3} + k\pi, \quad y = \mp\frac{\pi}{3} + k\pi, \quad z = \pi - 2k\pi$$

Although these number triples do satisfy the original system, many of the solutions have been lost. The reason is that in passing from the equations of system (1) to the equations for x and $x + y$, we must retain the independence of these equations by introducing distinct integral parameters k and n [as was done in (2) and (4)] and not connect them by introducing one and the same integer k.

3. *Solve the system of equations*

$$\cos x + \cos y = 1$$

$$\cos\frac{x}{2} + \cos\frac{y}{2} = \frac{\sqrt{2}}{2} - 1$$

In this system there is no equation that would permit us to express *directly* one of the unknowns in terms of the other and thus eliminate one of the unknowns. And so we will try to transform the equations of the system in order to obtain an algebraic system of equations in certain trigonometric functions of the unknowns x and y.

The cosine of any angle, it will be recalled, can be expressed in terms of the cosine of half an angle. This suggests denoting $\cos(x/2)$ by u and $\cos(y/2)$ by v and writing the first equation of the system in terms of u and v.* Obvious transformations lead at once to the system

$$u^2+v^2=\frac{3}{2}$$

$$u+v=\frac{\sqrt{2}}{2}-1$$

which is an ordinary algebraic system.

This system can easily be solved. Isolating in the left member of the first equation the perfect square of the sum of the unknowns u and v and taking into account the second equation, we get the value of the product uv. Thus, we obtain the following system:

$$u+v=\frac{\sqrt{2}}{2}-1$$

$$uv=-\frac{\sqrt{2}}{2}$$

which can be solved by the familiar reduction to a quadratic equation (via the converse of the Viète theorem). However, it is easier simply to guess the solution of the last system. It is evident that there are two pairs of such numbers,* that the sum of these numbers is equal to $(\sqrt{2}/2)-1$ and the product is equal to $-\sqrt{2}/2$:

$$u_1=\sqrt{2}/2, \quad v_1=-1, \quad u_2=-1, \quad v_2=\sqrt{2}/2$$

Hence, to determine the unknowns x and y we must now solve two systems of equations:

$$\cos\frac{x}{2}=\frac{\sqrt{2}}{2} \qquad\qquad \cos\frac{x}{2}=-1$$

and

$$\cos\frac{y}{2}=-1 \qquad\qquad \cos\frac{y}{2}=\frac{\sqrt{2}}{2}$$

whence we find two groups of solutions of the original system:

$$x_1=\pm\frac{\pi}{2}+4k\pi, \quad y_1=2\pi+4n\pi, \quad k, n=0, \pm1, \pm2, \ldots,$$

$$x_2=2\pi+4p\pi, \quad y_2=\pm\frac{\pi}{2}+4q\pi, \quad p, q=0, \pm1, \pm2, \ldots$$

In the next problem, again, elimination of one of the unknowns cannot be carried out directly, but the system can rather simply be

---

* Incidentally, we could express $\cos(x/2)$ in terms of $\cos(y/2)$ in the second equation and then, transforming the left member of the first equation to $\cos(x/2)$ and $\cos(y/2)$, eliminate $\cos(x/2)$. The result would be a quadratic equation in $\cos(y/2)$.

** Of course our guess does not mean that the system does not have any other solutions. For complete rigour we must refer to the fact that any system of the form $u+v=a$, $uv=b$ has at most two solutions.

reduced to a form that permits finding the solutions. The peculiarity of this problem is that we will not be interested in all the solutions of the trigonometric system but only in those that satisfy certain supplementary conditions.

4. *Find all the solutions of the system*

$$|\sin x|\ \sin y = -\frac{1}{4}$$

$$\cos(x+y)+\cos(x-y)=\frac{3}{2}$$

that satisfy the conditions $0 < x < 2\pi$, $\pi < y < 2\pi$.

To solve this system we employ the common technique of getting rid of the absolute-value sign. We consider two cases: $\sin x > 0$ and $\sin x < 0$. The former inequality is valid (with regard for the restriction on x imposed by the statement of the problem) for $0 < x < \pi$, the latter, for $\pi < x < 2\pi$. It is obvious that the equation $x = \pi$ is impossible.

To start with, let $\sin x > 0$, that is, $0 < x < \pi$. Then the first equation of the system can be rewritten as $\sin x \sin y = -1/4$, and we can expand the product of sines into a difference of cosines; the original system is thus reduced to the system

$$\cos(x-y)-\cos(x+y)=-\frac{1}{2}$$

$$\cos(x+y)+\cos(x-y)=\frac{3}{2}$$

whence we get a very simple system:

$$\begin{aligned}\cos(x-y)&=\frac{1}{2}\\ \cos(x+y)&=1\end{aligned}\qquad(7)$$

This system can clearly be reduced to an algebraic system of two linear equations in two unknowns:

$$x-y=\pm\frac{\pi}{3}+2k\pi,\qquad k=0,\ \pm1,\ \pm2,\ \ldots,$$

$$x+y=2n\pi,\qquad n=0,\ \pm1,\ \pm2,\ \ldots$$

which readily enables us to find all the solutions of system (7). However the subsequent choice of solutions that satisfy the supplementary restrictions would be very awkward because we would have to solve inequalities in integers and run through a variety of possibilities.

It is therefore easier to seek at once the needed solutions, that is, the solutions of system (7) lying in the intervals that interest us: $0 < x < \pi$ (we assumed that $\sin x > 0$) and $\pi < y < 2\pi$. From these inequalities it follows that

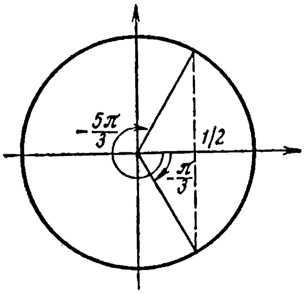
$$-2\pi < x-y < 0$$

$$\pi < x+y < 3\pi$$

In other words, we have to find only those solutions of system (7) for which these inequalities hold true.

But if the difference $x - y$ lies between -2π and 0, then the equation $\cos(x - y) = 1/2$ is only valid in two cases: when $x - y = -\pi/3$ and when $x - y = -5\pi/3$. This can be seen, say, by referring to the trigonometric circle in Fig. 59 or to the graph of the cosine. Furthermore, if the sum $x + y$ is situated between π and 3π, then the equation $\cos(x + y) = 1$ is valid in only one case, when $x + y = 2\pi$.

Fig. 59



Thus, we have two linear algebraic systems:

$$\begin{aligned} x - y &= -\pi/3 \\ x + y &= 2\pi \end{aligned} \qquad \text{and} \qquad \begin{aligned} x - y &= -5\pi/3 \\ x + y &= 2\pi \end{aligned}$$

whence we get, respectively, two solutions to our problem:

$$x_1 = 5\pi/6, \quad y_1 = 7\pi/6 \text{ and } x_2 = \pi/6, \quad y_2 = 11\pi/6$$

The case of $\sin x < 0$, which is valid for $\pi < x < 2\pi$, is considered in the same manner and yields another two solutions to the original system that satisfy the inequalities given in the statement of the problem:

$$x_3 = 7\pi/6, \quad y_3 = 7\pi/6 \text{ and } x_4 = 11\pi/6, \quad y_4 = 11\pi/6$$

5. *Find the solutions of the system*

$$\log_2 x \log_y 2 + 1 = 0$$
$$\sin x \cos y = 1 - \cos x \sin y$$

which satisfy the condition $x + y < 8$.

In solving this system, we again have to deal with a choice of solutions. But this time, besides the inequality $x + y < 8$ indicated in the statement of the problem there are other supplementary conditions necessitated by the fact that we have to ensure the existence of the logarithms in the system. Namely, we can only be interested in pairs of values of x and y such that for them $x > 0$, $y > 0$, $y \neq 1$.

Remembering these supplementary conditions, we can easily reduce the first equation of the system to the form $xy = 1$. The second equation of the system can be rewritten as $\sin(x+y)=1$, whence $x+y=$ $=(\pi/2)+2k\pi$, $k=0, \pm 1, \pm 2, \ldots$.

Thus, in place of the original trigonometric system we have the following infinite set of algebraic systems:

$$\begin{aligned} &x+y=\frac{\pi}{2}+2k\pi, \qquad k=0,\ \pm 1,\ \pm 2,\ \ldots \\ &xy=1 \end{aligned} \tag{8}$$

Here we are interested only in such solutions of each of these systems for which $x+y<8$, $x>0$, $y>0$, $y \neq 1$. Instead of solving all the systems of (8) and then choosing the solutions we need, we will start from the very beginning to seek the ones we need.

From the supplementary conditions $x+y<8$, $x>0$, $y>0$ it follows that $0<x+y<8$ and so it makes sense to consider only those systems of (8) which correspond to (integral) values of k that satisfy the inequality

$$0<\frac{\pi}{2}+2k\pi<8$$

A direct enumeration (see Problem 15 of Sec. 2.3) shows that this inequality is satisfied by only two (integral) values of k, namely, $k=0$ and $k=1$; the systems of (8) which correspond to the remaining values of k cannot contain solutions of interest to us.

It thus remains to solve the two algebraic systems

$$\begin{aligned} x+y&=\frac{\pi}{2} \quad \text{and} \quad & x+y&=\frac{5\pi}{2} \\ xy&=1 & xy&=1 \end{aligned} \tag{9}$$

the first of which does not have any real solutions and the second of which yields

$$x_1=\frac{5\pi+\sqrt{25\pi^2-16}}{4}, \quad y_1=\frac{5\pi-\sqrt{25\pi^2-16}}{4},$$
$$x_2=\frac{5\pi-\sqrt{25\pi^2-16}}{4}, \quad y_2=\frac{5\pi+\sqrt{25\pi^2-16}}{4}$$

We are convinced directly that for each of these solutions $x>0$, $y>0$, $y \neq 1$. The first two of these conditions are obvious, incidentally, from the second system of (9) (because both the sum and the product of the numbers x and y are positive). As to the condition $y \neq$ $\neq 1$, it can be obtained directly from the second system of (9). It does not have any solutions with $y=1$.

Let us consider yet another system of trigonometric equations whose solution requires a certain amount of ingenuity in carrying out trigonometric transformations. We give several modes of solution to illustrate a certain diversity of approaches that may prove useful in solving other trigonometric systems.

6. *Solve the system of equations*

$$\begin{aligned} \sin x + \sin y &= \sin \alpha \\ \cos x + \cos y &= \cos \alpha \end{aligned} \tag{10}$$

First mode. The underlying idea of this solution is to obtain a certain algebraic relationship between the unknowns x and y and then, via substitution, to eliminate one of the unknowns. In so doing, note that if we rewrite the original system as

$$\begin{aligned} 2\sin\frac{x+y}{2}\cos\frac{x-y}{2} &= \sin\alpha \\ 2\cos\frac{x+y}{2}\cos\frac{x-y}{2} &= \cos\alpha \end{aligned} \tag{11}$$

the left members of both equations contain one and the same factor $\cos[(x-y)/2]$. If we could eliminate this factor, the result would be a trigonometric equation in $x + y$ which would then yield a linear relationship between x and y.

Let us now carry through a full solution of system (10) by elimination of the factor $\cos[(x-y)/2]$ from the equations of (11).

First suppose that the angle α is such that $\cos\alpha \neq 0$ (the opposite case will be given special consideration below). Then, too, the left member of the second equation of (11) is nonzero and so we can divide the first equation termwise by the second to get

$$\tan\frac{x+y}{2} = \tan\alpha \tag{12}$$

whence

$$x + y = 2\alpha + 2k\pi, \quad k = 0, \pm 1, \pm 2, \dots \tag{13}$$

Note that the transition from (12) to (13) leads to an *equivalent* equation. Indeed, if $x + y = 2\alpha + 2k\pi$, then $(x+y)/2 = \alpha + k\pi$ and since $\tan(\alpha + k\pi) = \tan\alpha$ exists ($\cos\alpha \neq 0$ by assumption), it follows that $\tan[(x+y)/2]$ is also meaningful and equation (12) is valid.

What follows now is obvious. From relation (13) we find

$$y = 2\alpha + 2k\pi - x \tag{14}$$

which is substituted into the second equation of the system (10): $\cos x + \cos(2\alpha + 2k\pi - x) = \cos\alpha$. Now, making use of the periodicity of the cosine and transforming the sum of cosines into a product, we get the equation $2\cos\alpha\cos(x-\alpha) = \cos\alpha$.

Earlier we assumed that $\cos\alpha \neq 0$, and so from the last relation it follows at once that $\cos(x-\alpha) = 1/2$, i.e. $x = \alpha \pm (\pi/3) + 2n\pi$, $n = 0, \pm 1, \pm 2, \dots$. Substituting the values of x thus found into (14), we get the solutions of the original system (10):

$$x = \alpha \pm \frac{\pi}{3} + 2n\pi, \quad y = \alpha \mp \frac{\pi}{3} + 2(k-n)\pi \tag{15}$$

Here, k and n are arbitrary integers and in both formulas we take, simultaneously, either the upper or the lower signs.

We have thus completed consideration of the case $\cos\alpha \neq 0$. It must be borne in mind of course that all the foregoing relations (12), (13) and others and, in particular, the answer, (15), are valid *only for those values of* α *for which* $\cos\alpha \neq 0$ because they were all obtained on that assumption. For example, for the time being we cannot write formulas (15) for $\alpha = \pi/2$, although in themselves these formulas are meaningful for this value of α.

Now let us examine the case when $\cos\alpha = 0$. It can rather easily be reduced to the preceding cases. Indeed, if $\cos\alpha = 0$ then surely $\sin\alpha \neq 0$, and so we consider the original system (10) under the assumption that $\sin\alpha \neq 0$. But then we can divide the second equation of (11) termwise by the first to obtain

$$\cot\frac{x+y}{2} = \cot\alpha$$

whence again follows relation (13), but the difference is that previously this relation was obtained for all α except when $\cos\alpha = 0$. Now it turns out that it is valid also for those values of α for which $\cos\alpha = 0$, that is *for all* α.

Furthermore, in (13) expressing y in terms of x and substituting (14) into the first equation of (10), we now find (since $\sin\alpha \neq 0$) that $\cos(x-\alpha) = 1/2$ for arbitrary α. We can thus conclude that formulas (15) constitute the solution of the original system (10).

Second mode. The same idea (obtaining an algebraic relationship between the unknowns and eliminating one of them) can be effected differently. To do this, note that if the first equation of (10) is multiplied by $\cos\alpha$ and then $\sin\alpha$ times the second equation of the system is subtracted from it, we get the relation

$$\sin(x-\alpha) + \sin(y-\alpha) = 0$$

Obviously, if we *simultaneously* have $\cos\alpha \neq 0$ and $\sin\alpha \neq 0$ (these were the quantities we multiplied the equations of the system (10) by!), then in place of (10) we can, from now on, solve the *equivalent* system *

$$\begin{aligned} \sin(x-\alpha) + \sin(y-\alpha) &= 0 \\ \cos x + \cos y &= \cos\alpha \end{aligned} \qquad (16)$$

The first equation of this system rewritten in decomposable form,

$$\sin\left(\frac{x+y}{2} - \alpha\right)\cos\frac{x-y}{2} = 0,$$

* We might with equal justification consider the system

$$\begin{aligned} \sin(x-\alpha) + \sin(y-\alpha) &= 0 \\ \sin x + \sin y &= \sin\alpha \end{aligned}$$

makes it necessary to consider two possibilities. If

$$\sin\left(\frac{x+y}{2}-\alpha\right)=0, \text{ that is, } x+y=2\alpha+2k\pi$$

where k is an arbitrary integer, then we can express y in terms of x [see (14)] and substitute the resulting expression into the second equation of (16). We then get a trigonometric equation in the single unknown x:

$$2\cos\alpha\cos(x-\alpha)=\cos\alpha$$

or, since $\cos\alpha\neq 0$ by hypothesis, it follows that $\cos(x-\alpha)=1/2$, whence we readily find the solution (15) of the original system.

We still have to consider the second possibility, namely $\cos[(x-y)/2]=0$. The simplest way is to rewrite the second equation of the system (16) as

$$2\cos\frac{x+y}{2}\cos\frac{x-y}{2}=\cos\alpha$$

whence it is evident that, since $\cos\alpha\neq 0$, the equation $\cos[(x-y)/2]=0$ is impossible because then system (16) becomes inconsistent.*

It now remains to examine the cases $\sin\alpha=0$ or $\cos\alpha=0$. If $\sin\alpha=0$, then direct substitution of the resulting values (15) into the original system (10) shows that the system is satisfied, that is to say there are no extraneous solutions among the solutions (15). The very same thing goes for the case when $\cos\alpha=0$. Hence, the formulas (15) describe the solutions of system (10) *for all values of* α.

Third mode. The underlying concept here is simple. Using a fundamental trigonometric relation (the sum of the squares of the sine and cosine of any argument is equal to 1) attempt to obtain from (10) an equation involving only the unknown x and another equation involving only y.

To do this, rewrite the original system (10) in the form

$$\sin x=\sin\alpha-\sin y$$
$$\cos x=\cos\alpha-\cos y$$

square each of these equations and combine. Obvious transformations yield the equation $\cos(y-\alpha)=1/2$, whence $y-\alpha=\pm(\pi/3)+2k\pi$, k an integer. If (10) is rewritten so as to isolate $\sin y$ and $\cos y$ in the left members of the equations and carry through the same operations, we get the equation $\cos(x-\alpha)=1/2$, whence $x-\alpha=\pm(\pi/3)+2n\pi$.

---

* If from the relation $\cos[(x-y)/2]=0$ we found a relationship between x and y and then eliminated one of the unknowns by substituting the appropriate expression into the second equation of (16), we would get the relation $\cos\alpha=0$. Since this relation is invalid, the second possibility does not lead to any solutions of the original system.

We thus have *four* groups of values of the unknowns (note here that the relations $y-\alpha = \pm(\pi/3) + 2k\pi$ and $x-\alpha = \pm(\pi/3) + 2n\pi$ are absolutely independent, and so we have to take all possible combinations of signs):

$$x_1=\alpha+\frac{\pi}{3}+2n\pi, \qquad y_1=\alpha+\frac{\pi}{3}+2k\pi,$$

$$x_2=\alpha+\frac{\pi}{3}+2n\pi, \qquad y_2=\alpha-\frac{\pi}{3}+2k\pi,$$

$$x_3=\alpha-\frac{\pi}{3}+2n\pi, \qquad y_3=\alpha+\frac{\pi}{3}+2k\pi,$$

$$x_4=\alpha-\frac{\pi}{3}+2n\pi, \qquad y_4=\alpha-\frac{\pi}{3}+2k\pi$$

where k and n are arbitrary integers.

It would be premature to regard all four groups as solutions of the original system (10) since we squared the equations and this operation might have introduced extraneous solutions. A check is required.

The check only involves certain purely manipulative difficulties. For example, let us check the first group. Substitute the expressions x_1 and y_1 into the equations of system (10). Obvious manipulations yield the equations

$$2\sin\left(\alpha+\frac{\pi}{3}\right)=\sin\alpha \qquad \sqrt{3}\cos\alpha=0$$

or

$$2\cos\left(\alpha+\frac{\pi}{3}\right)=\cos\alpha \qquad \sqrt{3}\sin\alpha=0$$

which cannot be valid simultaneously. Hence, the first group is not a solution of system (10). In the same way we see that the fourth group is not a solution of the original system. Now the second and third groups do satisfy the original system [they coincide with solution (15)].

Fourth mode. This is yet another effective approach to the solution of system (10). It utilizes the peculiarities of the system and gets the result in a very simple manner. This time the underlying idea is to invoke the theory of complex numbers, more precisely, their geometric interpretation.

It may be noted that the system (10) is equivalent to the equation $z_1+z_2=w$ relating the complex numbers

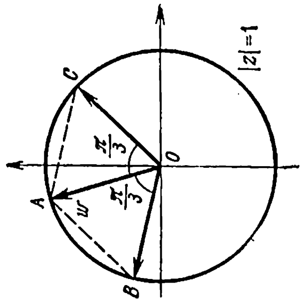
$$z_1=\cos x+i\sin x,\quad z_2=\cos y+i\sin y,\quad w=\cos\alpha+i\sin\alpha$$

where w is known (since its argument α is given), and z_1, z_2 are unknown complex numbers. It is further clear that $|z_1|=|z_2|=|w|=1$.

Thus the problem of solving system (10) reduces to this: for each complex number w with absolute value 1, find two complex numbers z_1 and z_2 with unit moduli whose sum is equal to w. Clearly, it suffices to determine only the arguments of the numbers z_1 and z_2, since their moduli have the specified value of 1.

Let point A represent the known number w; since $|w| = 1$, A will lie on the unit circle with centre at the origin (Fig. 60). Points B and C which depict the sought-for numbers z_1 and z_2 must also be located on this circle. The equation $z_1 + z_2 = w$ geometrically signifies that OA is a diagonal of a parallelogram whose nonparallel sides OB and OC are equal and are equal to this diagonal. But then the triangle OBA must be an equilateral triangle, or $\angle BOA = \angle AOC = \pi/3$.

Fig. 60



From this it is clear that since α is the argument of the given number w, then $\alpha + (\pi/3)$ is the argument of one of the desired numbers and $\alpha - (\pi/3)$ is the argument of the other. It will be recalled that all the arguments of a (nonzero) complex number are obtained from one of them by formula (3) of Sec. 1.5. Therefore all the arguments of one of the unknown numbers are of the form $\alpha + (\pi/3) + 2k\pi$, k any integer, and all the arguments of the other, of the form $\alpha - (\pi/3) + 2n\pi$, n any integer. If we now notice that the numbers z_1 and z_2 are of equal status (we may assume that point B represents the number z_1 and point C the number z_2, or vice versa, that B represents z_2 and C represents z_1), we get two groups of solutions defined by the formulas

$$x_1 = \alpha + \frac{\pi}{3} + 2k\pi, \qquad y_1 = \alpha - \frac{\pi}{3} + 2n\pi,$$

$$x_2 = \alpha - \frac{\pi}{3} + 2n\pi, \qquad y_2 = \alpha + \frac{\pi}{3} + 2k\pi$$

where k, n are arbitrary integers. This answer coincides with the answer (15).

We conclude this section with an example of the solution of a system of trigonometric equations involving a parameter.

7. *Find all the values of a for which the system*

$$\sin x \cos y = a^2$$
$$\sin y \cos x = a$$

has solutions and find these solutions.

Adding and subtracting the equations of the system, we get the system

$$\begin{aligned} \sin(x+y) &= a^2+a \\ \sin(x-y) &= a^2-a \end{aligned} \tag{17}$$

It is clear that this system has solutions if and only if the following two double inequalities are valid simultaneously:

$$\begin{aligned} -1 \leqslant a^2+a \leqslant 1 \\ -1 \leqslant a^2-a \leqslant 1 \end{aligned} \tag{18}$$

The first of these double inequalities can be rewritten as a system of two quadratic inequalities:

$$\begin{aligned} a^2+a+1 \geqslant 0 \\ a^2+a-1 \leqslant 0 \end{aligned}$$

the first being valid for arbitrary values of a, the second, for

$$\frac{-1-\sqrt{5}}{2} \leqslant a \leqslant \frac{-1+\sqrt{5}}{2} \tag{19}$$

It is this interval that serves as the solution of the first double inequality of (18).

The second double inequality of (18) can also be rewritten in the form of a system of two quadratic inequalities:

$$\begin{aligned} a^2-a+1 \geqslant 0 \\ a^2-a-1 \leqslant 0 \end{aligned}$$

the first being valid for arbitrary a, the second, for

$$\frac{1-\sqrt{5}}{2} \leqslant a \leqslant \frac{1+\sqrt{5}}{2} \tag{20}$$

This interval is the solution of the second double inequality of (18).

Consequently, the solution of the system (18) of double inequalities is the common portion (in set theory, it is called the intersection) of the intervals (19) and (20), namely,

$$\frac{1-\sqrt{5}}{2} \leqslant a \leqslant \frac{-1+\sqrt{5}}{2} \tag{21}$$

For these values of the parameter a, the original system of trigonometric equations has certain solutions; for all other values of a, it does not have any solutions.

It is not hard to find solutions of the original system, given condition (21). Indeed, for any value of the parameter a in the interval (21), system (17) can be rewritten as follows:

$$\begin{aligned} x+y &= (-1)^n \arcsin(a^2+a)+n\pi, \quad n=0, \pm1, \pm2, \ldots, \\ x-y &= (-1)^k \arcsin(a^2-a)+k\pi, \quad k=0, \pm1, \pm2, \ldots \end{aligned}$$

whence

$$x=\frac{1}{2}\left[(-1)^n \arcsin (a^2+a)+(-1)^k \arcsin (a^2-a)+(n+k)\pi\right],$$

$$y=\frac{1}{2}\left[(-1)^n \arcsin (a^2+a)-(-1)^k \arcsin (a^2-a)+(n-k)\pi\right]$$

where n and k are arbitrary integers.

Exercises

Solve the following systems of equations.

1. $x-y=\frac{\pi}{18}$

$\sin\left(x+\frac{\pi}{18}\right)\sin\left(y+\frac{\pi}{9}\right)=\frac{1}{2}.$

2. $x+y=\frac{2\pi}{3}$

$\frac{\sin x}{\sin y}=2.$

3. $\tan x+\cot y=3$

$|x-y|=\frac{\pi}{3}.$

4. $\tan x \tan z=3$

$\tan y \tan z=6$

$x+y+z=\pi.$

5. $\sin x+\sec y=2$

$\sin x \sec y=\frac{1}{2}.$

6. $\sin (x-y)=3 \sin x \cos y-1$

$\sin (x+y)=-2 \cos x \sin y.$

7. Find the solutions of the system

$$\frac{1}{2}\sin (1-y+x^2) \cos 2x=\cos (y-1-x^2) \sin x \cos x$$

$$\log_{2^v} \frac{2^{y+2x}}{2^{1+x^2}}=2-x$$

that satisfy the condition $y-1-x^2+2x \geqslant 0$.

8 Find the solutions of the system

$$\sin x=\sin 2y$$

$$\cos x=\sin y$$

that satisfy the conditions $0 \leqslant x \leqslant \pi$, $0 \leqslant y \leqslant \pi$.

9. Find the solutions of the system

$$|\sin x|=a \sin y \qquad (a>0)$$

$$\tan x=2 \tan y$$

that satisfy the conditions $0<x<2\pi$, $0<y<2\pi$.

10. Determine for which values of a and b the system

$$\tan x=a \cot x$$

$$\tan 2x=b \cos y$$

has solutions. Find these solutions.

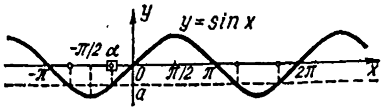
2.5 Inverse trigonometric functions

The student of secondary school does not have much to do with inverse trigonometric functions and, consequently, usually has a hazy idea of just what they are. The theory of these functions appears to be complicated and overloaded with cumbersome formulas that cannot be derived or memorized.

Actually, inverse trigonometric functions are not so complicated. The starting definitions are very simple and all one needs to handle them effectively is a firm grasp of trigonometry. As to the enormous formulas that some textbooks give, there is no need either to remember them or to be able to derive them.

First of all, it is important to clarify the difference between Arcsin a and arcsin a. From the properties of the function $y = \sin x$ it follows that for $-1 \leqslant a \leqslant 1$ there are an infinity of angles x which satisfy the equation $\sin x = a$. This infinite number of angles is symbolically denoted by Arcsin a. Of this infinite set of values, there is one which lies in the interval from $-\pi/2$ to $\pi/2$. This angle is sometimes called the *principal angle* and is denoted symbolically as arcsin a. All this is pictorially represented in Fig. 61, which displays the graph of the

Fig. 61



function $y = \sin x$. Points on the x-axis mark the angles included in Arcsin a; and the point α enclosed in a square between $-\pi/2$ and $\pi/2$, is arcsin a.

Thus, arcsin a *is an angle whose sine is equal to a and which is located between* $-\pi/2$ *and* $\pi/2$. This **definition** may be written compactly and formally as follows:

$$\alpha = \arcsin a \text{ if } (1)\ \sin \alpha = a \text{ and } (2) -\pi/2 \leqslant \alpha \leqslant \alpha/2.$$

A very common and serious mistake of students is this: for example, when they see an equation like $t = \sin \alpha$, they often immediately write $\alpha = \arcsin t$. This is not correct since from the equation $t = \sin \alpha$ it only follows that the angle α is in the set denoted by Arcsin t, but it does not in the least follow that it also satisfies the second condition $-\pi/2 \leqslant \alpha \leqslant \pi/2$.

Another common mistake is the incorrect statement of the second condition. Instead of the proper formulation, we hear things like this: "The angle lies in the first or fourth quadrant." But this phrase, the exact meaning of which is that the terminal side of the angle (radius vector) lies in the first or in the fourth quadrant, expresses something quite different from the second condition of the definition. For example, the radius vector $9\pi/4$ lies in the first quadrant but the inequality $-\pi/2 \leqslant 9\pi/4 \leqslant \pi/2$ is not valid.

Similarly, the definitions of the other inverse trigonometric functions merit careful scrutiny. These definitions are conveniently tabulated below.

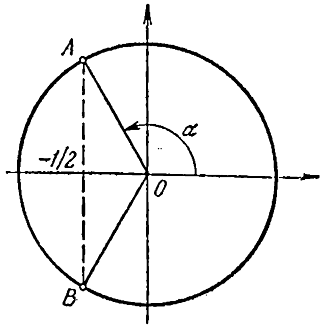
$\alpha = \arcsin a$	$\alpha = \arccos a$	$\alpha = \arctan a$	$\alpha = \operatorname{arccot} a$
(1) $\sin\alpha = a$, (2) $-\pi/2 \leqslant \alpha \leqslant \pi/2$	(1) $\cos\alpha = a$, (2) $0 \leqslant \alpha \leqslant \pi$	(1) $\tan\alpha = a$, (2) $-\pi/2 < \alpha < \pi/2$	(1) $\cot\alpha = a$, (2) $0 < \alpha < \pi$

The definitions and trigonometric formulas given above are quite sufficient to solve a variety of computational problems involving inverse trigonometric functions. We now take up a few typical examples of this kind. Some of them will yield very useful formulas.

1. *Compute* arccos $(-1/2)$.

According to the definition, the angle $\alpha = \arccos(-1/2)$ lies between 0 and π and its cosine is equal to $-1/2$. Consider the trigonometric circle in Fig. 62; let the positions OA and OB of the radius vector

Fig. 62



correspond to angles whose cosine is equal to $-1/2$ [these angles thus belong to Arccos $(-1/2)$]. The curved arrow depicts the angle α which lies between 0 and π; clearly, $\alpha = 2\pi/3$, that is, $\arccos(-1/2) = 2\pi/3$.

2. *Compute* cot [arccos $(-1/3)$].

First of all, do not be frightened by the aspect of expressions like this one. If the definitions are firmly grasped and the trigonometric formulas are at your finger tips, there is nothing difficult about this problem. What does it call for? We have to find the cotangent of the angle $\alpha = \arccos(-1/3)$. By the definition of the arc cosine, we can write that $\cos\alpha = -1/3$ and $0 \leqslant \alpha \leqslant \pi$; since the cosine is negative, the angle α lies in the second quadrant: $\pi/2 < \alpha < \pi$. The problem is thus reduced to the following simple one:

Given that $\pi/2 < \alpha < \pi$ *and* $\cos\alpha = -1/3$, *find* $\cot\alpha$.

This problem is solved by means of the fundamental relations between trigonometric functions. We have $\cot\alpha = \frac{\cos\alpha}{\sin\alpha} = \frac{\cos\alpha}{\sqrt{1-\cos^2\alpha}}$ (the sine in the second quadrant is positive; as usual, the radical sign denotes a nonnegative root), whence we finally get

$$\cot\alpha = \cot\left[\arccos\left(-\frac{1}{3}\right)\right] = -\frac{\sqrt{2}}{4}$$

3. *What is the solutions of* $\cos(\arcsin a)$, $|a| \leqslant 1$?

Let $\alpha = \arcsin a$, then (1) $\sin\alpha = a$ and (2) $-\pi/2 \leqslant \alpha \leqslant \pi/2$. We thus know the sine of the angle α and it is required to find the cosine of that angle. But $\cos^2\alpha = 1 - \sin^2\alpha$, that is, $\cos^2\alpha = 1 - a^2$. How do we find $\cos\alpha$ now? It is clearly sufficient to find the sign of $\cos\alpha$. But the angle α lies between $-\pi/2$ and $\pi/2$, and in this interval the cosine is nonnegative: $\cos\alpha \geqslant 0$, and so $\cos\alpha = \sqrt{1-a^2}$. Thus

$$\cos(\arcsin a) = \sqrt{1-a^2}$$

Similarly, it can be shown that the formula

$$\sin(\arccos a) = \sqrt{1-a^2}$$

holds true.

4. *Find the solution of* $\cos(2\arcsin 2/3)$.

Taking advantage of the formula for the cosine of a double angle, $\cos 2\alpha = 1 - 2\sin^2\alpha$, we get

$$\cos\left(2\arcsin\frac{2}{3}\right) = 1 - 2\sin^2\left(\arcsin\frac{2}{3}\right) = 1 - 2\cdot\left(\frac{2}{3}\right)^2 = \frac{1}{9}$$

Here, we also made use of the formula

$$\sin(\arcsin a) = a$$

which follows from the definition of the arcsine.

Similar formulas are valid for the other trigonometric functions as well:

$$\cos(\arccos a) = a, \quad \tan(\arctan a) = a, \quad \cot(\operatorname{arccot} a) = a$$

A separate group of problems have to do with different notations of an angle, that is, the notation of an angle through the use of different trigonometric functions. For instance, to each angle, say $\pi/6$, there corresponds a definite value of the sine: $\sin\frac{\pi}{6} = 1/2$, of the cosine: $\cos\frac{\pi}{6} = \sqrt{3}/2$, etc. Hence, this angle may be written in two distinct forms: as $\arcsin 1/2$ and as $\arccos\sqrt{3}/2$.

Another example of the same type is afforded by Problem 2 above. From the answer to this problem it is evident that the angle $\alpha = \arccos(-1/3)$ may be written in the form $\alpha = \operatorname{arccot}(-\sqrt{2}/4)$ since $\cot\alpha = -\sqrt{2}/4$ and $0 < \alpha < \pi$. Actually, it was demonstrated in this problem that both notations designate the same angle.

This aspect—establishing the fact that distinct notations designate one and the same angle—is of great importance in the solution of problems in geometry. Suppose, in a problem in which it is required to determine an unknown angle, we got the answer $\alpha = \arccos(-1/3)$. But the same problem can be solved differently, and the answer may, generally speaking, be quite different; say, $\alpha = \operatorname{arccot}(-\sqrt{2}/4)$. What this all amounts to is that different notational forms should not bother the student.

5. *Prove that*

$$\frac{1}{2}\arccos\frac{3}{5} = \arctan\frac{1}{2} = \frac{\pi}{4} - \frac{1}{2}\arccos\frac{4}{5}$$

Let us double each of the three expressions and then attempt to prove that

$$\arccos\frac{3}{5} = 2\arctan\frac{1}{2} = \frac{\pi}{2} - \arccos\frac{4}{5}$$

Note first of all that the three angles lie between 0 and $\pi/2$. For the first angle this is obvious since if $\alpha = \arccos 3/5$, then $0 \leqslant \alpha \leqslant \pi$ and $\cos\alpha = 3/5$ is positive, whence what is required follows. The same goes for the angle $\pi/2 - \arccos 4/5$. Finally, if $\beta = \arctan 1/2$ this signifies that $-\pi/2 < \beta < \pi/2$ and $\tan\beta = 1/2$. Hence, the angle β lies between 0 and $\pi/4$ and so the angle $2\arctan 1/2$ lies between 0 and $\pi/2$.

Since all three angles lie between 0 and $\pi/2$, to prove their equality it suffices to demonstrate that some trigonometric function, say the cosine of each of the angles, has one and the same value. Indeed, we easily find

$$\cos\left(\arccos\frac{3}{5}\right) = \frac{3}{5},$$

$$\cos\left(2\arctan\frac{1}{2}\right) = \frac{1 - \tan^2\left(\arctan\frac{1}{2}\right)}{1 + \tan^2\left(\arctan\frac{1}{2}\right)} = \frac{3}{5},$$

$$\cos\left(\frac{\pi}{2} - \arccos\frac{4}{5}\right) = \sin\left(\arccos\frac{4}{5}\right) = \sqrt{1 - \cos^2\left(\arccos\frac{4}{5}\right)} = \frac{3}{5}$$

(Note that $\arccos 4/5$ is an angle in the first quadrant and so we choose the plus sign in front of the radical!). The proof is complete.

Unfortunately, many mistakes are made in the solution of problems of this kind. The basic and worst mistake of this type is the following. In establishing the equality of the angles, a check is made to see that some trigonometric function of each of the angles has the same value; on this basis the conclusion is drawn that the angles are equal. But such a conclusion is totally unjustified. For example, the angles $\arcsin 1/2$ and $\arcsin(-1/2)$ are clearly equal, but

$$\cos\left(\arcsin\frac{1}{2}\right) = \cos\left[\arcsin\left(-\frac{1}{2}\right)\right] = \frac{\sqrt{3}}{2}$$

The whole point is that *it does not follow from the equality of the cosines (sines, etc.) of two angles that the angles are equal*. However, if we are able to demonstrate that the angles lie, say, between 0 and $\pi/2$ (as in the foregoing problem) then the equality of trigonometric functions implies the equality of the angles.

6. *Find the angle* $\arcsin 1/3 + \arcsin 3/4$.

Let $\alpha = \arcsin 1/3$, $\beta = \arcsin 3/4$, $\gamma = \alpha + \beta$. To find the angle γ it is first necessary to find one of its trigonometric functions. It would seem to be natural to compute $\sin\gamma$:

$$\sin\gamma = \sin(\alpha+\beta) = \sin\alpha\cos\beta + \sin\beta\cos\alpha$$

$$= \frac{1}{3}\sqrt{1-\frac{9}{16}} + \frac{3}{4}\sqrt{1-\frac{1}{9}} = \frac{\sqrt{7}+6\sqrt{2}}{12}$$

But how do we find the angle γ itself? As we already know, we cannot write $\gamma = \arcsin\frac{\sqrt{7}+6\sqrt{2}}{12}$ without having first found out whether γ lies in the interval from $-\pi/2$ to $\pi/2$. It is not so easy to determine this in the given case. Therefore let us try to solve this in a different way by first analyzing the condition. We know that the angles α and β lie in the interval from $-\pi/2$ to $\pi/2$, but their sines are positive so we can state more exactly that they lie between 0 and $\pi/2$, that is, $0 < \alpha < \pi/2$, $0 < \beta < \pi/2$. Adding these inequalities we get

$$0 < \gamma < \pi$$

It is now clear that the angle γ can be determined if we know the cosine of it:

$$\cos\gamma = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

$$= \sqrt{1-\frac{1}{9}}\sqrt{1-\frac{9}{16}} - \frac{1}{3}\cdot\frac{3}{4} = \frac{2\sqrt{14}-3}{12}$$

To summarize: (1) $\cos\gamma = 2\frac{2\sqrt{14}-3}{12}$, (2) $0 < \gamma < \pi$.*
Consequently

$$\gamma = \arccos\frac{2\sqrt{14}-3}{12}$$

7. *Find the angle* $2\arctan(-3)$.

The angle $\alpha = \arctan(-3)$ obviously satisfies the inequality $-\pi/2 < \alpha < 0$, whence $-\pi < 2\alpha < 0$. Now what function can be used to determine 2α? There are different ways of approaching this.

First way. Multiply the last inequality by -1 to get $\pi > -2\alpha > 0$. Then

$$\cos(-2\alpha) = \cos 2\alpha = \frac{1-\tan^2\alpha}{1+\tan^2\alpha} = \frac{1-9}{1+9} = -\frac{4}{5}$$

---

* It is clear that if $0 < \gamma < \pi$, then surely the inequality $0 \leqslant \gamma \leqslant \pi$ is valid (see Sec. 1.1).

Thus, (1) $\cos(-2\alpha) = -4/5$ and (2) $0 < -2\alpha < \pi$. Consequently, $-2\alpha = \arccos(-4/5)$, or $2\alpha = -\arccos(-4/5)$.

Second way. From the inequality $-\pi < 2\alpha < 0$ it follows that $0 < 2\alpha + \pi < \pi$. Furthermore, $\cos(2\alpha + \pi) = -\cos 2\alpha = 4/5$. But if these two conditions hold, then $2\alpha + \pi = \arccos 4/5$, that is, $2\alpha = -\pi + \arccos 4/5$.

Third way. From the inequality $-\pi < 2\alpha < 0$ it follows that $-\pi/2 < 2\alpha + \pi/2 < \pi/2$. Furthermore,

$$\tan\left(2\alpha + \frac{\pi}{2}\right) = -\cot 2\alpha = -\frac{1}{\tan 2\alpha} = -\frac{1 - \tan^2\alpha}{2\tan\alpha} = -\frac{4}{3}$$

And so $2\alpha + \pi/2 = \arctan(-4/3)$, whence we get yet another answer: $2\alpha = -\pi/2 + \operatorname{arccot}(-4/3)$.

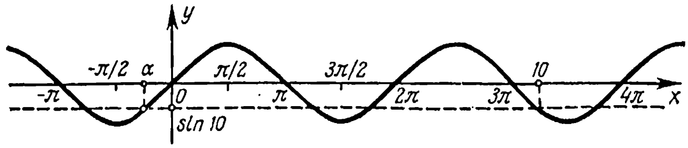
Three modes of solution have thus yielded three different answers. But all these answers only appear to be different, actually they are merely different forms of notation of *one and the same angle*, and this can be shown in the same way as was done in Problem 5.

The next problem is of a type that often causes considerable difficulty, although all that is needed to obtain a solution is a good knowledge of definitions.

8. *Find the angle* arcsin (sin 10).

By definition, $\alpha = \arcsin(\sin 10)$ is an angle that satisfies two conditions: $\sin\alpha = \sin 10$ and $-\pi/2 \leqslant \alpha \leqslant \pi/2$. The easiest way to determine this angle is by means of the graph of the function $y = \sin x$ (Fig. 63). Plot the number 10 on the x-axis and find sin 10 geometri-

Fig. 63



cally (this is the ordinate y of the point on the graph corresponding to $x = 10$) and then draw the horizontal line $y = \sin 10$. The abscissa of one of the points of intersection of this straight line with the graph lies in the interval from $-\pi/2$ to $\pi/2$. This abscissa is the desired angle, since by the construction it lies between $-\pi/2$ and $\pi/2$ and its sine is equal to sin 10. It can be computed by simple geometric reasoning. It is easy to see that the points α and 10 are symmetric about the point $3\pi/2$ so that $10 - 3\pi/2 = 3\pi/2 - \alpha$, whence $\alpha = 3\pi - 10$.

We now pass from these numerical examples to *inverse trigonometric functions* as such.

The expression $\arcsin\alpha$ is defined for numbers α that satisfy the condition $-1 \leqslant \alpha \leqslant 1$. We can consider the function

$$y = \arcsin x$$

which associates with every number a a number y equal to the radian measure of the angle arcsin a. This number is also denoted arcsin a. The domain of this function (that is, the set of values of x for which it is meaningful) is the set of numbers in the interval from —1 to 1.

Fig. 64

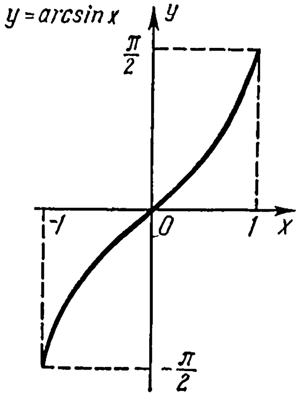
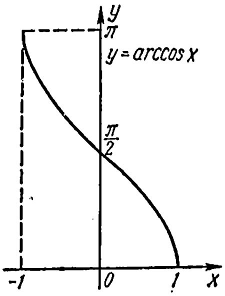
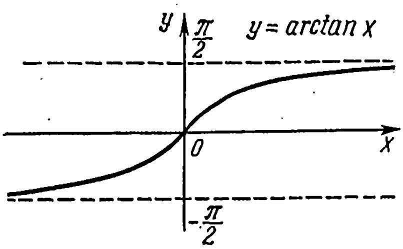


Fig. 65



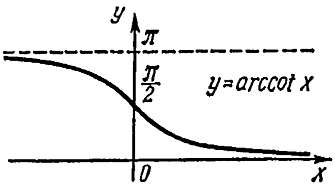
The other inverse trigonometric functions are defined in similar fashion. Their domains are: for the function arccos x, the set of numbers in the interval from —1 to 1, for the function arctan x and arccot x, the set of all real numbers.

Fig. 66



The graphs of the inverse trigonometric functions are shown in Figs. 64-67. Naturally, before they can be constructed the necessary properties of these functions have to be thoroughly investigated. For one thing, it must be proved that arcsin x and arctan x are increasing functions and that arccos x and arccot x are decreasing functions.

Fig. 67



These proofs are not very difficult, but we do not give them here because the graphs of inverse trigonometric functions are outside the scope of the secondary-school curriculum. We merely state them here for the sake of completeness. The interested student can carry out the proofs by himself. On the one hand, this is a useful exercise, and on the other, employing these properties frequently simplifies problem solving.

Let us consider some problems whose purpose is the proof of separate properties of inverse trigonometric functions.

9. *Prove the identity*

$$\arcsin(-x) = -\arcsin x, \quad -1 \leqslant x \leqslant 1$$

Rewriting this equation as $-\arcsin(-x) = \arcsin x$ and denoting $\alpha = \arcsin(-x)$, we get, by definition, (1) $\sin\alpha = -x$ and (2) $-\frac{\pi}{2} \leqslant \alpha \leqslant \frac{\pi}{2}$. However these conditions imply at once that

$$\sin(-\alpha) = -\sin\alpha = x \text{ and } \frac{\pi}{2} \geqslant -\alpha \geqslant -\frac{\pi}{2}$$

but this signifies that $-\alpha = \arcsin x$, which completes the proof.

The identity that has just been proved may be stated as follows: $y = \arcsin x$ is an odd function.

Note that we have **no** grounds to assert that $y = \arcsin x$ is an odd function **for the reason** that $y = \sin x$ is odd. Reasoning in that fashion we could then assert that the function $y = \cos x$ is even and so $y = \arccos x$ is also even. Yet this is obviously not true since, for example, $\arccos 1 = 0$ and $\arccos(-1) = \pi \neq 0$. The relationship between $\arccos x$ and $\arccos(-x)$ is more complicated.

10. *Prove the identity*

$$\arccos(-x) = \pi - \arccos x, \quad -1 \leqslant x \leqslant 1$$

The proof is analogous to the preceding case. Let $\alpha = \arccos(-x)$. This means that

$$(1)\ \cos\alpha = -x \text{ and } (2)\ 0 \leqslant \alpha \leqslant \pi$$

but then $\cos(\pi - \alpha) = -\cos\alpha = x$ and $0 \geqslant -\alpha \geqslant -\pi$, that is, $\pi \geqslant \pi - \alpha \geqslant 0$ so that $\pi - \alpha = \arccos x$ or $\pi - \arccos(-x) = \arccos x$, which is what we set out to prove.

Identities 9 and 10 permit us to simplify certain expressions. Also note that both these identities can be illustrated effectively on the trigonometric circle.

11. *Prove the identity*

$$\arcsin x + \arccos x = \frac{\pi}{2}, \quad -1 \leqslant x \leqslant 1$$

First method. The inequalities $-\pi/2 \leqslant \arcsin x \leqslant \pi/2$ and $0 \leqslant \arccos x \leqslant \pi$ imply that

$$-\frac{\pi}{2} \leqslant \arcsin x + \arccos x \leqslant \frac{3\pi}{2}$$

Besides (see Problem 3),

$$\sin(\arcsin x + \arccos x) = x^2 + \sqrt{1-x^2}\sqrt{1-x^2} = 1.$$

But in the interval between $-\pi/2$ and $3\pi/2$ there is only one angle whose sine is equal to 1, namely $\pi/2$, so that $\arcsin x + \arccos x = \pi/2$, and the proof is complete.

Second method. The given identity is equivalent to

$$\arcsin x = \frac{\pi}{2} - \arccos x$$

But $\sin\left(\frac{\pi}{2} - \arccos x\right) = \cos(\arccos x) = x$ and, besides,

$$-\frac{\pi}{2} \leqslant \frac{\pi}{2} - \arccos x \leqslant \frac{\pi}{2}$$

(this readily follows from the inequality $0 \leqslant \arccos x \leqslant \pi$). Therefore, $\pi/2 - \arccos x = \arcsin x$, which completes the proof.

This identity is well illustrated on the trigonometric circle (do not forget to consider two cases: $x \geqslant 0$ and $x < 0$).

Let us solve a few equations that involve inverse trigonometric functions.

12. *Solve the equation* $\arcsin x = \pi$.

This equation clearly does not have any solutions since $\arcsin x$ cannot exceed $\pi/2$.

Many students solve this equation as follows: "We take sines of both members: $\sin(\arcsin x) = \sin \pi$, that is, $x = \sin \pi = 0$. The answer is then $x = 0$." Where is the flaw in this chain of reasoning? The point is that, in the general case, the equations $f(x) = g(x)$ and $\sin f(x) = \sin g(x)$ are not at all equivalent. The second equation is a consequence of the first but, naturally, can have extraneous roots, which is what happened in the case at hand.

This should always be kept in mind because, when solving equations involving inverse trigonometric functions, one often has to take direct trigonometric functions of both members. Don't forget to check.* A check of this erroneous solution would have shown that $\arcsin 0 = 0 \neq \pi$, that is, the root $x = 0$ is extraneous.

13. *Solve the equation* $\arccos x\sqrt{3} + \arccos x = \pi/2$.

Rewrite the equation in the form

$$\arccos x\sqrt{3} = \frac{\pi}{2} - \arccos x$$

and take cosines of both members:

$$\cos(\arccos x\sqrt{3}) = \cos\left(\frac{\pi}{2} - \arccos x\right)$$

or $x\sqrt{3} = \sqrt{1-x^2}$. Now square both members of this equation (this may give rise to extraneous roots, but we already have to make

* Also remember that since the functions $\tan x$ and $\cot x$ are meaningless for certain values of x, roots can also be lost when taking these functions of both sides. In short, handle them with care.

a check since we took cosines of both members!): $3x^2 = 1 - x^2$, whence $4x^2 = 1$, or $x_{1,2} = \pm 1/2$.

Check. For $x = 1/2$, we have

$$\arccos\frac{\sqrt{3}}{2} + \arccos\frac{1}{2} = \frac{\pi}{6} + \frac{\pi}{3} = \frac{\pi}{2}$$

and, hence, $x_1 = 1/2$ is a root of the given equation. For $x = -1/2$ we have

$$\arccos\left(-\frac{\sqrt{3}}{2}\right) + \arccos\left(-\frac{1}{2}\right) = \frac{5\pi}{6} + \frac{2\pi}{3} = \frac{3\pi}{2}$$

which means $x = -1/2$ is an extraneous root.

14. *Solve the equation* $\arcsin\frac{3x}{5} + \arcsin\frac{4x}{5} = \arcsin x$.

Take sines of both members to get

$$\frac{3x}{5}\sqrt{1-\frac{16x^2}{25}} + \frac{4x}{5}\sqrt{1-\frac{9x^2}{25}} = x$$

or

$$x\left(3\sqrt{25-16x^2} + 4\sqrt{25-9x^2}\right) = 25x$$

One root is obvious: $x_1 = 0$. It remains to solve the equation

$$3\sqrt{25-16x^2} + 4\sqrt{25-9x^2} = 25$$

For the sake of simplicity, set $x^2 = y$ and solve the resulting irrational equation by the usual technique of squaring, first isolating one of the radicals,* to get $y = 1$.

Thus, $x^2 = 1$, or $x_{2,3} = \pm 1$. Now check. Since $\arcsin 0 + \arcsin 0 = \arcsin 0$, then $x_1 = 0$ is a root. It is more difficult to verify the second root. It reduces to proving or disproving the equation

$$\arcsin\frac{3}{5} + \arcsin\frac{4}{5} = \frac{\pi}{2}$$

Let us prove this equation. The angle $\arcsin 4/5$ is a positive acute angle and its cosine is equal to $\sqrt{1-\frac{16}{25}} = \frac{3}{5}$; therefore $\arcsin 4/5 = \arccos 3/5$. But then, on the basis of the identity of Problem 11,

$$\arcsin\frac{3}{5} + \arcsin\frac{4}{5} = \arcsin\frac{3}{5} + \arccos\frac{3}{5} = \frac{\pi}{2} = \arcsin 1$$

so that $x_2 = 1$ is a root of the original equation. Now, by Problem 9, we have the equations

$$\arcsin\left(-\frac{3}{5}\right) + \arcsin\left(-\frac{4}{5}\right) = -\arcsin\frac{3}{5} - \arcsin\frac{4}{5} = -\frac{\pi}{2} = \arcsin(-1)$$

Thus, $x_3 = -1$ is also a root of the original equation.

* This irrational equation is solved very simply if we note that $y = 1$ is a root and the left member is a decreasing function.

15. *Solve the equation* $\arcsin(1-x) - 2\arcsin x = \frac{\pi}{2}$.

First method. Rewrite the equation as $\arcsin(1-x) = \pi/2 + 2\arcsin x$ and take sines of both members to get, after obvious manipulations, the equation $x = 2x^2$, or $x_1 = 0$, $x_2 = 1/2$

A check shows that $x_2 = 1/2$ is an extraneous root.

Second method. The domain of the variable of our equation is defined by the inequalities

$$-1 \leqslant 1-x \leqslant 1, \quad -1 \leqslant x \leqslant 1$$

Solving these two inequalities, we get $0 \leqslant x \leqslant 1$. But if $x > 0$, then $2\arcsin x > 0$ and $1-x < 1$, and so $\arcsin(1-x) < \pi/2$. Hence if $x > 0$, then

$$\arcsin(1-x) - 2\arcsin x < \frac{\pi}{2}$$

so that $x > 0$ cannot be a root. It remains to check the value $x = 0$, which turns out to be a root.

Problems of this kind could be continued, but it is clear that there is nothing particularly difficult in the theory, and a good knowledge of trigonometry and of the basic definitions of inverse trigonometric functions is quite sufficient for the solution of many problems.

Exercises

Evaluate the following angles.

1. $\arcsin\left(-\frac{\sqrt{3}}{2}\right)$, $\arcsin 1$, $\arcsin(-1)$.

2. $\arccos\left(-\frac{\sqrt{3}}{2}\right)$, $\arccos(-1)$, $\arccos 0$.

3. $\arctan\left(-\frac{1}{\sqrt{3}}\right)$, $\arctan(-1)$.

4. $\operatorname{arccot}\left(-\frac{1}{\sqrt{3}}\right)$, $\operatorname{arccot}(-1)$, $\operatorname{arccot} 0$.

5. $\arcsin(\sin 5)$, $\arccos(\cos 10)$, $\arctan[\tan(-6)]$, $\operatorname{arccot}[\cot(-10)]$.

6. $\arccos\frac{4}{5} - \arccos\frac{1}{4}$, $\arctan\frac{1}{3} - \arctan\frac{1}{4}$, $\operatorname{arccot}(-2) - \arctan\left(-\frac{2}{3}\right)$.

Prove the following formulas.

7. $\tan(\arcsin a) = \frac{a}{\sqrt{1-a^2}}$, $-1 < a < 1$.

8. $\sin(\operatorname{arccot} a) = \frac{1}{\sqrt{1+a^2}}$.

9. $\cot(\arcsin a) = \frac{\sqrt{1-a^2}}{a}$, $-1 \leqslant a \leqslant 1$, $a \neq 0$.

Prove the identities:

10. $\arctan(-x) = -\arctan x$.

11. $\operatorname{arccot}(-x)=\pi-\operatorname{arccot} x$.

12. $\arctan x+\operatorname{arccot} x=\frac{\pi}{2}$.

State for which values of x the following equations are true:

13. $\arcsin x=\arccos \sqrt{1-x^2}$. **14.** $\operatorname{arccot} x=\arctan \frac{1}{x}$.

Solve the equations:

15. $\sin\left(\frac{1}{5}\arccos x\right)=1$.

16. $\arcsin \frac{1}{\sqrt{x}}-\arcsin \sqrt{1-x}=\frac{\pi}{2}$.

17. $\operatorname{arccot} x=\arccos x$.

18. $\arcsin x-\arccos x=\arccos\frac{\sqrt{3}}{2}$.

Chapter 3 GEOMETRY

3.1 General remarks on geometry

There seems to be a general dislike of geometry on the part of the student due, it would appear, to the fact that the student frequently regards geometry as a haphazard collection of definitions and theorems that have to be memorized. The result is that learning by rote is the practice, even down to the simplest designations. Yet geometric concepts and facts, if examined closely, are very logical, pictorial and natural.

Some students have the idea that the statements of geometry have to be memorized word for word as given in the standard textbook. This is not true at all. Any *statement* (formulation) will do as long as it is *clear-cut*, *exact* and *correct*.

It is a well-known fact that many theorems may be proved in a variety of ways. All proofs are admissible as long as they are *correct*. Choose the one that is easiest to understand and carry through. Bear in mind that it is not permissible to use a theorem which itself depends on the theorem to be proved. The result is a vicious circle that we must avoid under all circumstances.

Note in particular that the proofs of geometry must be exhaustive. All auxiliary theorems (lemmas) referred to in the proof of a theorem must be explicitly stated and, if necessary, proved.

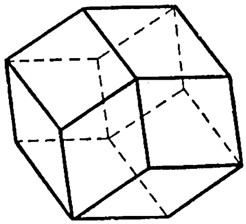
Some students have the habit of skipping a stage in a proof, stating that "it is obvious" or "quite clear", etc. The student should be able to explain why this is so. It is therefore advisable, in going through proofs, do strive towards a *complete* explanation of *every* assertion, *every* step in one's reasoning. In preparing for examinations, the student would do well to pose the question: Why? From what does that follow? Do not accept a single step of any proof on faith, do not leave any stage of a proof unclarified.

In the study of geometry one should never forget that even the simplest notions (except of course such primary ones as point, line, plane) have *definitions*. Be ready to answer questions like "What is a right angle?" "Why is it possible to draw a plane through two parallel lines in space?" and the like. It is well worth the effort to pay particular attention to definitions and not substitute geometric images for them. For instance, we all readily recognize a circle, prism, pyramid, cone, sphere, etc. but not everyone can give a proper definition for each.

To illustrate, let us examine the *prism*. We all have an intuitive idea of what sort of an object a prism is, but students often give faulty definitions of prism.

Some define the prism to be a polyhedron with two faces that are congruent polygons with respectively parallel sides and all other faces parallelograms. Firstly, one often loses sight of the fact that this only applies to *convex* polyhedrons. Secondly, even the convex polyhedron shown in Fig. 68 (it is called a *rhombic dodecahedron*; all faces are equal

Fig. 68



rhombuses, the planes of opposite faces are parallel, the edges of opposite faces are correspondingly parallel) fits this definition. It is quite obvious that this polyhedron does not fit the geometric conception of the prism, to say nothing of the fact that formulas for the surface area and volume that are proved in textbooks for the prism are invalid here.

The point is that in proving these formulas we assume that the prism has the form as we picture it geometrically, and so the definition should be made to fit this picture.

This can be done in different ways. For example, *a prism is a convex polyhedron two faces of which are equal convex polygons with correspondingly parallel sides, while the edges joining corresponding vertices of the polygons are equal and parallel.**

A different definition may be given by employing the concept of a cylindrical surface: *the prism is a solid bounded by a cylindrical surface, whose directrix is a convex polygon, and two parallel cutting planes not parallel to the generatrix.*

Geometry, like all other divisions of mathematics, requires a certain definite level of logical culture, so to speak. The ability on the part of the student to get a clear idea of what is given and what is required in a proof, to be able to state clearly and succinctly each mathe-

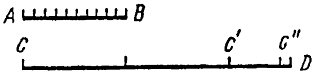
* Actually, this definition is somewhat redundant.

matical idea. This is what the student should aim for in his study of geometry.

"Commensurable and incommensurable segments" is a topic that causes a good deal of trouble since it is closely tied in with the new and logically difficult algebraic notion of irrational number.

The most common mistake is that made in the method of establishing whether two given segments are commensurable or not. Some students argue in this fashion: "Given two line segments AB and CD (Fig. 69). Take the smaller one (AB) and lay it off on the larger one

Fig. 69



(CD), starting from some point C, until the larger segment is completely exhausted, or a remainder (segment $C'D$) is left over, which is less than the segment AB. In the former case the segments AB and CD are commensurable. In the latter instance, we divide AB into 10 equal parts and lay off each tenth of AB on $C'D$ starting from point C' until it fits exactly or there is a new remainder—segment $C''D$, which is less than a tenth of AB. In the former case, AB and CD are commensurable, in the latter, they are not and we divide a tenth of AB into 10 equal parts and lay off the hundredth part of AB on $C''D$, beginning from C'', and so on. If the procedure just described comes to a halt at some stage, that is, if some part of AB fits the appropriate "segment" an integral number of times, then AB and CD are commensurable. If the process does not end, then the segments are incommensurable."

The mistake here is easy to see. Indeed, can we decide whether line segments of length 3 and 4 units are commensurable or not? First the smaller line segment fits the larger one once and has a remainder of length 1; then 1/10 of segment 3 fits the remainder 3 times and has a new remainder of length 0.1; 1/100th part of a segment of length 3 fits the remainder 3 times and has a new remainder of 0.01, etc. This process is clearly *without end*, and so, by the "rule", the segments of length 3 and 4 units are incommensurable. On the other hand, it is evident that these segments are commensurable after all, since their common measure is a line segment of length 1.

This "paradox" is resolved in a very simple manner: AB and CD are incommensurable not when the procedure described above continues indefinitely but when it leads to a *nonterminating nonperiodic (nonrepeating) fraction*. But if the process continues indefinitely but leads to a *terminating periodic (repeating) fraction*, then AB and CD *are commensurable*, just as in the case when the process comes to a halt at some *finite* stage. In the case of line segments of length 3 and 4 units, we get a nonterminating periodic fraction 1.(3), which means the segments are indeed commensurable.

Before reviewing questions connected with the length of the circumference of a circle, the area of a circle, and the surface areas and volumes of circular solids, the student should carefully examine the concept of the *limit of a sequence.*

For example, *the circumference of a circle is defined as the limit of the sequence of perimeters of regular inscribed polygons when the number of sides is increased without bound.*

Now the formula $C = 2\pi R$, which gives the numerical value of the circumference, is a *theorem* which is proved on the basis of the definition. This also holds true for the area of a circle, the surface area of a cone, and so on.

We bring this to the attention of the student because there is occasionally some confusion between the definition of the circumference of a circle and its evalution.*

A few words are in order concerning the proof of the formula for the circumference of a circle. Many students think that the formula $C = 2\pi R$ follows from the duplication formula. Actually, by proceeding from the definition of the circumference we prove that for any circles the ratio of the circumferences is equal to the ratio of the radii, whence it follows that for *all* circles the ratio of the circumference to the diameter is one and the same number. This number is π. Thus, the *equation* $\pi = C/2R$ or, what is the same thing, $C = 2\pi R$, *holds true by the definition of the number* π. Hence, the duplication formula has no connection here. It can merely serve as an approximate computation of π.

When proving geometric theorems and formulas, it is useful to make extensive use of trigonometry and algebraic methods. The trigonometric form of many geometric statements (the cosine law, the formula $S = 0.5ab \sin C$ for the area of a triangle, the sine law) simplifies the proofs and is more convenient in problem solving. Trigonometric functions permit stating many geometric facts more simply and with sufficient generality, which it is not always possible to do in purely geometric terms.

For instance, using trigonometric functions it is very easy to express the *side of a regular n-gon* in terms of the radius r of an inscribed circle or the radius R of a circumscribed circle. Since the central angle subtended by one side of a regular n-gon is equal to $2\pi/n$ radians, it is clear that the side is

$$a_n = 2r \tan \frac{\pi}{n} \tag{1}$$

Similarly, we are convinced that

$$a_n = 2R \sin \frac{\pi}{n} \tag{2}$$

* Logically, the situation here is similar to that which we encountered in defining and evaluating the sum of an infinitely decreasing geometric progression (see Sec. 1.7).

For $n = 3, 4, 6$, we get the familiar formulas for the sides of an equilateral triangle, a square and a regular hexagon.

A very useful relation is

$$R = \frac{a}{2 \sin A} \tag{3}$$

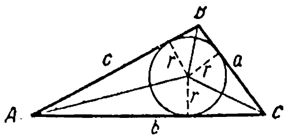
which expresses the *radius R of a circle circumscribed about a triangle in terms of a side and the opposite angle* (this relation is used in proving the sine law). A curious corollary is: to compute the radius of a circumscribed circle it suffices to know only one of the sides of the triangle and the opposite angle, yet these data do not fully define the triangle!

Another frequently used relation in problem solving is

$$S = pr \tag{4}$$

which relates the *area S of a triangle with the semiperimeter p and the radius r of an inscribed circle.* This relation is obtained directly from the obvious (Fig. 70) equation $0.5\,ar + 0.5\,br + 0.5\,cr = S$. It is easy to see that this formula holds true for any polygon with inscribed circle (where S is the area and p the semiperimeter of the polygon).

Fig. 70



It is interesting to note that an analogous formula may be written for solid geometry as well. If, for example, a sphere of radius r is inscribed in a pyramid, then the volume V may be computed from the formula

$$V = Sr/3 \tag{5}$$

where S is the total area of the pyramid. The proof of this formula is carried out in the same way as for the plane case. The centre of the sphere is joined to all vertices and then the pyramid can be conceived of as partitioned into several smaller pyramids. Then, noting that the radius of the inscribed sphere is the altitude of each of the subpyramids, we compute the volume of the pyramid as the sum of the volumes of the subpyramids.

The part played by drawings in geometry is regarded by students differently. Some students look upon the drawing as something superfluous; they give a rough sketch and argue without any reference to the drawing. Other students, on the contrary, take great pains with a drawing, consider it the decisive element in a solution, but do not even find it necessary to justify what, as they put it, is "obvious from the drawing".

Both of these extreme views are absurd. Naturally, even the neatest and most accurately drawn diagram cannot take the place of the logical proof of a geometric fact, because the drawing is merely *illustrative* and is intended to support the reasoning of the student (see Sec. 3.5 for more on this point). It is therefore necessary to justify every geometric fact that we "see" in a drawing. Only then can we be sure that this fact is true and is not merely the result of a correct (perhaps even incorrect) drawing.

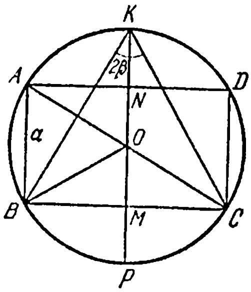
However, the role of the drawing is not confined to illustrating the student's reasoning in problem solving. It often happens that a properly constructed drawing helps to suggest an approach or a theorem or the necessity to make additional constructions. In most problems, a drawing plays a very important part, which ofttimes amounts to hinting at the idea of a solution. It is therefore well to put a good deal of effort into making drawings accurately and to learn to see geometric facts in them that may be extremely suggestive (see Sec. 3.6 on that same subject).

Sometimes a property extracted from the drawing permits reducing the solution of a problem to a couple of lines. The following problem, which has many solutions, is a good illustration of this point.

1. *A rectangle $ABCD$, side AB of which is equal to a, is inscribed in a circle. Draw diameter KP parallel to side AB, then side BC subtends at K an angle of 2β. Find the radius of the circle.*

Let O be the centre of the circle, $KP \parallel AB$, $\angle BKC = 2\beta$ and $AB = a$ (Fig. 71). It is required to find the radius R of the circle

Fig. 71



Here is a solution based on a superficial geometric consideration. Let M be the point of intersection of diameter KP and side BC. Since $R = KM - OM$, we compute KM and OM. From the right triangle KMB we find that $KM = BM \cot \beta$. Drawing the line OB and noting that BOK is an isosceles triangle, we can conclude that $\angle BOM = 2\beta$ (as an exterior angle of this triangle), and so from the right triangle BOM it follows that $OM = BM \cot 2\beta$. Thus, $R = BM/\sin 2\beta$.

Drawing the diagonal AC of rectangle $ABCD$, we get, by the Pythagorean theorem applied to the right triangle ABC, $(2R)^2 = a^2 + (2BM)^2$. Eliminating BM from this and the preceding relation, we get the desired radius of the circle: $R = a/(2|\cos 2\beta|)$.

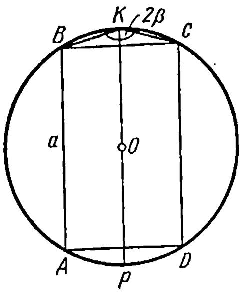
However, the shortest solution to this problem is this. Start by drawing the diagonal AC; it is then clear that the angles BKC and BAC are inscribed and intercept (cut off) the same arc BPC, and so the right triangle ABC yields the answer at once.

Of course, both solutions are quite admissible. In general, any correct and mathematically sensible solution is legitimate, irrespective of its length. At examinations it is best not to spend too much time looking for an elegant solution and to stick to one reliable solution and carry it through to the end, even though it may be rather long. True an elegant solution of only a few lines is undoubtedly valued highly as an indication not only of knowledge but of keen geometric intuition and observation on the part of the student.

Returning again to Fig. 71 which we used in solving this problem, an important point may be noted: *the statement of the problem does not determine the drawing in unambiguous fashion.* In Fig. 71 we took a rectangle $ABCD$ with side $AB = a$ the *smaller* side; again, at our own discretion, we chose the case when the angle BKC is *acute.* Actually, in carrying out the solution in accord with Fig. 71 we inserted *supplementary assumptions* that were absent in the statement of the problem.

Such indeterminateness in the condition of a problem is rather frequently encountered and it is left to the student to supply the necessary additional restrictions (which are ordinarily not indicated explicitly) when representing the geometric configuration in a drawing.

Fig. 72



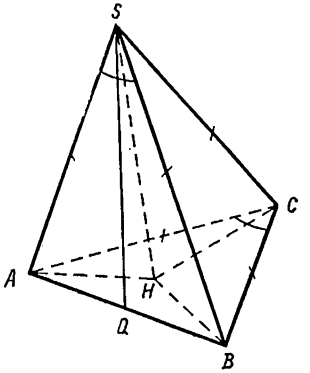
Strictly speaking, we should have considered *all possible* drawings and convinced ourselves that the choice of drawing does not in any way affect the answer. For instance, in Problem 1 we should have considered Fig. 72 (besides Fig. 71) and convinced ourselves that although the reasoning is somewhat altered, the final result is the same. This is not usually done because in a properly posed geometric problem the answer must be the same for all *possible* drawings and so it suffices to solve it for one of them.

The situation is more complicated when the supplmentary assumptions we make in the process of producing the drawing *are not compatible* with the statement of the problem. For example, in the next problem even a developed imagination is not enough to give us an idea of the proper configuration in space, and only in the course of a rigorous solution are we able to see the actual arrangement of the solids.

2. *Two faces of a triangular pyramid are equilateral triangles with side a, the other two faces are isosceles right triangles. Find the radius of a sphere inscribed in the pyramid.*

Make the traditional drawing: let $SABC$ be the pyramid (Fig. 73) and let the triangles ASC and BSC be equilateral triangles and $\angle ASB = \angle ACB = 90°$. Use formula (5) to compute the radius r of the inscribed sphere.

Fig. 73



The surface area of the pyramid $SABC$ is determined at once: $S = a^2(2+\sqrt{3})/2$. To find the volume, we have to compute the altitude SH. Join H to the vertices of the base. Since the lateral edges are equal, so are their projections $AH = BH = CH$. But this signifies that H is the centre of a circle circumscribed about the triangle ABC. But since $\angle ACB = 90°$ the centre of the circumscribed circle must lie at the midpoint on the hypotenuse AB, that is, point H must coincide with Q, the foot of the altitude of the lateral face ASB dropped from S onto AB.

Thus, the drawing in Fig. 73 is impossible, erroneous. Actually, the plane ASB is perpendicular to the plane ABC and the altitude of the pyramid coincides with the altitude SQ of the lateral face ASB.

With the proper drawing we can readily determine $V = a^3\sqrt{2}/12$ and then the desired radius of a sphere inscribed in the pyramid $r = a\sqrt{2}\,(2-\sqrt{3})/2$.

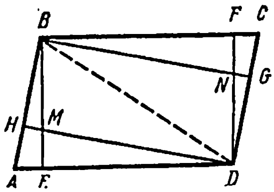
Here is another problem in which the customary drawing does not correspond to the statement of the problem and has to be replaced.

3. *Given a parallelogram $ABCD$ with $AB = 1$, $BC = 2$ and angle ABC obtuse. Two straight lines are drawn through each of the points B and D, one perpendicular to AB, the other perpendicular to BC. The intersections of these four lines yield a parallelogram that is similar to $ABCD$. Find the area of $ABCD$.*

Some students started the solution with the drawing shown in Fig. 74, where $BE \perp BC$, $DF \perp BC$, $BG \perp AB$, $DH \perp AB$. By hypothesis, the parallelogram $BNDM$ resulting from the intersections of the lines BE, DF, BG, and DH is similar to the parallelogram $ABCD$. To compute the area of $ABCD$ we have to find the angle BAD (because the sides of the parallelogram are given). This angle is acute, by the statement of the problem; denote it by α.

Let us first find the ratio $BM : MD$ of the sides of the parallelogram $BNDM$. To do this, it is necessary to establish which pairs of sides of the similar parallelograms $ABCD$ and $BNDM$ are corresponding

Fig. 74



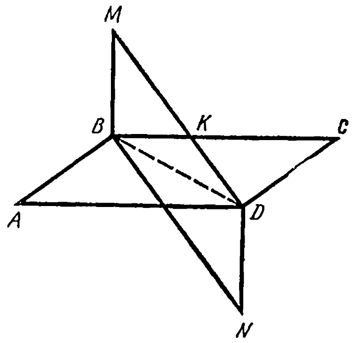
sides. Draw the diagonal BD and consider the triangles BAD and BMD. Since $\angle ABD > \angle ADB$ (because it is given that $AD > AB$ in the triangle BAD), and $\angle ABE = \angle ADH$ (acute angles with corresponding sides perpendicular), it therefore follows that $\angle MBD > \angle MDB$ and so the inequality $MD > BM$ is valid for the sides of the triangle BMD. Thus, in the parallelogram $BNDM$, side MD is *greater* than side BM, i.e., $MD : BM > 1$. Since in $ABCD$ it is given that $BC : AB = 2 > 1$, the pairs of sides AB and BM, BC and MD are similar, and so $MD : BM = 2$.

Let us now compute the angle α. From the similarity of right triangles MED and MHB we conclude that $ED : HB = MD : BM$, that is, $ED = 2HB$. But $ED = AD - AE = 2 - \cos\alpha$ (from the right triangle ABE it follows that $AE = AB\cos\alpha = \cos\alpha$) and $HB = AB - AH = 1 - 2\cos\alpha$ (from the right triangle AHD it follows that $AH = AD\cos\alpha = 2\cos\alpha$) and therefore $2 - \cos\alpha = 2(1 - 2\cos\alpha)$, whence $\cos\alpha = 0$, or $\alpha = 90°$.

Having obtained the value of the angle α (which, by hypothesis, is *acute*), many students could not find a way out of this contradiction. Some attempted to find a mistake in the computations (but in vain because for the configuration drawn in Fig. 74 all computations are perfectly legitimate), other students stated that the problem does not have a solution.

Few drew the correct conclusion: that the *given result indicates that Fig. 74 does not satisfy the conditions of the problem*. In other words, that under the conditions of the problem the configuration given in the drawing is *impossible*.* It is thus necessary to find out whether a different configuration of figures that fits the condition of the problem exists or not. Unfortunately, insufficient geometric imagination prevented most students from finding the configuration shown in Fig. 75, where the parallelogram $BNDM$ does not lie entirely inside the parallelogram $ABCD$.

Fig. 75



We will now solve the problem for the configuration shown in Fig. 75. Having established that $MD>BM$ (to do this it is sufficient to draw the diagonal BD and note that $\angle MBD > \angle MBC = 90^\circ$), we conclude, from the similarity of the parallelograms, that

$$\frac{BM}{DC}=\frac{MD}{BC} \tag{6}$$

Let K be the point of intersection of the straight line MD and side BC. From the similarity of the right triangles MBK and KDC it follows that

$$\frac{BM}{DC}=\frac{MK}{KC} \tag{7}$$

Comparing (6) and (7) and noting that $MD=MK+KD$, $BC=BK+KC$, we get the equation

$$\frac{MK}{KC}=\frac{MK+KD}{BK+KC} \quad \text{or} \quad \frac{MK}{KC}=\frac{KD}{BK} \tag{8}$$

However, it follows from the same similar triangles MBK and KDC that $MK:KC = BK:KD$. Comparing this equation with (8) we conclude that $MK = KC$. But then from (7) it follows that $BM = DC = 1$, and from (6) that $MD = BC = 2$. It thus turns out that the parallelograms $ABCD$ and $BMDN$ *are equal*.

Since $MK = KC$, from the right triangle MBK by the Pythagorean theorem we find that $MK^2 = 1 + (2 - MK)^2$, i.e., $MK=5/4$. Final-

* To be more exact, if in the parallelogram $ABCD$ (with sides $AB=1$, $BC=2$ and obtuse angle ABC) we draw perpendiculars BE, DF, BG and DH to get the parallelogram $BNDM$ lying *inside* $ABCD$ (Fig. 74), then these parallelograms *cannot be similar*.

ly, noting that $\angle BMK=\alpha$, from the same right triangle MBK we find that $\sin\alpha = BK/MK=(2-KC)/MK=(2-MK)/MK = 3/5$. Hence, the area of the parallelogram $ABCD$ is equal to $S=AB\times \times BC\cdot\sin\alpha=6/5$.

The problem is solved. It turned out that the only configurafion satisfying the statement of the problem is when parallelogram $BMDN$ does not lie entirely within parallelogram $ABCD$. It is interesting to note that some students drew the configuration of Fig. 75 from the very start and found the area of the parallelogram $ABCD$, but did not investigate the possibility, under the conditions of the problem, of the case of Fig. 74. Naturally, a complete solution of the problem is only that one in which both cases are considered.

Exercises

1. Define the following: (a) a convex polygon, (b) alternate-interior angles, (c) a circle inscribed in a triangle, (d) skew lines, (e) the angle between two intersecting planes, (f) a spherical sector.

2. State whether each of the following assertions is a definition, an axiom or a theorem: (a) two intersecting straight lines can have only one point in common, (b) a regular polygon is a polygon whose angles are equal and whose sides are equal, (c) any three points in space always lie in one plane, (d) a straight line is perpendicular to a plane if it is perpendicular to two intersecting straight lines lying in the plane.

3. Let a straight line l lie in a plane π and let L be an arbitrary straight line not in that plane. Consider the theorem: if L is parallel to l, then L is parallel to the plane π. State the converse theorem, the inverse theorem, and the contrapositive theorem. Which of these theorems are valid?

4. Prove that a triangle with sides 5, 13, 12 is a right triangle. How is this assertion related to the Pythagorean theorem?

5. Is it possible to define the circumference of a circle as the limit of a sequence of perimeters of inscribed polygons when (a) the number of sides of the polygons increases without limit, (b) the sequence of lengths of the largest sides of the polygons tends to zero?

6. Prove that $3<\pi<4$.

7. Consider the statement: "If in the triangles ABC and $A_1B_1C_1$ we have $AB = =A_1B_1$, $AC=A_1C_1$ and $\angle ABC=\angle A_1B_1C_1$, then $\triangle ABC=\triangle A_1B_1C_1$". Is this proof correct? Place triangle $A_1B_1C_1$ on ABC so that side A_1C_1 is coincident with side AC, and vertex B_1 lies on some point B_2 located on the other side of the straight line AC from vertex B. Join the points B and B_2. The triangle BAB_2 is an isosceles triangle (since $AB_2=A_1B_1=AB$) and so $\angle ABB_2=\angle AB_2B$. Since $\angle ABC=\angle AB_2C$ (because $\angle AB_2C=\angle A_1B_1C_1$), it follows that $\angle CBB_2=\angle CB_2B$, that is the triangle BCB_2 is isosceles and $BC=B_2C$. Hence, the triangles ABC and AB_2C have corresponding sides equal, or $\triangle ABC=\triangle A_1B_1C_1$. Is the original assertion true?

8. Given in space two triangles with corresponding sides parallel. What can be said about the straight lines connecting corresponding vertices of the first and second triangles?

9. Prove that all three bisecting planes of the dihedral angles of a trihedral angle intersect along one straight line.

3.2 Problems involving loci and construction problems

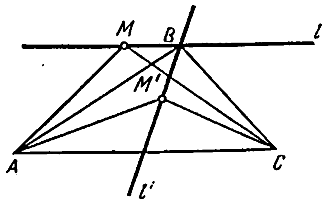
Descriptions of collections of points having a cartain property are of great importance in geometry. To handle questions of this nature, it is necessary to have a good understanding of the definition of a locus and to have a firm grasp of those loci encountered in the school curriculum.

It often happens however that the solution of *problems involving loci* stems from the apt use of some geometric fact.

1. *Given a triangle ABC. Find (in the plane of the triangle) a locus such that the areas of the triangles ABM and BMC are equal.*

Note that these triangles (Fig. 76) have a common side BM. Therefore their areas will be equal if the altitudes dropped onto this side are equal. We thus have to find a locus of points M such that the distances from the points A and C to the straight line BM are the same.

Fig. 76



Thus stated, the problem is simpler. We can at once state that any straight line parallel to AC has the property that the distances from it to the points A and C are equal, and so all the points of the straight line l which is parallel to AC and passes through vertex B of the triangle belong to the desired locus.

We see that it is possible merely to "guess" the points of the plane having the required property. The foregoing reasoning does not however constitute a solution to the problem, although some students at the examination thought so.

Indeed, we have the definition that a *collection of points in a plane (or in space, if the problem is one of solid geometry) having a certain property is called a locus if* (1) *every point of the set has the specified property,* (2) *every point not belonging to the set does not possess that property.*

The fact that all points M of the constructed line l (Fig. 76) possess the property stated in the problem is quite clear. But we are seeking *all* the points of the plane having the indicated property. We did not prove that there are no other such points with this property in the plane.

The question now is whether the line l exhausts the desired locus. No, it turns out, it does not. Any straight line passing through the midpoint of AC has the property that the distances from it to points A and C are the same. But we have to choose that straight line which passes through the vertex B. Hence, all the points M' of the straight line l'—the median of the triangle ABC drawn from vertex B (Fig. 76)—likewise belong to the desired locus. This second "guess" convincingly shows that we cannot dispense with rigorous proof, for we will never be sure that we have guessed everything.

Here is a proof carried out without any preliminary conjectures. First let us make a few remarks in order to establish the type of desired locux. Extend indefinitely the sides AB and BC of triangle ABC (Fig. 77) and then consider separately two possible cases.

(a) The point M *having the required property* lies inside the angle ABC (or inside the associated vertical angle $A'BC'$). Since the triangles AMB and BMC have the same area, the altitudes AH_1 and CH_2 are equal and so also are the right triangles AOH_1 and COH_2 (point O

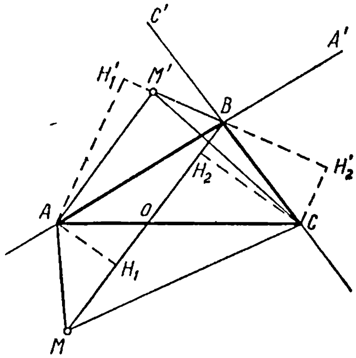


Fig. 77

is the intersection point of side AC and line segment MB or its extension; it is left to the reader to make the drawing for the case when point M lies inside the angle $A'BC'$) and, consequently, $AO=OC$. Thus, M lies on the median of triangle ABC drawn from the vertex B (the median is assumed to be extended indefinitely).

(b) Point M' *having the required property* lies inside the angle ABC' (or inside the associated vertical angle CBA'). Since triangles $AM'B$ and $BM'C$ have the same area, the altitudes AH'_1 and CH'_2 are equal and, hence, the quadrangle $AH'_1H'_2C$ is a rectangle ($AH'_1 \perp H'_1H'_2$, $CH'_2 \perp H'_1H'_2$, the sides AH'_1 and CH'_2 are equal and parallel; we leave it to the reader to make the drawing for the case where M' lies inside the angle CBA'). Thus, point M' lies on the straight line that is parallel to side AC and passes through the point B.

It is obvious that the points of the desired locus cannot lie on the straight lines AA' and CC', for then one of the triangles AMB or BMC would degenerate into a line segment.

The foregoing reasoning shows that the desired locus can only consist of two lines: the median (extended indefinitely) of triangle ABC drawn through the vertex B and the straight line parallel to side AC and passing through B (which is to say we again arrive at Fig. 76). We have thus proved the following: if a certain point M of the plane has the given property, then it must necessarily belong either to the line l or the line l'; in other words, every point of the plane that does not belong to these lines *does not possess* the given property.

The proof of the statement that every point of the straight lines l and l' belongs to the desired locus is simple enough. If M is an arbitrary point of l (Fig. 76), then triangles ABM and CMB have the same area since they have a common side MB and equal altitudes drawn to that side. If M' is an arbitrary point of l', then, inverting the arguments of case (a), we can prove that triangles AMB and CMB have the same area.

Hence, the straight lines l and l' are the desired locus.*

In the sequel we shall need two important loci in space (see Sec. 3.8).

The locus of points in space equidistant from the faces of a dihedral angle constitutes a plane that divides the dihedral angle into two equal dihedral angles (the reader can see the validity of this at once). This plane is called the *bisecting plane* of the dihedral angle (by analogy with the bisector of a plane angle).

Furthermore, it is easy to verify that *the locus of points in space equidistant from two given points A and B constitutes a plane perpendicular to the line segment AB and passing through the midpoint of the segment.*

A traditional type of problem is the *plane-geometry construction problem* involving the construction of a certain geometric configuration by means of two Euclidean tools: compass and straightedge. The main thing here is not a precise drawing neatly executed but the description of the algorithm, the sequence of operations necessary to perform the construction.

A knowledge of general methods of solving such problems (the method of symmetry, the method of similarity, etc.) is not provided for by the school curriculum and the student is called upon to solve only those construction problems that reduce directly to the basic techniques of construction taught in school.

In solving construction problems it is first necessary to perform an *analysis*, to find the idea that enables one to go forward with the construction. In the analysis, on the assumption that the problem has been solved, a drawing is made of the desired configuration and we attempt to find the relationships between the given data of the problem and the desired data that would permit employing the basic techniques of construction.

After the plan of the solution has been found, the student has to give a *detailed description of the sequence of operations in the construction* of the required configuration. It is then necessary to *prove* that the

---

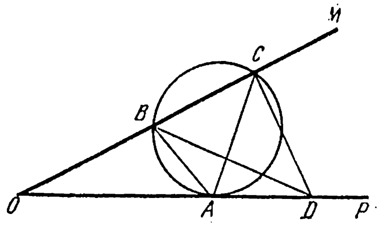
* However, a few words are in order concerning point B. If for M we choose the point B itself, then both triangles AMB and CMB degenerate into line segments, and on this basis we can eliminate B from the required locus. However, if we agree to consider such a degenerate triangle as having area zero, then, formally speaking, point B will possess the property required in the statement of the problem: both triangles AMB and CMB have the same area, equal to zero, when points M and B merge.

constructed configuration does indeed satisfy the requirements of the problem. It is finally necessary to *investigate* the problem, that is, to find out whether the construction is always possible, how many solutions the problem has, etc.

2. *On one side of a given acute angle lay off a line segment BC. On the other side of the angle construct a point from which BC is seen at the maximum angle.*

Let us begin with an *analysis* of the problem. Let the given line BC lie on side MO of angle MOP, and let point A lying on side OP be the desired point (Fig. 78). This means that for any other point D lying on ray OP the following condition holds true: $\angle BAC > \angle BDC$. To solve this problem we need to find a way of comparing the angles.

Fig. 78



A good way to compare angles is given by the theorems on the measurement of angles whose vertices lie on a circle, within it, and without it. If through points A, B and C we draw a circle, then $\angle BAC$ will be measured by half the arc it intercepts, while for any other point D on OP and outside the circle, the angle BDC will be measured by half the difference of the arcs intercepted by it, that is, the angle BDC will be *less* than angle BAC.

It is now obvious that to solve the problem we have to draw a circle through points B and C so that it is *tangent* to side OP of the angle MOP. Then the point of tangency will be the desired point because all other points of the ray OP will lie outside the constructed circle.

In order to make the *construction*, it is not at all necessary to build a circle passing through B and C and tangent to OP; it is sufficient to find the point A of tangency. To do this, we find the distance from this point to the vertex O of the angle MOP, using the familiar property of a tangent line and a secant line: $OA^2 = OC \cdot OB$.

And so in order to complete the construction we have to find the geometric mean of two known line segments OB and OC, and then lay off this line segment on the side OP of angle MOP from the vertex O. The other endpoint of that line segment is the point we desire.

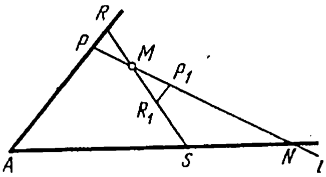
By carrying out all the arguments in reverse order, we can *prove* that the point thus constructed will be the desired one. An *investigation* shows that the problem is always solvable and has a unique solution.

3. *Given a point M inside an angle A. Draw through M a straight line l so that it cuts from the given angle a triangle of minimum area.*

Start with an *analysis*. For the sake of definiteness, we assume that the angle A is acute.

Let PN and RS be two lines passing through M (Fig. 79): Compare the areas of the triangles APN and ARS. Construct points P_1 and R_1 so that $P_1M = PM$ and $R_1M = RM$; then, passing from the triangle

Fig. 79

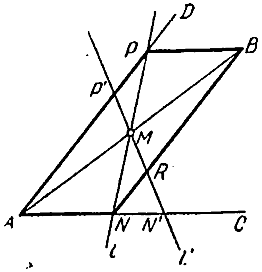


APN to the triangle ARS, we reduce the area since the discarded area of the triangle MNS is greater than the additional area of triangle PRM equal to triangle P_1R_1M. It is obvious that a reduction can only be executed if point M divides the line segment PN (between the sides of the angle) of line l into *unequal* parts.

And so through point M we have to draw a straight line l so that the equation $NM = MP$ holds; only in that case will the area of the triangle ANP be a minimum. But then it is obvious that AM is the median of triangle ANP.

Taking this triangle and building it up to a parallelogram, we at once have a method of *construction* (Fig. 80). Prolong AM beyond M

Fig. 80



and on it lay off $MB = AM$. Through point B draw a straight line parallel to one of the sides of the given angle, say AC, to intersection with side AD at P. The sought-for line l passes through P and M.

The *proof* consists essentially in repeating the steps taken in the analysis. Let l' be an arbitrary line passing through point M. Since $\triangle MPP' = \triangle MNR$ (by the second criterion for the equality of triangles), the triangle APN is equivalent to the quadrangle $AP'RN$, whose area is less than the area of the triangle $AP'N'$.

For the case when angle A is obtuse, the construction is analogous and is left to the reader. The *investigation* of the problem is simple, The problem is always uniquely solvable if angle A is less than a straight angle.

For additional remarks concerning the solution of construction problems see Sec. 3.3.

Exercises

1. Find the locus of points in the plane: (a) the difference in the distances of which to two given intersecting straight lines L and l of the plane is equal (in absolute value) to a given quantity $a>0$; (b) the sum of the squares of the distances of which to two given points A and B of the plane is equal to a given quantity a.

2. Given two fixed points A and B on a straight line. Two circles touch this line at points A and B respectively, and are tangent to each other at point M. Find the locus of points M.

3. Given in the plane three straight lines that intersect in pairs and do not pass through one point. Find the locus of the centres of the circles circumscribed about all possible triangles with vertices on these lines.

4. In a plane, find the locus of the feet of perpendiculars dropped from a given point A to straight lines passing through another given point B.

5. Find the locus of the midpoints of chords of a given circle passing through a given point A within the circle.

6. Find the locus of points in the plane such that the tangents drawn from these points to a given circle form an angle α.

7. Find the locus of the midpoints of line segments joining a given point A outside a given circle to points of the circle.

8. In a plane, given two fixed distinct points A and B. Find the locus of points M of the plane for which $AM \cdot BM \cdot \cos(\angle AMB)=3/4\, AB^2$.

9. Find the locus of points in space equidistant from three given points A, B, C (a) not lying on one straight line, (b) lying on one straight line.

10. Find the locus of points in space which lie at a given distance a from two given intersecting planes Π and π.

11. Find the locus of points in space which are equidistant from three planes intersecting in pairs and not passing through the same straight line and perpendicular to a plane π.

12. Find the locus of the feet of perpendiculars dropped from a given point A to all possible straight lines drawn in space through a fixed point B.

13. Find the locus of projections of a given point A on all possible planes passing through the given straight line l that does not contain A.

14. Find the locus of the midpoints of line segments AB where the points A and B lie on different faces of a given acute dihedral angle.

15. Find the locus of points in space through which it is impossible to draw a straight line intersecting the given skew lines L and l.

16. In the plane π lies a square with side a. Find the locus of points in space which lie at a given distance l from the vertices of the square.

17. Given a plane π and points A and B located to one side of it so that the straight line AB is not parallel to the plane π. Consider all possible spheres passing through the points A and B and tangent to the plane π. Find the locus of the points of tangency.

18. Given a cube with side a. Find the locus of the midpoints of line segments of given length l, one of the endpoints of which lies on the diagonal of the upper base of the cube, the other on a diagonal of the lower base not parallel to the first diagonal (or on prolongations of these diagonals).

19. Construct the line segment $\sqrt[4]{a^4+b^4}$ knowing the line segments a and b.

20. Construct a triangle, given side a, median m of the other side, and radius R of a circumscribed circle.

21. Construct a circle tangent to a given straight line l and passing through two given points A and B which lie to one side of the straight line.

22. Construct a circle tangent to a given circle and a given straight line l at a fixed point A.

23. Given a circle with diameter AB. Construct on the diameter AB a point C such that the sum of the areas of the circles constructed on the line segments AC and BC as diameters is 2/3 of the area of the given circle.

24. Given a triangle ABC. Construct a rectangle $ABDE$ one side of which is equal to side AB of the triangle, and the area together with the area of the square constructed on side BD is equal to the area of the given triangle.

25. Given a pyramid the base of which is a square; the altitude of the pyramid is equal to a side of the base. It is required to construct a trihedral prism whose altitude is equal to a given line segment h. The base of the prism is an isosceles right triangle. The volume of the prism is equal to the volume of the given pyramid. Describe a method for constructing the base of the prism using compass and straightedge.

3.3 Using trigonometry and algebra in geometry

A rather widespread view among students is that there are algebraic, geometric and trigonometric problems whose methods of solution are in no way related. That explains why geometric problems are so often tackled by purely geometric means.

This is a completely erroneous view that tends to separate the various branches of elementary mathematics, whereas many problems are best solved by invoking the whole range of knowledge from different divisions of elementary mathematics.* The employment of trigonometric or algebraic methods and facts in the solution of geometric problems is sometimes inevitable for the simple reason that no purely geometric approaches to the solution are possible.

In such problems, geometry, trigonometry and algebra must appear as a single whole, and successful solutions are only possible if the student combines facts from the diverse divisions of geometry, trigonometry and algebra, in other words, if he has a firm knowledge of the entire school course of mathematics.

How useful the employment of trigonometry is to the solution of so-called "computational problems" in geometry is familiar to all. Many problems would be beyond our scope if we did not invoke trigonometric relations between the sides and angles of different figures.

However, the use of trigonometry in solving geometric problems is not confined only to solving triangles and simplifying the resulting formulas. Its possibilities are much broader. For one thing, a very useful idea is to *find an angle from trigonometric relations*. Unfortunately, this method of solving geometric problems is not generally known. We will therefore illustrate it in the following two problems.

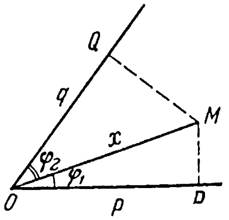
---

* See Problem 19 of Sec. 1.8 where geometric ideas were employed in the solution of an algebraic problem.

1. *A point M is taken inside an acute angle α. The feet P and Q of perpendiculars dropped from M on the sides of the angle are separated from the vertex O of the angle by $OP=p$ and $OQ=q$. Find the angles into which OM divides the angle α.*

Let us denote the (unknown) distance OM by x, the acute angles POM and QOM into which OM divides angle α by φ_1 and φ_2, respecti-

Fig. 81



vely (Fig. 81). The magnitude of angle φ_1 may be expressed in terms of p and x from the right triangle MPO:

$$\cos\varphi_1=\frac{p}{x}, \text{ or } \varphi_1=\arccos\frac{p}{x}$$

But then, from the right triangle MQO we get (noting that $\varphi_2=\alpha-\varphi_1$)

$$\frac{q}{x}=\cos\varphi_2=\cos(\alpha-\varphi_1)=\cos\alpha\cos\left(\arccos\frac{p}{x}\right)+\sin\alpha\sin\left(\arccos\frac{p}{x}\right)$$

Computing the resulting expressions (see Sec. 2.5), we find

$$q=p\cos\alpha+\sin\alpha\sqrt{x^2-p^2}$$

From this equation we could determine x and then from the right triangles MPO and MQO find $\cos\varphi_1$ and $\cos\varphi_2$. A simpler approach is this. Note that $\sqrt{x^2-p^2}=MP$ (from the right triangle MPO) and so

$$\tan\varphi_1=\frac{MP}{OP}=\frac{\sqrt{x^2-p^2}}{p}=\frac{q-p\cos\alpha}{p\sin\alpha}$$

And we finally get

$$\varphi_1=\arctan\frac{q-p\cos\alpha}{p\sin\alpha}, \quad \varphi_2=\alpha-\varphi_1{}^*$$

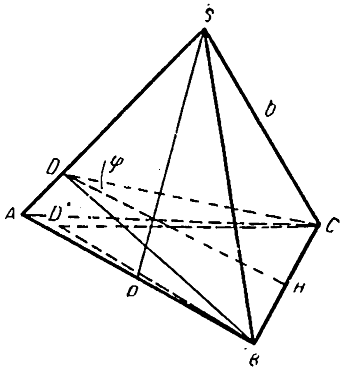
2. *The lateral edge of a regular triangular pyramid is equal to b and the angle between the lateral faces is φ. Find the side of the base.*

---

* The reader can perform some obvious manipulations and see for himself that $\varphi_2=\arctan\dfrac{p-q\cos\alpha}{q\sin\alpha}$.

Let the pyramid $SABC$ be regular: $AS = BS = CS = b$ (Fig. 82). Drop from points B and C the altitudes BD_1 and CD_2 of the triangles ASB and ASC. We will show that the feet of these altitudes are one and the same point $D_1 = D_2 = D$. Indeed, $\triangle SBD_1 = \triangle SCD_2$ and so the distance from vertex S to the feet of the altitudes dropped on side AS will be the same in each of these triangles.

Fig. 82



From this it follows that BDC is the plane angle of the dihedral angle between the lateral faces, or $\angle BDC = \varphi$.

Draw the slant height SP of the pyramid and the altitude DH of the isosceles triangle BDC. We then obviously have $AB = 2AP = 2b \times \times \cos(\angle DAP)$ and it is sufficient for us to determine the angle DAP. Since this is an acute angle (being the base angle of a lateral face of a regular triangular pyramid), we can write from the triangle BAD that

$$\angle DAP = \arcsin \frac{BD}{AB}$$

From triangle BDH we get

$$BD = \frac{BH}{\sin(\varphi/2)} = \frac{AB}{2\sin(\varphi/2)}$$

Hence

$$AB = 2b \cos\left[\arcsin\left(\frac{1}{2\sin(\varphi/2)}\right)\right]$$

Simplifying this expression (see Sec. 2.5), we find

$$AB = b\sqrt{4 - \frac{1}{\sin^2(\varphi/2)}} \qquad (1)$$

In connection with the result (1), note that usually in a geometric problem an attempt is made to reduce the answer to a form convenient for taking logs. The right member of (1) can readily be represented in the form

$$AB = \frac{2b}{\sin(\varphi/2)} \sqrt{\sin\left(\frac{\varphi}{2} + \frac{\pi}{6}\right) \sin\left(\frac{\varphi}{2} - \frac{\pi}{6}\right)} \qquad (2)$$

Such representation is not obligatory however.

Incidentally, this tradition of reducing to a form convenient for taking logarithms does not always by far result in the simplest form of the answer. Despite the widely held view that this form is the most convenient for computing concrete values of quantities indicated in the statement of a problem, it is not so in many cases. It is easy to see that for such computations the answer may be obtained faster and more simply by using formula (1) instead of (2).

When solving problems of this kind, many students do not only carry out the necessary manipulations and justify them, but also attempt to make a detailed study of the resulting formula, which as a rule consists in finding the domain of the variable.

This study goes somewhat as follows. It is clear that (1) is meaningful only if

$$4-\frac{1}{\sin^2(\varphi/2)}>0 \tag{3}$$

Since the angle φ between the lateral faces (due to obvious geometric considerations) must lie in the interval from 0° to 180°, so $0^\circ<\varphi/2<$ $<90^\circ$, and then $\sin(\varphi/2)>0$. Therefore, inequality (3) may be rewritten as $\sin(\varphi/2)>1/2$, whence (1) is meaningful for $60^\circ<\varphi<180^\circ$.

It is readily seen that this condition holds true for any regular triangular pyramid, since, by the familiar property of a perpendicular and inclined lines, $BD<BA$ (Fig. 82). For this reason, if in the plane of the base ABC of the pyramid we construct an isosceles triangle BD^*C with sides BC, $BD^*=BD$, $CD^*=CD$ (equal to the triangle BDC), then the point D^* will be located *inside* the triangle ABC. Whence, after an elementary computation of angles, we conclude that $\angle BDC=\angle BD^*C>\angle BAC=60^\circ$.

Consequently, *in a regular triangular pyramid, the dihedral angle at a lateral edge is always greater than* 60°. This in turn means that formula (1) is always meaningful for any regular triangular pyramid.

It should be stressed that such an investigation of the answer is not an obligatory element of the solution (if of course this is not explicitly required by the statement of the problem).

However, many students go into such investigations and make the following logical error: it is taken that the configuration stated in the problem exists exactly for those values of the letters for which the final formula is meaningful. In actuality, however, an investigation of the conditions of existence of a geometric configuration *is by no means equivalent* to a simple analysis of the answer.

This will be clear from the following problem.

3. *In the parallelogram $ABCD$ the larger side $AB=a$, the smaller side $BC=b$; the acute angle between the diagonals is equal to α. Find the distance between the parallel sides AB and DC*

Let us form two distinct expressions for the area S of the parallelogram (Fig. 83). On the one hand,

$$S = 4S_{\triangle BOC} = 2 \cdot OB \cdot OC \sin \alpha$$

where α is the acute angle between the diagonals. To determine the product $OB \cdot OC$ we apply the cosine law to the triangle BOC:

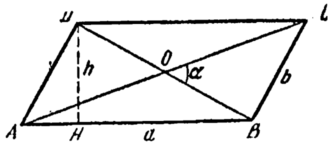
$$b^2 = OB^2 + OC^2 - 2 \cdot OB \cdot OC \cdot \cos \alpha$$

and the familiar property of a parallelogram

$$OB^2 + OC^2 = \frac{1}{4}(DB^2 + AC^2) = \frac{1}{2}(a^2 + b^2)$$

to get, finally, $S = \frac{1}{2}(a^2 - b^2) \tan \alpha$.

Fig. 83



On the other hand, $S = AB \cdot DH$, where $DH = h$ is the desired distance (the altitude of the parallelogram). Equating the two expressions for the area S, we get

$$h = \frac{a^2 - b^2}{2a} \tan \alpha \tag{4}$$

If we decided to seek the domain of this formula, we would see that the right member is meaningful, say, for all values of $a > 0$, $0 < b < a$ and $0° < \alpha < 90°$. However, does it follow from this that for all such values of the letters the geometric configuration mentioned in the problem exists? To put it differently, is it possible to construct a parallelogram with sides a and b and acute angle α between the diagonals for arbitrary values of $a > 0$, $0 < b < a$ and $0 < \alpha < 90°$?

If, say, we substitute in formula (4) the values $a = 3$, $b = 1$, $\alpha = 45°$, we get $h = 4/3$. This result can't but appear rather strange. Consider the right triangle AHD (Fig. 83). The hypotenuse $AD = 1$ must then be shorter than the side $DH = 4/3$. This can only mean one thing: that a parallelogram with the given values of a, b, α cannot exist.

This is also evident directly. Since the area of a parallelogram with sides a and b does not exceed the area of a rectangle with the same lengths of the sides, the inequality $\frac{1}{2}(a^2 - b^2) \tan \alpha \leqslant ab$ must hold true. It can be rewritten thus:

$$0 < \alpha \leqslant \arctan \frac{2ab}{a^2 - b^2} \tag{5}$$

In other words, the *quantities a, b, α cannot be specified in a totally arbitrary manner*, irrespective of one another. They must satisfy inequality (5).

To summarize, then, cases are possible where a geometric configuration of the problem does not exist for all values of the letters in the domain of the answer. For this reason, the investigation of a geometric problem (which amounts to determining the conditions under which the configuration exists) is a much more complicated undertaking and is not required of the student. The only thing required in all geometric problems is *to carry through the solution on the assumption that the geometric configuration given in the problem exists* (unless of course such an investigation is explicitly required).

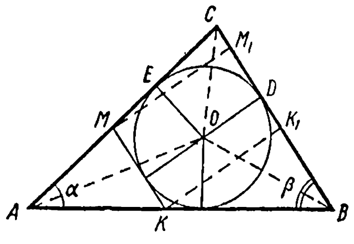
In the solution of geometric problems, one occasionally encounters relations of unusual form that result from the application of familiar trigonometric formulas. These exotic formulas cause much difficulty since the student is ordinarily unable to interpret them properly. Yet this is very important because such formulas often "think" for us and take into account cases that we have not indicated in the drawing or conditions that have eluded us.

Here is a problem where the formulas automatically make allowance for a condition which, it would seem, was not intended to be utilized explicitly and which many students did not even notice at the examination.

4. *A circle of radius r is inscribed in an acute-angle triangle ABC with angles α and β at the vertices A and B respectively. A tangent drawn to the circle parallel to BC cuts the sides AB and AC of the triangle at the points K and M respectively. Find the area of the trapezoid $BCMK$.*

By the formula for the area of a trapezoid, we have to find the bases BC and MK of the trapezoid and its altitude (Fig. 84). It is quite evident that the altitude of the trapezoid is equal to the diameter $2r$ of the inscribed circle.

Fig. 84



The first step in the solution is to find the base BC of the trapezoid, which base is a side of the given triangle. If we drop a perpendicular OD from the centre O of the inscribed circle (the point of intersection of the bisectors CO and BO) on the side BC, then $OD = r$. From the right triangles BOD and COD we then obtain

$$BD = r\cot\frac{\beta}{2}, \quad CD = r\cot\frac{\pi-\alpha-\beta}{2} = r\tan\frac{\alpha+\beta}{2}$$

respectively, and, hence,

$$BC = r\left(\cot\frac{\beta}{2} + \tan\frac{\alpha+\beta}{2}\right) \tag{6}$$

The second step in the solution is to find the base MK of the trapezoid. Drop perpendiculars MM_1 and KK_1 from points M and K on the side BC. Then clearly $MM_1 = KK_1 = 2r$. Since

$$MK = M_1K_1 = BC - BK_1 - CM_1 \tag{7}$$

it suffices to find BK_1 and CM_1. This can be done from the right triangles BKK_1 and CMM_1, respectively:

$$BK_1 = 2r\cot\beta,\ CM_1 = 2r\cot(\pi-\alpha-\beta) = -2r\cot(\alpha+\beta) \tag{8}$$

What is the geometric significance of this minus sign that has suddenly appeared in (8)? Quite naturally, it does not indicate any "negative" length. What is more, it is precisely this minus sign (which appeared by itself) that ensures us that the length of CM_1 is positive. The crux of the matter is that the triangle ABC is given as an *acute-angle triangle* and therefore $\alpha+\beta > \pi/2$, or $\cot(\alpha+\beta) < 0$. Thus, the formula takes into account the assumption concerning triangle ABC which was given in the statement of the problem and which we had forgotten about, had not stipulated and had not utilized explicitly.

How did this minus sign get here after all? How was this fact that the triangle is an acute triangle taken into account? To find out let us carefully go over all the arguments and manipulations. Quite naturally, when we started out we made the triangle ABC an acute triangle as indicated in the statement of the problem (Fig. 84). But did we make use of this fact later on? It is quite clear that we implicitly made use of it when we dropped the perpendicular MM_1 and considered it as obvious that the point M_1 lies on BC, for only then is (7) true. If the angle ACB were obtuse, the foot M_1 of the perpendicular would lie on the extension of BC beyond C and formula (7) would look like this: $MK = M_1K_1 = BC - BK_1 + CM_1$ (this is easy to see if an appropriate drawing is made).

Thus, there is nothing unusual in the appearance of the minus sign in the formula for CM_1. Simply the formula explicitly indicates what we took for granted in the drawing and did not even bother to stipulate or justify.

To complete the solution all we need to do is carry out the manipulations. Substituting expressions (6) and (8) into (7), we determine the base MK of the trapezoid and then find the area S of the trapezoid $BCMK$ from the familiar formula

$$S = 2r^2\left[\frac{1}{\sin\beta} + \frac{1}{\sin(\alpha+\beta)}\right] \tag{9}$$

It is well to bear in mind that the answer to a geometric problem can appear in a variety of very unlike forms, depending on the idea on which the solution is based or merely on the chosen sequence of manipulations.

For instance, if we seek the base MK in a different way, the answer will have a form unlike (9). Indeed, $\triangle ABC$ is similar to $\triangle AKM$ (Fig. 84), and so $BC : MK = H : h$, where H and h are the altitudes of triangles ABC and AKM, respectively, dropped from the common vertex A. Since $h = H - 2r$, from this proportion we get $MK = BC - (2r \cdot BC/H)$. Since the area of the triangle ABC is equal to $\frac{1}{2} H \times BC = pr$, where p is the semiperimeter [see formula (4) of Sec. 3.1], then $H = 2pr/BC$ and we only have p to determine. Drawing the bisector AO and the perpendiculars OE and OF to the sides of the triangle ABC, we find from the right triangle AOE that $AE = r\cot(\alpha/2)$, and so (by the property of tangents to a circle)

$$p = \frac{1}{2}(AB + BC + CA) = AE + BD + DC$$

Thus, we determine the base MK of the trapezoid and then its area:

$$S = r^2\left(\cot\frac{\beta}{2} + \tan\frac{\alpha+\beta}{2}\right)\left[1 + \frac{\cot\frac{\alpha}{2}}{\cot\frac{\alpha}{2} + \cot\frac{\beta}{2} + \tan\frac{\alpha+\beta}{2}}\right] \quad (10)$$

It is not essential of course in what form the answer is given, although it is preferable to have it in the simplest form possible. All these forms of the answer can be transformed into one another, which is proof that they are equivalent and that no mistakes have been made [the reader will easily see how formula (10) can be transformed to formula (9)]. For this reason, do not hurry to discard as incorrect an answer which differs from that given in the book.

In the next problem, the formula warns us that in different pyramids the centre of a circumscribed sphere can lie inside or outside the pyramid, or in the plane of a face. This fact (see Sec. 3.8). is familiar to many students, but some forget about it and fail to consider all possible configurations in the general case.

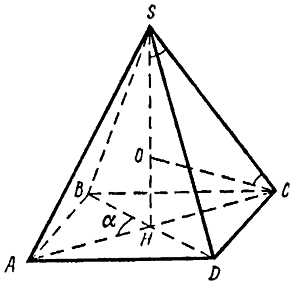
5. *The base of a pyramid is a rectangle with angle α between the diagonals, and all lateral edges form the same angle φ with the plane of the base. Determine the distance from the centre of a circumscribed sphere to the plane of the base and the volume of the pyramid if R is the radius of the sphere.*

Let the given pyramid be $SABCD$ (Fig. 85) with rectangular base. Draw the altitude SH. Then, by hypothesis, we have $\angle HAS = \angle HBS = \angle HCS = \angle HDS = \varphi$ and therefore (because $\triangle ASH = \triangle BSH = \ldots$) all the lateral edges are equal and H is the point of intersection of the diagonals of the rectangle. Since the centre

O of the circumsphere must be equidistant from all vertices, it lies on the perpendicular to the plane $ABCD$ erected in the centre of the rectangle, that is, on the altitude SH.

Consider the triangles CHS and COS. Since $\angle CSH = 90° - \varphi$ and $\triangle COS$ is an isosceles triangle, then, by the property of an exterior

Fig. 85



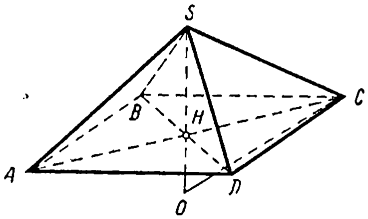
angle, we have $\angle COH = 180° - 2\varphi$. Solving triangle COH, we find the distance from the centre O to the plane of the base:

$$OH = R\cos(180° - 2\varphi) = -R\cos 2\varphi \tag{11}$$

The minus sign here should interest us. What does it stand for? The point is that having made the drawing in Fig. 85 we thus actually worked from the supposition that the centre of the circumscribed sphere lies *inside* the pyramid and we conducted the solution on that implicit assumption, which is not explicitly stated in the problem. But the centre of the circumscribed sphere need not necessarily lie inside the pyramid, and the formula is clear evidence that this is so.

The centre of a circumscribed sphere lies *inside* the pyramid if the altitude of the pyramid is greater than one half the diagonal of the rectangle (then there will be a point O on SH that is equidistant from C and from S), that is, if $\varphi > 45°$. But in this case, $\cos 2\varphi < 0$ and,

Fig. 86



consequently [see (11)], $OH > 0$. If the centre of the circumscribed sphere lies *on the base* of the pyramid, coinciding with point H, then the altitude of the pyramid is equal to one half the diagonal of the rectangle, and $\varphi = 45°$, and so $OH = 0$. Finally, if the centre of the circumscribed sphere lies *outside* the pyramid (Fig. 86), then $\varphi < 45°$ and we cannot use formula (11) to seek OH. In this case, obvious reasoning suggests that $\angle\, COH = 2\varphi$ and so the distance from point O to the plane of the base $OH = R \cos 2\varphi$ (a positive quantity).*

Let us now compute the volume. From the triangle COH we find $HC = R \sin (180° - 2\varphi) = R \sin 2\varphi$, an equation that holds true both for Fig. 85 and Fig. 86. Therefore the area of the base is

$$S_{ABCD} = 4S_{\triangle\, DHC} = 2R^2 \sin^2 2\varphi \sin \alpha$$

The angle α may be either acute or obtuse. We consider three cases in computing the altitude of the pyramid. If $\varphi > 45°$, that is, if the centre O lies inside the pyramid (Fig. 85), then

$$SH = SO + OH = R - R \cos 2\varphi = 2R \sin^2 \varphi$$

if $\varphi < 45°$, that is if the centre O lies outside the pyramid (Fig. 86), then

$$SH = SO - OH = R - R \cos 2\varphi = 2R \sin^2 \varphi$$

And if $\varphi = 45°$, then $SH = SO = R = 3R \sin^2 45°$.

Thus, for all values of the angle φ, $0° < \varphi < 90°$, we get the same result:

$$V = \frac{4}{3} R^3 \sin^2 2\varphi \sin^2 \varphi \sin \alpha$$

Very important in the solution of geometric problems are algebraic methods. Algebra, often linked up with trigonometry, enables the student to topple many a complicated problem.

The gist of the algebraic approach to geometric problems consists in setting up an equation for some quantity on the basis of geometric reasoning, and then in solving it or in investigating it by algebraic means. There of course remains the question of the geometric interpretation of the algebraic result.

Broad opportunities for using algebra in geometry are provided by the metric relations which obtain in the triangle and the circle, by the formulas for solving right triangles, the laws of sine and cosine, etc.

Note that problems amenable to solution by algebraic methods often require rather cumbersome computations. One should therefore get used to involved manipulations and awkward-looking answers. As a rule, the computations required in such problems have simple underlying ideas and are quite accessible to any student who has a firm grasp of the basic formulas of trigonometry and algebra, is facile with the techniques of algebraic and trigonometric transformations, and has trained himself in a careful execution of manipulations and works in a neat fashion generally.

* It is not hard to see that the distance from the centre O of the circumscribed sphere to the plane of the base $ABCD$ of the pyramid at hand is equal to $R \cdot |\cos 2\varphi|$, irrespective of the location of the centre O.

6. *The angle A in the triangle ABC is equal to α, the opposite side is equal to a. Find the other two sides if we know that side a is the geometric mean of the radii of the inscribed and circumscribed circles of the triangle.*

Let us reduce the solution of this problem to an algebraic system. To do this, set up the necessary number of independent equations. Denote by b and c the lengths of the sides of the triangle ABC which lie opposite to angles B and C respectively. These quantities b and c are our unknowns.

The first equation is provided by the cosine law:

$$b^2+c^2-2bc\cos\alpha=a^2 \tag{12}$$

To obtain the second equation we take advantage of the relation (given in the statement of the problem) $a=\sqrt{Rr}$, where R and r are the radii of the circumscribed and inscribed circles, respectively. Recalling the familiar expressions for these radii [formulas (3) and (4) of Sec. 3.1] and for the area of a triangle, we can write down the chain of equations

$$a^2=rR=\frac{2S}{a+b+c}\cdot\frac{a}{2\sin A}=\frac{a}{(a+b+c)\sin A}\cdot\frac{1}{2}bc\sin A=\frac{abc}{2(a+b+c)}$$

whence we find the second equation:

$$bc-2a(b+c)=2a^2 \tag{13}$$

We thus have a system of two equations (12) and (13) in two unknowns b and c. Our task now is purely algebraic— to solve the system.

Rewriting (12) as $(b+c)^2-2bc(1+\cos\alpha)=a^2$ and substituting the expression for bc from (13), we get a quadratic equation in $b+c$:

$$(b+c)^2-8a\cos^2\frac{\alpha}{2}\cdot(b+c)-\left(8a^2\cos^2\frac{\alpha}{2}+a^2\right)=0 \tag{14}$$

whence

$$b+c=a\left(1+8\cos^2\frac{\alpha}{2}\right) \tag{15}$$

(we discard the second—negative—root of (14) since it has no geometric meaning). From (13) we now determine

$$bc=4a^2\left(1+4\cos^2\frac{\alpha}{2}\right) \tag{16}$$

From the relations (15), (16), it is clear that b and c are the roots of the quadratic equation

$$z^2-a\left(1+8\cos^2\frac{\alpha}{2}\right)z+4a^2\left(1+4\cos^2\frac{\alpha}{2}\right)=0$$

Solving this equation, we get

$$z_{1,2}=\frac{a}{2}\left[5+4\cos\alpha\pm\sqrt{16\cos^2\alpha+8\cos\alpha-23}\right] \tag{17}$$

Now for b we can, say, take the plus sign and for c the minus sign; any other combination of signs yields virtually the same triangle but with the sides designated differently.

The solution is complete, the required lengths b and c of the sides of the triangle have been found. Although the problem does not require any supplementary investigation, some students at examinations determined the conditions under which formulas (17) are geometrically meaningful.

It is obvious that these formulas have geometric meaning only when the radicand is a nonnegative number and, besides, $z_1 > 0$, $z_2 > 0$. But the quadratic $16x^2 + 8x - 23$ (its roots are easy to find) is nonnegative when $x \leqslant (-1 - \sqrt{24})/4$ and when $x \geqslant (\sqrt{24} - 1)/4$. Now, when the students at the examination formulated the condition for the angle α, they made a diversity of mistakes which indicated their inability to handle trigonometric inequalities. Actually, the radicand in (17) is nonnegative if

$$0 < \alpha \leqslant \arccos \frac{\sqrt{24}-1}{4} \tag{18}$$

(Let us not forget that α is an angle of the triangle and for this reason we are only interested in the values $0 < \alpha < \pi$.) Furthermore, it is easy to see that for the values of α indicated in (18) both roots of (17) are positive, that is, are geometrically meaningful.*

7. *Given two concentric circles of radii r and R $(R > r)$. Find the side of a square two verticies of which lie on the circle of radius r and the other two on the circle of radius R. For what relationship between r and R:* (a) *is the solution possible,* (b) *is there only one solution?*

First notice that the geometric configuration stated in the condition of the problem does not exist for arbitrary radii of the given circles. Indeed, if R is "very much greater" than r, then it is easy to see that no square can exist two vertices of which lie on one circle and the other two vertices on the other circle. It is furthermore clear that for certain values of R and r there can be two squares** thus located (Fig. 87). Finally, there is no guarantee that there may not even be more than two such squares.

The above reasoning shows that one should not begin a solution with a drawing. Firstly, having made a drawing we thus at once confine ourselves to the case when the configuration *exists* and can say nothing about when it is impossible. Secondly, we already know that the configuration, is, generally speaking, not unique and we must

* It is to be stressed (cf. Problems 2 and 3) that we only determined those conditions under which formulas (17) *may be* meaningful geometrically; the question remains open as to whether for all values of α that satisfy condition (18) the geometric configuration under consideration *does indeed exist*.

** We mean that there exist two squares with *distinct sides*. There is no sense in distinguishing between squares which satisfy the condition of the problem but have sides of the same length.

therefore draw *all* possible cases; however, we do not yet know exactly how many such cases there are.

We shall therefore choose a different approach which does not make use of any specific drawing of the configuration but reduces to geometric constructions. It is quite obvious that if a square satisfying the condition of the problem exists, then one of its sides lies on the chord of the greater circle intersecting the smaller circle (for instance, such

Fig. 87

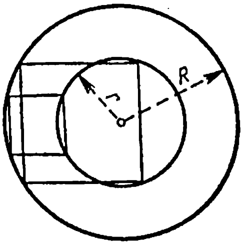
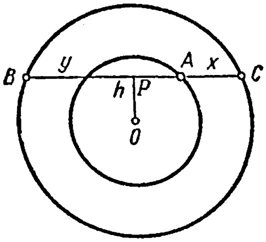


Fig. 88



a chord is the prolongation of a side of the square, one end of which lies on the larger circle, the other on the smaller one). Conversely, if we prove that the side of the desired square cannot lie on any chord of the larger circle that intersects the smaller circle, then this will mean that (for given R and r) the sought-for configuration is impossible.

We must therefore obtain *necessary* and *sufficient* conditions such that one of the sides of the square satisfying the condition of the problem will lie on a certain chord of the larger circle intersecting the smaller circle. Let BC (Fig. 88) be an arbitrary chord of the larger circle which intersects the smaller circle at some point A. This chord is uniquely determined by the length h of the perpendicular OP dropped on it from the centre O of the circles.

Denote the length of AC by x, that of AB by y. It is apparent that a side of the square of interest can lie on chord BC if and only if either $x=2h$ or $y=2h$. Consequently, if (for given R and r) there is at least one value of h which satisfies either the condition $x=2h$ or $y=2h$, then the required configuration is possible. But if there is no such value of h, then (for the given R and r) the problem does not have a solution.

Since

$$x=CP-AP=\sqrt{R^2-h^2}-\sqrt{r^2-h^2},$$
$$y=BP+PA=\sqrt{R^2-h^2}+\sqrt{r^2-h^2}$$

the necessary and sufficient condition thus obtained for solvability of the problem may be stated: the configuration we need occurs for (positive) values of h which satisfy at least one of the equations

$$\begin{aligned}\sqrt{R^2-h^2}-\sqrt{r^2-h^2}&=2h,\\ \sqrt{R^2-h^2}+\sqrt{r^2-h^2}&=2h\end{aligned}\tag{19}$$

and does not occur for a single other value of h. In particular, if neither of the equations of (19) has positive roots, then the required configuration does not exist.*

The irrational equations (19) are conveniently solved in parallel. The domain of the variable (see Sec. 1.9) consists of values of h such that $h \leqslant r$ (because $R > r$). The condition $R > r$ implies that $\sqrt{R^2-h^2} > \sqrt{r^2-h^2}$ so that both members of both equations are nonnegative. Therefore, when squaring we obtain (after obvious manipulations) two equations:

$$\begin{aligned} -2\sqrt{(R^2-h^2)(r^2-h^2)} &= 4h^2-(R^2+r^2) \\ 2\sqrt{(R^2-h^2)(r^2-h^2)} &= 4h^2-(R^2+r^2) \end{aligned} \tag{20}$$

which, in the domain, are equivalent to the corresponding equations (19). Again squaring each of the equations of (20), we obtain in both cases (after simple transformations) *one and the same equation*:

$$32h^4-8(R^2+r^2)h^2+(R^2-r^2)^2=0 \tag{21}$$

Now what we have to do, essentially, is find the roots of this biquadratic equation and, in accord with the general procedure for solving irrational equations, choose those that satisfy the first equation of (20)—the roots for which $h \leqslant r$ and $4h^2-(R^2+r^2) \leqslant 0$—and those that satisfy the second equation of (20)—the roots for which $h \leqslant r$ and $4h^2-(R^2+r^2) \geqslant 0$.

This very long selection procedure in our specific case need not, it turns out, be performed if we note one obvious thing: we are not in the least interested in which root of equation (21) is a root of a particular equation of (20), because in either case that root leads to the required configuration. Therefore no selection procedure needs to be carried out and our problem has as many solutions as there are positive roots of the biquadratic equation (21) which satisfy the condition $h \leqslant r$.

Let us now study and solve the biquadratic equation (21). If the discriminant $D = 16(6R^2r^2-R^4-r^4) < 0$, then (21) has no real roots and the configuration we desire does not exist.

Let us determine for what relationship between R and r this occurs. To do so, we have to establish the relationship between R and r for which the inequality $6R^2r^2-R^4-r^4<0$ is valid. Denoting R^2/r^2 by ρ, we reduce this inequality to the form $\rho^2-6\rho+1>0$, whence it is clear that it occurs when $\rho < 3-2\sqrt{2}$ and when $\rho > 3+2\sqrt{2}$. Since we are only interested in values of $\rho > 1$ (because $R > r$, by hypo-

* Every positive root of the first equation of (19) corresponds to a configuration such that the square does not contain within it the centre O, and every positive root of the second equation of (19) is associated with a square that contains within it the centre O.

thesis), there remains only $\rho > 3+2\sqrt{2}=(1+\sqrt{2})^2$. Thus, *if* $R>(1+\sqrt{2})\,r$, *then the geometric configuration* envisioned by the problem *is impossible* and *the problem does not have a solution.*

In analogous fashion it is established that the inequality $D \geqslant 0$ holds true when $r < R \leqslant (1+\sqrt{2})\,r$, and the equation $D = 0$ occurs when $R=(1+\sqrt{2})\,r$. Denoting h^2 by z and thus reducing the equation (21) to a quadratic equation in z, we get

$$z_{1,2}=\frac{(R^2+r^2)\pm\sqrt{6R^2r^2-R^4-r^4}}{8} \tag{22}$$

Now from the Viète theorem it follows that both these roots are positive (for $D = 0$ the roots coincide).

Before computing the roots of (21) themselves, let us check to see if they lie in the domain of the original equations (19). To do this, we determine whether the inequalities $z_1 \leqslant r^2$, $z_2 \leqslant r^2$ hold; since $0 < z_2 \leqslant z_1$, it suffices to verify the inequality $z_1 \leqslant r^2$, that is,

$$\frac{R^2+r^2+\sqrt{6R^2r^2-R^4-r^4}}{8}\leqslant r^2$$

or

$$\sqrt{6R^2r^2-R^4-r^4}\leqslant 7r^2-R^2 \tag{23}$$

Since $D \geqslant 0$ in the case at hand, that is, $R \leqslant (1+\sqrt{2})r$, then $7r^2-R^2 \geqslant 7r^2-(1+\sqrt{2})^2r^2>0$, and so, squaring inequality (23), we get (after a few simple manipulations) the equivalent inequality $2\,(5r^2-R^2)^2 \geqslant 0$, which clearly holds. Thus, the roots of the biquadratic equation (21) (for $D \geqslant 0$) lie in the domain of the equations (19).

Now we recall that we are only interested in the positive roots of (21) and from (22) we find that *for* $r < R < (1+\sqrt{2})\,r$ *the problem has two solutions corresponding to the values*

$$h_{1,2}=\sqrt{\frac{(R^2+r^2)\pm\sqrt{6R^2r^2-R^4-r^4}}{8}}$$

and for $R=(1+\sqrt{2})r$ *it has a unique solution that corresponds to the value*

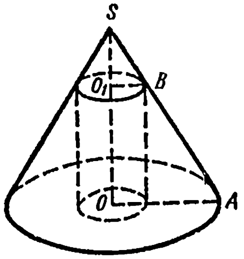
$$h=\sqrt{\frac{R^2+r^2}{8}}$$

Then, in the case of two solutions, the sides of the squares are equal to $2h_1$ and $2h_2$ and, in the case of the unique solution, the side of the square is equal to $2h$.*

* Note that the approach we discarded of a *separate* solution to each of the irrational equations of (19) would have led to a geometrically more interesting result. Namely, when $r<R<r\sqrt{5}$, both existing squares fail to contain the centre O; for $r\sqrt{5}<R<r(1+\sqrt{2})$, one of the squares contains the centre and the other does not; for $R=r\sqrt{5}$, one of the squares does not contain the centre, while the side of the other square passes right through this centre.

Algebra and trigonometry are widely used in the solution of *geometric problems involving maxima and minima.* In such problems, one ordinarily considers a solid (or figure) and seeks to find the dimensions so that some quantity associated with the solid assumes a maximum (or minimum) value. For an algebraic solution to such a problem, the procedure is, as a rule, to write down the *function* relating the quantity of interest to the dimensions of the solid and then to investigate the function. A geometric problem is thus reduced to an algebraic one, an investigation of the properties of functions.

Fig. 89



8. *The radius of the base of a right circular cone is R, its altitude is equal to H. Which cylinder inscribed in this cone has the largest lateral surface area?*

Let a cylinder with base radius r and altitude h be inscribed in the cone (see Sec. 3.8). From the similarity of the triangles AOS and BO_1S (Fig. 89) it follows that $r=R\,(H-h)/H$. The lateral surface area of the cylinder $s=2\pi rh$, or

$$s=\frac{2\pi R}{H}h\,(H-h)$$

This is the function connecting the quantity s, which we are interested in, with the altitude of the cylinder h, which we have at our disposal. From geometric considerations, the independent variable h varies in the interval $0<h<H$.

We now have to find the maximum of this function on the indicated range of the argument h. The function is a polynomial of degree two and so we shall proceed as one ordinarily does when seeking the maximum (or minimum) of a quadratic trinomial. Namely, isolating the perfect square, we rewrite the formula for s in the form

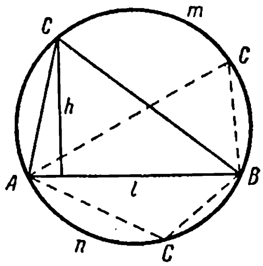
$$s=\frac{\pi RH}{2}-\frac{2\pi R}{H}\left(h-\frac{H}{2}\right)^2$$

From this expression we can see that the function $s=s\,(h)$ assumes a maximum value equal to $\frac{1}{2}\pi RH$ when $h=\frac{1}{2}H$. Since the value $h=\frac{1}{2}H$ lies in the interval $0<h<H$, the maximum value of the la-

teral surface area $s_{max}=\frac{1}{2}\pi RH$ is obtained when the altitude of the inscribed cylinder is equal to one half the altitude of the cone.

9. *Given on a circle of radius R two points A and B separated by a distance of l. What is the maximum value that the sum AC^2+BC^2 can assume if point C also lies on the circle?*

Fig. 90



Let C be an arbitrary point of the circle on which are given A and B (Fig. 90). If we denote angle ACB by α, then by the cosine law we have, from the triangle ABC,

$$AC^2+BC^2=l^2+2AC\cdot BC\cos\alpha$$

Hence, we have to determine the position of C on the circle for which the product $AC\cdot BC\cos\alpha$ assumes the *largest* possible value.

The chord AB divides the circle into two arcs: arc AmB, which exceeds the semicircle, and arc AnB, which is smaller than the semicircle.* It is readily seen that for any position of point C on the arc AmB (except for the endpoints A and B themselves of the arc) the angle α has one and the same value lying between 0° and 90° (because for any other point C the angle ACB is an inscribed angle intercepting the arc AnB, which is less than the semicircle), and for any position of C on the arc AnB the angle α has one and the same value lying in the interval from 90° to 180°. Since the cosine of an obtuse angle is negative, we can only be interested in points C on the arc AmB: in this case the product $AC\cdot BC\cos\alpha$ *is positive*.

Since, as we have already noted, the angle α is the *same* for any point C on the arc AmB, we then have to determine the position of C (on this arc) so that the product $AC\cdot BC$ is the largest possible one. Using formula $S=\frac{1}{2}AC\cdot BC\sin\alpha$ for the area S of the triangle ABC, we can write the equation $AC\cdot BC=2S/\sin\alpha$, whence it is evident that the product $AC\cdot BC$ takes on the largest value when C on arc AmB is such that the triangle ABC has the largest possible area.

---

* The reader will have no difficulty in considering the case when AB is the diameter of the circle, that is, when $l=2R$. If $l>2R$, then the configuration indicated in the problem is impossible.

However, by a familiar formula, $S=\frac{1}{2}lh$, where h is the length of the perpendicular dropped from C on the chord AB. For this reason, of all possible triangles ABC with vertex C on arc AmB the triangle in which this perpendicular is longest will have the largest area, and this perpendicular is longest when the point C is chosen as the midpoint of the arc AmB.

Thus, the sum AC^2+BC^2 assumes the maximum value when C *is located at the midpoint of the largest* of the two arcs into which the points A and B partition the circle. This maximum value is equal to $4R^2+2R\sqrt{4R^2-l^2}$.

10. *Required: to construct a kite in the form of a right prism having as base a right triangle with hypotenuse equal to* 50 cm. *The lateral surface area of this prism is equal to* 0.96 m². *What should the sides of the base of the triangle be so that the sum of the lengths of all edges of the prism is a minimum?*

Denote the legs of the base of the prism by x and y and its lateral edge by z. We then set up two relations connecting the dimensions of the prism:

$$\begin{gathered} x^2+y^2=0.25 \\ (x+y+0.5)z=0.96 \end{gathered} \tag{24}$$

We are interested in the smallest value of the sum $l=2(x+y+0.5)+$ $+3z$ of all edges of the prism.

The quantity l is a function of three variables x, y, z which are connected by the equations of (24). Substituting in place of the sum $x+$ $+y+0.5$ its expression in terms of z obtained from the second equation of (24), we can represent l as a function of *one* variable:

$$l=\frac{1.92}{z}+3z \tag{25}$$

We determine the minimum of this function. Applying the inequality between the arithmetic and the geometric mean (see Sec. 1.8), we can state that for arbitrary $z>0$

$$l=\frac{1.92}{z}+3z\geqslant 2\sqrt{\frac{1.92}{z}\cdot 3z}=4.8$$

In other words, the function (25) has a minimum value of 4.8. It attains this minimum value when $1.92/z=3z$, that is, for $z=0.8$ (we are naturally interested only in *positive* values of z).

There is a logical subtlety at this point. We proved that *function* (25), as a function of the variable z, assumes its smallest value 4.8 when $z=0.8$. But to be assured that this is the smallest value of the *geometric quantity* (the sum of the edges of the prism) we have to be sure that there actually *exists* a prism which satisfies the condition of the problem and which has a lateral edge $z=0.8$.

In other words, we still have to determine whether system (24) has a solution for $z=0.8$. If it does, then the corresponding prism (or prisms, if the system has several solutions) will be the solution to the problem. But if system (24) does not have a solution for $z=0.8$, then the sum of the lateral edges cannot be equal to 4.8 and we will have to reexamine the situation.

Substituting $z=0.8$ into (24), we get

$$x^2+y^2=0.25$$
$$x+y=0.7$$

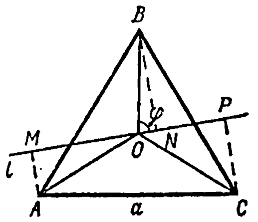
whence $x_1=0.4$, $y_1=0.3$, $x_2=0,3$, $y_2=0.4$. These two solutions geometrically correspond to one and the same prism.*

Thus, the sum of the lengths of all edges of the prism will be least if the legs of the triangle of the base are equal to 30 cm and 40 cm.

Though students make extensive use of trigonometry and algebra in solving computational problems in geometry, they rarely invoke algebraic and trigonometric methods to prove geometric facts, to find loci, or to perform constructions. Yet in all such problems the significance of such methods is hard to overestimate.

Here is a problem in which the proof of the assertion we want is obtained in a purely analytical fashion without any geometrical reasoning whatsoever. There are any number of instances when the simplest proof is obtained by *direct computation*.

Fig. 91



11. *An arbitrary straight line is drawn through the centre of an equilateral triangle in the plane of the triangle. Prove that the sum of the squares of the distances from the vertices of the triangle to this straight line is independent of the choice of the line.*

Suppose the straight line l passes through O, the centre of an equilateral triangle ABC (Fig. 91). Drop perpendiculars onto l from vertices

* One should not think that the remark concerning the logical subtlety was superfluous for the reason that everything was so simple. If the length of the hypotenuse was, say, 40 cm, then the corresponding system

$$x^2+y^2=0.16$$
$$x+y=0.7$$

would not have had any real solutions. But this does not yet mean that there is no prism satisfying the condition of the problem. To solve this more complicated problem we would have to seek, from among all z such that system (24) has a solution, the solution associated with the smallest possible value of the function (25).

A, B, and C. The feet of these perpendiculars will be denoted M, N and P respectively. Join O to the vertices of ABC and denote angle BOP by φ. Then $\angle AOM = \angle AOB - \angle BOM = 120° - (180 - \varphi) = \varphi - 60°$ and $\angle COP = 120° - \varphi$. Furthermore, let $OA = OB = OC = R$, then $BN = R \sin \varphi$, $AM = R \sin(\varphi - 60°)$, $CP = R \sin(120° - \varphi) = R \sin(\varphi + 60°)$. It is now easy to verify that

$$AM^2 + BN^2 + CP^2 = R^2 [\sin^2 \varphi + \sin^2(\varphi - 60°) + \sin^2(\varphi + 60°)] = \frac{3}{2} R^2 = \frac{1}{2} a^2$$

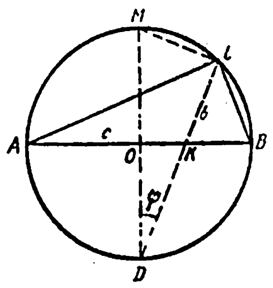
where a is a side of the given equilateral triangle.

In conclusion we give an example of the use of trigonometry and algebra in the solution of *construction problems*. Construction problems are often viewed by students as brain teasers. True, when worked by geometric tools alone they give rise to considerable difficulties. But trigonometry and algebra can substantially simplify the solution of such problems.

12. *Construct a right triangle using the hypotenuse c and the bisector b of the right angle.*

Suppose that the triangle ABC is the desired one (Fig. 92). Circumscribe about the triangle a circle with diameter equal to c. It is easy

Fig. 92



to demonstrate that the point D of intersection of the extension of the bisector CK with the circle is the endpoint of a diameter DM that is perpendicular to AB. It is now clear at once that the problem can be solved as soon as we find the line segment CD, which is what we will do.

Denoting angle CDM by φ, we can write (from the right triangle MDC) that $CD = c \cos \varphi$, and, from the triangle DKO, that $KD = c/(2 \cos \varphi)$ since $DM = 2\,DO = c$. Then set up the quadratic equation

$$c \cos \varphi - \frac{c}{2 \cos \varphi} = b$$

only one root of which

$$\cos \varphi = \frac{b + \sqrt{b^2 + 2c^2}}{2c}$$

is geometrically meaningful. From this we have

$$CD = c\cos\varphi = \frac{b}{2} + \frac{\sqrt{b^2+2c^2}}{2}$$

Now the construction is clear. Construct on $AB = c$ (diameter) a circle (with centre O) and locate the point D such that $DO \perp AB$. Then separately construct $a = CD$. To do this, first construct the line segment $c\sqrt{2}$ as the geometric mean of the line segments c and $2c$.* Then construct $x = \sqrt{b^2 + (c\sqrt{2})^2}$ as the hypotenuse of a right triangle with legs b and $c\sqrt{2}$. Finally $a = \frac{1}{2}(b + x)$. Obtain point C as the vertex of the right angle of the desired triangle by striking an arc equal to a with centre at point D.

Exercises

1. Given two sides b and c of a triangle and the bisector, of length l, of the angle between them. Compute the third side of the triangle.
2. In a cylinder the altitude h is equal to the diameter of the base circle. A point of the upper circle is connected with a point of the lower circle. The straight line joining these points forms an angle α with the base plane of the cylinder. Determine the shortest distance between this line and the axis of the cylinder.
3. In a triangle ABC, the tangents of the angles stand in the ratio: $\tan A : \tan B : \tan C = 1 : 2 : 3$. Find the ratio of the sines of these angles.
4. In an equilateral triangle ABC each of the sides is divided into three equal parts: side AB by points D and E (so that $AD = DE = EB$), side BC by points F and G (so that $BF = FG = GC$), side CA by the points H and I (so that $CH = HI = IA$). L, M, and N denote, respectively, the points of intersection of pairs of lines BI and CD, AF and CE, AG and BH. Find the ratio of the areas of the triangles LMN and ABC.
5. The volume of a regular many-sided pyramid is v and a side of the base is equal to a. Determine the angle of inclination of a lateral edge of the pyramid to the base.
6. Given sides a and b of a parallelogram. Determine the ratio of the volumes of the solids obtained by rotating the parallelogram about the sides a and b, respectively.
7. Given in a triangle the base a and adjoining angles α and $\alpha + 90°$. Determine the volume of the solid generated by rotating this triangle about the altitude dropped on side a.
8. The area of a triangle ABC satisfies the relation $S = a^2 - (b - c)^2$, where a, b and c are the sides of the triangle opposite the angles A, B and C, respectively. Find the angle A.
9. Given sides b and c of a triangle and angle A between them. The triangle rotates on an axis that does not intersect it, passes through vertex A, and forms equal angles with sides b and c. Find the volume of the solid of revolution.
10. The radius of a sector is equal to R, the radius of the circle inscribed in the sector is r. Compute the area of the sector.

---

* It is useful to bear in mind that if the line segments a and ka, where k is a numerical coefficient, are given, then to construct $x = a\sqrt{k}$ it is sufficient to construct the geometric mean of the given line segments $x = \sqrt{a \cdot ka} = a\sqrt{k}$. Incidentally, this device enables us to construct line segments of length $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$ and so on if unit length is given.

11. Given a triangle whose base is a and the angle at the vertex is α. A circle is drawn through the centre of another circle inscribed in the triangle and through the endpoints of the base. Find the radius of the first circle.

12. The base of a pyramid is a square. Two lateral edges are perpendicular to the plane of the base and the other two form with it angles equal to α. Determine the dihedral angle between these latter lateral faces.

13. Inscribed in a right triangle with legs b and c is a square having a common right angle with the triangle (that is, two sides of the square lie on the legs, and one vertex on the hypotenuse). Find the area of the square.

14. The base angle of an isosceles triangle is α. The altitude dropped to the base is greater by m than the radius of an inscribed circle. Determine the base of the triangle and the radius of a circumscribed circle.

15. Two chords of a circle of radius R intersect at right angles. The portions of one chord are as $4:5$, those of the other as $5:16$. Find the lengths of the four arcs into which the endpoints of the chords divide the circle.

16. Inscribed in a circle are an isosceles triangle and a trapezoid. The lateral sides of the triangle are parallel to the nonparallel sides of the trapezoid. One of the bases of the trapezoid is the diameter of the circle. Find the altitude of the trapezoid if the midline is l and the area of the triangle is S.

17. Given a circle of radius r. From an exterior point A are drawn a tangent AP and a secant AQ passing through the centre of the circle, $AQ=2AP$. Find the radius of the circle that touches the secant, the tangent line exterior to AP and the radius (drawn to point P) of the given circle.

18. Given a circle of radius R. A chord AB is drawn through point M of the diameter at an angle φ to the diameter, $BM : AM=p : q$. Through point B is drawn a chord BC perpendicular to the given diameter, and point C is joined to point A. Find the area of the triangle ABC.

19. Given a circle of radius 1 with centre at point O and a straight line tangent to the circle at point E. A point M is taken on the circle. Find the radius of the circle that is tangent to the given circle at M and to the given line if the angle EOM is equal to φ, $0° < \varphi < 90°$.

20. The distance between two parallel lines is 1. A point A lies between the lines at a distance a from one of them. Find the length of a side of an equilateral triangle ABC, vertex B of which lies on one of the parallel lines and vertex C on the other.

21. An isosceles trapezoid with base angle 60° is such that it is possible to inscribe in it two circles tangent to each other, each one of which is tangent to the bases of the trapezoid and to one of the nonparallel sides. Each nonparallel side of the trapezoid is equal to 2. Find the area of the trapezoid.

22. In a trapezoid with bases a and b a straight line parallel to the bases is drawn through the point of intersection of the diagonals. Find the portion of this line lying between the nonparallel sides of the trapezoid.

23. Given angles α, β, γ of triangle ABC. Let A', B', C' be points of intersection of the bisectors of the interior angles of ABC with a circle circumscribed about the triangle. Find the ratio of the area of triangle ABC to the area of triangle $A'B'C'$.

24. Given the altitudes $AA'=h_a$ and $BB'=h_b$ of a triangle ABC and the length $CD=l$ of the bisector of angle C. Find angle C.

25. Inscribed in an isosceles triangle is a square of unit area whose side lies on the base of the triangle. Find the area of the triangle if the centroids of the triangle and the square are known to coincide.

26. Compute the area of the common portion of two rhombuses: the diagonals of the first are 2 and 3, and the second is obtained by turning the first through 90° about its centre.

27. Given in an isosceles triangle a lateral side b and the base angle α. Compute the distance from the centre of the inscribed circle to the centre of the circumscribed circle.

28. From point A exterior to a circle of radius r is drawn a secant that does not pass through the centre O of the circle. Let B and C be points at which the secant cuts the circle. Find the value of $\tan\left(\frac{1}{2}\angle AOB\right)\cdot\tan\left(\frac{1}{2}\angle AOC\right)$ if $OA=a$.

29. A square is inscribed in a circular sector bounded by radii OA and OB with central angle α ($\alpha<\pi/2$) so that two adjacent vertices lie on radius OA, the third vertex on the radius OB, and the fourth vertex on the arc AB. Find the ratio of the areas of the square and the sector.

30. Two equal spheres are tangent to each other and to the faces of a dihedral angle. A third sphere of smaller radius is also tangent to the faces of the dihedral angle and to both given spheres. Given the ratio m of the radius of the smaller sphere to the radius of one of the equal spheres. Find the dihedral angle. Over what interval can m vary?

31. The base angle of an isosceles triangle ABC is α ($\alpha>45°$), the area of which is S. Find the area of a triangle whose vertices are the feet of the altitudes of the triangle ABC.

32. In a triangle ABC, $\angle A=\angle B=\alpha$, $AB=a$; AH is the altitude, BE a bisector (point H lies on side BC, point E on AC). Point H is joined by a line segment to point E. Find the areas of the triangle CHE.

33. A point M is taken inside an angle AOB; $\angle MOA=\alpha$, $\angle MOB=\beta$, $OM=a$ ($\alpha+\beta<\pi$). Find the radius of a circle passing through M and cutting off, on sides OA and OB of the given angle, chords of length $2a$.

34. Given an angle AOB equal to α ($\alpha<\pi$). A point C is taken on side OA, a point D on side OB, $OC=a\neq 0$, $OD=b\neq 0$. A circle is constructed tangent to the side OA at C and passing through D. Let this circle cut side OB a second time, at point E. Compute the radius r of the circle thus constructed and the length of the chord DE.

35. A pedestal in the form of a rectangular parallelepiped is to be cut out of granit. The altitude must be equal to the diagonal of the base, and the area of the base must be 4 square metres. For which sides of the base will the total surface area of the pedestal be a minimum?

36. A box is to be built in the form of a rectangular parallelepiped with base area 1 cm^2. The sum of the lengths of all edges of the parallelepiped must equal 20 cm. For which box dimensions is the total surface area a maximum?

37. A hole is cut in a square sheet of plywood with side equal to 10 units of length. The hole is in the form of a rectangle whose diagonal is equal to 5 units of length. The opening is bordered with a fine wire frame. A unit of area (that is, the area of a square with side equal to unit length) of the plywood weighs 2 grams, a unit of length of the wire weighs 7 grams. What must the sides of the rectangle be so that the weight of the resulting sheet with the bordered opening is a maximum?

38. In a triangle ABC, a straight line is drawn from vertex A to intersection with side BC at point D which lies between B and C; $CD : BC=\alpha$ (where $\alpha<1/2$). A point E is taken on side BC between B and D and through it is drawn a line parallel to AC and intersecting AB at F. Find the ratio of areas of the trapezoid $ACEF$ and the triangle ADC if we know that $CD=DE$.

39. A point D is taken on side AB of a triangle ABC between A and B so that $AD : AB=\alpha$ (where $\alpha<1$). Take a point E on BC between B and C so that $BE : BC=\beta$ (where $\beta<1$). Through point E draw a line parallel to AC and intersecting side AB at F. Find the ratio of the areas of the triangles BDE and BEF.

40. Indicate possible types of triangles in which the sides constitute a geometric progression and the angles an arithmetic progression.

41. Will it suffice to buy 4000 pieces of glazed tile of size 10 cm by 10 cm to face the walls of the above-water portion of a swimming pool having the shape of a rhombus of area 450 square metres with sides 0.5 metre above the water level?

42. The base of a quadrangular pyramid is a rectangle. The pyramid has all lateral edges equal and altitude $\sqrt{2}$ cm. A bug is crawling along the edges of the pyramid at a rate of 1 cm/sec. Will 2 seconds be sufficient for it to go down from the vertex of the pyramid along a lateral edge to a vertex of the base if it can traverse the perimeter of the base in 8 seconds?

43. Find the cosine of the base angle α of an isosceles triangle if it is given that the point of intersection of the altitudes lies on a circle inscribed in the triangle.

44. Drawn in a triangle ABC are: bisector AD of angle BAC and bisector CF of angle ACB (point D lies on side BC, point F on side AB). Find the ratio of the areas of the triangles ABC and AFD if it is given that $AB=21$, $AC=28$, $CB=20$.

45. Inscribed in a circle is an isosceles triangle ABC in which $AB=BC$ and $\angle ABC=\alpha$. From vertex A draw a bisector of the angle BAC to intersection with side BC at D, and to intersection with the circle at point E. Vertex B is joined by a straight line to point E. Find the ratio of the areas of the triangles ABE and BDE.

46. In a trapezoid $ABCD$ the angles at the large base a are equal to α and β and the altitude of the trapezoid is h. Let O_1, O_2, O_3, O_4 be centres of circles circumscribed, respectively, about triangles ABC, BCD, CDA, DAB. Find the area of the rectangle $O_1O_2O_3O_4$.

47. In a right triangle, the ratio of the product of the lengths of the bisectors of the interior acute angles to the square of the length of the hypotenuse is equal to 1/2. Find the acute angles of the triangle.

48. From an endpoint of a diameter of a sphere draw a chord so that the surface generated by a rotation about this diameter divides the volume of the sphere into two equal parts. Determine the angle between the chord and the diameter.

49. In a convex quadrangle $ABCD$ the bisector of the angle ABC intersects side AD at point M, and the perpendicular dropped from vertex A to side BC cuts BC at point N so that $BN=NC$ and $AM=2\cdot MD$. Find the sides and the area S of quadrangle $ABCD$ if the perimeter is equal to $5+\sqrt{3}$, $\angle BAD=90°$, $\angle ABC=60°$.

50. The base of a quadrangular pyramid $OABCD$ is a trapezoid $ABCD$ and the lateral faces OAD and OBC are perpendicular to the plane of the base. Find the volume of the pyramid knowing that $AB=3$, $CD=5$, the area of face OAB is equal to 9 and the area of face OCD is equal to 20.

3.4 Lines and planes in space

This section of solid geometry is of fundamental significance for a thorough mastering of all subsequent material and the development of spatial imagery in general. The theorems are in themselves simple but require a certain level of logical culture, the ability to reason rigorously on the basis of given definitions, and to reduce the proof of any theorem to the application solely of the starting axioms and earlier proved theorems. This in turn requires a very precise knowledge of the initial definitions and theorems. Unfortunately, the attitude of students is often somewhat slipshod, they believe that geometric intuition will always help them to get at the proper definition and statement of the axioms. As a result, in the best of cases we hear equivalent formulations, which lead to unexpected (for the students) complications in the formulations and proofs of theorems.

Crucial to the study of this section is a *precise understanding and firm memorizing* of definitions. The parallelism and perpendicularity of lines and planes, the angle between skew lines and the distance between them, the angle between a line and a plane, the angle between two planes—all these notions must be firmly rooted in the mind of the student.

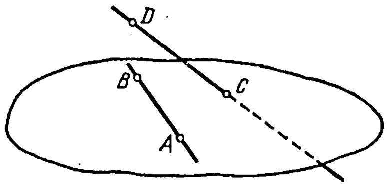
True, one should not go to extremes. Firstly, one should not learn by rote the definitions without grasping their geometric meaning or visualizing an associated geometric configuration. Secondly, (this is typical and therefore more dangerous) the student must not confine himself to geometric imagery alone and dismiss precise definitions.

To illustrate, every one can visualize parallel lines in space, yet many students start out by proving the "completely obvious" assertion that a plane can be drawn through two parallel straight lines. They lose sight of the fact that the existence of such a plane is part of the very *definition* of parallel lines.

Another similar case has to do with proving the *existence* of skew lines. There are some who confine themselves to pointing to the floor and the ceiling. It is of course a good thing to have ready such a neat example of skew lines but this does not relieve them of the necessity of giving a rigorous proof of their existence.

Take any three points A, B, C and, exterior to the plane defined by them, a fourth point D. Then the straight lines AB and CD are skew lines. Indeed, if they were in one plane, then all their points would be in that plane as well, but then that would contradict the choice of point D (Fig. 93).

Fig. 93



Then there are cases where the student, since he doesn't know the proper definition, makes do with such erroneous analogies and fantastic inventions as "a straight line is parallel to a plane if it is parallel to any line in the plane"; "a straight line is perpendicular to a plane if it is perpendicular to some straight line in the plane"; "the angle between a straight line and a plane is the angle which this line forms with the straight lines of the given plane", and others in the same vein. Their absurdity becomes obvious as soon as we attempt to picture the geometric meaning of such definitions.

At the same time, there is no necessity to memorize definitions word for word; certain departures from the statements given in standard textbooks are permissible. It is best, however, to keep such deviations within bounds. For instance, rather often the student takes the *criterion* for the perpendicularity of a straight line and a plane for the *definition* of the perpendicularity of a straight line and plane, Unfortunately, they fail to grasp the fact that in this new definition, the earlier *definition* becomes a *theorem* that has to be proved. So here too such loose handling of definitions can lead to added complications.

Many mistakes are made due to *incomplete* formulations of definitions and, particularly, theorems. Here is a typical answer to a request, to formulate the criterion of parallelism of two planes: "If two straight lines lying in a plane are parallel to two straight lines lying in the other plane, then the planes are parallel." Just a single word was left out; the student should have said, two *intersecting* straight lines. So the student's statement of the theorem was incorrect and the answer cannot be considered exact.

One of the most important notions dealt with in this section is that of the *angle between skew lines*. Let us recall it. To construct an angle between skew lines one usually does as follows: he takes an arbitrary point in space and draws through it two straight lines parallel to the given lines. The angle between the constructed skew lines is, by definition, the angle between the given skew lines.

This definition may give rise to a natural question: does not the angle between the skew lines depend on the point chosen as its vertex? It turns out that the choice of the point does not affect the size of the angle. To justify this assertion, it is necessary to refer to the familiar criterion of parallelism of two planes and to the theorem on angles with parallel sides in space.

The concept of an angle between skew lines can, when systematically employed, simplify many definitions and theorems. For example, it is possible at once to define *perpendicular lines* as lines the angle between which is a right angle, irrespective of whether the lines lie in one plane or not. In turn, this definition greatly simplifies the solution of many problems.

In the same way it is easy to see that a *line perpendicular to a plane is perpendicular to any line in the plane*, including such that does not pass through the point of intersection of the given straight line and the plane. This follows directly from the definition of a straight line perpendicular to a plane and from the definition of perpendicular lines.

Of particular interest in this respect is the so-called *strong criterion of the perpendicularity* of a straight line and a plane: *if a straight line is perpendicular to two arbitrary nonparallel (intersecting) straight lines lying in some plane, then it is perpendicular to the plane itself.*

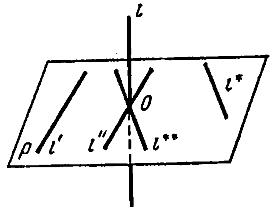
This statement, unlike the ordinary one, does not require that two straight lines necessarily pass through the point of intersection of the given straight line and the plane. What would seem to be a minor detail is seen to play an essential role in problem solving.

Proof of the strong criterion is exceedingly simple (Fig. 94). If the given straight line l is perpendicular to two intersecting straight lines l' and l^* of plane P, then it is perpendicular to the parallel lines l'' and l^{**} drawn through the point O of its intersection with the plane P.

Hence, the line l, by the ordinary criterion for the perpendicularity of a straight line and a plane, is perpendicular to the entire plane P. The proof is complete.*

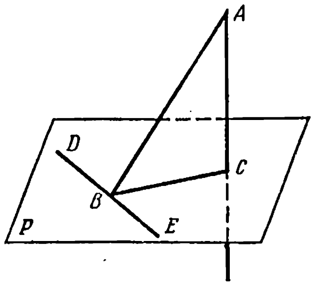
The foregoing strong criterion permits, say, reducing to a few words the proof of the so-called theorem on three perpendiculars, which, pro-

Fig. 94



perly speaking, is now merely a special case. Indeed (Fig. 95), let AC be perpendicular to plane P, let AB be inclined, and let BC be the projection of the inclined line on the P plane. If DE is drawn in the P plane perpendicular to AB, then DE is perpendicular to the plane ACB

Fig. 95



(since $DE \perp AC$ and $DE \perp AB$), and so DE is perpendicular to any straight line in this plane, in particular, to the projection BC of the inclined line. Conversely, if DE is drawn in the P plane perpendicular to the projection BC, then again DE is perpendicular to the plane ACB, and therefore $DE \perp AB$.

Note besides that the use of the notion of the angle between skew lines permits us, in the theorem on three perpendiculars, not to assume that a straight line in a given plane must necessarily pass through the foot of the inclined line.

The concepts we consider here are particularly important for a proper understanding of geometric facts when solving a variety of problems involving pyramids. The following theorem, for example, is very useful. It has to do with an arbitrary triangular pyramid (Fig. 96): *the altitude SH of a triangular pyramid $SABC$ passes through the altitude AD of the base if and only if the lateral edge SA is perpendicular to the*

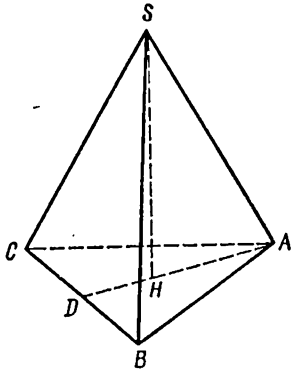
---

* Quite naturally, the fact that a straight line is perpendicular to two *parallel* lines in a plane does not imply the perpendicularity of the straight line to the plane itself (give an example!).

base edge BC. It is easy to see that this assertion is simply a different formulation of the theorem concerning three perpendiculars: SH and SA are the perpendicular and inclined lines, respectively, to the plane of the base ABC, and BC is the third straight line.

The next problem, which many students find particularly difficult, is solved at once with the aid of this theorem.

Fig. 96



1. *In a triangular pyramid, the altitude drawn from the vertex falls in the point of intersection of the altitudes of the triangle lying in the base. Show that all altitudes of the pyramid dropped from the vertices of the base on the lateral faces possess this property.*

Indeed, since the altitude of the pyramid passes through all altitudes of the base, it follows that each lateral edge of the pyramid is perpendicular to the opposite edge of the base. By the same theorem, only in reverse, any altitude of the pyramid has this property.

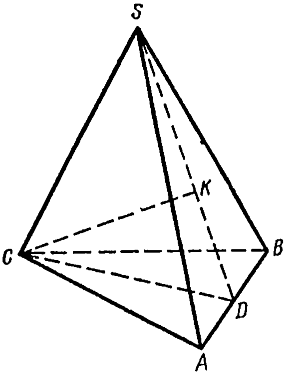
From this same theorem it follows, for example, that the lateral edges of a *regular* triangular pyramid are perpendicular to the opposite edges of the base. This statement may for instance be needed in the construction of the plane angle of the dihedral angle at the lateral edge of the pyramid. Students often suggest the following construction: "We pass a plane through a side of the base perpendicular to a lateral edge and then the angle obtained in the section will be the required angle." In principle the idea is correct but there is one very essential detail lacking—why such a plane can be drawn; because in the case of arbitrary skew lines it is not, generally speaking, possible to draw such a plane. It is possible when and only when the lines are perpendicular (prove this!). For this reason, the construction is indeed possible in a *regular* triangular pyramid.

We note another typical mistake. Many students drop a perpendicular from a vertex of the base of a regular triangular pyramid on a lateral face and, without any hesitation, consider that this altitude falls on the altitude of the lateral face. This assertion is of course true but it requires justification. It can readily be seen that it follows from the theorem we formulated above and also from the result of Problem 1. However, when solving a problem in which this statement is used, one should not rely on the indicated theorem (or, all the more so, on Problem 1) but must provide a separate independent proof.

The most unpromising is the following "brute force" approach to a proof of this statement: draw the altitude SD of the lateral face ASB of the regular pyramid $SABC$, drop the altitude CK of the pyramid from vertex C of the base on face ASB and prove that SD and CK intersect (Fig. 97). Unfortunately many students reason in this fashion, whereas it is much simpler to do otherwise.

Let $SABC$ be a regular pyramid and CK a perpendicular dropped from vertex C on face ASB (Fig. 97). Join K and D, the midpoint of

Fig. 97



edge AB. The line DC is perpendicular to AB, being the median of the equilateral triangle ABC. The line CK is perpendicular to the plane ASB and hence to any straight line of this plane, in particular, to edge AB. Hence, edge AB is perpendicular to the plane DCK since it is perpendicular to two intersecting straight lines CD and CK lying in the plane. Therefore, AB is perpendicular to any straight line in the plane DCK, to DK for one.

Thus, $DK \perp AB$ and since D is the midpoint of the base AB of the isosceles triangle ASB, the point K lies on the altitude SD of this triangle.

The strong criterion of perpendicularity is a good worker in many problems where for given dihedral angles it is required to construct plane angles.

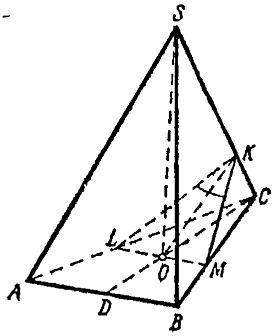
2. *A perpendicular equal to p is drawn from the foot of the altitude of a regular triangular pyramid to a lateral edge. Find the volume of the pyramid if the dihedral angle between its lateral faces is equal to 2α.*

Let $SABC$ be the given pyramid (Fig. 98), OK the perpendicular equal to p, and let the dihedral angle at edge SC be equal to 2α.

Let us first construct the plane angle of this dihedral angle. It is natural to take point K for the vertex of the plane angle. It is then necessary, in the planes of the faces ASC and BSC, to draw perpendiculars KL and KM from K to edge SC.

At this point, many students at the examination made mistakes. Some figured that these perpendiculars must necessarily pass through points A and B, others believed the points L and M to be the midpoints of sides AC and BC of the triangle ABC. Neither view is correct. The reasoning should be as follows.

Fig. 98



The plane containing KL and KM must be perpendicular to edge SC (since $SC \perp KL$ and $SC \perp KM$). To find this plane that is perpendicular to SC, it suffices to find any two nonparallel lines perpendicular to SC. One such line is given; this is OK, since $OK \perp SC$. For the second such line we can take edge AB. Indeed, SO is a line perpendicular to plane ABC, SC is an inclined line, CO is its projection; but $AB \perp CO$ and, hence, by the theorem on three perpendiculars, $AB \perp SC$.

We have thus found two straight lines, OK and AB, which are perpendicular to SC. But unfortunately they do not intersect, they are skew lines. However, there is a way out of this situation: through O draw $LM \parallel AB$. Then LM is also perpendicular to SC, and LM and OK then intersect. Consequently, SC is perpendicular to the plane in which these lines lie, that is, to the plane of the triangle KLM. But then $SC \perp KL$ and $SC \perp KM$, that is, $\angle LKM$ is the plane angle of the dihedral angle at the edge SC. Thus, $\angle LKM = 2\alpha$, and points L and M of perpendiculars KL and KM (to edge SC) are located on the sides AC and BC of triangle ABC in such a manner that LM passes through the centre of triangle ABC parallel to side AB.

Only computations remain now. Having proved that $\angle OKM = \angle OKL = \alpha$, we find $OM = p \tan \alpha$, and so (from the right triangle OCM) we have $OC = p\sqrt{3} \tan \alpha$. But OC is the radius of a circle circumscribed about the equilateral triangle ABC; hence, $AB = 3p \times \tan \alpha$, and the area $S_{\triangle ABC}$ of the base of the pyramid follows at once.

On the other hand, from the right triangle KOC

$$\sin(\angle OCK) = \frac{OK}{OC} = \frac{1}{\sqrt{3}} \cot\alpha$$

Since $\angle KOS = \angle OCK$ (as acute angles with perpendicular sides), the altitude of the pyramid is

$$SO = \frac{p}{\cos(\angle OCK)} = \frac{p\sqrt{3}}{\sqrt{3-\cot^2\alpha}}$$

Thus the sought-for volume is

$$V = \frac{1}{3} SO \cdot S_{\triangle ABC} = \frac{9}{4} p^3 \frac{\tan^2\alpha}{\sqrt{3-\cot^2\alpha}}$$

3. *The base of a pyramid is an equilateral triangle with side a. One of the faces of the pyramid is perpendicular to the plane of the base. This face is an isosceles triangle with side $b \neq a$. Find the area of a section through the pyramid that is square.*

Let $KLMN$ be the square stipulated in the problem (Fig. 99). We do not as yet make any assumptions as to which particular lateral face, ASB, BSC or ASC, is perpendicular to the plane of the base.

Fig. 99

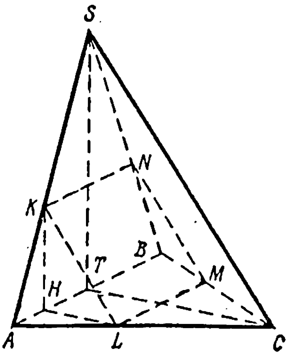
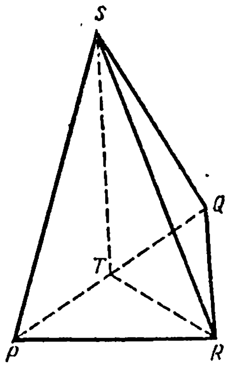


Fig. 100



Since $KN \parallel LM$, it follows that KN is parallel to the plane of the base ABC, and so the plane ASB that passes through KN intersects the plane ABC along the edge AB which is parallel to KN. The proof that $SC \parallel KL$ is analogous.

From this it follows that, first, the plane of the section is parallel to the skew edges AB and SC and, second, these edges must be mutually perpendicular since the (right) angle LKN of the square $KLMN$ is the angle between these skew lines.

Now notice that the pyramid $SABC$ has only one pair of mutually perpendicular skew edges. We will prove this on another drawing (Fig. 100).

Let $SPQR$ be the given pyramid in which the face PSQ is perpendicular to the base PQR. Draw the altitude ST of the pyramid. It is obvious that ST lies in the plane of the face PSQ and is at the same time the altitude of the isosceles triangle PSQ and, hence, the median as well, since RT is the altitude of the equilateral triangle PQR. By the theorem on three perpendiculars, the inclined line SR is perpendicular to PQ, which is the line perpendicular to its projection PT. On the other hand, the line in plane PQR perpendicular to edge SQ must be perpendicular to its projection PQ. But PR does not have this property, and so edges SQ and PR are not perpendicular. Similarly, edges SP and QR are not perpendicular. Thus, only edges SR and PQ are mutually perpendicular. These are: the lateral edge not lying in the face perpendicular to the base, and the edge of the right dihedral angle.

Returning to Fig. 99 we conclude that it is precisely AB that is the edge of the right dihedral angle, that is, face ASB is perpendicular to the plane of the base.

To carry out the computations, draw $KH \perp AB$ and join H and L. The line KH is perpendicular to HL, since it lies in one of two mutually perpendicular planes and is perpendicular to the line of their intersection. Also, draw the altitude ST of the isosceles triangle ASB and the altitude CT of the base ABC. Denote the side of the square $KLMN$ by x. Then

$$ST=\sqrt{b^2-\frac{a^2}{4}},\quad AT=\frac{a}{2},\quad AH=\frac{a-x}{2}$$

Since $\triangle AKH$ is similar to $\triangle AST$ and $\triangle ALH$ is similar to $\triangle ACT$, it follows that

$$KH=\frac{a-x}{a}\sqrt{b^2-\frac{a^2}{4}},\quad HL=\frac{a-x}{a}\sqrt{a^2-\frac{a^2}{4}}$$

and, by the Pythagorean theorem, from the right triangle KHL

$$x^2=\left(\frac{a-x}{a}\right)^2\left(\frac{a^2}{2}+b^2\right)$$

whence, after extracting the root (with allowance made for $a>x$) and performing the necessary manipulations, we get the side x of the square and then its area:

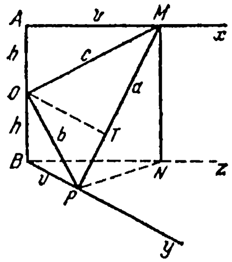
$$S=\left(\frac{a\sqrt{2a^2+4b^2}}{2a+\sqrt{2a^2+4b^2}}\right)^2$$

Of great importance in solving problems of solid geometry is a properly constructed drawing. If it is done properly, the solution is frequently obvious at a glance.

However, a plane representation of spatial configurations is only possible with certain distortions, and therefore the drawing, no matter how carefully made, will require some effort to grasp properly, necessitating that the student indicate right angles which appear acute in the drawing, skew lines which in the drawing intersect, etc.

Every fact, no matter how obvious it looks on paper, must be justified by rigorous logic so as to avoid the confusion caused by the peculiarities of a plane representation of spatial configurations. This is particularly true of problems involving proof. Moreover, in these problems the particularities of representation (distortion of angles and the like) frequently obscure the true picture and complicate the proof.

Fig. 101



4. *Given in space two rays Ax and By not lying in the same plane and forming an angle of $90°$; AB is their common perpendicular. Consider point M on Ax and point P on By such that $2AM \cdot BP = AB^2$. Prove that the distance from the midpoint O of AB to the line MP is equal to $AB/2$.*

Let Ax and By (see Fig. 101) be the rays given in the problem. In our drawing they will appear as two rays of intersecting straight lines, but we will bear in mind that this is merely an illusion. The rays do not lie in one plane.

The following construction will be found useful in many problems involving skew lines. Draw through a point B ray $Bz \parallel Ax$ and lay off line segment $BN = AM$. By the definition of the angle between skew lines AM and BP, we have $\angle PBN = 90°$, which means PBN is a right triangle with hypotenuse PN. It is well to stress this fact because some students fell into the trap of the drawing and regarded BPN as the right angle.

By the Pythagorean theorem, from the right triangle we find $PN^2 = BN^2 + BP^2 = u^2 + v^2$ (the designations are as given in Fig. 101). The line MN is perpendicular to the plane yBz, and so from the right triangle PNM we have $PM^2 = 4h^2 + u^2 + v^2$. But it is given that $(2h)^2 = 2uv$, hence $PM^2 = (u+v)^2$, or $PM = u + v$.

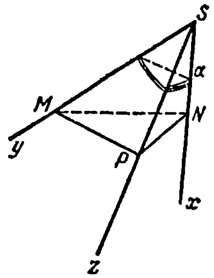
Draw the altitude OT of triangle POM, the length of which is the desired distance. When applying the Pythagorean theorem to the right triangles PTO and MTO, we can write the equation $b^2 - PT^2 = c^2 - (u + v - PT)^2$. Noting that $b^2 = u^2 + h^2$ (from the right triangle OBP) and $c^2 = v^2 + h^2$ (from the right triangle OAM), we get the equation $2u(u+v) = 2 \cdot PT \cdot (u+v)$ or $PT = u$. Hence $OT^2 = b^2 - PT^2 = h^2$ or $OT = h = AB/2$.

Ordinarily, problems involving the computation of polyhedral angles cause some difficulty that is due largely to an insufficiently developed geometrical imagination and an inability to make a proper drawing, whereas all the elements of a trihedral angle can rather easily be determined with the aid of trigonometry by proceeding from simple geometric reasoning.

5. *Let the acute angles A, B and C be the plane angles of a trihedral angle. Prove that if the dihedral angle opposite the plane angle C is a right angle, then* $\cos C = \cos A \cdot \cos B$.

Let $Sxyz$ be a trihedral angle (Fig. 102) in which $\angle xSy = A$, $\angle zSx = B$, $\angle ySz = C$, and the plane of the face zSx is perpendicular to the plane of face ySx. Draw plane MNP perpendicular to Sx;

Fig. 102



then $MN \perp Sx$, $PN \perp Sx$, $\angle MNP = 90°$. It may be noted that the remaining angles of the triangle MNP, namely $\angle PMN$ and $\angle MPN$, *are not* the plane angles of the dihedral angles at the edges Sy and Sz.

Denoting the length of SN by a, we get, from the right triangles PSN and MSN,

$$PN = a \tan B, \quad MN = a \tan A,$$
$$PS = \frac{a}{\cos B}, \qquad MS = \frac{a}{\cos A}$$

We now calculate in two ways (from the right triangle MNP and from the oblique triangle MSP) the length of MP and equate the results:

$$a^2 \tan^2 B + a^2 \tan^2 A = \frac{a^2}{\cos^2 B} + \frac{a^2}{\cos^2 A} - \frac{2a^2 \cos C}{\cos B \cos A}$$

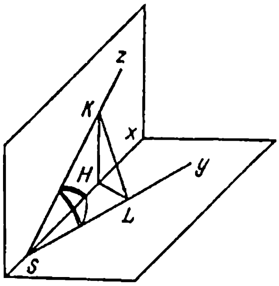
Cancelling out a^2 and performing the obvious manipulations, we get the relation required.

Note that both Fig. 102 and the foregoing solution were carried through under the assumption that all angles A, B, C are *acute*. However, the formula which we derived holds true without any supplementary assumption. We leave it to the reader to consider all possible cases and to justify the formula.

Note that if we depict the given trihedral angle not vertex upwards as in Fig. 102 (and as most students do), but more conveniently, the solution may be obtained in a much simpler fashion. Indeed, the right dihedral angle is more reasonably depicted as one ordinarily represents a pair of mutually perpendicular planes (Fig. 103).

Let Sx be the edge of the right dihedral angle, the edges Sy and Sz lie in different planes of this angle. From an arbitrary point K of edge Sz drop perpendiculars KH and KL, respectively, on Sx and Sy.

Fig. 103



Since the dihedral angle at edge Sx is a right angle, the line KH is perpendicular to the whole plane xSy and, in particular, $KH \perp SL$. Hence, by the theorem on three perpendiculars, $HL \perp SL$.

From the right triangle KLS we now obtain $SL = SK \cos C$, and from the right triangles SLH and SHK we have $SL = SH \cos A = SK \cos B \cos A$. Comparing these two expressions for SL, we see that the required equation is valid.

6. *The plane angles of a trihedral angle are equal to* 45°, 45° *and* 60°. *Through the vertex is drawn a line perpendicular to one of the faces whose plane angle is equal to* 45°. *Find the angle between this line and the edge (of the trihedral angle) that does not lie in the indicated face.*

In constructing the drawing we will again deviate from the standard procedure. Since we are dealing (as stated) with a perpendicular to two straight lines, that is, to the plane defined by them, it is convenient to represent this plane horizontally and the perpendicular vertically.

Thus, let S be the vertex of the trihedral angle (Fig. 104); from it issue the edges Sl and Sm which form an angle of 45°. Sp is a perpendicular to these lines; Sn is the third edge of the trihedral angle, and $\angle nSm = 45°$ while $\angle nSl = 60°$.

On Sn take a line segment SA of length a and from point A drop a perpendicular AK on the plane mSl and perpendiculars AB and AC on the lines Sl and Sm, respectively. Joining points B and C to K, by the theorem on three perpendiculars we find that BK and CK are perpendicular, respectively, to Sl and Sm.

From the right triangles ACS and ABS we readily find $SC = a\sqrt{2}/2$, $SB = a/2$ However if we join B and C and consider the triangle CBS, then by the cosine law we can determine $CB = a/2$, or $CB = SB$. Since $\angle CSB = 45°$, the isosceles triangle CBS has a *right* angle at the vertex B. But it was already noted that $\angle KBS = 90°$, and we quite unexpectedly come up with a contradiction.

Very few of the students were able to find a way out of this contradiction, although the matter at hand is not so complicated. From the foregoing reasoning it follows that the configuration represented in Fig. 104 actually does not occur (see Sec. 3.2): point K must lie on

Fig. 104

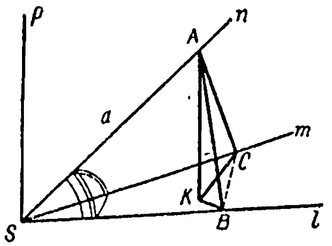
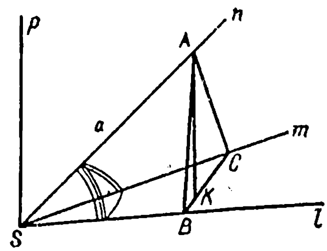


Fig. 105

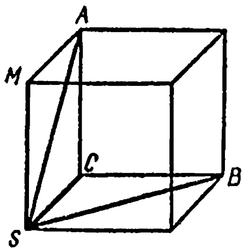


line CB (Fig. 105). But as soon as we put K on CB, we immediately reveal another contradiction: since $CK \perp Sm$ and $BK \perp Sl$, there are two right angles in triangle SBC. How do we resolve this contradiction? From right triangles ACS and ABS we find $AC = a\sqrt{2}/2$, $AB = a\sqrt{3}/2$. Recalling that $CB = a/2$, we easily find, by the converse of the Pythagorean theorem, that ABC is a right triangle with right angle ACB.

Thus, neither does the configuration given in Fig. 105 occur. Points K and C coincide, which is to say that the perpendicular AC to the straight line Sm is at the same time perpendicular to the entire plane mSl.

But in this case it is obvious that $AC \parallel Sp$, that is, the lines Sp and AC lie in one plane. Since $\angle ASC = 45°$, the sought-for angle between Sp and Sn, or angle nSp, is equal to 45°.

Fig. 106



It is interesting to note that with a sufficiently developed geometrical imagination it is possible to find a very short, purely geometric, solution to this problem that does not require any computations at all. Consider the cube (Fig. 106). If one notes that the trihedral angle $SABC$ is precisely the trihedral angle mentioned in the problem and the edge SM of this cube is perpendicular to face BSC, then, clearly, angle MSA is the desired angle and it is equal to 45°.

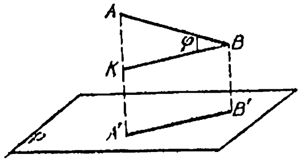
We conclude this section with an investigation of the concept of *orthogonal projection*.

If in space a line segment AB is orthogonally projected on some plane P, then the length of the projection, $A'B'$, is connected with the length of AB by the relation

$$A'B' = AB \cos \varphi \tag{1}$$

where φ is the angle between AB and the P plane (Fig. 107).* If through B we draw BK parallel to the projection $A'B'$, then this formula follows in obvious fashion from a consideration of the right triangle AKB because angle ABK is the angle between the straight line AB and the P plane.

Fig. 107



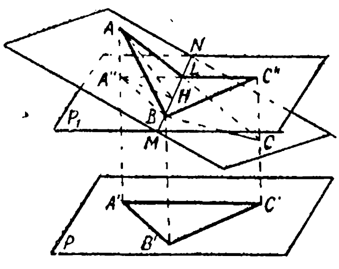
Formula (1) shows that the length of the projection of a line segment can never exceed the length of the segment itself and is equal to it when the line segment is parallel to the plane on which the projection is performed. Now if the line segment is perpendicular to this plane, then its projection is a single point.

The following proposition very often comes in handy in problem solving. It describes the relationship between the areas of a plane figure and its orthogonal projection. *If S is the area of a plane polygon and S_{pr} is the area of its projection on some plane P, then*

$$S_{pr} = S \cos \alpha \tag{2}$$

*where α is the angle between the P plane and the plane of the polygon at hand.***

Fig. 108



We first prove this formula for a triangle (Fig. 108). Let triangle ABC have an orthogonal projection $A'B'C'$ on plane P. (If the projec-

* If a straight line and a plane are parallel, then the angle between them is taken (by definition) to be zero.

** The angle between two (intersecting) planes is measured by the plane angle of any one of the dihedral angles formed by these planes. If the planes are parallel, then the angle between them is taken to be (by definition) zero.

tion is a line segment, then this means that the plane of triangle ABC is perpendicular to the P plane and the formula is obvious if we take into account that the area of a triangle that degenerates into a line is equal to zero.) Of the three vertices of triangle ABC let vertex B have the property that $AA' > BB' > CC'$. (We leave it to the reader to examine the case $AA' = BB' \neq CC'$; if $AA' = BB' = CC$, then the plane of triangle ABC is parallel to the P plane.)

Through point B draw plane P_1 parallel to the P plane. The projection of triangle ABC on the plane P_1 is the triangle $A''BC''$ equal to triangle $A'B'C'$. The line MN of intersection of the plane P_1 with the plane of the triangle ABC divides this triangle into two triangles: $\triangle ABL$ and $\triangle BLC$.

In the plane of triangle ABC, drop a perpendicular AH from point A on side BL; then, by the theorem on three perpendiculars, $A''H \perp BL$ and so

$$S_{\triangle ABL} = \frac{1}{2} BL \cdot AH, \quad S_{\triangle A''BL} = \frac{1}{2} BL \cdot A''H$$

But the angle between AH and plane P_1 is equal to the angle α between the plane of triangle ABC and the plane P_1 (or plane P). And so $A''H = AH \cdot \cos\alpha$ and, hence,

$$S_{\triangle A''BL} = \frac{1}{2} BL \cdot A''H = \frac{1}{2} BL \cdot AH \cdot \cos\alpha = S_{\triangle ABL} \cdot \cos\alpha$$

The reasoning is exactly the same for the triangles BCL and $BC''L$; thus, we finally have

$$S_{pr} = S_{\triangle A'B'C'} = S_{\triangle A''BC''} = S_{\triangle ABC} \cdot \cos\alpha = S \cdot \cos\alpha$$

If the figure being projected is a polygon, then by dividing it and its projection into triangles and summing the results for each pair of triangles, we obtain the formula to be proved.

Formula (2) is conveniently used to compute the areas of sections of various solids, lateral surface areas, angles between planes, and the like. For example, if we know the area s of the base of a *regular* many-sided pyramid and the angle α of inclination of a lateral face to the plane of the base, then the lateral surface area is

$$S = \frac{s}{\cos\alpha}$$

But if the solid is a truncated pyramid (parallel bases; also called frustum), then for s we have to take the difference of the areas of the bottom base and top base. It is easy to see that this fact is valid also for any irregular pyramid as long as all its lateral faces are inclined to the plane of the base *at one and the same angle* α.

Exercises

1. Let $ABCDA_1B_1C_1D_1$ be a cube ($ABCD$ is the square of the bottom base; AA_1, BB_1, CC_1, DD_1 are the lateral edges). What is the size of the angle between the lines: (a) AA_1 and B_1D, (b) AD_1 and D_1C, (c) AD_1 and B_1D? Find the distance between the lines: (d) AA_1 and B_1D, (e) AD_1 and DC_1 if an edge of the cube is equal to a.

2. In a regular tetrahedron find the angle between skew edges and the dihedral angle between faces. Find the distance between skew edges if the side of the tetrahedron is equal to a.

3. Is the statement true that if L and l are skew lines, then through L it is possible to pass a unique plane parallel to the straight line l? Determine whether the following proof of it is correct.

Take a point A on line L and draw through it a line L^* parallel to l. The plane π passing through the intersecting lines L and L^* is obviously parallel to l. Since it is possible to draw only one line L^* parallel to l through the point A, and since it is possible, through two intersecting lines L and L^*, to pass only one plane, it follows that through L it is possible to pass a unique plane parallel to the straight line l.

4. L and l are two skew lines. Is it always possible to construct a plane containing L that is perpendicular to l?

5. Does there always exist a line perpendicular to three given pairwise skew lines?

6. Does there always exist a line that intersects all three given pairwise skew lines?

7. Can there exist in space four pairwise mutually perpendicular skew lines?

8. In a cube with edge a is drawn a common perpendicular of two skew diagonals of adjoining faces. Find the lengths of the line segments into which it divides the given diagonals of the faces.

9. Prove that if a straight line forms equal angles with each of three pairwise nonparallel lines lying in one plane, then it is perpendicular to that plane.

10. Can two nonadjacent lateral faces of a many-sided pyramid be perpendicular to the plane of the base?

11. A plane angle is formed as the result of cutting a given dihedral angle by all possible planes intersecting its edge. Within what limits can the plane angle vary?

12. Let A, B, C, D be four arbitrary points in space. Prove that the midpoints of the line segments AB, BC, CD, DA lie in one plane. What figure is formed if the midpoints of these line segments are joined in succession?

13. Prove that all four altitudes of a regular tetrahedron intersect in one point.

14. A line AB is parallel to a plane π. CD intersects AB at an acute angle α and forms with the plane π an angle φ. Determine the angle between the π plane and the plane in which the lines AB and CD lie.

15. The plane passing through one of the edges of a regular tetrahedron divides the volume of the tetrahedron in the ratio 3 : 5. Find the tangents of the angles into which this plane divides the dihedral angle of the tetrahedron.

16. In a regular triangular pyramid the side of the base is equal to a and the lateral edges are equal to b. Determine the dihedral angle at a lateral edge.

17. A right triangle is located so that its hypotenuse lies in a plane π and the legs form with this plane the angles α and β, respectively. Find the angle between the plane of the triangle and the π plane.

18. From a point of the edge of a dihedral angle α ($0 < \alpha < 90°$) issue two rays lying in different faces. One of the rays is perpendicular to the edge of the dihedral angle and the other one forms with the edge an acute angle β. Find the angle between these rays.

19. The line segments of two straight lines contained between parallel planes stand in the ratio 2 : 3 and their angles with one of the planes stand in the ratio 2 : 1. Determine the angles.

20. The plane angles of a trihedral angle are α, β and γ. A point is taken at a distance l from the vertex of the trihedral angle on the edge adjoined by the plane angles β and γ. Determine the distance from this point to the plane of angle α.

21. A point M is taken on a sphere of radius R and from it are drawn three equal chords MP, MQ and MR so that $\angle PMQ = \angle QMR = \angle RMP = \alpha$. Find the lengths of the chords.

22. In a right triangle, a plane is drawn through the bisector of the right angle. The plane forms with the plane of the triangle an angle α. What angles does it form with the legs of the triangle?

23. A straight line tangent to a cone forms with its generatrix, at the point of tangency, an acute angle α and is inclined to the plane of the base of the cone at an angle β. Determine the angle between the generatrix and the base plane.

24. The altitude of a triangular pyramid $ABCD$ dropped from the vertex D passes through the point of intersection of the altitudes of the triangle ABC. It is also given that $DB=b$, $DC=c$, $\angle BDC=90°$. Find the ratio of the areas of the faces ADB and ADC.

25. Two line segments, AB of length a and CD of length b, lie on skew lines, the angle between which is α. The feet O and O' of a common perpendicular (of length c) to these lines divide AB and CD so that $OA : OB = 2 : 3$ and $CO' : O'D = 3 : 2$. Find the lengths of the line segments BD and BC.

26. Let $ABCDA_1B_1C_1D_1$ be a right parallelepiped ($ABCD$ and $A_1B_1C_1D_1$ are parallelograms and the lateral (parallel) edges AA_1, BB_1, CC_1, DD_1 are perpendicular to the planes of these parallelograms), in which $AD_1 = A_1D = a$, $BA_1 = AB_1 = b$, $AC = c$, $BD = d$. Compute the dihedral angle between the faces AA_1B_1B and AA_1D_1D.

27. A straight line passes inside a trihedral angle, all plane angles of which are equal to α. The line is inclined identically to all the edges. Find the angle of inclination of this line to each edge of the trihedral angle.

28. An equilateral triangle ABC with side a is given in a plane P. A line segment $AS=a$ is laid off on a perpendicular to plane P at point A. Find the tangent of the acute angle between AB and SC and the shortest distance between them.

29. A perpendicular, equal to a, is dropped from the foot of the altitude of a regular triangular pyramid to a lateral face. Find the volume of the pyramid, given that the plane angle at the vertex of the pyramid is equal to α.

30. A perpendicular, equal to p, is drawn from the foot of the altitude of a regular triangular pyramid to a lateral edge. Find the volume of the pyramid if the dihedral angle between a lateral face and the base is equal to α.

31. Determine the volume of a parallelepiped, all edges of which are equal to unity and the plane angles at one of the vertices are $\varphi < 90°$.

32. Determine the sine of the angle between two altitudes dropped from two vertices of a regular tetrahedron on opposite faces.

33. In a trihedral angle $OABC$ the angle between the faces OAB and OBC is a right angle and each of the other dihedral angles is equal to γ. Find the plane angle AOC.

34. The angle between two skew lines is 60°. A point A lies on one line, a point B on the other, the distances from each of these points to the common perpendicular of the skew lines are the same and are equal to the distance between the lines. Find the angle between the common perpendicular and the line AB. Note the possibility of an ambiguous solution to this problem.

35. A line segment AB lies on the edge of a dihedral angle. M is the point of one of the faces of the angle and is at a distance l from the edge. A perpendicular dropped from M on the other face of the dihedral angle is seen from A at an angle of α and from point B at an angle of β. The distance from the centroid of triangle ABM to the second face of the dihedral angle is m. Find the length of the line segment AB. Note the possibility of an ambiguous solution to this problem.

36. Three equal right circular cones with angle α ($\alpha \leqslant 2\pi/3$) in the axial section have a common vertex and are externally tangent to one another along the generatrices l_1, l_2, l_3. Find the angle between l_1 and l_2.

3.5 Proofs in geometry

Geometric proofs have always plagued students. The various problems involving proofs that come up are ordinarily considered to be very complicated and harder than computational problems.

Geometric proofs are difficult because they require an ability to reason logically and express one's ideas precisely, a clear understanding of what is given and what it is required to prove. It is precisely a lack of properly developed habits in carrying through a logical argument that accounts for the numerous mistakes made in the solving of problems involving proofs. What it means to *prove* a fact is already a stumbling block to many.

The ability to reason logically, rigorously justifying geometric facts can only be developed by practice, by solving a sufficient number of problems. There are of course no general procedures for finding a proof of some assertion or solving a specific problem involving proof. The following selected problems will help the student to see how logical arguments are conducted and to overcome some typical mistakes.

When taking up the proof of some geometric fact or solving a problem involving a proof, it is first of all necessary to find the underlying idea, with the aid of which it is possible to construct a rigorous justification of the statement at hand. This requires ingenuity, the ability to "see" the applicability of certain familiar theorems, to seek various supplementary constructions and to notice a specific property of the geometric configuration under study.*

The exploratory search need not of course be presented in the final version of the solution, which must simply contain the rigorous proof. It is of no importance how the student guessed and what avenue he proceeded by to start off the proper chain of logic. What is definitely required in the final version is an exhaustive, logically correct justification of all arguments.

One should not, however, overstate the case for ingenuity, since at examinations the student is not required to do more than can be expected of one who has thoroughly and actively worked through the school course of mathematics.

Let us examine a number of examples of how one might approach the solution of geometric problems involving proof.

1. *Draw through the centre of an equilateral triangle a line parallel to the base. On this line take an arbitrary point M inside the triangle. Prove that the distance from M to the base of the triangle is the arithmetic mean of the distances from M to the sides of the triangle.*

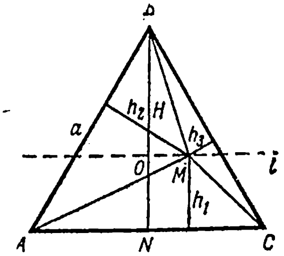
---

* The method of grouping in algebra is an analogy. In the general form it is hard to say anything definite about what will lead to an appropriate grouping, but experience in problem solving facilitates seeking out the right approaches after a few attempts and some probing.

Let us first try to find a starting idea of the proof. Let h_1 be the distance from M to the base AC, and h_2 and h_3, the distances from this point to the sides AB and BC, respectively (Fig. 109).

We need the equation $h_1 = \frac{1}{2}(h_2 + h_3)$. Since the distance from l to the base AC is equal to one third the altitude H of the triangle, then $h_1 = \frac{1}{3}H$. It is therefore sufficient to prove that $h_2 + h_3 = \frac{2}{3}H$.

Fig. 109



It is easy to notice that in place of this one we can prove the equation $h_1 + h_2 + h_3 = H$. Since each of the line segments h_1, h_2 and h_3 is perpendicular to the corresponding side of the triangle, it is natural to attempt to prove the last equation by arguments involving areas.

Now let us carry through a rigorous proof. Join M to the vertices of the triangle. Then $S_{ABC} = S_{AMB} + S_{MBC} + S_{CMA}$ or, what is the same thing, $\frac{1}{2}aH = \frac{1}{2}ah_2 + \frac{1}{2}ah_3 + \frac{1}{2}ah_1$ where a is the length of a side of the triangle. From this it follows that $h_1 + h_2 + h_3 = H$. On the other hand, $h_1 = ON = \frac{1}{3}H$, since O—the centre of an equilateral triangle—is the point of intersection of the medians. Therefore, naturally, $h_2 + h_3 = H - h_1 = \frac{2}{3}H$. Hence, we indeed have $h_2 + h_3 = 2h_1$, or h_1 is the arithmetic mean of the line segments h_2 and h_3.

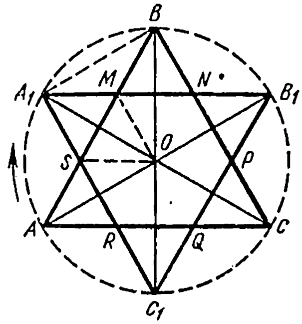
Here is another plane problem involving proof. The chief difficulty of this problem consists apparently in that the statement to be proved is simply too obvious from the drawing and so it is not very clear how one should substantiate it rigorously.

2. *An equilateral triangle is turned through* 60° *about its centre. Prove that the hexagon obtained at the intersection of the original and new positions is regular.*

Let ABC be the given triangle (Fig. 110). A rotation through 60° takes A into A_1, B into B_1, C into C_1. By hypothesis, $\angle AOA_1 = \angle BOB_1 = \angle COC_1 = 60°$. But, on the other hand, $\angle AOC = \angle AOB = \angle BOC = 120°$ and, hence, $\angle AOA_1 = \angle A_1OB = \angle BOB_1 = \angle B_1OC = \angle COC_1 = \angle C_1OA = 60°$. Besides, all line segments OA, OA_1, OB, OB_1, OC, OC_1 have the same length. From these two facts it follows that points A, A_1, B, B_1, C, C_1 are the vertices of a regular hexagon inscribed in a circle with centre at O.

Furthermore, since $\angle AOA_1 + \angle A_1OB + \angle BOB_1 = 180°$, it follows that the radii AO and OB_1 constitute one straight line. This line AB_1 is the transversal of the two lines A_1B_1 and AC, and the alternate interior angles A_1B_1A and B_1AC are equal (namely, $\angle A_1B_1A = 30°$, since it is an inscribed angle subtended by arc AA_1, the central angle of which, A_1OA, is equal to $60°$; similarly $\angle B_1AC = 30°$). Hence, $A_1B_1 \parallel AC$. In the same way we prove that $BC \parallel A_1C_1$ and $B_1C_1 \parallel BA$.

Fig. 110



But then $MN \parallel AC$ so that $\triangle MBN$ is similar to $\triangle ABC$, whence it follows that $\triangle MBN$ is an equilateral triangle. In similar fashion it can be proved that the remaining small triangles with vertices B_1, C, C_1, A, A_1 are equilateral.

Draw A_1B. The angle A_1BA is inscribed in a circle and intercepts an arc of $60°$. Angle B_1A_1B is also inscribed in a circle and also intercepts an arc of $60°$. Hence, $\angle A_1BM = \angle MA_1B$. Therefore $\triangle A_1MB$ is an isosceles triangle: $A_1M = BM$. But the triangles A_1SM and BMN are equilateral so that from this equation it may be deduced that $SM = MN$. In similar fashion we can prove the equations $MN = NP = PQ = QR = RS$ of all the other sides of the hexagon $MNPQRS$. Finally, all the angles of this hexagon form straight angles with the $60°$ angles and are thus equal to $120°$ each. Hence, the hexagon $MNPQRS$ has equal angles and equal sides, which means it is a regular hexagon.

A different proof can be given if we begin from the other end. Divide the sides of triangle ABC into three equal parts. The resulting hexagon $MNPQRS$ is clearly regular since from the similarity of triangles it is readily deduced that all the sides are equal to $1/3\ AB$ and all angles are equal to $120°$.

It now remains to prove that precisely this hexagon is obtained in the intersection of triangle ABC and the rotated triangle. To do this, extend sides MN, RS and PQ of the hexagon to mutual pairwise intersectons and consider the newly formed triangle $A_1B_1C_1$. Clearly, $\angle A_1MS = 60°$ (as the supplement of $\angle SMN$, which is equal to $120°$) and $\angle A_1SM = 60°$. Therefore $\angle SA_1M = 60°$ and so also $\angle NB_1P = \angle QC_1R = 60°$. Consequently, $\triangle A_1B_1C_1$ is an equilateral triangle.

Consider the quadrangle A_1MOS. Its opposite angles are equal in pairs ($\angle SA_1M = \angle MOS = 60°$, $\angle A_1SO = \angle A_1MO = 120°$). Hence, A_1MOS is a parallelogram, more precisely a rhombus, since $OS = OM$. Therefore A_1O is the bisector of angle SA_1M and, similarly, B_1O is the bisector of angle NB_1P. It then follows that O is the centre of the triangle $A_1B_1C_1$. Besides, $A_1B_1 = A_1M + MN + NB_1 = SM + MN + NP = AB$, which is to say $\triangle ABC$ and $\triangle A_1B_1C_1$ are equal triangles.

Finally, $\angle A_1OS = 30°$ (since A_1O is the bisector of angle MOS) and similarly $\angle AOS = 30°$, whence $\angle AOA_1 = 60°$.

Thus, the equilateral triangles ABC and $A_1B_1C_1$ are equal, have a common centre and the vertices of the second one are obtained from the vertices of the first by means of a rotation through 60°. Hence, the triangle $A_1B_1C_1$ is the second position of the given triangle, and $MNPQRS$ is the regular hexagon we are interested in.

In this problem the statement to be proved was quite obvious. In the next one, we have an extremely cumbersome condition, numerous constructions and a conclusion that is not at all evident from any scrutiny of the drawing. However, if one goes into the problem painstakingly, it will be found that there is nothing involved about the constructions, and the proof of the required assertion is extremely natural.

3. *The angles C, A, B of triangle ABC form (in that order) a geometric progression with ratio 2. Let O be the centre of a circle inscribed in triangle ABC, K the centre of the circle tangent to the side AC and the extensions of sides BC and BA beyond points C and A, and let L be the centre of a circle tangent to the side BC and the extensions of the sides AC and AB of the triangle beyond points C and B. Prove that the triangles ABC and OKL are similar.*

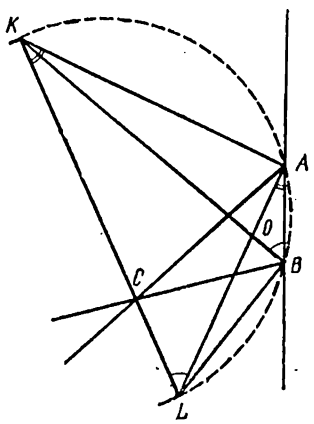


Fig. 111

From the statement of the problem it follows that in triangle ABC (Fig. 111) we have the relations: $\angle CAB = 2\angle BCA$, $\angle CBA = 2\angle CAB$. Point K is clearly equidistant from the lines AC, AB, and CB and is therefore the point of intersection of the bisectors BK and AK. Similarly, AL and BL and also AO and BO are bisectors of the corresponding angles. Then the angles LAK and LBK are right angles as being between the bisectors of supplementary angles.

Therefore if we construct circle on KL (as diameter), then points A and B will lie on the circle. But angles KLA and KBA will then be subtended by arc AK. Consequently, they are equal. Similarly,

$\angle LKB = \angle LAB$. But $\angle KBA = 1/2 \angle CBA = \angle CAB$ and $\angle LAB = 1/2 \angle CAB = \angle BCA$ so that $\angle KLA = \angle CAB$, $\angle LKB = \angle BCA$.

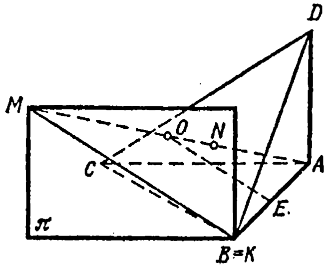
Hence, $\triangle ABC$ is similar to $\triangle OKL$ (two corresponding angles in each are equal).

There can be no question that proof problems in solid geometry are more difficult than those in plane geometry. This is due firstly to having to imagine spatial configurations that are often rather difficult to visualize on a flat drawing (see Sec. 3.6 where this is discussed). Also, it is more difficult in solid-geometry problems to "see" a useful auxiliary construction that might lead to a solution. What is more, the logical level of reasoning for proofs in space is much higher than for proofs in the plane.

4. *In a triangular pyramid, all the plane angles at the vertex are right angles. Prove that the vertex of the pyramid, the point of intersection of the medians of the base, and the centre of a sphere circumscribed about the pyramid lie on one straight line.*

Denote the vertex of the pyramid by A, its base by BCD (Fig. 112).

Fig. 112



It is convenient, for the solution of this problem, to depict the pyramid $ABCD$ as lying on one of the lateral faces (say CAB). It is given that $\angle BAD = \angle DAC = \angle CAB = 90°$.

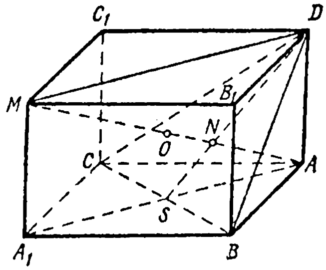
It will be recalled that the centre of a sphere circumscribed about a pyramid is the point of intersection of all planes drawn through the midpoints of the edges of the pyramid perpendicular to these edges (see Sec. 3.8). It then follows, for one thing, that the *straight lines connecting the centre O of the sphere circumscribed about the pyramid $ABCD$ with the midpoints of the edges AB, AC and AD are respectively perpendicular to these edges.*

Join O to vertex A. We have to prove that the point N of intersection of the line OA with the plane BCD of the base is the point of intersection of the medians of triangle BCD. To do this, we will have to make some more constructions in space and outside the given pyramid.

Construct on AO, beyond O, a point M such that $AM = 2AO$. Through this point, pass a plane π perpendicular to edge AB. Let K be the point of intersection of the plane π and the edge AB. Then $MK \perp AB$. Furthermore, let E be the midpoint of line segment AB, As was mentioned earlier, $OE \perp AB$. The right triangles AEO and AKM lie in one plane (drawn through the intersecting lines AB and AM) and have a common angle at the vertex A, and so they are similar. From their similarity it follows that $AK = 2AE$, that is, $AK = AB$. This means that points K and B coincide.

Thus, *the plane π passing through point M perpendicular to edge AB intersects this edge at the point B.* It can similarly be demonstrated that the planes passing through point M and perpendicular to edges AC and AD intersect these edges in points C and D, respectively. But then a *rectangular parallelepiped* $ABA_1CDB_1MC_1$ is formed with diagonal AM (Fig. 113) at the intersection of these three planes with the planes ABC, ABD and ACD of the lateral faces of the pyramid.

Fig. 113



It now remains to determine at what point the diagonal AM of the parallelepiped intersects the plane of triangle BCD. The diagonal AM lies in the plane $ADMA_1$. This plane intersects BC in the midpoint S (because the diagonals of the parallelogram ABA_1C are divided in half by their intersection). For this reason, the plane $ADMA_1$ intersects plane BCD along a straight line joining D to the midpoint S of BC, which is to say, along the median SD of triangle BCD. Hence, the point of intersection N of diagonal AM with the plane of the base BCD of the pyramid lies on the median of the triangle BCD drawn from vertex D to side BC.

A similar proof is given (by considering the plane BMC_1A not depicted in Fig. 113) that the point N of intersection of the diagonal AM with the plane of triangle BCD lies on the median of this triangle drawn from vertex B to side CD.

Thus, the point of intersection of OA with the plane of the base BCD of the pyramid $ABCD$ is indeed the point of intersection of the medians of the triangle BCD.*

* Note that in carrying out the constructions of Figs. 112 and 113 we actually assumed that the centre O of the circumscribed sphere lies outside the pyramid, that is, is located on line AM at a greater distance from A than N. This fact might be proved in rigorous fashion (*if in a triangular pyramid the angles at the vertex are right angles, then the centre of a circumscribed sphere lies outside the pyramid, on the other side of the plane of the base*) however it is not necessary to the solution of our problem, for in our proof we did not use the order of the points A, N, O, M on the diagonal AM.

The most typical mistake of students is that the *logical justification of their proofs is frequently incomplete and insufficient.* They often prove less than is required and their conclusions lack the proper justification. An illustration is the following problem.

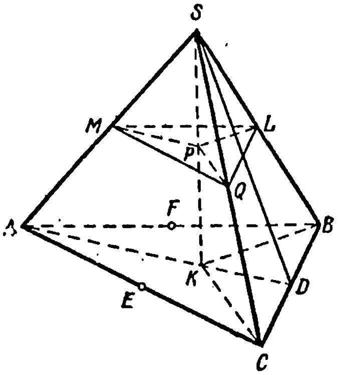
5. *The lateral surface area of a triangular pyramid is equal to s, the perimeter of the base is 3a. A sphere is tangent to all three sides of the base at their midpoints and intersects the lateral edges at their midpoints as well. Prove that the pyramid is regular. Find the radius of the sphere.*

Note first of all that the sphere in question is "pierced" by the lateral edges of the pyramid (at the midpoints). Yet, despite the clear-cut formulation of the conditions of the problem, some students regarded the sphere as *tangent* to the lateral edges of the pyramid at their midpoints. This would be quite a different problem. Thus, careless reading of the conditions may give rise to an error.

Considerable difficulties were also caused in this problem by the attempt, on the part of many students, to draw the sphere and depict the configuration. Actually, all one needs is to imagine it, since its representation is not needed in solving the problem (see Sec. 3.8).

With these preliminary remarks, let us now take up the proof—that the pyramid in question, call it $SABC$ (Fig. 114), is a regular pyramid. The sphere touches the sides AB, BC and CA of the base at their midpoints, at points F, D and E. By the property of tangents drawn to a sphere from one point, $AF = AE$, $BF = BD$, $CE = CD$. But by hypothesis $AE = EC$, $CD = DB$, $AF = FB$, and therefore it is clear that the base of the pyramid is an equilateral triangle.

Fig. 114



The plane of the base ABC cuts out a circle from the sphere (the circle is inscribed in this equilateral triangle). Obviously the centre of the circle, point K, coincides with the centre of the equilateral triangle ABC.

Let M, L, Q be the midpoints of the lateral edges AS, BS, CS of the pyramid. Many students wrote at once that the plane drawn through these points is parallel to the base ABC and did not take the pains to justify this statement. But why indeed are these planes parallel? Since MQ is the midline of the triangle ASC, it follows that $MQ \parallel AC$; for the same reasons, $QL \parallel CB$, whence, by the criterion of parallelism of two planes, our assertion follows.

By the property of parallel sections in a pyramid, the triangle MLQ, being similar to the equilateral triangle ABC, is itself equilateral. The plane of the triangle MLQ cuts out of the sphere a circle circumscribed about this triangle, and its centre coincides with the centre of the equilateral triangle MLQ.

A good many students denoted this centre by P and drew a straight line through points S, P and K and called this line the altitude of the pyramid. However, we do not have any grounds yet for such a conclusion, since we have not yet proved that these three points are collinear or that the line is perpendicular to the plane of the base.

We give a rigorous proof. Draw SK through the vertex of the pyramid and the centre of the base; let this line intersect the plane MLQ in some point P. Join this point to the vertices of triangle MLQ, and point K to the vertices of triangle ABC. PL and KB are parallel (as the lines of intersection of two parallel planes MLQ and ABC with the plane KSB), and so $\triangle KSB$ is similar to $\triangle PSL$ and, hence, $PL = \frac{1}{2} KB$. Similarly, we demonstrate that $PQ = \frac{1}{2} KC$, $PM = \frac{1}{2} KA$. Since $KB = KC = KA$, point P—the point of intersection of KS and the plane MLQ—is equidistant from all three vertices of the triangle MLQ and is therefore the centre of the circumscribed circle. We have thus proved that the vertex S and the centres P and K of triangles MLQ and ABC lie on one straight line (are collinear).

This line connects the centres of the two circles cut out of the sphere by the two parallel planes. We know that such a line is perpendicular to these planes and passes through the centre of the sphere. Hence, SK is perpendicular to the plane ABC, i.e., it is the altitude of the pyramid. It is also shown that the centre of the sphere in question lies on the altitude of the pyramid.

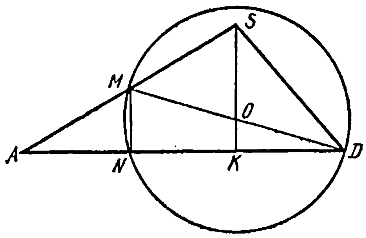
Thus, the base of the pyramid $SABC$ is an equilateral triangle, and the altitude SK of the pyramid passes through the centre K of the base. This means, by definition, that the pyramid $SABC$ is a regular pyramid.

To compute the desired radius of the sphere, it is desirable to make a separate drawing in the plane of the triangle ASD (Fig. 115). Let point O lying on altitude SK be the centre of the sphere. OM and OD are clearly equal to the radius r of the sphere.

Many students took MOD to be the diameter of the sphere and based the subsequent solution on that fact. But this cannot yet be justified for it does not follow directly from the earlier reasoning and requires a separate proof.

Denote by N the point of intersection, with AD, of the circle cut out of the sphere by plane ASD. Since $KD = 1/3\ AD$ (the point K is the centre of the equilateral triangle ABC) and $NK = KD$ (because K is the foot of the perpendicular dropped from centre O on the chord ND), it follows that $AN = NK = KD$. But then MN is the midline of the triangle ASK. Hence, $MN \parallel SK$, that is $\angle MND = 90°$. It then follows that MD—the chord on which stands an inscribed right angle—is the diameter of the circle.

Fig. 115



From this point on the computations are easy. We get the length of the slant height SD by the formula for the lateral surface area of the pyramid and then have to find the altitude of the pyramid from the right triangle SKD. Then, applying the Pythagorean theorem to triangle MND, we finally get

$$r = \frac{\sqrt{16s^2 + 45a^4}}{24a}$$

The above problem is a good illustration of how each geometric fact can be proved by carrying out the necessary additional constructions and using familiar theorems. As a rule, these facts are not in the least so obvious that they can be left without proper justification. Without such substantiation, not one of these facts can be considered established with full logical rigor and therefore a solution of the problem lacking such justification cannot be regarded as exhaustive.

The question sometimes arises as to whether the student should state formulation in full when proving theorems. It is up to the student to write out in full the formulation he finds necessary for the arguments pertaining to his theorem or simply to make a brief reference to it. The important thing is that all geometric facts of any assertion or construction should be clearly described, logically stated and convincingly justified.

When speaking about the necessity to give logically rigorous proofs of geometric statements, we wish to point out that the student often makes use of such expressions as "it is quite obvious from the drawing" or "it is clear from the drawing that...", and the like. Remember that geometric proof has to derive any required fact not from pictorialness, which can easily be illusory, but from the axioms of geometry, from definitions and familiar theorems of the school course.

Here is an example of a problem in which many students obtained the correct answer, but it was not considered complete due to logical flaws in the solution.

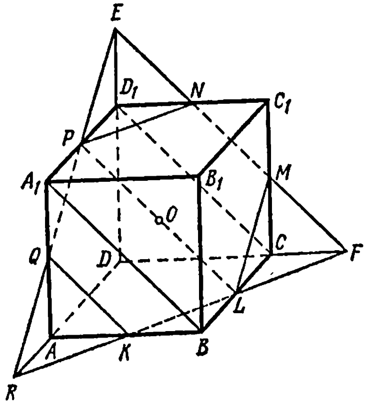
6. *Given a cube $ABCDA_1B_1C_1D_1$, where AA_1, BB_1, CC_1 and DD_1 are lateral edges. Find the area of a hexagon obtained in the section of this cube by a plane passing through the centre of the cube and the midpoints of the edges AB and BC. The cube has edge equal to* 1.

It is easy to picture this hexagon since the shape of the section is explicitly given in the statement of the problem. Apparently, that was why many students immediately guessed that the hexagon was regular and suggested a solution along the following lines:

"Let K and L be midpoints of edges AB and BC. Due to symmetry, the section passes through P and N, the midpoints of A_1D_1 and D_1C_1. It is quite obvious that the section also passes through the midpoints M and Q of the edges CC_1 and AA_1. From the triangles KBL, LCM, etc. by the Pythagorean theorem, it is easy to determine the sides of the hexagon: $KL = LM = MN = NP = PQ = QK = \sqrt{1/2}$. Since all sides of the hexagon are equal, it is a regular hexagon, and by the known length of a side, its area is equal to $3\sqrt{3}/4$."

The answer is correct but the reasoning is faulty and so this cannot be considered a *complete solution*. There are many unjustified geometric statements and one is even erroneous: a hexagon with all sides equal need not necessarily be regular because the definition of a regular polygon speaks of equality of sides and also the requirement that the angles be equal; this however does not follow from the equality of the sides.

Fig. 116



However, the equality of angles of the hexagon that is a section is not justified in the foregoing "solution" and so the conclusion that the hexagon is regular contains a logical flaw.

The following is a possible complete solution of the problem (use is made of a method that is developed in detail in Sec. 3.7).

Let point O be the centre of the given cube (Fig. 116). This point lies at the intersection of the diagonals BD_1 and A_1C of the cube (not indicated in the drawing); the plane passing through these two diagonals cuts out of the cube a rectangle BA_1D_1C. This plane has two points in common with the section that interests us: O and the midpoint L of edge BC. Therefore their line of intersection is the line OL. But point O in the rectangle BA_1D_1C is the centre of symmetry and so LO intersects the line A_1D_1 at its midpoint. Consequently, it is proved that the point P—the midpoint of edge A_1D_1—belongs to the section under consideration.

Literally the very same arguments demonstrate that the midpoint N of edge D_1C_1 belongs to the section at hand.

KL lying in the plane of face $ABCD$ (and belonging to the section) intersects the extensions of the edges AD and CD at points R and F, respectively. Join N and F which lie in the plane of face CC_1D_1D on different sides of the line segment CC_1. This straight line will intersect CC_1 at some point M, which also belongs to the section. Similarly, we see that the section passes through a point Q on edge AA_1. Finally it is clear that the lines NF and RP, which belong to the plane of the section and are the lines of intersection of this plane with the planes of the faces CC_1D_1D and AA_1D_1D respectively, intersect in some point E lying on the line of intersection of the planes of indicated faces, i.e., on DD_1.

The points of intersection of the desired section with all edges of the cube have been constructed. Let us now prove that the points Q and M are the midpoints of the appropriate edges. Comparing the right triangles RAK, KBL and LCF, we see they are equal and therefore $RA = FC = 1/2\,AB$. Comparing the right triangles RAQ, QA_1P and PD_1E, we see they are equal (for instance, $\triangle RAQ = \triangle QA_1P$, since $\angle RQA = \angle A_1QP$ being vertical angles, and $RA = 1/2\,AB = 1/2\,A_1D_1 = A_1P$ since P is the midpoint of the edge). But it then follows that Q is the midpoint of edge AA_1, and $D_1E = 1/2\,AA_1$. The proof is the same that point M is the midpoint of edge CC_1.

It is thus proved that the section $KLMNPQ$ given in the problem passes through the midpoints of the edges AB, BC, CC_1, C_1D_1, D_1A_1, A_1A of the cube. It is quite obvious that each side of the hexagon is equal to half the diagonal of the face of the cube, that is, all sides of the hexagon are equal.

It remains to establish the equality of all angles of the hexagon. It will then be proved that the hexagon obtained in the section is regular. From the equality of the right triangles RAK, RAQ and QAK it follows that triangle RKQ is equilateral; hence, $\angle RKQ = 60°$ and so $\angle QKL = 120°$. Similarly we find that all the remaining angles of the hexagon are equal to 120°.

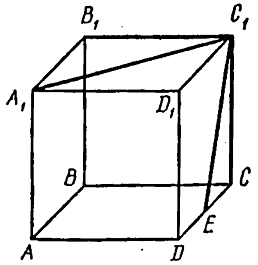
Now, to complete the solution, we have every right to take advantage of the expression for the area of a regular hexagon in terms of the length of its side.

Some students attempt to justify facts of solid geometry by reference to similar statements in plane geometry. For example, in solving Problem 5 we used the following theorem: *if tangents are drawn to a sphere from an exterior point, the line segments of each of the tangent lines from this point to the point of tangency are equal*. This was taken by some to mean that the theorem is true because in the plane a similar property of tangent lines is valid.

However, any analogy between three-dimensional (spatial) and two-dimensional statements cannot be regarded as proof. Every spatial statement must be justified in its own right. As a particular instance, the property mentioned above about tangents to a sphere requires a special proof (which the reader can carry out himself).

It is well to keep in mind that analogies between two-dimensional and three-dimensional statements can even lead to erroneous conclusions. It will be recalled that in the plane two acute angles with corresponding perpendicular sides are equal. But *will two plane acute angles in space having corresponding perpendicular sides be equal?* It is easy to construct a case in which such angles need not be equal (Fig. 117 depicts a cube; although $B_1C_1 \perp C_1E$ and $A_1C_1 \perp C_1C$, the angles $B_1C_1A_1$ and CC_1E are not equal).

Fig. 117



When carrying out geometric proofs, students frequently substitute a converse statement for the direct statement. Suppose, in a chain of reasoning, it is necessary to justify some fact, i.e., *to prove a theorem* (or *refer to* a theorem of the school course) *that asserts the validity of the fact at hand* by proceeding from what is given or known. Well, in place of this direct theorem, the student refers to the *converse*, that is, to a statement *that is valid on the assumption that the fact of interest is valid*.

This is clearly a crude logical error and the fact at hand cannot then be regarded as proved. The roots of such a mistake apparently lie in a hazy understanding of *what is given* at each stage of the proof and what *has to be justified*. Therefore, if in the course of a proof it is necessary to refer to a statement, it is advisable to recall the *exact* formulation and be sure that the assumptions under which the statement is proved are valid.

This kind of logical error caused a lot of trouble in solving the following problem.

7. *In a trapezoid ABCD, point E is the midpoint of the base BC and F is the midpoint of base AD. Denote by P the point of intersection of BF and AE and by Q the point of intersection of ED and CF. Prove that PQ is parallel to the bases of the trapezoid.*

Since $\angle BEP = \angle PAF$ and $\angle PBE = \angle PFA$ (Fig. 118), it follows that $\triangle BPE$ is similar to $\triangle APF$ and so

$$\frac{EP}{AP}=\frac{BE}{AF}$$

Similarly, it may be proved that $\triangle EQC$ is similar to $\triangle FQD$, whence follows the equation

$$\frac{EQ}{QD}=\frac{EC}{FD}$$

It is given that $BE = EC$ and $AF = FD$. Therefore the right members of the two proportions are equal, but then so are the left members:

$$\frac{AP}{EP}=\frac{DQ}{EQ}$$

Forming the derived proportion

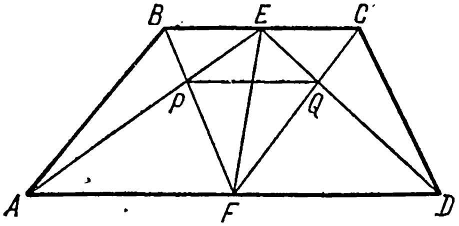
$$\frac{AP+EP}{EP}=\frac{DQ+EQ}{EQ}$$

we get the equation

$$\frac{AE}{EP}=\frac{DE}{EQ} \tag{1}$$

Hence, triangles AED and PEQ are similar, for they have a common angle E and the corresponding sides are proportional.

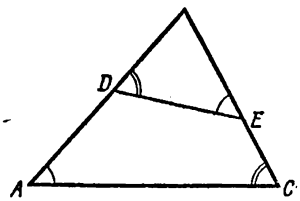
Fig. 118



At this point, many students drew the following conclusion: "Since $\triangle AED$ is similar to $\triangle PEQ$, then $PQ \parallel AD$ because a line parallel to the base of a triangle cuts off a triangle that is similar to the original one." Yet this theorem *cannot be applied here* because it is *assumed* (in the statement) that the line is parallel to the base of the triangle, and it is being *proved* that the triangle thus cut off is similar to the given triangle. Our situation is the reverse of this: we know that $\triangle AED$ is similar to $\triangle PEQ$ and we have to prove that $PQ \parallel AD$. The logical error of this deduction is apparent.

For the deduction to be logically justified, we should have referred to the *converse* of the theorem: "If a line intersecting the sides of a triangle cuts off a triangle similar to the given one, then it is parallel to the base." But this statement is *incorrect* (see Fig. 119 where $\triangle ABC$ is similar to $\triangle DBE$ since $\angle BED = \angle BAC$, $\angle BDE = \angle BCA$, but DE is not parallel to AC). We see that the parallelism of PQ and AD requires rigorous justification.

Fig. 119



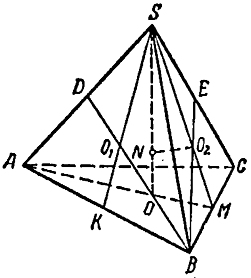
This can be done as follows. Corresponding angles in similar triangles are equal and so, proceeding from proportion (1), we can conclude that $\angle EPQ = \angle EAD$. But these angles are corresponding angles when the lines PQ and AD are cut by the transversal EA; it then follows, by the familiar criterion of parallelism, that $PQ \parallel AD$

Here is another problem.

8. *A sphere of radius r touches the lateral faces of a triangular pyramid at the points of intersection of their altitudes. The sum of three plane angles at the vertex of the pyramid is equal to* 3α. *Prove that the pyramid is regular. Find the length of a lateral edge of the pyramid.*

We start by proving that the pyramid $SABC$ is regular (Fig. 120). Let O_1 be the point of intersection of the altitudes SK and BD of the lateral face BSA and O_2, the point of intersection of the altitudes SM and BE of the lateral face BSC.

Fig. 120



It is given that the sphere is tangent to the planes BSA and BSC, respectively, at the points O_1 and O_2. This means that any straight line passing through O_1 (or O_2) and lying in the plane BSA (or BSC) will be tangent to the sphere. In particular, the lines SK, BD, SM, BE are tangent to the sphere. But then $SO_1 = SO_2$ and $BO_1 = BO_2$, by the property of tangent lines drawn to a sphere from one and the same point.

Considering the triangles SBO_1 and SBO_2, we see that they have three pairwise equal sides. From the equality of these triangles it follows that $\angle BSO_1 = \angle BSO_2$, $\angle SBO_1 = \angle SBO_2$.

It is now obvious that $\triangle BSE = \triangle BSD$ (they are right triangles and have a common hypotenuse BS and equal acute angles); $\triangle SBM = \triangle SBK$ (for the same reasons). From the equality of these triangles we conclude that $\angle BSD = \angle BSE$, $\angle SBM = \angle SBK$.

We finally consider triangles ASB and BSC. They are equal since they have a common side BS and two pairwise equal angles adjoining this side, whence $AB = BC$ and $AS = CS$.

Reasoning analogously with respect to faces ASB and ASC, we get the equations $AB = AC$ and $BS = CS$. We thus prove that in our pyramid all lateral edges are equal and all sides of the base are equal.

At this point, many students concluded: "Hence, the pyramid is regular since a regular triangular pyramid has equal lateral edges and the base is an equilateral triangle." The assertion that a regular triangular pyramid has equal lateral edges and an equilateral triangle for the base is true, yet this has nothing to do with the case at hand. What we need is the *converse*: if in a certain pyramid all lateral edges are equal and the base is an equilateral triangle, then the pyramid is regular. And this statement, which is different from the earlier one, has to be proved;* that is, we have to demonstrate that under these assumptions all the requirements of the *definition* of a regular triangular pyramid hold true.

This can easily be shown to be so: from the equality of the lateral edges it immediately follows that the altitude SO of the pyramid passes through the centre of the equilateral triangle ABC, and so, in accordance with the definition, $SABC$ is a regular pyramid.

We now undertake the second, computational, step in the solution. For this we need the fact that the *centre of a sphere tangent to the lateral edges of a regular triangular pyramid lies on the altitude of the pyramid.* Although we speak here about a sphere tangent only to the lateral faces of a regular pyramid (and not about an inscribed sphere), the proof of this assertion coincides word for word with the arguments that establish the location of the centre of a sphere inscribed in a regular triangular pyramid (see Sec. 3.8). We therefore leave this proof to the reader.

However, it must be observed that this proof is a necessary element of the solution. Unfortunately, at the examination, many students did not give this proof and confined themselves to the phrase: "This fact

---

* This is sometimes countered by the remark that we can *define* a regular triangular pyramid as a pyramid having an equilateral triangle for the base and all lateral edges equal. This definition is true of course, and accordingly the given pyramid can at once be termed regular. But then all arguments must be made on the basis of this definition (see Sec. 3.4). For one thing, we will have to prove (in the sequel of the solution this *has* to be done) that in a thus defined regular pyramid the altitude passes through the centre of the base. To avoid confusion and logical errors, it is best to proceed from the generally accepted definitions given in standard textbooks.

is an obvious consequence of the drawing because of symmetry." True, the drawing does clearly indicate this statement due to symmetry, but such a phrase cannot be regarded as an exhaustive proof. When asked to give a more detailed justification, it was a rare student who gave a proper logical proof.

Thus, let point *N lying on the altitude SO of the pyramid SABC* be the centre of a sphere tangent to the lateral faces. Draw plane SOM and join points N and O_2. Since O_2 is a point of tangency, the radius of the sphere $NO_2 = r$ is perpendicular to the plane BSC and, hence, $\angle NO_2S = 90°$. The right triangles NO_2S and MOS are clearly similar and so

$$NO_2 : OM = SO_2 : SO \tag{2}$$

Let us denote the required length of the lateral edge of the pyramid by x. Since $\angle BSM = \alpha/2$, it is easy to determine the side of the base of the pyramid and, hence, the line segment OM. Then the altitude SO of the pyramid is found from triangle SOM by the Pythagorean theorem. Finally, from the right triangle SEB we get $\angle SBE = 90° - \alpha$ and so the sine law applied to triangle SBO_2 enables us to determine SO_2.

Substituting the resulting expressions for NO_2, OM, SO_2 and SO into (2), we get

$$x = \frac{r \cot(\alpha/2)}{\cos \alpha} \sqrt{1 + 2 \cos \alpha}$$

A common view held by students is that the most important thing in a computational problem of geometry is to get the right answer. As a rule they cope well with that portion of the solution which involves computations (even when they are rather cumbersome) of the required quantity. But many leave untouched the other portion of the solution, which may be considered more important still: they disregard or fail to understand the necessity of justifying the legitimacy of the computations, of proving the geometric facts that underlie these computations. What is more, it often happens that a student can freely handle a variety of formulas but is quite helpless when asked to give a rigorous justification for the geometric assertions used in such computations.

The proof of geometric facts employed in computations is an unalienable and fundamentally important part of the solution of any computational problem. Why, in a given concrete pyramid, does the centre of the inscribed sphere lie on the altitude? Why is a given straight line perpendicular to the constructed plane? Why is the sphere under consideration tangent to the given plane precisely at the indicated point? These and similar assertions which are essential to the solution of a computational problem must be stated and also proved. One should first and foremost give a complete and well formulated justification for the solution of the problem and not merely the manipulations, for without proof the computations are groundless and for this reason the problem cannot, in the full sense of the word, be considered solved.

Generally speaking, the division of geometrical problems into computational problems and proof problems is a pure convention. The foregoing problems (see Problems 1 and 2) show that only by performing certain computations is it possible to prove what is required. On the other hand, there are many computational problems in which the proof of a fact is more essential to the solution than the required manipulations. That precisely was what we encountered in Problem 6. Again, in the next problem we could not even carry out the computations without a full justification of the required assertion, because only in the course of the proof are we able to find the approach which enables us to compute the answer.

9. *Given a triangle of area s. The medians of the given triangle are used to form a second triangle, then the medians of that one are used to form a third triangle, etc. Generally, the* $(n+1)$*th triangle is constructed from the medians of the nth triangle. Find the sum of the areas of all triangles in that sequence.*

Let us first prove that, using the medians of an *arbitrary* triangle, it is indeed possible to construct a triangle.

Let AK, BM and CN be the medians of the triangle ABC. (Fig. 121). Through point A draw a line parallel to CN and lay off $AP=CN$. We will have the proof if we can demonstrate that $KP=BM$.

Since $ANCP$ is a parallelogram (by construction) and $AM=MC$, the points N, M and P are collinear (they lie on the diagonal of this

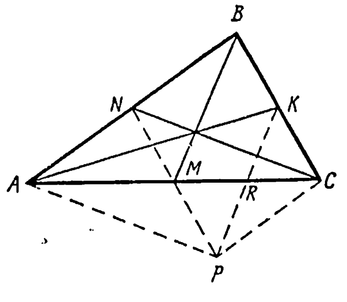


Fig. 121

parellelogram) and $NM=MP$. However, NM is the midline of the triangle ABC and so $NP \| BC$ and $NM=1/2\ BC=BK$. Therefore, in the quadrangle $BMPK$ the sides MP and BK are equal and parallel, that is, $BMPK$ is a parallelogram and, hence, $KP=BM$.

Thus, a triangle can be constructed from the medians of *any* triangle. Now let ABC be the given triangle of area s and furthermore let triangle AKP be built from its medians. We find $S_{\triangle AKP}=s_1$ if $S_{\triangle ABC}=s$. Note that AR is a median of triangle AKP (the equality $PR=RK$ follows from the equality $\triangle PMR=\triangle RKC$), and we recall that a median divides the area of a triangle in half. Therefore $S_{\triangle AKR} = s_1/2$. On the other hand, $S_{\triangle AKR} = 3s/8$ since AR, the base of triangle AKR, is equal to 3/4 AC and the altitude of this triangle is half the altitude of triangle ABC, whence we get $s_1 = 3/4\, s$.

If we use the medians of triangle AKP to construct another triangle, its area will be $s_2 = 3/4\, s_1$, etc. The problem reduces to finding the sum $S = s + s_1 + s_2 + \ldots + s_n + \ldots$. It is not hard to realize that this is the sum of a nonterminating decreasing geometric progression:

$$S = s + \frac{3}{4} s + \left(\frac{3}{4}\right)^2 s + \ldots = 4s$$

The justification of geometric facts on which computations are based is particularly important in solid-geometry problems. It very often happens that the crux of a solution to such a problem lies in the proof and not in the computations, which frequently amount to a rather simple application of familiar formulas.

10. *The altitude of a regular triangular pyramid is equal to h. The points of intersection of the altitudes of each of the lateral faces and the vertex of the pyramid lie on the surface of a sphere of radius r. Find the volume of the pyramid.*

At an examination, many students took it for granted that the centre of the sphere lies on the altitude of the pyramid (this fact is utilized in the computations) and, noting that fact without any sub-

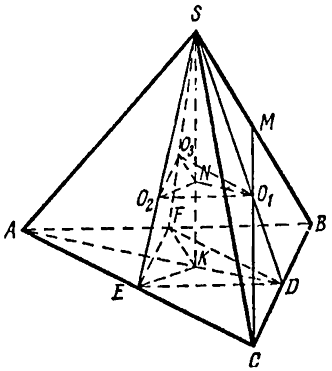


Fig. 122

stantiation, they immediately went on to the computations. Some, when the mistake (the lack of a proof) was pointed out to them, countered with the remark that no one ever required of them a detailed description and explanation of a drawing. Yet it is not a question here of explaining a drawing, but of missing an essential portion of a solution!

The plane π drawn through the points O_1, O_2, O_3—these are the points of intersection of the altitudes of the lateral faces of the pyramid $SABC$ (see Fig. 122)—cuts out of the sphere a circle that passes through these three points. The centre of the sphere lies on the perpendicular (to the π plane) erected from the centre of the circle circumscribed about the triangle $O_1O_2O_3$. We will prove that this perpendicular coincides with the altitude of the pyramid.

Since the pyramid is regular, all the lateral faces are congruent triangles and so the points of intersection of their altitudes (the points O_1, O_2, O_3) lie at the same distance from the vertex S. Thus $SO_1 = SO_2 = SO_3$.*

Draw the altitude SK of the pyramid and denote by N the point of intersection of SK with the π plane. From the equality of the triangles SKD, SKF and SKE it follows that $\angle DSK = \angle FSK = \angle ESK$ and from the equality of these angles and the equality of the line segments SO_1, SO_2 and SO_3 it follows that $\triangle SNO_1 = \triangle SNO_2 = \triangle SNO_3$. Therefore $NO_1 = NO_2 = NO_3$, or the point N is the centre of a circle circumscribed about the triangle $O_1O_2O_3$.

Thus the altitude of the pyramid does indeed pass through the centre of the circle circumsribed about triangle $O_1O_2O_3$. It has not yet been proved however that this altitude is perpendicular to the π plane.

From the similarity of the isosceles triangles ESD and O_2SO_1 (their sides, containing a common angle at the vertex, are proportional) it follows that $O_1O_2 \| ED$ (see Problem 7). In the same way it can be shown that $O_1O_3 \| DF$. Hence the π plane is parallel to the plane of the base of the pyramid and for this reason the altitude SK is perpendicular to the π plane. This completes the proof that the centre of the sphere lies on the altitude of the pyramid.

Now let us take up the computations. To determine the volume of the pyramid we have to find a side of the base. Consider face BSC. Since the right triangles SDB and BMC have a common acute angle B, $\angle DSB = \angle MCB$ and so the right triangles SDB and O_1DC are similar. From their similarity we conclude that

$$O_1D = \frac{x^2}{4SD} \qquad (3)$$

where x is the length of a side of the base of the pyramid.

Now let us make a drawing (Fig. 123) and consider the plane ADS. Let the point O lying on the altitude SK of triangle ADS be the centre of the sphere, that is, the centre of a circle passing through the points S and O_1; then $OS = OO_1 = r$. Drawing the altitude OL of the isosceles triangle SOO_1 and noting that $\triangle SLO$ is similar to $\triangle SKD$, we find

$$SL = \frac{rh}{SD} \qquad (4)$$

* We stress that this fact is a property of congruent triangles. It occurs irrespective of the sphere in the problem. Some students attempted to obtain these equations by using the property of tangents to a sphere, but SE, SD and SF "pierce" the sphere and for this reason are not tangents to the sphere.

But $SD=2SL+O_1D$, or, with regard for (4) and (3),

$$SD^2=2rh+\frac{x^2}{4}$$

On the other hand, from the right triangle SKD we have $SD^2=h^2+KD^2$, where $KD=1/3\ AD$ (because K is the centre of the equilateral triangle ABC, see Fig. 122). The two resulting equations for SD^2 yield an equation that permits determining x, after which we at once find the volume $V=\frac{1}{2}\sqrt{3}h^2(h-2r)$.

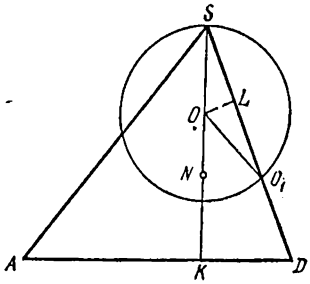


Fig. 123

We note in conclusion that in the process of solving this problem we did not prove all the geometric facts that were used. Some were so obvious that we did not even state them. For example, without any commentary, we asserted that the altitude SK (Fig. 122) *intersects* the π plane of triangle $O_1O_2O_3$ and is not parallel to it. For the computations, we made use of Fig. 123, in which point O lies on altitude SK *above* point K, and so forth.

Of course, it is logically necessary to prove such statements, but in reality there is simply not time enough to try to justify absolutely all assertions. What is more, many geometric statements actually cannot be rigorously proved without a preliminary construction of a complete system of axioms of geometry, which is not done in school.

The student is therefore required to isolate the basic statements in each problem and give the proofs.

Exercises

1. Prove that the lines joining successive centres of squares that are constructed on the sides of a parallelogram and adjoin it from the outside form a square.

2. Prove that if all dihedral angles of a triangular pyramid are equal, then all the edges are also equal.

3. Prove that if the opposite edges of a triangular pyramid are perpendicular in pairs, then all altitudes intersect in one point.

4. An arbitrary point O is taken inside a triangle ABC. Through O are drawn lines parallel to the sides of the triangle: $EK \parallel BC$, $PM \parallel AC$, $TX \parallel AB$. Points E and P lie on side AB, points K and T on side AC, M and X on BC. Prove that $\frac{AP}{AB}+\frac{BX}{BC}+\frac{CK}{CA}=1$.

5. The sides a, b and c of a triangle ABC lie, respectively, opposite the angles A, B and C. Prove that the bisector of angle A, $\beta_a = \frac{2bc\cos(A/2)}{b+c}$. Using this formula, prove that a triangle with two equal bisectors is an isosceles triangle.

6. Squares are constructed exteriorly on the sides of an isosceles triangle. Prove that the distance between the centres of the squares constructed on the sides is equal to that from the centre of the square constructed on the base to the opposite vertex of the triangle.

7. Prove that if the median and the altitude drawn from vertex B of a triangle ABC trisect angle B, then ABC is a right triangle.

8. Given that the vertices of the bottom base of a right triangular prism lie on the surface of a sphere, while the sides of the top base are tangent to the sphere. Prove that the prism is then regular.

9. Medians BD and CE are drawn in a triangle ABC; G is their point of intersection. Prove that the triangle BCG and the quadrangle $ADGE$ have equal areas.

10. Two circles are concentric, the smaller one dividing the larger one into two parts of the same size. Prove that the portion of the annulus between parallel tangents to the circle of smaller radius has an area equal to that of a square inscribed in the smaller circle.

11. A circle is inscribed in an isosceles trapezoid. Prove that the ratio of the area of the circle to the area of the trapezoid is equal to the ratio of the circumference of the circle to the perimeter of the trapezoid.

12. A sphere is inscribed in a truncated cone (frustum of a cone). Prove that the surface area of the sphere is less than the lateral surface area of the cone.

13. The angle at the vertex of an isosceles triangle is equal to 10°. Prove that the lateral side a and the base b of the triangle are connected by the relation $1/2\times \times(\sqrt{6}-\sqrt{2})\,a^3+b^3=3ba^2$.

14. Prove that the radius of a circle passing through the midpoints of the sides of a triangle ABC is half the radius of a circle circumscribed about the triangle.

15. A quadrangular frustum of a pyramid is circumscribed about a sphere. Prove that the volumes of the sphere and the frustum of the pyramid stand in the same ratio as their total surface areas.

16. Choose points E, F, G, H on the sides AB, BC, CD, DA of a parallelogram $ABCD$ so that $AE:EB=CF:FB=CG:GD=AH:HD=1:2$. Prove that the quadrangle $EFGH$ is a parallelogram and find the ratio of its area to the area of the parallelogram $ABCD$.

17. Prove that in an equilateral triangle the sum of the distances from any point to the three sides is a constant.

18. Prove that in an isosceles triangle the sum of the distances from any point of the base to the sides is a constant.

19. Given, in a triangle ABC, angles $A=\pi/7$, $B=2\pi/7$, $C=4\pi/7$. Prove that the lengths a, b, c, of sides BC, CA and AB are connected by the relation $a^{-1}=b^{-1}+c^{-1}$.

20. Given a tetrahedron $ABCD$. Prove that its edges AD and BC are mutually perpendicular if and only if the equation

$$AB^2+DC^2=AC^2+DB^2$$

holds true.

21. Given in a plane P an equilateral triangle ABC with side a. Line segments $BD=a/\sqrt{2}$ and $CE=a\sqrt{2}$ are laid off on lines perpendicular, at points B and C, to the P plane on one side of the plane. Prove that the triangle DAE is a right triangle. Compute the area of the triangle and find the cosine of the dihedral angle formed by plane DAE and plane P.

22. AB and CD are two mutually perpendicular diameters of a circle S_1. A circle S_2 has centre D and radius DA. From D are drawn two rays that cut S_1 in the points P and Q and cut the arc AB of circle S_2, which arc lies inside circle S_1,

at the points M and N. Let P_1 and Q_1 be projections of P and Q on the diameter AB. Prove that the figure bounded by the arcs PQ and MN and the line segments MP and NQ is equivalent in area to the triangle DP_1Q_1.

23. Prove that in a triangle ABC whose sides $AB=4$ cm, $BC=3$ cm and $AC=\sqrt{5}$ cm, the medians AK and CL are mutually perpendicular.

24. Prove that the projection of a regular tetrahedron on a plane will have maximum area when the plane is parallel to two skew edges of the tetrahedron.

25. Prove that the sum of the squares of the lengths of projections of the edges of a unit cube on a plane does not depend on the mutual positions of the cube and the plane and is equal to 8.

26. The dihedral angle between planes P and Q is α. A square with side 1 lies in the P plane. Prove that the perimeter of the projection of the square on the Q plane is a maximum when the diagonal of the square is parallel to the Q plane.

27. The dihedral angle between the planes P and Q is α. An equilateral triangle with side 1 lies in the P plane. Prove that the sum of the squares of the projections of its sides in the Q plane is independent of its position in the P plane.

28. Prove that if the lengths of the sides of a triangle form an arithmetic progression, then the centre of a circle inscribed in this triangle and the point of intersection of the medians lie on a line parallel to the side of middle length of the triangle.

29. A sphere is tangent to all lateral faces of a triangular pyramid at the centres of circumscribed circles. Each of the three plane angles at the vertex of the pyramid is equal to α. The sum of the lengths of the lateral edges is equal to $3b$. Prove that the pyramid is regular and find the radius of the sphere.

30. The altitude of a triangular pyramid is equal to h, the sum of the nine plane angles at the vertices of the base is equal to α. Given a sphere that is tangent to all lateral faces at the points of intersection of their medians. Prove that it is a regular pyramid and find the radius of the sphere.

31. A sphere touches all three lateral faces of a triangular pyramid $SABC$ at the points of intersection of their bisectors. From the vertex S are drawn bisectors SD and SE of the lateral faces SAB and SAC. Angle DSE is equal to α, the volume of the pyramid is V. Prove that it is a regular pyramid and find the perimeter of the base.

32. Given that the sides of a triangle ABC satisfy the relation $AC\cdot AB=BC^2-AC^2$, prove that the angle A is twice the angle B.

3.6 Geometric imagination

Particularly difficult to the student are geometric problems in which it is required not merely to use a certain formula or prove a fact, but to visualize the geometric configurations.

Pictorial geometric visualization is developed gradually through constant practice. It is extremely important to be able to visualize an object or figure "from different angles" and to make a proper and accurate drawing.

The following problem is an illustration of how a student's inability to visualize a configuration in space and to properly grasp the true interrelationship of solids shown in a drawing give rise to erroneous solutions.

1. *The bottom base $ABCD$ of the right prism $ABCDA_1B_1C_1D_1$ (where A_1A, B_1B, C_1C, D_1D are lateral edges) is a rhombus with acute angle φ. It is given that a sphere of diameter d can be inscribed in the*

prism and will be tangent internally to all faces. Find the area of the section of the prism by a plane passing through the edges BC and A_1D_1.

Let us determine the area S of the rectangle A_1D_1CB (Fig. 124). Since it is possible to inscribe a sphere (not shown in the drawing) in a right prism, it is possible to inscribe a circle in the rhombus $ABCD$. The centre of the sphere is equidistant from all lateral faces of the right prism, and so the orthogonal projection of the centre on the plane $ABCD$ is equidistant from all sides of the rhombus.

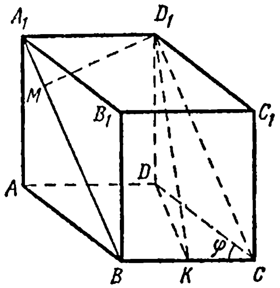


Fig. 124

Thus, the altitude DK of the rhombus is equal to the diameter of the sphere, that is, the side (of the rhombus) $DC=d/\sin\varphi$. It is easy to see that the altitude of the prism is equal to the diameter of the inscribed sphere, or $DD_1=d$.

Some students at an examination proceeded from the drawing and regarded A_1D_1CB as a rectangle and so they sought the area of the section by the formula $S=BC\cdot D_1C$. Actually, however, $S=BC\times \times D_1K$, where D_1K is the altitude of the parallelogram A_1D_1CB, that is, $D_1K\perp BC$. By the Pythagorean theorem, from the triangle D_1DK we find $D_1K=d\sqrt{2}$ and, hence, $S=d^2\sqrt{2}/\sin\varphi$.

Another incorrect solution was this. From point D_1 the altitude D_1M of this parallelogram was dropped on side A_1B of the parallelogram A_1D_1CB. The side A_1B is at once found from triangle A_1AB by the Pythagorean theorem: $A_1B=d\sqrt{1+\sin^2\varphi}/\sin\varphi$. Furthermore, since the sphere touches the faces AA_1B_1B and DD_1C_1C, the conclusion is drawn that $D_1M=d$ and so $S=A_1B\cdot d$. But here too the spatial visualization was faulty; actually $D_1M\neq d$;* in other words the altitude D_1M of the section under consideration is not equal to the distance between the parallel planes AA_1B_1B and DD_1C_1C, or, simply, D_1M is not perpendicular to the *plane* AA_1B_1B. Also note that A_1B and D_1C are not tangent to the sphere inscribed in the prism.

In these erroneous solutions, the mistakes were of course due to a faulty understanding of the drawing, to insufficient geometrical

* Using the expression found above for S and the formula $S=A_1B\cdot D_1M$, it is easy to compute the actual altitude D_1M.

imagination. However, these mistakes could have been avoided if the students had not merely used the fact (which was incorrect) which they "saw" in the drawing but had attempted to justify it in rigorous terms. It would then be evident at once that this fact is nonexistent.

It is well to bear in mind that a good spatial visualization cannot be separated from a complete logical demonstration of all the geometric facts used to support the solution. No matter how clearly the student "sees" (visualizes) a given spatial configuration and no matter how carefully the drawing has been made, he should supply rigorous proof of all assertions, even those that appear to be "obvious" from the drawing.

Figuratively speaking, geometric imagination can suggest an approach to a solution, it enables us to work the solution out in the rough. Here we give free rein to our intuition and only attempt to check to see if we are moving in the right direction. But this must be followed by a "final" solution in which the hazy intuitive reasoning and conjectures are replaced by exhaustive proofs.

The next two problems will serve to illustrate the importance of geometric imagination. It will be hard to chart a path towards the solution without some geometric imagination. Incidentally, even a good drawing does not always hint at the basic fact that can lead to a solution.

2. *Given a regular quadrangular pyramid $SABCD$ with vertex S. A plane is passed through points A and B and the midpoint of edge SC. In what ratio does this plane divide the volume of the pyramid?*

First solution. First of all note that the line AB is parallel to the plane of the lateral face DSC (Fig. 125) because edge AB is parallel

Fig. 125

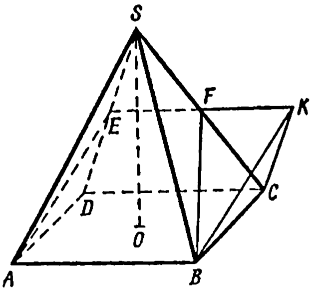
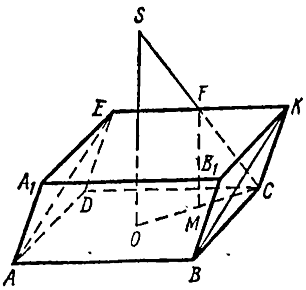


Fig. 126



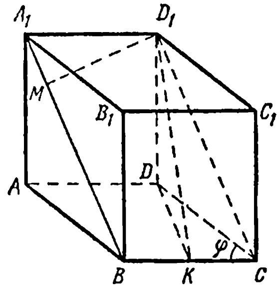
to edge DC. Therefore the cutting plane passing through edge AB and point F—the midpoint of edge SC—intersects the plane of the face DSC along the straight line EF, which is parallel to AB and, hence, to DC. From this it is clear that EF is the midline of triangle DSC.

We have to compare the volumes of two solids, one under the cutting plane, the other above it; the solid underneath is of irregular shape. This comparison requires some additional construction that should lead to a more "natural" solid: lay off line segment $FK=EF$

prism and will be tangent internally to all faces. Find the area of the section of the prism by a plane passing through the edges BC and A_1D_1.

Let us determine the area S of the rectangle A_1D_1CB (Fig. 124). Since it is possible to inscribe a sphere (not shown in the drawing) in a right prism, it is possible to inscribe a circle in the rhombus $ABCD$. The centre of the sphere is equidistant from all lateral faces of the right prism, and so the orthogonal projection of the centre on the plane $ABCD$ is equidistant from all sides of the rhombus.

Fig. 124



Thus, the altitude DK of the rhombus is equal to the diameter of the sphere, that is, the side (of the rhombus) $DC=d/\sin\varphi$. It is easy to see that the altitude of the prism is equal to the diameter of the inscribed sphere, or $DD_1=d$.

Some students at an examination proceeded from the drawing and regarded A_1D_1CB as a rectangle and so they sought the area of the section by the formula $S=BC\cdot D_1C$. Actually, however, $S=BC\times \times D_1K$, where D_1K is the altitude of the parallelogram A_1D_1CB, that is, $D_1K \perp BC$. By the Pythagorean theorem, from the triangle D_1DK we find $D_1K=d\sqrt{2}$ and, hence, $S=d^2\sqrt{2}/\sin\varphi$.

Another incorrect solution was this. From point D_1 the altitude D_1M of this parallelogram was dropped on side A_1B of the parallelogram A_1D_1CB. The side A_1B is at once found from triangle A_1AB by the Pythagorean theorem: $A_1B=d\sqrt{1+\sin^2\varphi}/\sin\varphi$. Furthermore, since the sphere touches the faces AA_1B_1B and DD_1C_1C, the conclusion is drawn that $D_1M=d$ and so $S=A_1B\cdot d$. But here too the spatial visualization was faulty; actually $D_1M\neq d$;* in other words the altitude D_1M of the section under consideration is not equal to the distance between the parallel planes AA_1B_1B and DD_1C_1C, or, simply, D_1M is not perpendicular to the *plane* AA_1B_1B. Also note that A_1B and D_1C are not tangent to the sphere inscribed in the prism.

In these erroneous solutions, the mistakes were of course due to a faulty understanding of the drawing, to insufficient geometrical

* Using the expression found above for S and the formula $S=A_1B\cdot D_1M$, it is easy to compute the actual altitude D_1M.

imagination. However, these mistakes could have been avoided if the students had not merely used the fact (which was incorrect) which they "saw" in the drawing but had attempted to justify it in rigorous terms. It would then be evident at once that this fact is nonexistent.

It is well to bear in mind that a good spatial visualization cannot be separated from a complete logical demonstration of all the geometric facts used to support the solution. No matter how clearly the student "sees" (visualizes) a given spatial configuration and no matter how carefully the drawing has been made, he should supply rigorous proof of all assertions, even those that appear to be "obvious" from the drawing.

Figuratively speaking, geometric imagination can suggest an approach to a solution, it enables us to work the solution out in the rough. Here we give free rein to our intuition and only attempt to check to see if we are moving in the right direction. But this must be followed by a "final" solution in which the hazy intuitive reasoning and conjectures are replaced by exhaustive proofs.

The next two problems will serve to illustrate the importance of geometric imagination. It will be hard to chart a path towards the solution without some geometric imagination. Incidentally, even a good drawing does not always hint at the basic fact that can lead to a solution.

2. *Given a regular quadrangular pyramid SABCD with vertex S. A plane is passed through points A and B and the midpoint of edge SC. In what ratio does this plane divide the volume of the pyramid?*

First solution. First of all note that the line AB is parallel to the plane of the lateral face DSC (Fig. 125) because edge AB is parallel

Fig. 125

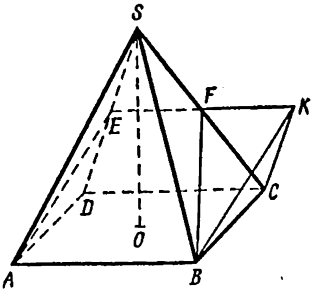
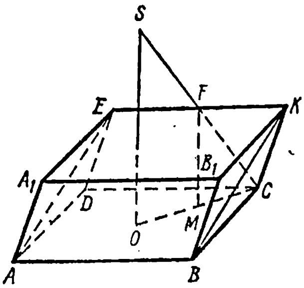


Fig. 126



to edge DC. Therefore the cutting plane passing through edge AB and point F—the midpoint of edge SC—intersects the plane of the face DSC along the straight line EF, which is parallel to AB and, hence, to DC. From this it is clear that EF is the midline of triangle DSC.

We have to compare the volumes of two solids, one under the cutting plane, the other above it; the solid underneath is of irregular shape. This comparison requires some additional construction that should lead to a more "natural" solid: lay off line segment $FK=EF$

on the extension of EF beyond point F, and then join K to the vertices B and C of the base of the pyramid.

Consider the quadrangle $DCKE$. This is a parallelogram since the opposite sides DC and KE are equal and parallel and so the opposite sides DE and CK are also equal and parallel. It is shown in similar fashion that the quadrangle $ABKE$ is a parallelogram and therefore its opposite sides AE and BK are equal and parallel. Since the base of the regular pyramid $SABCD$ is a square, the sides DA and CB of the square $ABCD$ are also equal and parallel.

It follows from what has been said that in triangles ADE and BCK the corresponding sides are equal and parallel and this means that the triangles ADE and BCK are equal and their planes are parallel. Since, besides, $DC=EK=AB$ and the straight lines DC, EK and AB are parallel, the solid $CBKDAE$ is a triangular (oblique) prism with bases CBK and DAE.

This prism can also be regarded as one half the parallelepiped $ABCDA_1B_1KE$ in which the base is a square $ABCD$ and one of the lateral faces is a parallelogram $DCKE$ (Fig. 126). Therefore the volume of prism $CBKDAE$ is equal to one half the volume of this parallelepiped, that is, it is equal to one half the product of the area of the square $ABCD$ by the altitude of the parallelepiped, or, say, by the length of the perpendicular FM dropped from F to the plane $ABCD$.

If SO is the altitude of pyramid $SABCD$, then from the similarity of the triangles OSC and MFC it follows that the altitude FM of our parallelepiped is equal to half the altitude of the pyramid.

Let V be the volume of the pyramid, V_1 the volume of the prism $CBKDAE$, Q the area of the square $ABCD$, and H the altitude of the pyramid. Then $V=1/3\ QH$, $V_1=1/4\ QH$. Thus, $V_1=3/4\ V$, or the volume of the prism $CBKDAE$ constitutes 3/4 of the volume of the pyramid $SABCD$.

However, we are not interested in the volume of the prism $CBKDAE$, but in that of the polyhedron $CFBDEA$ (Fig. 125), which is equal to the difference of the volumes of the prism $CBKDAE$ and the pyramid $BCKF$. And so we seek the volume V_2 of the pyramid $BCKF$.

Let Q_1 be the area of triangle CKF and h the altitude dropped from vertex B on the plane of triangle CKF. Thus, the volume of pyramid $BCKF$ is $V_2=1/3\ Q_1h$. But the prism $CBKDAE$ may be regarded as half the parallelepiped $DCKEABB_1A_1$ whose base is the parallelogram $DCKE$, and the square $ABCD$ is one of the lateral faces (Fig. 126). Then the altitude of this parallelepiped is equal to h and the area of the base $DCKE$ is four times the area of triangle CKF. The volume of prism $CBKDAE$ can thus also be written as $V_1=2Q_1h$, whence, taking into account the expression for V_2, we find that $V_2=1/6\ V_1$, and so the volume, V_3, of the polyhedron $SFBDEA$ is $V_3=V_1-V_2=5/6\ V_1$. But since $V_1=3/4\ V$, we finally get $V_3=5/8\ V$.

Hence, the volume of the polyhedron $CFBDEA$ constitues 5/8 of the volume of the pyramid $SABCD$; that is, the cutting plane divides the volume of this pyramid in the ratio 3 : 5.

Second solution. As in the first solution, we first prove that the cutting plane intersects the lateral face DSC along the midline EF. We then consider the volumes resulting from the section of two solids. But this time we concentrate on the quadrangular pyramid $SABFE$, the base of which is the section—the trapezoid $ABFE$ (Fig. 127).

Denote the side of the square $ABCD$ by a and the altitude SO of the pyramid $SABCD$ by h. Then the volume V of the pyramid $SABCD$ is clearly equal to $V=1/3\ a^2h$.

Fig. 127

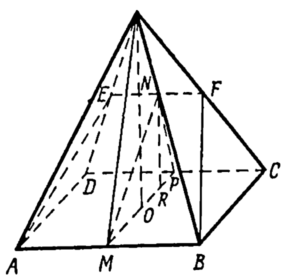
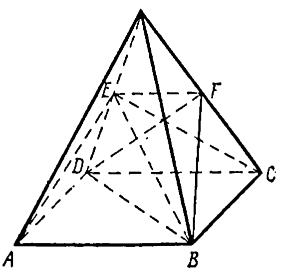


Fig. 128



The volume of the pyramid $SABFE$ is equal to a third of the product of the area of the trapezoid $ABFE$ by the length of the perpendicular dropped from S to the cutting plane.

Let us begin by computing the area of the trapezoid $ABFE$. The bases are a and $a/2$, respectively, so we have to find the altitude. Draw through altitude SO of the pyramid a plane MSP perpendicular to the edge AB. The straight line MN, which is the line of intersection of this plane with the plane of the section, is the altitude of the trapezoid $ABFE$ since it connects the midpoints of the bases of an isosceles trapezoid. Dropping a perpendicular NR from N on plane $ABCD$ and considering the similar triangles SOP and NRP, we readily find that $NR=h/2$. Then, from the right triangle NRM we determine $MN=(1/4)\sqrt{4h^2+9a^2}$ and for this reason the area of the trapezoid $ABFE$ is equal to $(3/16)\,a\sqrt{4h^2+9a^2}$.

Since the plane MSN is perpendicular to the plane $ABFE$ (because the plane $ABFE$ contains the straight line AB which is perpendicular to plane MSN), then, consequently, the perpendicular dropped from vertex S onto plane $ABFE$ lies in the plane MSN and forms with MN a right angle. In other words, the altitude of the pyramid $SABFE$ coincides with the altitude of the triangle MSN dropped from vertex S. However, it is easy to find the altitude of this triangle by comparing two expressions for the area of the triangle: the expression in terms of three sides by Hero's formula (it will be noted that all three sides of the triangle MSN are easily determined) and the expression in terms of the product of half the side MN by the corresponding altitude.

Simple computations show that the altitude of triangle MSN dropped from vertex S, that is, the altitude of the pyramid $SABFE$, is equal to $2ah/\sqrt{4h^2+9a^2}$.

But then the volume of pyramid $SABFE$ is equal to $1/8\ a^2h$, in other words, it constitutes 3/8 of the volume of the pyramid $SABCD$.

Third solution. Having proved that the cutting plane passes through the midline EF of triangle DSC, we consider the solid under the cutting plane. Unlike the first solution, we will attempt to compute the volume of this solid as the sum of the volumes of specially constructed pyramids.

Passing a plane through the points B, E and C (Fig. 128), we partition the polyhedron $CFBDEA$ of interest into two pyramids: the quadrangular pyramid $EABCD$ with vertex E and base $ABCD$ and the triangular pyramid $FBCE$ with vertex F and base BCE.

Since the altitude of pyramid $EABCD$ is half the altitude of pyramid $SABCD$, the volume of pyramid $EABCD$ is equal to half the volume of pyramid $SABCD$.

We can readily compute the volume of pyramid $FBCE$ if we consider E as the vertex and take the triangle BCE for the base: it turns out that the volume of this pyramid is equal to half the volume of the triangular pyramid $DBCF$ (to prove this, it suffices to drop perpendiculars from points D and E on the plane BSC and, having considered the corresponding similar triangles, to be assured that the altitude of the pyramid $DBCF$ is twice that of the pyramid $EBCF$). Now, in pyramid $DBCF$ take point F as vertex and the triangle DBC as the base. Comparing the pyramids $FDBC$ and $SABCD$, we see that the volume of the former is one fourth that of the latter.

Hence, the volume of pyramid $FBCE$ is equal to 1/8 the volume of the given pyramid $SABCD$ and so the volume of the polyhedron $CFBDEA$ is 5/8 the volume of this pyramid.

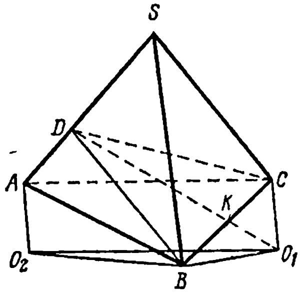
3. *Given a regular triangular pyramid $SABC$, with vertex S, with side of base a and lateral edge b. A first sphere with centre at O_1 is tangent to planes SAB and SAC at points B and C and a second sphere with centre at O_2 is tangent to planes SAC and SBC at points A and B. Find the volume of the pyramid SBO_1O_2.*

To find the volume of any triangular pyramid, first of all decide which face is to be taken for the base in order to simplify computing the area of the base and the altitude of the pyramid dropped on the plane of the base.

The student with some geometric imagination will at once see that the edge SB (Fig. 129) is perpendicular to the plane O_1BO_2 and, besides, that $O_1O_2=AC$. Thus, geometric visualization suggests right off a simple solution: to prove the foregoing facts and compute the radii O_1B and O_2B.

Here is the solution. Since BO_1 is perpendicular to the plane ABS, then $BO_1 \perp SB$; since BO_2 is perpendicular to plane BCS, then $BO_2 \perp SB$ and hence (see Sec. 3.4), SB is perpendicular to the plane O_1BO_2, that is, SB is the altitude of the pyramid.

Fig. 129



To compute the area of the triangle O_1BO_2, we have to find the radii O_1B and O_2B. Let us first find radius O_1B. Draw a plane through the three points B, C and O_1; it will cut the edge AS at point D. Since BO_1 is perpendicular to the plane ABS, then, for one thing, $BO_1 \perp AS$; similarly, $CO_1 \perp AS$; hence, AS is perpendicular to plane BCO_1. But then BD and CD are the altitudes of the lateral faces ABS and ACS and are readily found: $BD = CD = (a/b)\sqrt{b^2 - 1/4\ a^2}$. Having proved the similarity of the triangles BDK and BDO_1, we find $BO_1 = a\sqrt{(b^2 - 1/4\ a^2)/(3b^2 - a^2)}$. Similarly, we find $BO_2 = BO_1$.

It remains to find O_1O_2, i.e., to prove that $O_1O_2 = AC$.

Noticing that AO_2 is perpendicular to the plane ACS and CO_1 is perpendicular to the same plane ACS, we get $AO_2 \| CO_1$. Hence the points A, C, O_1, and O_2 are coplanar. And so the quadrangle ACO_1O_2 is plane and $AO_2 \| CO_1$, $AO_2 = CO_1$ so that ACO_1O_2 is a parallelogram and therefore $O_1O_2 = AC = a$. Now, using three sides, we find the area of the triangle BO_1O_2 and then the desired volume $V = a^2b^2/(12\sqrt{3b^2 - a^2})$.

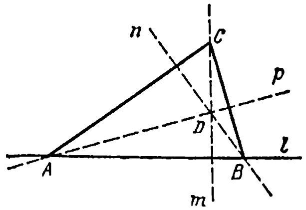
One should not get the impression however that geometric imagination is only needed in solid geometry. The most important thing in the next problem of plane geometry is a proper visualization of the drawing and the student's ability to consider and explain all possible cases.

4. *Given, in a plane, four distinct points A, B, C and D, $AB \perp CD$ and $AC \perp BD$. Prove that $AD \perp BC$.*

The positions of points A and B may be chosen in the plane in arbitrary fashion (Fig. 130). Now, let C be located so that the foot of a perpendicular dropped from it on the straight line l on which AB lies is *inside* this line segment.

It is clear that the point D must lie on line m, which is perpendicular to AB and passes through point C. Since, by hypothesis, $AC \perp BD$, point D must lie on line n that passes through point B perpendicular to AC. Since AC and AB intersect, the lines m and n, perpendicular to them, intersect as well. Their point of intersection is D. Let us now consider the triangle ABC. Suppose that D lies *inside* it; we draw line

Fig. 130



p through A and D. Obviously, m and n are the altitudes of this triangle and so p is also an altitude. Indeed, if we drop from vertex A an altitude on BC, then the altitude (by a familiar theorem on the properties of the altitudes of a triangle) will have to pass through D and so will coincide with p (they have two points in common). It is then clear that $AD \perp BC$.

The cases when point D lies *outside* the triangle ABC or when m intersects l *outside* AB are considered similarly. Point C cannot lie on the line l itself. The proof of this is left to the reader. Do not forget that a solution cannot be considered satisfactory without an exhaustive examination of all possible cases.

The part played by geometric imagination is well illustrated by "shadow problems" where one can conjecture the expected result and can get a "feeling" of the solution before substantiating it rigorously.

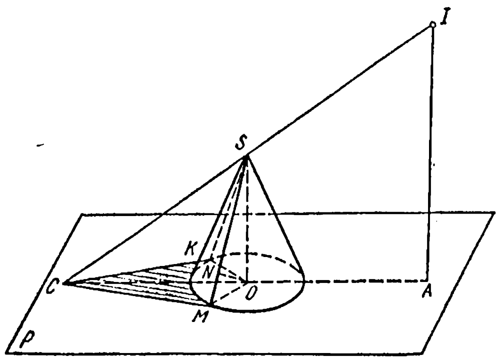
5. *Given in the plane a right circular cone and a vertical pole (line segment). The radius of the base of the cone is* 1 *metre, the altitude,* 2 *metres. The base of the pole is* 2 *metres from the centre of the base of the cone, and the altitude of the pole is* 4 *metres. At the top of the pole is a source of light. Find the area of the shadow cast by the cone on the plane (disregarding the area of the base).*

To construct the shadow of the cone, join point I, the light source (Fig. 131), to all points of the cone. The points of intersection of these rays with the plane P on which the cone rests will constitute the shadow cast by the cone. In a practical construction, it is of course unnecessary to draw all the rays; a few will suffice merely to give the general outline of the shadow. Therefore the most important thing here is to determine which rays form the boundary of the shadow.

It is easy to see that the boundary points of the shadow are obtained from rays which have *exactly one* common point with the cone: if a ray intersects the surface of the cone in two points, then the point of intersection with the plane will lie inside the shadow, and if a ray does not intersect the cone, then the corresponding point of intersection with the plane does not belong to the shadow at all. But how do we find these rays having one point in common with the cone?

Let us take a look at the cone from point I. First we see the shadow cast by S, it is point C. Then moving down the left side of the cone we get the appropriate points of the plane moving along a line CM.

Fig. 131



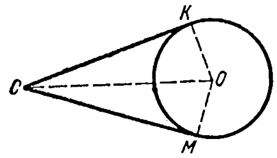
It is readily seen that CM will not pass by the base of the cone, for this would mean that our view deviated from the cone, and it will not intersect the base, for this would mean that we were looking through the cone. Hence, CM is tangent to the circle of the cone base. The shadow has thus been fully defined: it is the curvilinear figure CKM, where CK is the second tangent line to the circle of the cone base.

It now remains "merely" to prove this in rigorous fashion. No hazy statements like "seeing" the shadow and the like are admissible any more, they have played their intuitive part.

Now for the proof. Note first of all that if ON is drawn perpendicular to the generatrix SM, then the plane passing through SM perpendicular to ON does not have any common points with the cone except those lying on SM. (This statement also has to be proved, but it is quite simple and is left to the reader.) For this reason, not a single straight line of this plane can have more than one common point with the cone.

We construct the shadow of the vertex. Since $SO \parallel IA$, through these lines it is possible to pass a vertical plane, and the shadow C of vertex S is the point of intersection of the lines IS and AO lying in this plane. Now draw from point C lines CK and CM tangent to the circle of the cone base. Then $CM \perp MO$, $CM \perp SO$ and so the line CM is perpendicular to the plane SOM. Therefore $CM \perp ON$. But by construction ON is perpendicular to SM and, hence, it is perpendicular to the plane CSM. Thus, plane CSM is a plane of just the type that was mentioned at the beginning of the proof. But then every ray emanating from point I and lying in this plane has exactly one common point with the cone (provided of course that it is not parallel to SM; but such a ray does not interest us) and, consequently, the boundary of the shadow is determined precisely by these rays and is the line of intersection of the plane CSM with the P plane; that is, it coincides with the tangent line CM. The required statement is thus proved.

Fig. 132

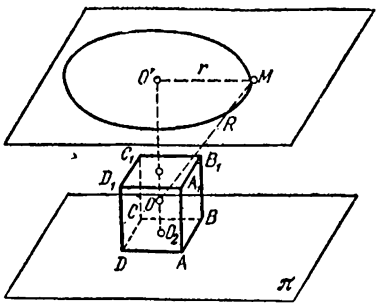


Our problem has thus been reduced to one of plane geometry: to find the area of the figure CKM (Fig. 132). True, it is still necessary to find the distance CO, but this can easily be determined from the similarity of the triangles SCO and ICA. This is a simple task, and the desired area is found to be $(3\sqrt{3}-\pi)/3$ square metres.

Some of the students stated that the boundary points K and M of the shadow are the endpoints of the diameter drawn perpendicular to the straight line AO. These students were not disturbed even by the fact that the tangents at the endpoints of the diameter would be parallel and, hence, could not pass through point C.

6. *A light source is placed at point M, at a distance $2h$ from the plane of the base of a cube with edge h and at a distance R ($R>3h$) from the centre of the cube. Prove that the shadow cast by the cube on the plane of the base will have a maximum area when the plane passing through the centre of the cube, point M and one of the vertices is perpendicular to the plane of the base.*

Fig. 133



Let the cube be $ABCDA_1B_1C_1D_1$ (Fig. 133) and such that edges AA_1, BB_1, CC_1, DD_1 are perpendicular to the plane π of the base $ABCD$ and point O is the centre of the cube. It is then easy to imagine that all points M that satisfy the condition of the problem lie on a circle in a plane parallel to the π plane and distant $2h$ from the π plane. The centre of this circle lies on the perpendicular OO' to the π plane.

The radius r of this circle can readily be computed:

$$r=\sqrt{R^2-\left(\frac{3}{2}h\right)^2}>\frac{3\sqrt{3}}{2}h \qquad (1)$$

It is clear that if, as in the preceding problem, we draw from point M straight lines to all points of the cube, then the points of intersection of these lines with the π plane yield the shadow cast by the cube. To simplify subsequent computations we include in the shadow the square $ABCD$ of the base.

It is also quite obvious that for a practical construction of the shadow it is sufficient to draw lines joining point M and the vertices of the cube, to find their points of intersection with the π plane, and connect these points with straight lines. We will then have the outline of the shadow.

If we project on the π plane the circle on which M lies and also the cube, we get (Fig. 134) a circle of radius r, inside which is the square $ABCD$ with side h.

Take an arbitrary point K in the π plane and a point K_1 on the perpendicular drawn from K to the π plane at a distance h above the

Fig. 134

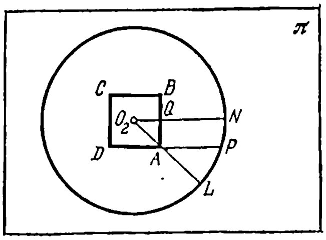
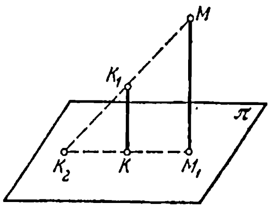


Fig. 135



π plane. It is then easy to demonstrate that if K_2 is the point of intersection of MK_1 with the π plane, then $M_1K_2=2KM_1$ (Fig. 135). Now joining A_1, B_1, C_1, D_1 to M and denoting the points of intersection of these lines with the π plane by A_2, B_2, C_2, D_2, we get the equations $A_2M_1=2AM_1$, $B_2M_1=2BM_1$, $C_2M_1=2CM_1$, $D_2M=2DM_1$.

From the foregoing arguments it is clear how the shadow is to be constructed and we can now pass from the spatial problem to the plane problem. Further note (Fig. 134) that because of symmetry we find it sufficient to consider the shadows only when point M_1 lies on arc LN ($O_2N \perp AB$ and points O_2, A, L are collinear).

Regarding shadows for various positions of point M_1 on arc LN, we assure ourselves that the point P of this arc (P lies on line DA) plays a special role. And so in the sequel we will have to consider two cases: (a) when point M_1 lies on arc PN, (b) when point M_1 lies on arc PL.

(a) Let the point M_1 lie on the arc PN (Fig. 134). We want to determine for which position of the point M_1 the shadow will have a maximum area. If M_1 does not coincide with P, then the shadow is the hexagon $AA_2D_2C_2B_2B$ (Fig. 136) in which, as can readily be shown, the sides $B_2C_2=A_2D_2=C_2D_2=2h$ and, besides, $A_2D_2 \parallel AD$, $B_2C_2 \parallel BC$, $D_2C_2 \parallel DC$. Joining points A_2 and B_2, we find that our shadow consists of the square $A_2B_2C_2D_2$ with side $2h$ and the trapezoid AA_2B_2B with bases h and $2h$ and altitude equal to the distance from M_1 to the line AB. This altitude will be a maximum at point N, that is, when the line M_1O_2 is perpendicular to edge AB.

Fig. 136

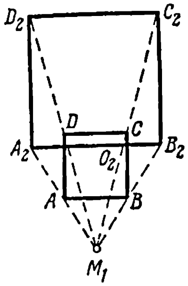
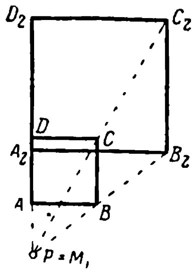


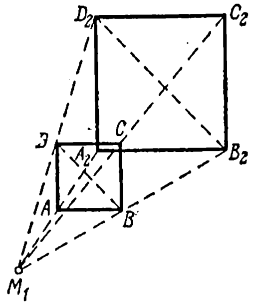
Fig. 137



This means that when M_1 moves along arc NP from N to P, the area of the shadow diminishes and becomes smallest when M_1 coincides with P. Then the shadow is the pentagon $AD_2C_2B_2B$ (Fig. 137), but its area can also be found as the area of the square $A_2B_2C_2D_2$ with side $2h$ and trapezoid AA_2B_2B with bases h and $2h$ and altitude equal to the distance from P to AB.

(b) Let us see how the area of the shadow will vary if point M_1 moves along the arc PL. In this case, the shadow will also be a hexagon, but this time the hexagon $ABB_2C_2D_2D$ (Fig. 138). This shadow is

Fig. 138



more conveniently considered thus: it consists of two right triangles (ABD with leg h and $B_2C_2D_2$ with leg $2h$) and the trapezoid BB_2D_2D with bases $h\sqrt{2}$ and $2h\sqrt{2}$ and altitude equal to the distance from M_1 to BD.

It is clear that the perpendiculars dropped from points of the arc PL on the line BD have different lengths, and it is easy to find that the largest occurs when M_1 coincides with L.

Thus, as M_1 moves along arc PL from P to L, the shadow increases and its area reaches a maximum when M_1 coincides with L.

To summarize, at point P we have the smallest shadow; at points L of arc LP and N of arc PN the shadows are the largest for their arcs. To choose the largest of all possible shadows, it remains to compare the areas of these two shadows.

Let us first compute the area of the shadow when M_1 coincides with N. Then the shadow consists of the square $A_2B_2C_2D_2$ with area $4h^2$ and the trapezoid AA_2B_2B with area equal to $(3/2)\ h\cdot NQ$. Since $NQ=$ $=r-(h/2)$, the area of the shadow is in this case equal to $(13h^2+$ $+6\,hr)/4$.

When point M_1 coincides with point L the shadow consists of the triangle ABD with area $1/2\ h^2$, the triangle $B_2C_2D_2$ with area $2h^2$ and the trapezoid BB_2D_2D with area $1/2\ (BD+B_2D_2)LO_2$. Since $LO_2=r$, $BD=h\sqrt{2}$, $B_2D_2=2h\sqrt{2}$, the area of the shadow in this case is $1/2\ (5h^2+3\sqrt{2}\ hr)$.

The inequality

$$\frac{5}{2}h^2+\frac{3}{2}\sqrt{2}\,hr>\frac{13}{4}h^2+\frac{3}{2}hr \qquad (2)$$

is valid if the equivalent inequality $r>(\sqrt{2}+1)\,h/2$ likewise holds true. This last inequality is apparently valid by virtue of (1) and so inequality (2) holds true as well.

This means that the area of the shadow is a maximum when M_1 coincides with L. To complete the proof, it only remains to note that the point L, the centre O of the cube, and the vertex A lie in a plane that is perpendicular to the plane of the base.

To solve this problem we needed six drawings instead of the usual single drawing. We could of course make do with a smaller number of drawings, but it is much better to have a fresh one for each stage in the solution, if even for the simple reason that too many lines accumulate in one drawing and it is hard to disentangle them.

Exercises

1. Given three pairwise skew lines not parallel to the same plane. Prove that there is a parallelepiped, the skew diagonals of three faces of which lie on the given straight lines.
2. The square of the diagonal in a rectangular parallelepiped is known to be equal to the sum of the squares of the three dimensions. Does the converse theorem hold true?
3. We know that in any triangle the foot of at least one altitude lies on the side itself and not on its extension. Is it true that in any triangular pyramid the foot of at least one altitude lies on the face itself and not on its extension?
4. Is it possible to cut an arbitrary tetrahedral angle with a plane so that the resulting section is a parallelogram?
5. Is it possible to construct, in a triangular pyramid, a cutting plane that intersects five edges of the pyramid?

6. Name the possible figures that result from cutting a regular tetrahedron with a plane.

7. Is it possible to cut an arbitraty regular triangular pyramid with a plane so that the section obtained is: (a) a parallelogram, (b) a rhombus, (c) a rectangle?

8. Determine the shape of the projection of a regular tetrahedron on a plane parallel to two nonadjacent edges.

9. Find the shadow cast by a cube on a plane perpendicular to a diagonal from a beam of rays parallel to that diagonal.

10. Is it possible, in a wooden cube, to bore a hole large enough to pass a wooden cube equal to the original one?

11. Fit a square with side a into an equilateral triangle of minimum dimensions. Find the side of the triangle.

12. Place inside a square with side a an equilateral triangle of maximum dimensions. Find the side of the triangle.

13. Place a square of maximum dimensions inside a regular hexagon with side a. Find the side of the square.

14. Place a regular hexagon of maximum dimensions inside a square with side a. Find the side of the hexagon.

15. Is it possible to cut a cube with a plane so that the resulting section is: (a) a square, (b) a pentagon, (c) a hexagon, (d) a regular hexagon?

16. A hemisphere is inscribed in a regular tetrahedron, the edge of which is equal to unity, so that three faces of the tetrahedron are tangent to its spherical surface, and the fourth serves as the plane of the diameter. Determine the total surface area of the hemisphere.

17. In a triangular pyramid, two faces are right isosceles triangles whose hypotenuses adjoin and form a dihedral angle α. Determine the dihedral angle, in this pyramid, the edge of which is a leg of a right triangle.

18. The side of a regular tetrahedron is equal to a. Determine the radius of a sphere tangent to the lateral edges of the tetrahedron at the vertices of the base.

19. The radius of a sphere is equal to R. From point A, distant l from the centre of the sphere, are drawn n tangents to the sphere so that all plane angles of the polyhedral angle at the vertex A are equal. Determine the distance between the points of tangency of two adjacent rays and the angle between the rays.

20. Two mutually perpendicular generatrices of a right circular cone are known to divide the circle of the base into two arcs, one of which is half the other. Find the volume of the cone if its height is h.

21. The edge of a cube is equal to a. A sphere with centre O intersects three edges (at their midpoints) converging at vertex A. A perpendicular is dropped from the point B of intersection of the sphere with one of the edges of the cube on a diagonal of the cube that passes through vertex A, the angle between the perpendicular and the radius OB is divided in half by the edge of the cube. Find the radius of the sphere.

22. The midpoint of the altitude of a right cone with generatrix l and vertex angle α is taken as the centre of a sphere passing through the vertex. Determine the radius of the circle resulting from the intersection of the surfaces of cone and sphere.

23. The edges of a triangular pyramid issuing from vertex O are pairwise perpendicular and their lengths are equal to a, b, c. Find the volume of the cube inscribed in this pyramid such that one vertex coincides with vertex O.

24. A regular pyramid whose base is a square with side a is rotated about a straight line passing through the vertex of the pyramid parallel to one of the sides of the base. Compute the volume of the solid of revolution if the plane angle at the vertex of the pyramid is equal to α.

25. Three spheres touch the plane of triangle ABC at its vertices and touch each other in pairs. Find the radii of the spheres if we know the length c of side AB and the adjoining angles A and B.

26. Given an isosceles triangle ABC, $AB=AC=b$, and angle $BAC=\alpha$. This triangle is revolved about an axis passing through the vertex A so that the angle between the axis and the plane of the triangle is equal to β and the base of the triangle is perpendicular to the axis. Compute the volume of the solid obtained by revolving the triangle ABC.

27. Two equal cones have a common vertex and are tangent along a common generatrix. The angle of the axial section of the cone is equal to 2α. Find the dihedral angle between two planes, each of which touches both cones but does not pass through their common generatrix.

28. Given an angle α $(\alpha < \pi/2)$ of the axial section of a right circular cone with vertex S and generatrix of length l. Through point A, taken on the base of the cone, is passed a plane P, which is perpendicular to the generatrix SA. Through the vertex of the cone is passed a plane Q, which is perpendicular to the plane of the axial section of the cone passing through SA. The Q plane forms with the generatrix SA of the cone an angle β $(\beta < \alpha/2)$ and cuts the cone along two generatrices. Let the extensions of these generatrices intersect the P plane at two points: C_1 and C_2. Find the length of C_1C_2.

29. A sphere is tangent to all lateral edges of a regular right hexagonal prism whose base lies outside the sphere. Find the ratio of that portion of the area of the lateral surface of the prism which lies inside the sphere to the portion of the surface of the sphere that lies outside the prism.

3.7 Cutting polyhedrons with planes

One frequently encounters geometric problems in which a cutting plane is passed through a given polyhedron and it is required to compute, for instance, the area of the section or the ratio in which the cutting plane divides the volume of the polyhedron.

Such a problem consists of two parts: the construction of the section and the computation of what is required. Each part involves certain difficulties.

Without the first part, the second part cannot of course be solved. Ordinarily, once the geometric considerations involving construction of the required section are successfully overcome, the remainder of the problem is rather simple. Thus, the crucial part of problems involving sections by planes does not lie in trigonometric computations or the solution of triangles but in the geometry in the proper sense of the word.

Experience has shown that students usually indicate the proper shape of the section and carry through the subsequent computations but find difficulty in substantiating the geometric aspect of the solution. Some students do not even undertake to justify their actions and take too much for granted. Naturally, in such cases the problem cannot be considered solved. It will not be amiss then to stress once again that in problem solving the student should justify his every step and not leave anything in the form of "quite obvious".

We give here a rather general method for constructing sections of polyhedrons. To construct a section means to indicate the points of intersection of the cutting plane with the edges of the polyhedron (in some cases, these points may turn out to be the vertices of the polyhedron) and to join these points with straight lines lying in the faces. To do this, it suffices to indicate, in the plane of the face of the polyhedron, two points belonging to the section and connect them with a straight line and find the points of intersection of this line with the edges of the polyhedron.

This is a very natural construction but it is not always made sufficient use of because it is frequently hard to find two points of a section that lie in the face itself. Yet we would be quite satisfied with two points of the section lying in the *plane* of the face (not necessarily in the face itself), and this can often be done with relative ease. As a rule, however, one has to get outside the polyhedron. This is where the student balks. He dislikes additional constructions outside the polyhedron he is working with.

In some problems, such additional constructions are unavoidable. What is more, they permit carrying through the requisite proofs and computations with relative ease.

The following concrete problems will illustrate how this method of constructing sections may be applied.

1. *Given a cube $ABCDA_1B_1C_1D_1$, where AA_1, BB_1, CC_1, DD_1 are lateral edges. Find the area of a section of the cube cut by a plane passing through the vertex A and the midpoints of the edges B_1C_1 and C_1D_1. The cube has edge equal to* 1.

First construct the section. Let points K and L be the midpoints of edges D_1C_1 and B_1C_1 (Fig. 139). KL lies in the plane of face $A_1B_1C_1D_1$ and so it intersects the extensions of edges A_1B_1 and A_1D_1 at points F and E; it is easy to compute that $B_1F=\frac{1}{2}A_1B_1=\frac{1}{3}A_1F$ and $D_1E=\frac{1}{2}A_1D_1=\frac{1}{3}A_1E$.

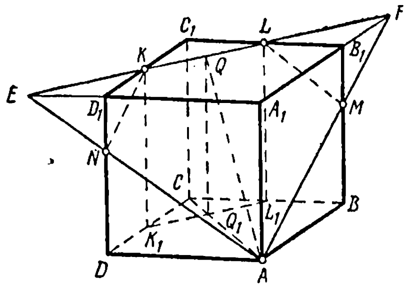


Fig. 139

Points A and F lie in the plane of face AA_1B_1B and so AF intersects edge BB_1 at some point M. Considering the triangles AA_1F and MB_1F, we see that they are similar. From their similarity follows $MB_1=\frac{1}{3}AA_1$. Noting that $AA_1=BB_1$, we finally get that point M divides the edge BB_1 in the ratio 2 to 1.

Points A and E lie in the plane of face ADD_1A_1 and so AE intersects edge DD_1 at point N. As before, it can be shown that N divides edge DD_1 in the ratio 2 to 1.

The section has thus been constructed. It passes through vertex A, the midpoints of the edges B_1C_1 and C_1D_1 and through the points that divide the edges BB_1 and DD_1 in the ratio 2 : 1. It constitutes a pentagon, $AMLKN$.

Let us compute the area of the pentagon, which is obtained from the triangle AEF by discarding two congruent triangles EKN and FLM. From the foregoing construction it is easy to obtain the lengths of the sides of these triangles and to compute their areas. The final answer is: the area of the section is equal to $7\sqrt{17}/24$.

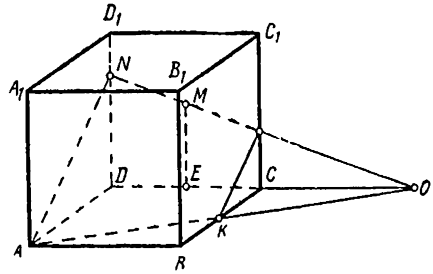
Another approach is this. Consider the pentagon $AMLKN$ as composed of the triangle AMN and the quadrangle $KLMN$. Then it is necessary to demonstrate that $KLMN$ is a trapezoid.

Both approaches require considerable computations, but these do not involve any fundamental difficulties. Actually, however, we can dispense with cumbersome computations if we take advantage of the formula for the area of a trapezoid (see Sec. 3.4). The projection of the pentagon $AMLKN$ on the bottom base of the cube will obviously be the pentagon ABL_1K_1D whose area is equal to 7/8 and from $\triangle AQQ_1$ it is easy to find $\cos(\angle QAQ_1)=3/\sqrt{17}$. But the most important thing remains, that is, to prove that $\angle QAQ_1$ is precisely the angle between the cutting plane and the bottom face of the cube.

2. *Given a cube $ABCDA_1B_1C_1D_1$, where AA_1, BB_1, CC_1, DD_1 are lateral edges. In what ratio does a plane passing through the vertex A, the midpoint of edge BC and the centre of face DCC_1D_1 divide the volume of the cube?*

Let K be the midpoint of the edge BC (Fig. 140). Since A and K lie in the plane of the bottom face, AK cuts the extension of edge

Fig. 140



DC at a certain point O. We consider the triangles ABK and KCO and are assured that they are equal, but then $CO=AB=DC$ and $DO=2DC$.

Let point M be the centre of face DCC_1D_1. Points M and O lie in the plane of face DCC_1D_1 and so MO cuts edges C_1C and D_1D at points L and N. Thus, the section is in the shape of a quadrangle, $AKLN$.

Dropping a perpendicular from M on DC, we find that the foot—point E—is the midpoint of edge DC. We consider triangles DNO, EMO and CLO. From the similarity of these triangles it follows that $CL=2/3\ ME$ and $DN=4/3\ ME$. Taking into account that $ME=1/2\ CC_1=1/2\ DD_1$, we get that $DN=2/3\ DD_1$ and $CL=1/3\ CC_1$.

We have thus determined the positions of all the points of intersection of the plane of the section with the edges of the cube.

We now determine in what ratio the cutting plane divides the volume of the cube. Denoting the length of the edge of the cube by a, we compute the volume of the polyhedron beneath the cutting plane.

It will be noted that this polyhedron is obtained from the triangular pyramid $NADO$ (with vertex N) by discarding the pyramid $LKCO$ (with vertex L). The volumes of these pyramids can readily be determined since ND and LC—their altitudes—have already been computed. We find that the volume of polyhedron $ADCKNL$ is equal to $7a^3/36$ and so the cutting plane divides the volume of the cube in the ratio $7:29$.

Most students at an examination made the section differently by employing an intuitive, purely geometric, image. They placed points N and L at random on the appropriate edges.

Those who possessed a good geometric imagination and visualized that point L lies *below* the midpoint of edge CC_1 were able to construct a correct drawing and notice (and then prove, of course) that the part of the cube under the cutting plane is a truncated pyramid (frustum) lying on its side with bases AND and KLC. The other students who placed L above the midpoint of edge CC_1 obtained a distorted drawing and naturally were unable even to begin computations.

Yet the earlier described standard method permitted us automatically to construct the required section and indicate the positions of the points of intersection of the cutting plane with the edges of the cube and to carry out the necessary computations in a simple manner.

3. *Given a regular quadrangular pyramid $SABCD$ with vertex S. A plane is drawn through the midpoints of the edges AB, AD and CS. In what ratio does the plane divide the volume of the pyramid?*

Let points K and F be the midpoints of edges AB and AD (Fig. 141). Joining them with a straight line, we see that it cuts the extensions of the edges CB and CD at points M and E. Comparing triangles BKM, AKF and DEF, we get $MB=1/2\ BC$ and $ED=1/2\ DC$.

Let point N be the midpoint of edge CS. Points N and M lie in the plane of face SBC and so MN cuts edge BS at some point L.

Now let us determine the ratio in which L divides the edge SB. To facilitate computations, it is useful to construct an auxiliary plane-geometry drawing by taking out the plane of face SBC into Fig. 142. Drawing the midline NQ in the triangle CBS, we get two equal triangles NQL and LBM, from which we find that $BL=1/2\ BQ=1/4\ BS$.

In similar fashion it is shown that $DP=1/4\ SD$, where P is the point of intersection of line EN with edge SD.

We have thus constructed the needed section. The cutting plane intersects the pyramid along the pentagon $LKFPN$.

Fig. 141

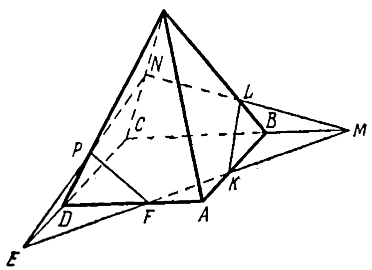
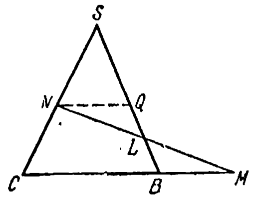


Fig. 142



This plane divides our pyramid into two polyhedrons such that we are not able to compute their volumes directly. To compute the volume of at least one of these polyhedrons requires additional constructions and the consideration of several pyramids.

But from Fig. 141 we can see that the volume of the polyhedron $CDFKBLNP$ lying under the cutting plane is equal to the volume of the triangular pyramid $NECM$ minus the volume of two triangular pyramids $LKBM$ and $PEDF$. Let us compute the volumes of these pyramids.

Let the altitude of the pyramid $SABCD$ be equal to H and the edge of the base be equal to a. Then the volume $V=1/3\ a^2H$. Since N is the midpoint of edge CS, the perpendicular dropped from this point on the plane $ABCD$ is equal to $1/2\ H$. It is just as easy to show that the perpendiculars dropped from points L and P on the plane $ABCD$ are equal to $1/4\ H$. The areas of the bases of these pyramids are readily computed: they are equal to $9a^2/8$, $a^2/8$ and $a^2/8$, respectively. The volume of our polyhedron $CDFKBLNP$ is now

$$V_1=\frac{9}{48}a^2H-2\cdot\frac{a^2H}{96}=\frac{V}{2}$$

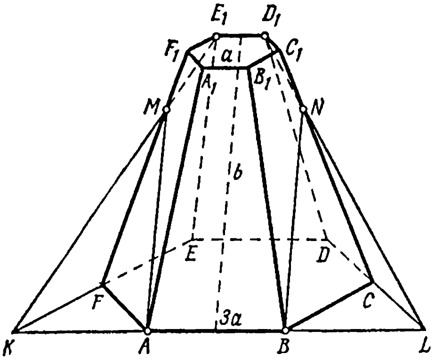
Hence, the cutting plane divides the volume of the pyramid in the ratio 1 : 1.

In this problem, it is rather easy to determine the shape of the section without applying the method under consideration because it is quite clear that the section obtained is a pentagon, $NLKFP$. If we had not gone outside the pyramid, we would not have found such an elegant method for computing the volume of the polyhedron lying under the cutting plane.

Many students attempted to find the volume by splitting up the polyhedron $CDFKBLNP$ into pyramids. But this requires excellent geometric visualization, something quite beyond what is needed for the solution that we gave. Moreover, this method involves extraordinarily cumbersome computations, whereas in the above solution they are simple.

4. *The sides of the bases of a regular hexagonal frustum of a pyramid are equal to a and $3a$. The distance between two parallel edges lying in the*

Fig. 143



planes of the different bases and in distinct lateral faces is equal to b. Compute the area of the section of the pyramid made by a plane passing through the indicated parallel edges.

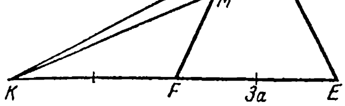
Pass the cutting plane through the parallel edges AB and D_1E_1 (Fig. 143). To construct the section, find the points K and L of intersection of line AB with the extensions of edges EF and DC. Since points K and E_1 lie in the plane of the lateral face EFF_1E_1, then KE_1 cuts the edge FF_1 at some point M. The line LD_1 lies in the plane of the lateral face CDD_1C_1 and intersects the edge CC_1 at some point N.

Thus, the shape of the section has been found to be a hexagon, $ABND_1E_1M$.

For subsequent computations note that AKF and BLC are equilateral triangles with side equal to $3a$, whence, for one thing, it follows that $KF=LC=3a$.

To facilitate computations, it is useful to construct auxiliary plane drawings. Depicting the plane of the face EE_1F_1F in Fig. 144 and

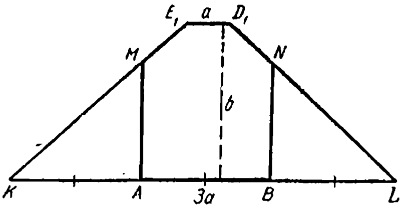
Fig. 144



comparing the similar triangles ME_1F_1 and MKF, we find that the point M divides the edge F_1F in the ratio $1:3$. It can be shown similarly that N divides edge C_1C in the ratio $1:3$ as well. Besides, it then follows that $E_1M:MK=1:3$ and $D_1N:NL=1:3$.

Depict the plane of the section in Fig. 145. Our hexagon $ABND_1E_1M$ is obtained from the trapezoid KLD_1E_1 by discarding two equal triangles KAM and LBN. Since $KL=9a$, $E_1D_1=a$, and the altitude of the trapezoid, which is the distance between the lines AB and E_1D_1, is equal to b, the area of the trapezoid is $5ab$. Since M divides E_1K in the ratio 1 : 3, the altitude of the triangle KMA dropped from M

Fig. 145



on AK is equal to 3/4 b. (We may note in passing that MA and NB are perpendicular to KL.) For this reason the area of triangle AKM is equal to $9ab/8$.

The area of the section is now easily found to be $11ab/4$.

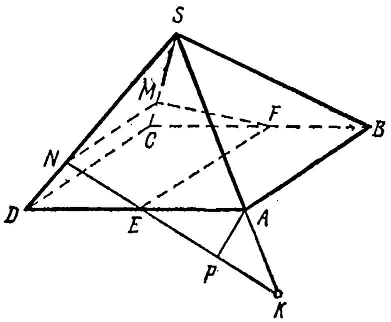
There are problems in which the cutting plane is given by three points lying on the edges (or on their extensions) of the polyhedron under consideration, and nevertheless, the formal execution of the foregoing method of constructing a section fails.

This occurs when the straight line connecting the two given points of the section turns out to be parallel to an edge of the polyhedron.

In such cases, use the theorem that if two planes are parallel to a straight line, then the line of their intersection is also parallel to this line.

5. *In a regular quadrangular pyramid SABCD all lateral faces are equilateral triangles with side equal to 1 metre. A point K is taken on the extension of edge AS beyond point A so that AK=1/2 metre. A plane*

Fig. 146



is drawn through K and the midpoints of the edges BC and AD. Find the area of the resulting section.

Let E be the midpoint of the edge AD and F the midpoint of edge BC (Fig. 146). Since the points K and E lie in the plane of the face ADS, the line KE cuts edge DS at point N. Drawing $AP \parallel DS$ and considering the similar triangles KAP and KSN, we find that $AP=1/3\ SN$. Noting that $AP=DN$, we see that N divides edge SD in the ratio 3 : 1.

We now need another point (other than the given point F) of any lateral face. We can only do this with the aid of the theorem stated above. Since $EF \parallel CD$, the plane of the section is parallel to edge CD, and taking into account that edge CD lies in the face CDS, we find that the plane of the section cuts face CDS along the line MN parallel to edge CD.

The section thus constructed is the trapezoid $EFMN$. Now let us find the area. Since $SN=3/4\ DS$ and since $\triangle DSC$ is similar to $\triangle MNS$, then $SN=3/4$ metre. Considering $\triangle DNE$ and using the cosine law, we find that $NE=\sqrt{3}/4$ metre. Similarly we find $FM=\sqrt{3}/4$ metre. Now with all the sides of the trapezoid $EFMN$, we find the area: $7\sqrt{11}/64$ square metres.

6. *The altitude of a right prism is* 1, *the base is a rhombus with side* 2 *and acute angle* 30°. *A plane is passed through a side of the base cutting the prism at an angle of* 60° *to the plane of the base. Find the area of the section.*

Here again we have to take advantage of the theorem on the intersection of planes parallel to a line. Some students gave this "solution" to the problem. "Let the cutting plane pass through edge AB of the prism $ABCDA_1B_1C_1D_1$; then it cuts the plane of the face DCC_1D_1 along MN (Fig. 147). Dropping a perpendicular from M on AB, we get the right triangles ADK and MKD because by the theorem on

Fig. 147

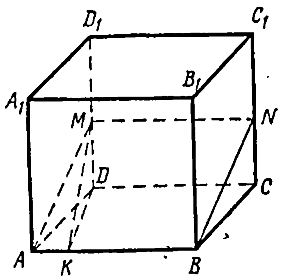
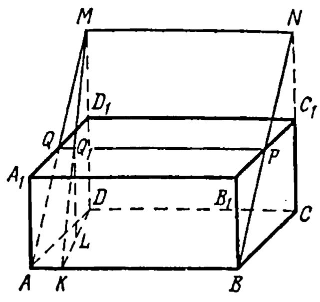


Fig. 148



three perpendiculars, $KD \perp AB$. From $\triangle ADK$ we have $DK=1$. Since $AB \perp KD$ and $AB \perp KM$, angle MKD is the plane angle of the dihedral angle between the plane of the base and the cutting plane, that is $\angle MKD=60°$. From $\triangle MKD$ we find $MK=2$. Since MK is the altitude of the parallelogram $AMNB$, its area is 4. Hence, the area of the section is 4."

Everything is correct in this solution except the last phrase to the effect that the section is the parallelogram $AMNB$. Actually, in the foregoing solution the section was not constructed because that requires knowing the exact positions of the points M and N on the edges DD_1 and CC_1. But this was not done. To solve this problem properly we first have to determine the position of M. Reasoning as above, we find from $\triangle KMD$ that $MD=\sqrt{3}$. It turns out that MD is longer than edge DD_1, which means our drawing (Fig. 147) is faulty and we have to make another drawing, in which M lies above point D_1 (Fig. 148).* Joining A and M and B and N, we find that the cutting plane also intersects face $A_1B_1C_1D_1$ along the line $QP \| AB$. Hence, the section is the parallelogram $ABPQ$. We find the altitude by drawing $MK \perp AB$ and from point Q_1 (the point of intersection of the lines MK and QP) dropping a perpendicular on KD. From the right triangle Q_1KL, in which $LQ_1=1$ and $\angle Q_1KL=60°$, we find $KQ_1=2/\sqrt{3}$. The desired area of the section is now easily found to be $4/\sqrt{3}$.

In the preceding problems we have always had two points of the desired section in the plane of at least one face of the polyhedron. Knowing these points, we found one or two more points lying on the edges and, hence, in other faces. We now also have two points of the section in the plane of a new face, etc.

However, this does not occur in all problems by far. It often happens that one of the points defining the section lies inside the polyhedron or all points are specified in different faces. In such problems, it is first necessary to make additional constructions. Ordinarily, an auxiliary plane is drawn containing some line from the plane of the section and some line lying in the plane of one of the faces of the polyhedron. Then in the auxiliary plane we seek the point of intersection of these lines and thus find yet another point (of the section) lying in the plane of a face. The subsequent construction then follows the scheme given above.

7. *Given a cube $ABCDA_1B_1C_1D_1$, where AA_1, BB_1, CC_1, DD_1 are lateral edges. In what ratio is edge B_1C_1 divided by the point E, which belongs to the plane passing through vertex A and the centres of the faces $A_1B_1C_1D_1$ and B_1C_1CB?*

In this problem, no three points of the section lie in the same face of the cube.

Let the points S and R be the centres of the faces BCC_1B_1 and $A_1B_1C_1D_1$ (Fig. 149). Through these two points draw a plane α perpendicular to the edge B_1C_1. This plane obviously cuts out of the cube a square $NLMQ$ whose vertices are the midpoints of the corresponding edges of the cube.

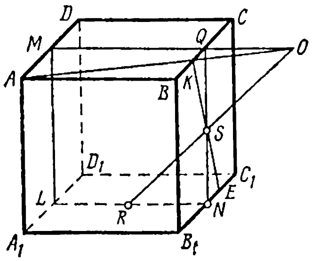
Since the points R and S lie in this plane, the entire line RS also does. But then RS cuts MQ at some point O. Then $\triangle RNS = \triangle OQS$, whence it follows that OQ is equal to half an edge of the cube.

---

* Since in determining the length of MD we did not make use of the condition that point M lies above or below point D_1, the length of MD can be computed by using either Fig. 147 or Fig. 148.

We now have two points of the section in the plane of face $ABCD$: A and O. The line AO intersects edge BC in some point K. From the similarity of the triangles ABK and KQO we get $BK=2QK$. Since BQ is equal to one half edge BC, we then get $BK=1/3\ BC$.

Fig. 149



The line KS lies in the plane of the face BCC_1B_1 and cuts edge B_1C_1 at point E. This is the point mentioned in the statement of the problem. Since S is the centre of square BCC_1B_1, then $EC_1=BK=1/3\times\times B_1C_1$ and point E divides edge B_1C_1 in the ratio 2 : 1.

Many students solved this problem on the basis of purely geometric intuition. Placing E at random closer to vertex C_1 or to vertex B_1, they obtained either Fig. 150 or Fig. 151 (E cannot lie exactly at the midpoint of B_1C_1, for then the plane passing through E, S and R would be parallel to the plane of the face ABB_1A_1 and would not pass through point A).

It is intuitively clear that the quadrangle $AKEL$ in Fig. 151 is not plane, but to prove that this drawing is impossible is no easy job. Yet a few students were only able to picture this drawing (Fig. 151) and precisely for this reason were unable to solve the problem.

Fig. 150

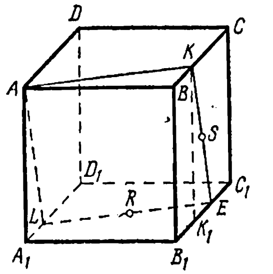
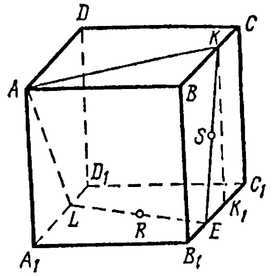


Fig. 151



Let us reason by using both drawings at the same time. This will yield a rigorous solution and at the same time will convince us that Fig. 150 is correct.

The cutting plane mentioned in the statement of the problem—denote it by β—passes through the vertex A and the centres R and S of the faces $A_1B_1C_1D_1$ and B_1C_1CB and, hence, is defined uniquely. The line l along which the β plane intersects face $A_1B_1C_1D_1$-cannot be parallel to edge A_1D_1 because otherwise the β plane would not be intersected by face B_1C_1CB and could not, therefore, pass through point S. Hence, the straight line l intersects A_1D_1, that is, the β plane intersects this line and so also the parallel straight lines B_1C_1 and BC. Thus, the β plane cuts the edges A_1D_1, B_1C_1, BC of the cube (or their extensions). Denote the corresponding points of intersection by L, E, K.

The quadrangle $ALEK$ is plane (all four vertices lie in the β plane); $AL \parallel KE$ and $AK \parallel LE$, as the lines of intersection of a pair of parallel planes by a third one. Consequently, $ALEK$ is a parallelogram. We have thus established the shape of the section.

In the plane of face B_1C_1CB draw $KK_1 \perp B_1C_1$. Since $\triangle K_1KE = \triangle A_1AL$, then $K_1E = A_1L$. Since R and S are the centres of the faces, it follows that $EC_1 = BK$ and $EC_1 = A_1L$. Besides, it is clear that $BK = B_1K_1$, whence $B_1K_1 = K_1E = EC_1$.

It is now clear that the points B_1, K_1, E, C are located precisely as indicated in Fig. 150, and Fig. 151 is impossible because there $B_1K_1 > K_1E$.

We thus automatically have the solution: $B_1E : EC_1 = 2$.

8. *Given a regular triangular prism $ABCA_1B_1C_1$ with lateral edges AA_1, BB_1 and CC_1. Let the point P divide the axis OO_1 of the prism in the ratio 5 : 1. Draw a plane through P and the midpoints of edges AB and A_1C_1. In what ratio does this plane divide the volume of the prism?*

Let E and N be the midpoints of the edges A_1C_1 and AB (Fig. 152). To construct the section, we find another point lying in the same face

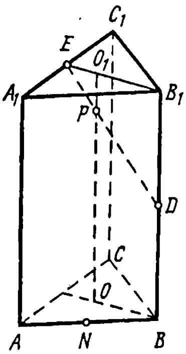


Fig. 152

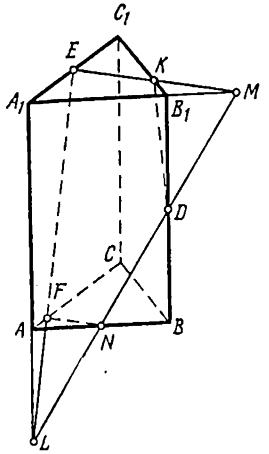
as N, for example, the point of intersection of EP and face ABB_1A_1. It is clear that this line lies, in particular, in the α plane passing through E, O_1 and P. Line EO_1 and hence point B_1 lie in this plane.

Since PO_1 and BB_1 are parallel and since PO_1 lies in the α plane, then so also does BB_1. Hence, the line EP cuts the edge BB_1 at some point D. Since the triangles EDB_1 and PEO_1 are similar and since

$EO_1 = 1/3\, B_1E$, $PO_1 = 1/6\, BB_1$, it follows that $DB_1 = 1/2\, BB_1$. We have thus found another point of the section, the midpoint of edge BB_1. From here on, the reasoning is similar to that of the preceding problems.

Since the points N and D (Fig. 153) lie in the plane of face ABB_1A_1, then, by drawing ND, we find that it cuts the extensions of edges A_1B_1 and A_1A at points M and L. It can easily be shown that the point L is distant from A by half of edge AA_1, and M is distant from B_1 by half of edge A_1B_1.

Fig. 153



Joining point E to M and L, we get K and F, which finally determine the desired section, which is the pentagon $EKDNF$.

From the similarity of the triangles ALF and A_1LE we get $AF = 1/6\, AC$. To determine B_1K we can reason as in Problem 3 to get $B_1K = 1/4\, B_1C_1$. Thus the positions of the points E, K, D, N, F are fully defined.

Let us compute the volume of the polyhedron A_1EKB_1DNAF, denoting the lateral edge of the prism by h and the area of its base by s. This polyhedron is obtained from the pyramid LA_1EM (L a vertex) by discarding the pyramids $LAFN$ and DB_1KM (L and D vertices).

A simple calculation shows that the volume of our polyhedron is equal to $49hs/144$, whence it follows that the cutting plane divides the volume of the prism in the ratio 49 : 95.

It sometimes happens that the cutting plane is given not by three points but by other conditions, say, one point and the conditions that the cutting plane is parallel to some plane, or by a point and the condition that the cutting plane is parallel to two skew lines. In such problems, one has to use these conditions to find some points lying in the planes of the faces and then continue the solution by the standard method described earlier.

9. *Given in a rectangular parallelepiped* $ABCDA_1B_1C_1D_1$ *(*$ABCD$ *and* $A_1B_1C_1D_1$ *are bases,* $AA_1 \| BB_1 \| CC_1 \| DD_1$*) the lengths of edges* $AB = a$, $AD = b$, $AA_1 = c$. *Let* O *be the centre of base* $ABCD$, O_1 *the centre of base* $A_1B_1C_1D_1$, *and* S *a point that divides the line segment* O_1O *in the ratio* 1 : 3, *that is,* $O_1S : SO = 1 : 3$. *Find the area of the section of the given parallelepiped by a plane passing through point* S *parallel to the diagonal* AC_1 *of the parallelepiped and to the diagonal* BD *of its base* $ABCD$.

Since S lies in the diagonal plane BDD_1B_1, the cutting plane parallel to the diagonal BD intersects this plane along EF, which is parallel to BD (Fig. 154). It is then quite clear that $D_1E=1/4\ D_1D=c/4$ and $B_1F=c/4$. Thus, using the condition of parallelism of the cutting plane to the diagonal BD, we have found two points of the section that lie on the edges of the parallelepiped.

Now let us take advantage of the second condition: that the cutting plane is parallel to the diagonal AC_1. S also lies in the diagonal plane

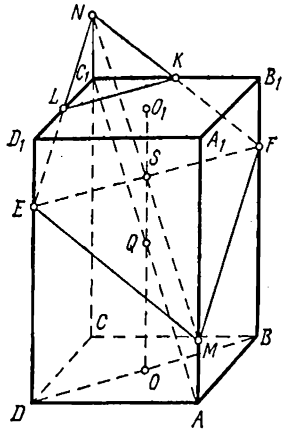


Fig. 154

ACC_1A_1 and so the cutting plane intersects this plane along the straight line MN, which is parallel to the diagonal AC_1. Let point Q be the midpoint of the diagonal AC_1. Since the line OO_1 lies in the plane ACC_1A_1, it is clear that $SQ=1/4\ O_1O=1/4\ AA_1=c/4$. Since $MN \parallel AC_1$, $AM \parallel SQ$ and $NC_1 \parallel SQ$, we get $MA = NC_1 = SQ = c/4$. We have thus found four points which belong to the section and lie either on the edges of the parallelepiped or on the extensions of these lines.

Points E and N lie in the plane of face DCC_1D_1 and so EN intersects edge D_1C_1 at some point L. Since $ED_1 = NC_1$, it is clear that L is the midpoint of edge D_1C_1 and, besides, $EL = LN$. We similarly find that FN intersects edge B_1C_1 at its midpoint, K, and that $KN = KF$.

The section is now completely defined; it is the pentagon $MFKLE$. To find the area, note that the pentagon is obtained from the quadrangle $MFNE$ by discarding the triangle KNL. The quadrangle $MFNE$ is a parallelogram because $ME \parallel FN$ and $MF \parallel EN$ (the cutting plane cuts parallel planes along parallel lines) and so the area is equal to two areas of triangle MFE.

Noting that $LN=1/2\ EN$ and $KN=1/2\ FN$, we find the area of the triangle LNK to be equal to a quarter of the area of triangle EFN or a quarter of the area of the equivalent triangle MFE. This means that the area of the section is equal to 7/4 the area of the triangle MFE.

Now let us find the area S of the triangle MFE. It is easy to see that $EF = DB = \sqrt{a^2+b^2}$, $MF = \sqrt{a^2+(c^2/4)}$, $ME = \sqrt{b^2+(c^2/4)}$. By Hero's formula, after some manipulations, we find the area of triangle MFE. We then immediately get the area of the section: $\frac{7}{16}\sqrt{4a^2b^2+a^2c^2+b^2c^2}$.

To summarize: in all the problems considered above the standard method permitted constructing the section almost automatically, and the additional constructions exterior to the polyhedron made it possible to carry out the necessary computations very simply.

Also, it is well to note that supplementary constructions exterior to a given polyhedron can be successfully applied in the solution of other problems as well (see, for instance, Problem 4 of Sec. 3.5 and Problem 2 of Sec. 3.6).

Exercises

1. A regular quadrangular pyramid with base side a and dihedral angle at the base equal to 2α is cut by a plane that bisects the dihedral angle at the base. Find the area of the section.

2. The plane angle at the vertex of a regular triangular pyramid is equal to 2α. Find the area of the section of the pyramid by a plane passing through one of the sides of the base perpendicular to the opposite lateral edge. The pyramid has base side equal to a.

3. Given a regular quadrangular pyramid. Through a side of the base is passed a plane perpendicular to the opposite lateral face. Compute the area of the section if the side of the base of the pyramid is equal to a and the dihedral angle at the base is α.

4. Find the ratio of the volumes of two solids obtained in passing a plane through a regular quadrangular pyramid so that it passes through the midpoints of two adjacent sides of the base perpendicular to the base.

5. In a regular quadrangular pyramid, the plane passing through a side of the base and the midline of the opposite lateral face forms with the base an angle of 60°. Find the volume of the pyramid if the base has side a.

6. A regular triangular pyramid is cut by a plane parallel to the base so that the lateral surface is cut in half. In what ratio is the altitude divided?

7. A regular triangular pyramid has altitude h and lateral edge l. Find the area of a section that is parallel to the base and distant a from it.

8. A regular quadrangular pyramid is cut by a plane parallel to the base. In what ratio is the volume of the pyramid divided if the area of the section is one third the area of the base?

9. In a regular quadrangular pyramid, a plane is passed through a vertex of the base perpendicular to the opposite lateral edge. Determine the area of the section if the edge at the base is equal to 1 and the lateral edge is 2.

10. A plane is drawn through one of the sides of the base of a regular right triangular prism at an angle α to the base and cuts from the prism a pyramid of volume V. Determine the area of the section.

11. In a regular quadrangular pyramid $PABCD$ with vertex P and side of the base equal to a, the angle of inclination of a lateral face to the base is equal to φ. Through this pyramid is passed a cutting plane that bisects the dihedral angle at edge CD. Find the length of the line segment along which this plane intersects face APB.

12. The bottom base $ABCD$ of a right prism $ABCDA_1B_1C_1D_1$ (where AA_1, BB_1, CC_1, DD_1 are lateral edges) is a rhombus with acute angle φ. A sphere of diameter d

can be inscribed in this prism so as to be internally tangent to all faces. Find the area of the section of the prism cut by a plane passing through edges BC and A_1D_1.

13. Given a regular quadrangular pyramid. Another pyramid is cut from this one by a plane passing through edge AB of the base. The ratio of the lateral surfaces of these pyramids is equal to two. Determine the ratio of the area of the triangle cut, by the indicated section, from the face opposite edge AB to the area of this face.

14. Given a right prism with an equilateral triangle for the base. A plane is drawn through one of the sides of the bottom base and the opposite vertex of the top base. The angle between this plane and the base of the prism is equal to α, the area of the section of the prism is equal to S. Determine the volume of the prism.

15. Given a regular triangular prism $ABCA'B'C'$ with lateral edges AA', BB' and CC'. Take on the extension of edge BA a point M so that $MA=AB$ ($MB=2AB$). A plane is drawn through M, B', and the midpoint of edge AC. In what ratio does this plane divide the volume of the prism?

16. Given a cube $ABCDA'B'C'D'$ with lateral edges AA', BB', CC', DD'. On the extensions of the edges AB, AA', AD are laid off, respectively, line segments BP, $A'Q$, DR of length $1.5\,AB$ ($AP=AQ=AR=2.5\,AB$). A plane is drawn through points P, Q and R. In what ratio does this plane divide the volume of the cube?

17. The base of a regular quadrangular pyramid is a square with side a. The altitude of the pyramid is equal to the diagonal of the square. The pyramid is cut by a plane parallel to the altitude and to two opposite sides of the base. Find the perimeter of the resulting section if it is known that a circle can be inscribed in the section.

18. A plane is drawn through the midpoint of a diagonal of a cube perpendicular to the diagonal. Determine the area of the figure resulting from the section of the cube cut by this plane if the edge of the cube is equal to a.

19. Through the edge of the base of a regular quadrangular pyramid is drawn a plane which cuts from the opposite face a triangle of area a^2. Find the lateral surface area of a pyramid that is cut out of the given pyramid by the given plane if the lateral surface area of the original pyramid is equal to b^2.

20. In a regular triangular pyramid $SABC$, a plane passing through the side AC of the base perpendicular to the edge SB cuts out a pyramid $DABC$ whose volume is 1.5 times less than that of the pyramid $SABC$. Find the lateral surface area of the pyramid $SABC$ if $AC=a$.

21. In a regular quadrangular pyramid $SABCD$, a plane drawn through side AD of the base perpendicular to face BSC divides this face into two parts of equal area. Find the total surface area of the pyramid if $AD=a$.

22. In a regular triangular pyramid $SABC$ with vertex S and volume V, a plane is drawn parallel to the median of the base BN and intersects the lateral edge SA at K, the lateral edge SB at L, and $SK=1/2\ SA$, $SL=1/3\ SB$. Find the volume of that part of the pyramid which lies beneath this plane.

23. A plane cuts off from the lateral edges SA, SB and SC of a regular quadrangular pyramid $SABCD$ with vertex S line segments $SK=2/3\ SA$, $SL=1/2\ SB$, $SM=1/3\ SC$, respectively. The length of a lateral edge of the pyramid is equal to a. Find the length of line segment SN cut off by this plane on edge SD.

24. In a regular quadrangular pyramid $SABCD$ with vertex S and altitude h, a plane is passed through the centre of the base parallel to face SAB. The area of the resulting section is equal to the area of the base. Find the volume of that part of the pyramid which lies under the plane.

25. In a regular hexagonal pyramid, a side of the base is a and the altitude is h. Compute the area of a section passing through the midpoints of two nonadjacent and nonparallel sides of the base and through the midpoint of the altitude of the pyramid.

26. The area of a section drawn through a diagonal of the base of a regular quadrangular pyramid parallel to the lateral edge that does not intersect this diagonal is equal to S. Find the area of the section passing through the midpoints

of two adjacent sides of the base and the midpoint of the altitude of the pyramid.

27. The area of a lateral face of a regular hexagonal pyramid is equal to S. Compute the area of the section passing through the midpoint of the altitude of the pyramid parallel to a lateral edge.

28. The angle between a lateral edge and the plane of the base of a regular triangular pyramid $SABC$ with vertex S is equal to $60°$. A plane is drawn through point A perpendicular to the bisector of angle S of triangle BSC. In what ratio does the line of intersection of this plane with the plane BSC divide the area of the face BSC?

29. A plane is drawn through a side of the base of a regular quadrangular pyramid perpendicular to the opposite face. Find the area of the resulting section if the altitude of the pyramid is equal to h and the ratio of the lateral surface area to the area of the base is equal to $\sqrt{3}$.

30. In a triangular pyramid $SABC$ with vertex S, $SA=SC$, $SB=2AC$, $AB=BC=1.5\,AC$. A plane is drawn through AC and the midpoint D of edge SB. The total surface area of the pyramid $SADC$ is greater than the total area of pyramid $DABC$ by a quantity equal to the area of the base of the pyramid $SABC$. Determine the angle of inclination of face ASC to the plane of the base of the pyramid $SABC$.

31. In a regular triangular pyramid $SABC$ with vertex S, base side a, and plane angles at the vertex equal to $30°$, a plane is drawn through point A and the midpoint of edge SB. The plane divides the pyramid into two equal parts. Find the line segments into which the plane divides the altitude of the pyramid.

32. In a regular quadrangular pyramid $SABCD$ with vertex S, a cutting plane P is passed parallel to side AB (through the point of tangency of an inscribed sphere) parallel to face SAB and through the point of this sphere that is closest to the vertex S. Find the area of the section of the pyramid cut by the P plane, provided $AB=1$, $SA=\sqrt{5}/2$.

33. Inscribed in a regular quadrangular pyramid $SABCD$ with vertex S and base side $d\sqrt{3}$ is a sphere with radius $d/2$. The P plane, which forms an angle of $30°$ with the plane of the base, is tangent to the sphere and cuts the base of the pyramid along a line parallel to the side AB (the point of tangency of the P plane and the sphere lies below the centre of the sphere). Find the area of the section of the pyramid cut by the P plane.

34. The base of a right prism $ABCA_1B_1C_1$ is an isosceles right triangle ABC with legs $AB=BC=1$ decimetre. A plane is drawn through the midpoints of the edges AB and BC and through point P lying on the extension of edge B_1B (beyond B). Find the area of the resulting section if $BP=1/2$ decimetre and $B_1B=1$ decimetre.

35. Given a cube $ABCDA_1B_1C_1D_1$ with edge equal to 1 decimetre. A point P is taken on the extension of edge D_1D, beyond D, such that $DP=1/2$ decimetre. A plane is drawn through P and the midpoints of edges AA_1 and CC_1. Find the area of the resulting section.

3.8 Combinations of solids

A common problem in solid geometry is one that involves combinations of solids. To solve a problem of this kind, one has to give careful thought to the mutual positions of the solids in space, make a neat drawing, and prove all the assertions.

Most drawing difficulties have to do with cases where one of the solids is a sphere. It often happens that there is no need to actually depict the sphere, it being sufficient to indicate the centre and the points of tangency with various planes and straight lines.

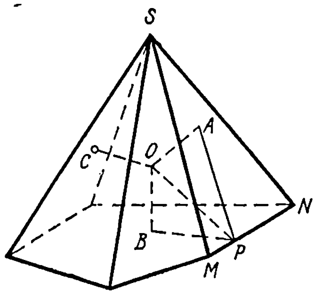
When solving problems involving combinations of solids, it is very useful to make auxiliary plane drawings and to extract plane configurations which are distorted by spatial perspective.

Let us consider in more detail a case combining a pyramid and a sphere. We first examine the configuration consisting of the pyramid and an *inscribed* sphere.

Definition. *A sphere is termed inscribed in an (arbitrary) pyramid if it is tangent to all faces of the pyramid (both lateral and base).*

Thus, the centre O of an *inscribed* sphere (Fig. 155) is a point equidistant from all *faces of the pyramid.* This means that if from the centre

Fig. 155



of the inscribed sphere we drop perpendiculars OA, OB, OC, ... on the faces of the pyramid, then all these perpendiculars will be of the same length. The points A, B, C, . . .— the feet of these perpendiculars—are the points of tangency of the inscribed sphere with the faces of the pyramid.

Also note that if from points A and B we drop perpendiculars to edge MN, then one and the same point will serve as the foot of all the perpendiculars. Indeed, the line segment OB is perpendicular to the plane of the base and so $OB \perp MN$ (see Sec. 3.4), line segment OA is perpendicular to the plane MSN and so $OA \perp MN$. Hence, edge MN is perpendicular to the plane drawn through the intersecting lines OA and OB. If P is the point of intersection of this plane with the edge MN (note that this point may lie on the *extension* of the edge), then $AP \perp MN$, $BP \perp MN$ and so the point P is the common foot of the perpendiculars dropped from A and B on edge MN. Of course all these arguments hold true if in place of MN we take any other edge of the pyramid (as an exercise, carry out the reasoning for edge MS).

Theorem. *If a sphere is inscribed in a pyramid, the centre of the sphere is the point of intersection of the bisecting planes of all dihedral angles of the pyramid.*

Indeed, any point equidistant from both faces of a dihedral angle lies in the bisecting plane of the dihedral angle (see Sec. 3.2). Therefore, the centre of the inscribed sphere, since it is equidistant from all faces of the pyramid, must lie in each of the bisecting planes, that is, it is the point of intersection of the bisecting planes of all dihedral angles.*

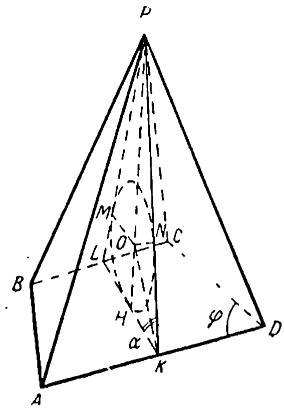
The centre of a sphere inscribed in a pyramid always lies inside the pyramid; this follows from the fact that all points of the bisecting plane lie *between* the faces of the dihedral angle.

It is also necessary to stress that if a sphere inscribed in a pyramid is projected on the plane of the base of the pyramid, the resulting circle will not be inscribed in the polygon of the base of the pyramid.

The next problem suggests how to carry through an exhaustive solution to a problem involving a combination of pyramid and sphere (and to other similar problems). It is of added interest since the pyramid is a "bad" one, the base being a trapezoid and not the usual regular polygon.

1. *A sphere is inscribed in a quadrangular pyramid $PABCD$ tangent to all its faces. The base of the pyramid is an isosceles trapezoid $ABCD$ with nonparallel side $AB = l$ and acute angle φ, and lateral faces APD*

Fig. 156



and BPC are isosceles triangles ($AP = PD$, $BP = PC$) forming with the base of the pyramid one and the same angle α. Find the radius of the inscribed sphere.

Let $PABCD$ be a pyramid with an isosceles trapezoid $ABCD$ as base; $AB = CD = l$, $\angle BAD = \angle ADC = \varphi$. The lateral faces of the pyramid, APD and BPC are isosceles triangles: $AP = PD$, $BP = PC$. Their bases are also the bases of the trapezoid (Fig. 156).

* It is easy to see that the converse is also true; *a sphere can be inscribed in a pyramid if the bisecting planes of all dihedral angles intersect in one point.* However, in the general case it does not follow from any fact that all bisecting planes intersect in a single point. An instance can easily be thought up of a quadrangular pyramid which does not have a point *common* to all eight bisecting planes. It is therefore impossible to inscribe a sphere in such a pyramid. It can be proved (and we advise the reader to try to do so) that (1) *a sphere can be inscribed in any triangular pyramid,* (2) *a sphere can be inscribed in a regular many-sided pyramid.*

Draw the slant height PK of face APD and the altitude PH of the pyramid. The edge AD, since it is perpendicular to two intersecting lines PH and PK, is perpendicular to the plane HPK; $BC \parallel AD$ and therefore the edge BC is perpendicular to the plane HPK and, in particular, to the line PL of intersection of this plane with the face BPC (PL is the slant height). Therefore $\angle PLK$ and $\angle PKL$ are the plane angles of the dihedral angles between the base and the lateral faces BPC and APD, respectively. It is given that $\angle PLK = \angle PKL = \alpha$.

By hypothesis, a sphere is inscribed in the given pyramid. It is required to find the radius of the sphere. The centre of the inscribed sphere lies on the bisecting plane of the dihedral angle between the planes BPC and APD since it is equidistant from both faces. Furthermore, by the same reasoning, the centre also lies on the bisecting plane of the dihedral angle between the planes APB and CPD. It therefore lies on the line of intersection of these two bisecting planes. We will prove that this line of intersection is the altitude PH of the pyramid. To do this, it suffices to demonstrate that the altitude PH lies in *each* of these bisecting planes.

First of all, we will prove that PH lies in the bisecting plane of the dihedral angle between the planes BPC and APD. Since plane LPK is perpendicular to plane APD (because plane APD passes through AD which is perpendicular to the plane LPK) and plane LPK is perpendicular to plane BPC (for similar reasons), it follows that the plane LPK is perpendicular to the line of intersection of the planes APD and BPC and, hence, the angle LPK is the plane angle of the dihedral angle between the faces BPC and APD. But $\angle PLK = \angle PKL$ and so $\triangle PKL$ is an isosceles triangle and the altitude PH is at the same time the bisector. As we know, the bisector of a plane angle corresponding to a dihedral angle lies in the bisecting plane of the dihedral angle.

Secondly, we will prove that PH lies in the bisecting plane of the dihedral angle between the planes APB and DPC. We wish to assure ourselves that the plane LPK is this bisecting plane. To do so, we will show that the distances from an arbitrary point S of this plane to the faces APB and DPC are the same. Mentally join S to the vertices P, A, B, C, and D of the pyramid. Since $V_{PABLK} = V_{PDCLK}$ (these pyramids have the same altitude PH and congruent bases $ABLK$ and $DCLK$) and (for the same reasons) $V_{SABLK} = V_{SDCLK}$, $V_{SAPK} = V_{SDPK}$, $V_{SBPL} = V_{SCPL}$, it then follows that $V_{SABP} = V_{SDCP}$. But the equality of volumes of the pyramids $SABP$ and $SDCP$, by virtue of the equality $\triangle ABP = \triangle DCP$ (by three sides) implies the equality of the altitudes dropped from S on the faces ABP and DCP, respectively. Thus, LPK is a bisecting plane and the altitude PH lies in this plane.

Thus, the centre of the inscribed sphere—denote it by O—does indeed lie on the altitude PH of the pyramid.

We now consider the isosceles triangle LPK. Inscribed in it is a great circle of our sphere because this plane passes through the centre of the sphere. (From the foregoing it follows that the sphere is tangent to the pyramid at H and at the points M and N lying on the slant heights PL and PK.) The problem thus reduces to the following problem of plane geometry: find the radius of a circle inscribed in the isosceles triangle LPK (base LK) with base angle α. LK is the altitude of the isosceles trapezoid $ABCD$ with acute angle φ and nonparallel side l and for this reason has length $l \sin \varphi$.

Since the centre of the inscribed circle lies at the point of intersection of the bisectors, $\angle OKH = \alpha/2$ and so the desired radius

$$r = OH = KH \tan\left(\frac{\alpha}{2}\right) = \frac{1}{2} l \sin\varphi \tan\left(\frac{\alpha}{2}\right)$$

Note that in the statement of the problem the bases of the trapezoid are not given. This does not however mean that the lengths of these bases can be arbitrary. The points of the line PH are of course equidistant from the planes APD and BPC, just as they are equidistant from the planes APB and CPD. But by no means for just any dimensions of the bases of the trapezoid is the distance from a point on the line PH to the plane APD equal to the distance from this point to the plane APB. To put it differently, not for arbitrary lengths of the bases AD and BC on the altitude PH is it possible to find a point equidistant from all five faces of the pyramid. Using the fact of the existence of an inscribed sphere, that is, assuming that such a point exists, we could, if we so desired, determine the dimensions of the trapezoid of the base.

Now let us take up the case when a sphere is *circumscribed* about a pyramid.

Definition. *A sphere is said to be circumscribed about an (arbitrary) pyramid if all vertices of the pyramid lie on the surface of the sphere.*

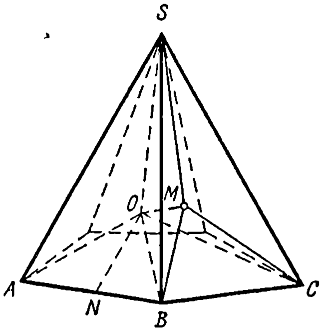


Fig. 157

Thus, the centre O of a circumscribed sphere (Fig. 157) is a point equidistant from all vertices of the pyramid. This means that if the centre of a circumscribed sphere is joined to the vertices S, A, B,

$C, \ldots$ of the pyramid, then all line segments $OS, OA, OB, OC, \ldots$ will have the same length.

Note the following. If from the centre O of the circumscribed sphere we drop a perpendicular OM on face BSC, the foot of the perpendicular will be the centre of the circle circumscribed about this face. Indeed, $MS = MB = MC$, being the projections of equal inclined lines OS, OB and OC on the plane BSC. (Repeating this argument for the base of the pyramid, it is easy to see that a circle can be circumscribed about the polygon of the base.) Thus, if a sphere is circumscribed about a pyramid, then the centre lies at the intersection of the perpendiculars erected to each of the faces of the pyramid at the centre of the circle circumscribed about that face. It will also be noted that if from point O we drop a perpendicular to any edge, say AB, then the foot N of the perpendicular will be the midpoint of the edge AB.

Theorem. *If a sphere is circumscribed about a pyramid, then its centre is the point of intersection of all planes drawn through the midpoints of the edges of the pyramid perpendicular to these edges.*

Indeed, any point equidistant from two vertices of the pyramid adjoining one edge lies in the plane drawn perpendicular to this edge through its midpoint (see Sec. 3.2). Therefore, the centre of a circumscribed sphere must, since it is equidistant from all vertices of the pyramid, lie in each of these planes, which is to say it is the point of intersection of all these planes.*

When making the drawing, the student often places the centre of the circumscribed sphere at random without proper visualization of the given spatial configuration and, all the more so, without any reasoning as to just what position the centre should occupy. As a rule, the centre is placed inside the pyramid. Yet the centre of a circumscribed sphere can lie inside, outside and on the surface of the pyramid (all depending on the type of pyramid). For example, in the next problem the centre of the circumscribed sphere turns out to be exterior to the pyramid, and this readily follows from the computations carried out in the solution.

2. *In a triangular pyramid $SABC$, edge BC is equal to a, $AB = AC$, edge SA is perpendicular to base ABC of the pyramid, the dihedral angle at the edge SA is equal to 2α, and that at the edge BC is equal to β. Find the radius of the circumscribed sphere.*

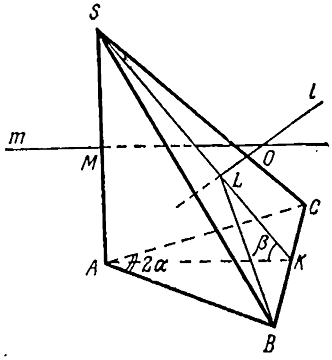
Consider the pyramid $SABC$ (Fig. 158). Since edge SA is perpendicular to the plane of the base, $\angle BAS = \angle CAS = 90°$, and so angle BAC is precisely the plane angle of the dihedral angle at edge

* A sphere cannot be circumscribed about an arbitrary pyramid. It is not hard to think up an example of a quadrangular pyramid about which a sphere cannot be circumscribed. We leave it to the reader to prove that *a sphere can be circumscribed about a pyramid if it is possible to circumscribe a circle about the base polygon of the pyramid.* From this it follows that it is possible (1) *to circumscribe a sphere about a triangular pyramid,* (2) *to circumscribe a sphere about a regular many-sided pyramid.*

SA. Thus, the base of the pyramid is an isosceles triangle with angle 2α at the vertex, and the altitude of the pyramid coincides with the edge SA.

Since the projections of the lateral edges SB and SC on the plane of the base are equal, the edges themselves are equal. Therefore face BSC is an isosceles triangle and its altitude dropped from vertex S

Fig. 158



hits the midpoint K of edge BC. By the theorem on three perpendiculars, AK is the altitude of triangle BAC, whence it is evident that the angle SKA is the plane angle of the dihedral angle at edge BC, i. e., $\angle SKA = \beta$.

The centre of a circumscribed sphere lies at the intersection of the straight line l, which passes through the centre of the circle circumscribed about triangle BSC, with the plane passing through the midpoint of edge AS perpendicular to AS. Line l lies in the plane ASK. Indeed, plane BSC passes through the line BC which is perpendicular to the plane ASK, that is, the planes BSC and ASK are perpendicular; and l is perpendicular to plane BSC and passes through the line of their intersection so that it lies in the plane ASK.

Fig. 159

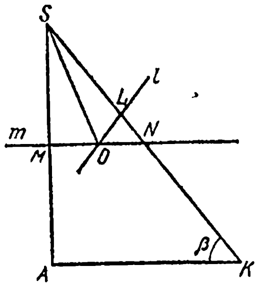
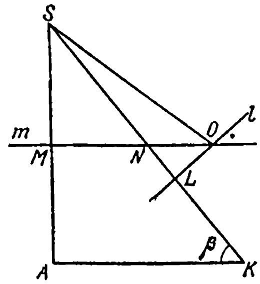


Fig. 160

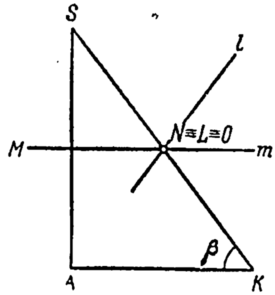


Thus, the centre of the sphere lies in the plane ASK. Let us examine this plane in a separate drawing. The centre O of the sphere will then lie at the intersection of l and line m, which is perpendicular to AS and passes through its midpoint. But there can, generally speaking, be three possibilities: the straight lines l and m intersect inside or outside of triangle ASK or on a side of it, and we will have to consider all three possibilities (Figs. 159, 160, 161). In the computations given below it will be shown that two of the three cases are invalid.

We are interested in the radius R of the circumscribed sphere, that is, the distance from point O (the point of intersection of the perpendiculars m and l to the sides of the angle KSA) to the points S, the vertex of the angle KSA.

Let us first seek SL, the projection of the desired distance on the side SK of triangle KAS. Since we know that in triangle AKB (Fig. 158)

Fig. 161



leg $BK = 1/2\ a$ and angle $KAB = \alpha$, it follows that $AK = 1/2 \times \times a \cot \alpha$. Furthermore, from triangle KAS we have

$$SK = \frac{a \cot \alpha}{2 \cos \beta}$$

Since L is the centre of the circle circumscribed about triangle BSC, it follows that $LS = LB$ and so from triangle BKL we find $(SK - SL)^2 + KB^2 = SL^2$, that is,

$$SL = \frac{a(\cot^2 \alpha + \cos^2 \beta)}{4 \cot \alpha \cos \beta}$$

Noting that the above computations of SL are in no way dependent on the location of the centre O of the circumscribed sphere, we return to Figs. 159, 160, 161. Denote by N the point of intersection of the line m with side SK. It is quite clear that l and m intersect *outside* triangle KAS if $SN < SL$ (Fig. 160); but if $SN > SL$, then O lies inside this triangle (Fig. 159); finally, if $SN = SL$, then O lies on the side SK of the triangle (Fig. 161). Let us try to find out which of these locations actually occurs.

Since MN is the midline of triangle KAS, $SN = 1/2\ SK$. Comparing the lengths of SN and SL, we easily prove that for arbitrary a, α and β

$$\frac{a \cot \alpha}{4 \cos \beta} < \frac{a(\cot^2 \alpha + \cos^2 \beta)}{4 \cot \alpha \cos \beta}$$

(from geometric arguments it follows that $a > 0$, $0° < \alpha < 90°$ and $0° < \beta < 90°$). Hence, no matter what the dimensions of a, α and β

of the pyramid $SABC$, the centre O of the circumscribed sphere will always be exterior to the pyramid. This in turn means that the plane configuration in plane KAS can only be of the form indicated in Fig. 160, and that the arrangements shown in Figs. 159 and 161 do not occur.

Considering Fig. 160, it is easy to demonstrate that $\angle ONL = \beta$ and so $LO = NL \tan\beta = (SL - SN)\tan\beta$. Substituting into this equation the expressions obtained above for SL and SN, we get (after obvious manipulations) $LO = 1/4\, a \tan\alpha \sin\beta$. Finally, from the right triangle OLS we find

$$R=\sqrt{LO^2+SL^2}=\frac{a}{2\sin 2\alpha\cos\beta}\sqrt{\cos^2\beta+\sin^2\beta\cos^4\alpha}$$

The computations in this problem were thus very simple, the chief difficulty being in the arguments that establish the location of the centre of the circumscribed sphere.

Combinations of pyramid and inscribed or circumscribed sphere are usually solved by students in a satisfactory manner, but other cases of pyramid and sphere give rise to nearly insuperable difficulties. It is a fact that few are able to visualize a spatial configuration, say, in the case of a sphere tangent to all edges of a triangular pyramid, a sphere tangent to the base and passing through the vertex, of a sphere touching two skew edges of a triangular pyramid, and the like. In all these cases, the crux of the problem lies in establishing the common properties (needed for the computations) of the centre of the sphere and the elements of the pyramid.

There is no possibility of considering and investigating all possible mutual positions of sphere and pyramid. We therefore confine ourselves to two examples.

3. *A sphere of radius r is tangent to all edges of a triangular pyramid. The centre of the sphere lies inside the pyramid on its altitude at a distance of $r\sqrt{3}$ from the vertex. Prove that the pyramid is regular and find the altitude of the pyramid.*

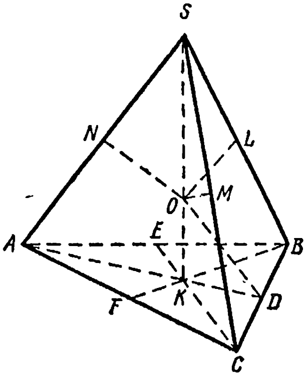
Let the points M, N and L be the points of tangency of the sphere and the lateral edges of the pyramid $SABC$, and D, E and F the points of tangency of the sphere and all sides of the base (Fig. 162). Let O denote the centre of the sphere, which is stated to be on the altitude SK of the pyramid.

By the definition of tangent lines to a sphere, $ON \perp AS$, $OL \perp BS$, $OM \perp CS$; $ON = OM = OL = r$, whence it follows that the right triangles SNO, SLO and SMO are equal and so $SN = SL = SM$ and $\angle NSO = \angle LSO = \angle MSO$.

The last equality of the angles permits us to conclude that triangles AKS, BKS and CKS are equal; hence, $AS = BS = CS$, which means that all the lateral edges of our pyramid are equal. This of course does not yet let us affirm that the pyramid is a regular pyramid.

The equality of lateral edges and the equality of line segments SN, SL and SM show that $AN = BL = CM$. Now take advantage of the fact that the tangents to a sphere drawn from a single point are equal: $AN = AF = AE$, $BL = BE = BD$, $CM = CF = CD$. Therefore, $AN = BL = CM = AF = AE = BE = BD = CF = CD$, whence it follows that the points D, E and F are the midpoints

Fig. 162



of the edges of the base, and $AB = BC = CA$, which means the triangle ABC is equilateral. The equality of the lateral edges of the pyramid implies equality of their projections: $AK = BK = CK$, which means K—the foot of the altitude—is the centre of an equilateral triangle ABC.

We have thus proved that the pyramid $SABC$ is regular. Let us find the length of the altitude SK.

The right triangles AKS and ONS have a common acute angle at the vertex S and are similar, and so $SK = AK \cdot NS/NO$. For the sake of brevity, denote the altitude SK by h. We have seen that $ON = r$. From the right triangle SNO we get $NS = \sqrt{SO^2 - NO^2} = r\sqrt{2}$ and so $h = AK \cdot \sqrt{2}$ and it only remains to determine one more relation between h and AK.

This relation can be obtained from the right triangle OKD. In this triangle, $KD = 1/2\ AK$, $OD = r$ and $OK = h - r\sqrt{3}$; by the Pythagorean theorem we have $1/4\ AK^2 = r^2 - (h - r\sqrt{3})^2$. Substituting $AK = h/\sqrt{2}$ into this equation, we get a quadratic equation in h having two positive roots:

$$h_1 = \frac{4}{3} r\sqrt{3}, \quad h_2 = \frac{4}{9} r\sqrt{3}$$

But the second root does not satisfy the condition of the problem. The centre of the sphere has to lie *inside* the pyramid on its altitude at a distance of $r\sqrt{3}$ from the vertex. This means the altitude of the pyramid must *exceed* $r\sqrt{3}$, whereas the second root h_2 is less than

$r\sqrt{3}$. The first root h_1 yields the desired altitude of the pyramid, $h=4/3\ r\sqrt{3}$.

Note that we did not depict the sphere in the drawing. The important thing in this case is that the sphere in the problem goes beyond the limits of the pyramid and intersects the faces of the pyramid (some students attempted to represent the sphere in the drawing as tangent to the faces!).

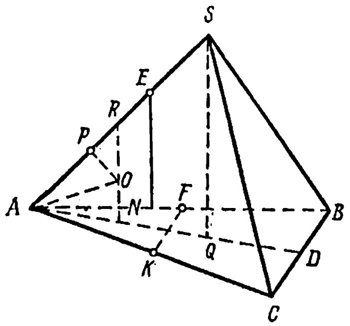
4. *Given a regular triangular pyramid SABC (S the vertex) with base side a and lateral edge $a\sqrt{2}$. A sphere passes through point A and is tangent to the lateral edges SB and SC at their midpoints. Find the radius of the sphere.*

In this problem, too, any attempt to visualize the entire configuration does not simplify the solution. We have to take advantage solely of the fact of tangency.

This problem is most easily solved by means of a theorem of plane geometry: the square of a tangent to a circle is equal to the product of the secant by its exterior portion.

Knowing that the sphere is tangent to the edge SC at its midpoint and passes through vertex A, we find another two points in which the sphere intersects edges AC and AS (Fig. 163). Indeed, the plane ACS

Fig. 163



intersects the sphere along a circle which is tangent to the straight line CS (at the midpoint of CS) and which passes through A. Using the foregoing theorem, we find that this circle cuts edge AC at the midpoint (point K) and edge AS at E such that $SE=1/2\ a\sqrt{2}$.

Similar arguments concerning plane ABS permit finding another two points of intersection of the sphere with the edges AB and AS: F, the midpoint of edge AB and the same point E of edge AS. We can now find the radius of the sphere by taking advantage solely of the points A, K, F and E.

Since the sphere passes through A, K and F, the centre of the sphere lies on the perpendicular to the plane AKF erected from the centre of $\triangle AKF$. Since point N (the centre of $\triangle AKF$) lies on the median AD of the triangle ABC, and $AN = 2/3\ (AD/2) = 1/3\ AD$, the centre of the sphere lies in the plane ADS, since, as we know, the altitude SQ of the pyramid also lies in this plane. Since $AQ = 2/3\ AD$, it follows that $AN = NQ$ and, hence, the centre of the sphere lies on the midline NR of triangle ASQ.

But the centre of the sphere also lies in the plane perpendicular to AE and passing through the midpoint of AE, i.e. point P. Hence, the centre of the sphere lies on the straight line lying in the plane ASD perpendicular to edge AS and passing through point P. Thus, the centre of the sphere lies in the plane ASQ at the intersection of the following straight lines: the midline of $\triangle ASQ$ and the perpendicular to the line segment AE at its midpoint.

Now let us take up the computations:

$$R = AO = \sqrt{AP^2 + PO^2}$$

Since $AP = 1/2\ AE = 1/2\ (AS - ES) = 3/8\ a\sqrt{2}$, it remains to find PO. From the similarity of the triangles RPO and RAN we find $PO = RP \cdot AN/RN$. Since $RP = AR - AP = 1/8\ a\sqrt{2}$, $AN = 1/3\ AD = 1/6\ a\sqrt{3}$, $RN = 1/2\ SQ = a\sqrt{15}/6$, then $PO = a\sqrt{10}/40$, whence $R = a\sqrt{115}/20$.

It is of course impossible to provide for all kinds of other combinations of geometric solids. Below is a list of definitions of some of the more or less ordinary configurations; we also consider a problem involving tangency of sphere and cone. There is no need to memorize these definitions, the important thing is to get a clear-cut understanding of the geometric picture. In our solution of the problem we strive to show that the crux lies in visualizing the specific mutual arrangement of the solids.

A sphere is inscribed in a prism if it touches all faces of the prism. If the prism is a right prism, then the orthogonal projection of the sphere on the plane of the base of the prism is a circle inscribed in a polygon (the base of the prism). This does not hold true for an oblique prism. In any case, the altitude of the prism is equal to the diameter of the sphere.

A sphere is inscribed in a right circular cone if it is tangent to the base of the cone and also to the lateral surface. The point of tangency of the sphere and the base is the centre of the base; the sphere is tangent to the lateral surface along a circle (which is not a great circle!), the plane of which is parallel to the plane of the base. The centre of the sphere lies on the altitude of the cone.

A sphere is inscribed in a right circular cylinder if it is tangent both to the bases of the cylinder and to its lateral surface. The points of tangency of the sphere with the bases are the centres of the bases. The sphere is tangent to the lateral surface along a great circle of the sphere parallel to the bases. The centre of the sphere lies on the axis of the cylinder.

The diameter of the base of the cylinder is equal to the diameter of the sphere and is equal to the altitude of the cylinder.

*A sphere is inscribed in a truncated pyramid (with parallel bases)** *(or a truncated right circular cone) if it is tangent to the bases and to the lateral surface.* The diameter of the sphere is equal to the altitude of the truncated pyramid (truncated cone).

A right circular cylinder is inscribed in a prism if its lateral surface is tangent to the lateral faces of the prism, and the bases (circles inscribed in polygons) are the bases of the prism. If a right circular cylinder is inscribed in a prism, then the prism is a right prism. The lines along which the lateral surface of the cylinder is tangent to the lateral faces of the prism are straight lines perpendicular to the bases of the prism.

A right circular cylinder is inscribed in a pyramid if the circle of one of its bases is tangent to all lateral faces of the pyramid and the other base lies on the base of the pyramid. It is to be noted that the pyramid need not be a regular pyramid. If the cylinder is inscribed in the pyramid, then, firstly, the foot of the altitude of the pyramid lies inside (or on the sides of) the polygon forming the base of the pyramid, and, secondly, the base of the pyramid is a polygon in which it is possible to inscribe a circle (however, the base of the cylinder lying on the base of the pyramid is not a circle inscribed in the base of the pyramid!).

A right circular cylinder is inscribed in a right circular cone if the circle of one of the bases of the cylinder lies on the lateral surface of the cone and the other base of the cylinder lies on the base of the cone. The axis of the cylinder lies on the altitude of the cone. It is possible to inscribe an infinity of cylinders in every cone.

A right circular cylinder is inscribed in a sphere if the circles of the bases lie on the sphere. The bases of the cylinder are small circles of the sphere; the centre of the sphere coincides with the midpoint of the axis of the cylinder.

A right circular cone is inscribed in a prism if its vertex lies on the top base of the prism and its base (a circle inscribed in a polygon) is the bottom base of the prism. If a cone is inscribed in a prism, then, firstly, the bases of the prism are polygons in which it is possible to inscribe a circle and, secondly, the straight line perpendicular to the bottom base and passing through the centre of the circle inscribed in the polygon (the bottom base) intersects the top base (and not its extension). The altitude of the cone is equal to the altitude of the prism.

A right circular cone is inscribed in a pyramid if the vertex of the cone coincides with the vertex of the pyramid, and the base of the pyramid is a polygon circumscribed about the circle of the base of the cone. If the cone is inscribed in a pyramid, then, firstly, the base of the pyramid is a polygon in which we can inscribe a circle and, secondly, the

* Also termed a frustum of a pyramid.

altitude of the pyramid passes through the centre of this circle. The altitudes of the pyramid and cone coincide.

A right circular cone is inscribed in a sphere if the vertex and the circle of the base lie on the sphere. The base of the cone is a small (or great) circle of the sphere. The centre of the sphere lies on the altitude of the cone.

A pyramid is inscribed in a cone if the vertex of the pyramid coincides with the vertex of the cone and the base of the pyramid is a polygon inscribed in the circle of the base of the cone. The altitudes of the pyramid and cone coincide; the lateral edges of the pyramid lie on the lateral surface of the cone.

A prism is inscribed in a right circular cone if all vertices of the top base of the prism lie on the lateral surface of the cone and the bottom base of the prism lies on the base of the cone. The base of the prism is a polygon about which we can circumscribe a circle (but the bottom base of the prism is not inscribed in the circle of the base of the cone!).

A prism is inscribed in a cylinder if its bases are polygons inscribed in the circles of the bases of the cylinder. The base of a prism is a polygon about which a circle can be circumscribed; the prism is a right prism and its altitude is equal to the altitude of the cylinder.

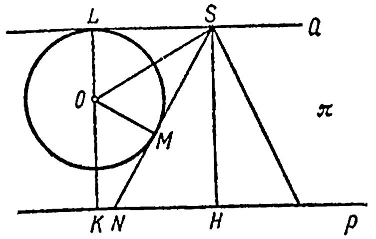
A prism is inscribed in a sphere if all its vertices lie on the sphere. The prism is a right prism; its base is a polygon which can be inscribed in a circle.

5. *Given three right circular cones with angle α ($\alpha < 2\pi/3$) in the axial section and radius of base equal to r. The bases of these cones lie in one plane and are externally tangent to one another in pairs. Find the radius of the sphere tangent to all three cones and to the plane passing through their vertices.*

The configuration of solids in the problem is rather easy to visualize but extremely difficult to depict. Now it is not really necessary to get a general view of this configuration; it will be enough to visualize it in one's mind's eye. Descriptively, we can say that the sphere is embedded in a funnel between three equal cones (standing on one and the same horizontal plane P so that their bases touch one another externally in pairwise fashion) and the size of the sphere is such that it is tangent to the "roof" of the plane Q that lies on the three vertices of the cones.

However, before solving the problem, we have to make precise the meaning of the words "a sphere embedded in the funnel between cones" in strict mathematical terms. It is clear that the sphere has only one common point with the lateral surface of each of the cones, or, as we say, is tangent to the lateral surface of the cone externally. What this means is that if we pass a cutting plane through the altitude of the cone and the centre of the sphere, then the resulting great circle in the section of the sphere is externally tangent to a side of the isosceles triangle that is the section of the cone (Fig. 164). Such, precisely, is the *definition* of tangency of a sphere, exterior to a cone, to the lateral surface of the cone. In other words, if M is the point of tangency of the sphere and the cone, then the sphere touches the generatrix SMN of the cone (with vertex S) at point M and the radius OM of the sphere is perpendicular to the generatrix.

Fig. 164



Now let Fig. 164 depict the section of one of the cones of our configuration, of the sphere and of the planes P and Q (the planes of the bases of the cones and the planes of their vertices) cut by the plane π passing through the altitude SH of this cone and the centre O of the sphere (the other two cones are not depicted). The fact that the cutting plane π will be perpendicular to the parallel planes P and Q follows from the fact that it passes through altitude SH of the cone, which altitude is perpendicular to the P plane of its base and so is perpendicular to the Q plane (the P and Q planes are parallel).

Let M be the point of tangency of sphere and cone; then, by virtue of what has already been said, the great circle obtained in the section of the sphere by the π plane is tangent to the generatrix NS of the cone at point M. However, the fact that this same circle is tangent to the straight line along which the planes π and Q intersect requires special proof.

Denote by L the point of tangency of the sphere and the plane Q. Any straight line in the Q plane passing through L and, in particular, LS, will be tangent to the sphere, that is to say, it will be perpendicular to the radius OL and so $OL \perp LS$. Since the radius OL of the sphere drawn to the point of tangency is perpendicular to the Q plane, it follows that $OL \parallel SH$, these being two perpendiculars to one and the same plane Q. But two parallel lines lie in one plane; this implies that the radius OL lies in the π plane drawn through SH and the point O. In other words, the point L of tangency of the sphere and the Q plane lies in the π plane. This in turn means that LS (the line of intersection of the planes π and Q) is tangent to the great circle obtained in the section of the sphere cut by the π plane. Indeed, the straight line LS passes through the endpoint of the radius OL of this great circle perpendicular to OL.

Extend LO to intersection with NH at point K. It is clear that $LSHK$ is a rectangle, whence, for one thing, it follows that $\angle LSH = 90°$ and $KH = LS$. Since $\angle NSH = \alpha/2$, by hypothesis, then, joining the points O and S, we easily find that $\angle OSM = (\pi - \alpha)/4$ and so, from the triangle OSM,

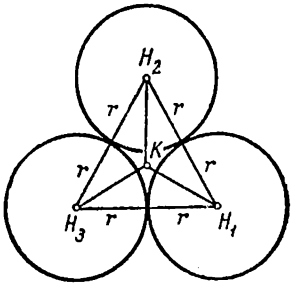
$$SM = R\cot\frac{\pi-\alpha}{4}$$

where R is the desired radius of the sphere. Taking into account that $LS = SM$ (two tangents drawn from one point to one and the same circle), we get

$$KH = R\cot\frac{\pi-\alpha}{4}$$

From this it follows that the distance from the projection K of the centre O of the sphere on the P plane to the centre of the cone does not depend on which of the three given cones we consider. In other words, K is equidistant from all three centres of the bases of the cones

Fig. 165



and so KH is readily found (Fig. 165) as the radius of the circle circumscribed about an equilateral triangle with side $2r$, namely, $KH = 2r/\sqrt{3}$. We now have

$$R = \frac{2r}{\sqrt{3}}\tan\frac{\pi-\alpha}{4}\,*$$

Exercises

1. A right circular cone is inscribed in a sphere of radius R. Find the lateral surface area of the cone if the altitude is h.

2. The side of a regular tetrahedron is equal to a. Determine the radius of a sphere tangent to the lateral faces of the tetrahedron at points lying on the sides of the base.

---

* Note, as a supplement to this solution, that the angle α in the axial section of the cone always satisfies the inequalities $0 < \alpha < \pi$ and for all such α the formula we obtained for the radius of the sphere is obviously meaningful. This does not however mean that for any α the geometric configuration of the problem exists: it is rather easy to see that if the angle α is great, that is, the cones are broad and low, then the radius of the sphere tangent to them will be great and it will pass above the plane passing through the vertices of the cones. For a sphere having the requisite properties to exist, it is obviously necessary and sufficient that the point of intersection of the bisector SO and the straight line KL be projected on the generatrix of the cone and not on its extension; that is, that the inequality $SM < l$ or, what is the same thing, $2r/\sqrt{3} < r$ cosec $(\alpha/2)$, hold true, whence $\sin(\alpha/2) < \sqrt{3}/2$, i.e., $\alpha < 2\pi/3$: It is precisely this restriction that is imposed on α in the statement of the problem.

3. A sphere is inscribed in a right circular cone. The radius of the circle of tangency of the surface of the sphere and the lateral surface of the cone is equal to r. The radius of the base of the cone is equal to R. Determine the lateral surface of the sphere.

4. A sphere is inscribed in a truncated cone (parallel bases) of volume V. The sphere (of radius R) is tangent both to the lateral surface and to both bases. What is the angle formed by the generatrix of the cone and the greater base?

5. Inscribed in a right circular cone is a sphere whose surface is equal to the area of the base of the cone. In what ratio is the lateral surface of the cone divided by the line of tangency of the sphere and cone?

6. A pentagonal pyramid is circumscribed about a right circular cone of altitude equal to the radius of the base of the cone. The total surface area of the pyramid is twice that of the cone. Find the volume of the pyramid if the lateral surface area of the cone is equal to $\pi\sqrt{2}$.

7. The base of a pyramid is a right triangle and the lateral faces passing through the legs form angles of 30° and 60° with the base. A right circular cone is circumscribed about the pyramid. Find the volume of the cone if the altitude of the pyramid is equal to h.

8. A right circular cylinder is inscribed in a sphere. How many times does the volume of the sphere exceed that of the cylinder if it is known that the ratio of the radius of the sphere to the radius of the base of the cylinder is half the ratio of the surface of the sphere to the lateral surface of the cylinder?

9. Inscribed in a sphere of radius a is a regular tetrahedron. Find the volume of the tetrahedron.

10. In a regular triangular pyramid the ratio of the radii of a circumscribed and an inscribed sphere is 3. Find the ratio of the volume of the pyramid to the volume of the inscribed sphere.

11. The base of a pyramid is an isosceles triangle, the two sides of which are equal to b; the corresponding lateral faces are perpendicular to the plane of the base and form, between each other, an angle of α. The angle between the third lateral face and the plane of the base is also equal to α. Find the radius of the sphere inscribed in the pyramid.

12. Find the volume of a regular triangular pyramid, knowing the radius r of an inscribed sphere and the angle α of inclination of its lateral face to the base.

13. A sphere is inscribed in a truncated cone (parallel bases) whose generatrix is of length l and forms with the base an angle α. Determine the radius of the circle along which the sphere is tangent to the truncated cone.

14. A sphere is inscribed in a spherical sector. The radius of the circle along which the sphere is tangent to the sector is equal to r. A diametral plane of the spherical sector cuts out of it a circular sector with central angle 2φ. Determine the radius of the spherical sector.

15. The edge of a cube is given as a. A sphere with centre O intersects three edges (at their midpoints) converging to the vertex A. A perpendicular is dropped from the point B of intersection of the sphere with one of the edges of the cube onto the diagonal of the cube passing through the vertex A, the angle between the perpendicular and the radius OB being divided in half by the edge of the cube. Find the radius of the sphere.

16. Five equal spheres are placed in a cone. Four of them lie on the base of the cone, and each of these four spheres is tangent to two others lying on the base and also to the lateral surface of the cone. The fifth sphere is tangent to the lateral surface and to the other four spheres. Determine the volume of the cone if the radius of each sphere is equal to R.

17. Two equal spheres of radius r are tangent to each other and to the faces of a dihedral angle equal to α. Find the radius of the sphere that is tangent to the faces of the dihedral angle and to both given spheres.

18. The altitude of a cone is four times the radius of a sphere inscribed in the cone. The generatrix of the cone is equal to l. Find the lateral surface area of the cone and the radius of the sphere circumscribed about the cone.

19. The lateral surface area of a regular triangular pyramid with base side a is 5 times the area of the base. Find the volume of the cone inscribed in the pyramid.

20. A sphere of radius r is tangent to all edges of a regular quadrangular pyramid with base side a. Find the volume of the pyramid.

21. A sphere is inscribed in a cube with edge a. Determine the radius of another sphere tangent to three faces of the cube and to the first sphere.

22. Given in a pyramid $SABC$ (S vertex, ABC base): $AB=AC=a$, $BC=b$. The altitude of the pyramid passes through the midpoint of the altitude AD of the base; the dihedral angle τ formed by the faces SBC and ABC is equal to $\pi/4$. A cylinder whose altitude is equal to the diameter of the base is inscribed in this pyramid so that one base is tangent to the faces of the dihedral angle τ, the other, to the faces of a trihedral angle A; the axis of the cylinder is parallel to AD. Find the radius of the base of the cylinder.

23. The edge of a cube is equal to a. Find the volume of a right circular cylinder inscribed in the cube so that the axis is the diagonal l of the cube and the circles of the bases are tangent to those diagonals of the faces of the cube which do not have any points in common with the diagonal l of the cube.

24. Three spheres of radius r lie on the base of a right circular cone. On top of them lies a fourth sphere of radius r. Each of the four spheres is tangent to the lateral surface of the cone and to the three other spheres. Find the altitude of the cone.

25. Three spheres of radius r lie on the bottom base of a regular triangular prism, each of them is tangent to the two other spheres and to two lateral faces of the prism. On these spheres lies a fourth one which is tangent to all lateral faces and to the top base of the prism. Determine the altitude of the prism.

26. Four equal spheres of radius r are externally tangent to each other so that each is tangent to the three others. Find the radius of a sphere tangent to all four spheres and containing them within it.

27. A sphere is inscribed in a right circular cone. The ratio of the volumes of the cone and sphere is equal to two. Find the ratio of the total surface area of the cone to the surface area of the sphere.

28. In a regular triangular pyramid the altitude is h and the side of the base is a. One of the vertices of the base is the centre of a sphere that is tangent to the opposite face of the pyramid. Find the area of those portions of lateral faces of the pyramid that are located inside the sphere.

29. Given a regular triangular pyramid $SABC$ (S vertex) with base side a and lateral edge b ($b > a$). A sphere lies above the plane of the base ABC, is tangent to this plane at point A, and, besides, is tangent to a lateral edge SB. Find the radius of the sphere.

30. Given a regular quadrangular pyramid $SABCD$ (S vertex) with base side a and lateral edge a. A sphere with centre at O passes through A and is tangent to edges SB and SD at their midpoints. Find the volume of the pyramid $OSCD$.

Chapter 4 NONSTANDARD PROBLEMS

4.1 Introduction

Nonstandard problems appear in a number of types. Some of them apper to be very unusual, so much so that one does not know how to attack them. Others are in the form of quite an ordinary equation, which however is not solved by standard procedures. A third type involves very subtle and clear-cut logical reasoning. A fourth type ... well, actually, there are many different kinds of nonstandard problems and it is rather hopeless to strive to embrace them all.

Nonstandard problems require a definite degree of ingenuity and mastery of various sections of mathematics and a high degree of logical culture, so to speak. And besides all this, a proper frame of mind. Time and again, a problem that essentially is not complicated but is formulated in an unusual manner has caused insuperable difficulties, whereas the solution only required a few words.

It is of course impossible to indicate all possible methods of working nonstandard problems. They require the use of graphs and diverse properties of functions, inequalities and, last but not least, *logic*.

A word about the term *nonstandard*. We believe that this term refers not to complexity but rather to the unusual nature and aspect of the problem. However, we hope that after the student has worked through this chapter he will regard the methods and the problems given here as normal and standard.

We will examine the solutions of a number of nonstandard problems and in some cases will offer a variety of approaches to their solution. The reader will notice that many of the solutions are not really involved at all and can be grasped at once. What is more complicated is the finding of these solutions. We therefore suggest that the student make an attempt to solve each problem before going on to read the solution given in the book.

Another thing is to differentiate strictly in these problems between the rough solution and the final version. The point is that during the exploratory stage, the student seeks an approach to the problem and carries out various computations and runs down chains of logic without yet knowing whether they will actually be of any use or not. All this results in what we call a rough solution. Here the logic is often incomplete, with gaps and hazy statements, extra, frequently unneeded, statements, and the like. The teacher is not interested in how the solution was arrived at so long as the problem is solved correctly and every step is justified and substantiated. This constitutes the final version of the solution, and that is what counts in the final analysis. This final version need not of course include every single step of the rough solution, and so the task of the student consists in polishing off his rough solution to a final solution.

In some of the cases given below we include both types but in most cases we confine ourselves only to the rough solution and leave it to the reader to perform the exceedingly useful exercise involved in rewording the rough version to a final version of the solution.

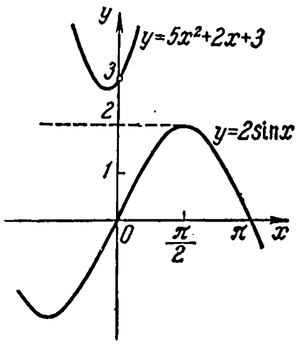
4.2 Problems that are nonstandard in aspect

A cursory glance at the problems offered in this section will give the reader an idea of what is meant by the title. It is clear from the very start that ordinary manipulations, the use of algebraic or trigonometric formulas will not yield the desired result if certain othei reasoning is not invoked.

This other reasoning ordinarily involves inequalities, graphs and, generally, the various properties of functions.

1. *Solve the equation* $2 \sin x = 5x^2 + 2x + 3$.

Fig. 166



To approach this equation, let us first see how the graphs of the left and right members behave. If the curves intersect, then the abscissas of the points of intersection will be the roots of the equation, and if they do not, then the equation does not have any roots. From Fig. 166 it is evident that the equation has no roots.

This completes the rough solution, but it is by no means the final version, for a graph is not a proof and it still must be demonstrated formally that the equation has no roots. Our graphical solution is of course no loss because we will need it to find a rigorous proof.

Let us take a closer look to see what properties of the graphs indicate that the graphs do not intersect. This follows from the fact that the graph of the function $y = 5x^2 + 2x + 3$ lies everywhere *above* the graph of the function $y = 2 \sin x$. Going from geometric to algebraic terms, this means that for any x the inequality

$$5x^2 + 2x + 3 > 2 \sin x$$

holds true.

This is prec sely the inequality that we must prove in rigorous fash ion. Indeed, on the one hand, for any x,

$$5x^2 + 2x + 3 = 5\left(x + \frac{1}{5}\right)^2 + \frac{14}{5} \geqslant \frac{14}{5} > 2$$

yet, on the other, $2 \sin x \leqslant 2$ also for arbitrary values of x; and therefore for any x the inequality to be proved is valid.

But now that we have a rigorous solution and an exact proof that the given equation has no roots, why do we need to go through with all the previous arguments? They have already performed their task, that of helping us to find a rigorous proof. The final version of the solution can then be written as follows:

"The given equation has no roots; indeed, for arbitrary x we have the inequalities

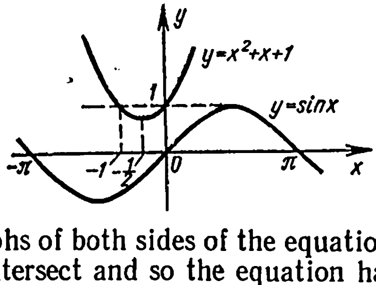
$$5x^2 + 2x + 3 = 5\left(x + \frac{1}{5}\right)^2 + \frac{14}{5} \geqslant \frac{14}{5} > 2 \text{ and } 2 \geqslant 2 \sin x$$

that is, $5x^2 + 2x + 3 > 2 \sin x$, which completes the proof."

The hidden element of this solution is that we first made use of non-rigorous reasoning and solved the problem by means of graphs.

2. *Solve the equation* $\sin x = x^2 + x + 1$.

Fig. 167



Construct the graphs of both sides of the equation (Fig. 167). Again, the curves do not intersect and so the equation has no roots. But the same proof is not applicable here because the parabola dips below the straight line $y = 1$.

Still, the graph suggests how we are to find a rigorous proof. From the property of a quadratic trinomial it follows that for $x > 0$ and for $x < -1$, the inequality $x^2 + x + 1 > 1$ holds true, so that the inequality $x^2 + x + 1 > \sin x$ is valid for all these values of x. And on the remaining interval $-1 \leqslant x \leqslant 0$, the inequalities $x^2 + x + 1 > 0$ and $\sin x \leqslant 0$ are valid, so that here too (on this interval) the equation has no roots.

It is not always necessary to construct graphs in solving such equations. Occasionally the very construction of the graphs may turn out to be complicated. Actually, in the solution we will take advantage not so much of graphs as of various inequalities, the graphs merely hinting at the approach to a proof. In more complicated instances, we have to seek the proof in a strictly formal manner without having before us a visual (geometric) image. Naturally, this is somewhat more complicated.

3. *Solve the equation* $2\cos^2 \frac{x^2+x}{6} = 2^x + 2^{-x}$.

Clearly, it is best not to try to construct the graph of the left member, all hope resting on inequalities. On the one hand, for any x we have the inequality $2\cos^2 \frac{x^2+x}{6} \leqslant 2$; on the other, $2^x + 2^{-x} \geqslant 2$— the sum of positive reciprocal quantities. Therefore, the left and right members of the original equation are equal if and only if they are both equal to 2.

In other words, the following system of *two* equations in *one* unknown must hold:

$$2\cos^2 \frac{x^2+x}{6} = 2$$

$$2^x + 2^{-x} = 2.$$

The second equation of this system has the unique root $x = 0$. This root satisfies the first equation too, that is, it is the sole solution of the system and, hence, of the original equation.

The next few problems also reduce to systems of two equations in one unknown, thanks to the use of the properties of trigonometric functions.

4. *Solve the equation* $\cos^7 x + \sin^4 x = 1$.

Since $\cos^7 x \leqslant \cos^2 x$ and $\sin^4 x \leqslant \sin^2 x$, the left member of the given equation does not exceed unity and is equal to unity only when equality occurs in both the above weak inequalities; in other words, when the following system of equations holds true:

$$\cos^7 x = \cos^2 x$$

$$\sin^4 x = \sin^2 x$$

The first equation is satisfied for $\cos x = 0$ and for $\cos x = 1$. But the second equation is also satisfied for these values of x: if $\cos x = 0$, then $\sin^2 x = 1$, and if $\cos x = 1$, then $\sin x = 0$. Therefore, the solutions of the system and, hence, of the original equation, are $x = \pi/2 + k\pi$ and $x = 2k\pi$, where k is any integer.

5. *Solve the equation* $\sin^4 x - \cos^7 x = 1$.

The reasoning of Problem 4 does not directly apply in this case but it can be modified to some extent. It clearly follows from this equation that $\cos^7 x \leqslant 0$ (otherwise $\sin^4 x > 1$), that is $\cos x \leqslant 0$. But then $|\cos^7 x| = -\cos^7 x$, and the equation is rewritten as

$$\sin^4 x + |\cos x|^7 = 1$$

Now we can argue as in the preceding case:

$$\sin^4 x \leqslant \sin^2 x, \quad |\cos x|^7 \leqslant |\cos x|^2 = \cos^2 x$$

and we finally get the system of equations

$$\sin^4 x = \sin^2 x$$
$$|\cos x|^7 = |\cos x|^2$$

The second equation is satisfied for $|\cos x| = 0$ and for $|\cos x| = 1$. From $|\cos x| = 0$ we get the solutions $x = \pi/2 + k\pi$, $k = 0, \pm 1, \pm 2, \ldots$ which also satisfy the first equation. But if $|\cos x| = 1$, then $\cos x = -1$ (since $\cos x \leqslant 0$) and $x = (2k+1)\pi$. These values also satisfy the first equation. Hence, the solution of the original equation is given by two sets (groups):

$$x = \frac{\pi}{2} + k\pi \text{ and } x = (2k+1)\pi \quad (k \text{ any integer})$$

Here is another method of solution that merely reduces the given equation to the previous one. Replacing x by $\pi - y$, we get the equation $\sin^4(\pi - y) - \cos^7(\pi - y) = 1$ or, using the reduction formulas, the equation

$$\sin^4 y + \cos^7 y = 1$$

which was analyzed above. Its solutions are: $y = \pi/2 + k\pi$ and $y = 2k\pi$, whence, since $x = \pi - y$, we have

$$x = \frac{\pi}{2} - k\pi \text{ and } x = \pi - 2k\pi \quad (k \text{ any integer})$$

6. *Solve the equation*

$$\cos^2\left[\frac{\pi}{4}(\sin x + \sqrt{2}\cos^2 x)\right] - \tan^2\left(x + \frac{\pi}{4}\tan^2 x\right) = 1$$

Since the square of the cosine of any argument does not exceed 1, the given equation holds true if and only if we have, simultaneously,

$$\cos^2\left[\frac{\pi}{4}(\sin x + \sqrt{2}\cos^2 x)\right] = 1 \text{ and } \tan\left(x + \frac{\pi}{4}\tan^2 x\right) = 0$$

We thus have a system of two equations in one unknown. To solve it, we find the roots of the first equation and then substitute them into the second equation, and then choose those that satisfy the second equation and, hence, the system.

From the first equation we at once get

$$\sin x + \sqrt{2} \cos^2 x = 4k$$

where k is an integer. However,

$$|\sin x + \sqrt{2} \cos^2 x| \leqslant |\sin x| + \sqrt{2}\, |\cos^2 x| \leqslant 1 + \sqrt{2} < 4$$

and so the last equation has no solution for $k \neq 0$. We consider $k = 0$, that is, we solve the equation $\sin x + \sqrt{2} \cos^2 x = 0$. Replacing $\cos^2 x$ by $1 - \sin^2 x$, we get the quadratic equation $\sin x + \sqrt{2} - \sqrt{2} \times \sin^2 x = 0$, whence $\sin x = -1/\sqrt{2}$ (the second root of the quadratic exceeds 1). Therefore the solutions of the first equation are:

$$x_1 = -\frac{\pi}{4} + 2n\pi, \quad x_2 = \frac{5\pi}{4} + 2n\pi, \quad n = 0, \pm 1, \pm 2, \dots$$

For the sake of convenience we wrote the solution in the form of two sets instead of the common notation $x = (-1)^{n+1} \frac{\pi}{4} + n\pi$. Generally speaking, the contracted notation for the solution in the form of a compact single group (set) is convenient for representing the final answer, but is less convenient than the two-set notation when we still have computations to perform.

Consider the first group $x = -\pi/4 + 2n\pi$: Substitute these values of x into the second equation to get $\tan(-\pi/4 + 2n\pi + \pi/4) = 0$ (since obviously $\tan^2 x = 1$), which is an identity. Thus, all values of x of the first group are solutions of the system and, hence, of the original equation.

Considering the second group in similar fashion, we are convinced that not a single value is a root of the second equation and none satisfies the original equation.

Thus, the solutions of the original equation are given by the formula

$$x = -\frac{\pi}{4} + 2n\pi$$

where n is any integer.

A common type of problem unusual in aspect is the single equation or inequality in two or more unknowns and a system of equations in which the number of unknowns is not equal to the number of equations. The preceding problems (systems of two equations in one unknown) are of this type. A considerable psychological barrier to the student is a problem where there are more unknowns than equations or inequalities. This is apparently due to the view, explicitly expressed or latent, that a large number of unknowns cannot be determined from a small number of conditions. The following problems show that this is not so.

7. *Solve the system of equations*

$$\tan^2 x+\cot^2 x=2\sin^2 y$$
$$\sin^2 y+\cos^2 z=1$$

This system involves three unknowns and only two equations, but it is clear at once that in the first equation the left member $\geqslant 2$ because we have a sum of positive reciprocals, while the right member $\leqslant 2$. Therefore, the first equation is equivalent to a system of two equations:

$$\tan^2 x+\cot^2 x=2$$
$$\sin^2 y=1$$

The system is no longer unusual for we have an ordinary system of three equations in three unknowns. What is more, it is exceedingly simple. From the two new equations and the second given one we immediately get $\tan^2 x=1$, $\cos^2 z=0$. And so the solutions of the given system are given by the formulas

$$x=\frac{\pi}{4}(2k+1),\quad y=\frac{\pi}{2}+l\pi,\quad z=\frac{\pi}{2}+m\pi$$

where k, l, m are arbitrary integers.

8. *Solve the inequality*

$$-|y|+x-\sqrt{x^2+y^2-1}\geqslant 1$$

To begin with, *to solve an inequality in two unknowns* x and y means to indicate all number pairs x, y such that when they are substituted into the given inequality it becomes a true statement.

Such number pairs can obviously be indicated either directly, as given in the solution of ordinary systems of equations (in which case there may be an infinitude of such solutions, say, for trigonometric systems) or geometrically, by representing the region composed of appropriate points in the plane. This geometric method was employed in solving Problem 30 of Sec. 1.13. Here too it is natural to reduce the given inequality to a form such that its solutions can readily be represented in the plane.

Rewrite the inequality as

$$x-|y|\geqslant\sqrt{x^2+y^2-1}+1$$

It is now evident that any solution will satisfy the condition $x-|y|\geqslant 0$, and when this condition is fulfilled, both members of the inequality are nonnegative (in the domain of the variables) and we can square them to obtain an equivalent (in the same domain) inequality

$$-x|y| \geqslant \sqrt{x^2+y^2-1}$$

But in the region under consideration, $x \geqslant |y| \geqslant 0$, so that the left member of the resulting inequality is nonpositive, while the right member is nonnegative. It is therefore satisfied if and only if both members are zero:

$$x|y|=0 \text{ and } x^2+y^2-1=0$$

The first equation means that either x or y is 0. If $x=0$, then from the condition $x \geqslant |y|$ it follows that $y=0$, and the pair $x=0$, $y=0$ evidently does not enter into the domain of the original inequality. Consequently, $y=0$, and from the second equation we get (noting that $x \geqslant 0$) $x=1$.

Direct substitution into the original inequality shows that the resulting pair $x=1$, $y=0$ satisfies it. It is interesting to note that although we were well set from the very beginning on a geometric representation of the solution, it turned out to be unnecessary for recording the answer.

9. *Find all pairs of numbers x, y that satisfy the equation*

$$\cos x+\cos y-\cos(x+y)=\frac{3}{2}$$

First solution. The problem could of course be stated simply: solve the equation. This equation admits an almost standard solution which reduces, as we will presently see, to the solution of a trigonometric inequality. The first thing that comes to mind when we are confronted by a single equation in two unknowns is: try to express one unknown in terms of the other. Then, assigning values to one unknown, we would obtain the corresponding values of the other and the resulting pairs would be the solutions of the equation. The problem would then be solved.

Let us try to express y in terms of x. Our equation can be written in the form

$$\cos x+2\sin\left(\frac{x}{2}+y\right)\sin\frac{x}{2}=\frac{3}{2}$$

Clearly, $\sin(x/2) \neq 0$ (or $\cos x=3/2$) and therefore

$$\sin\left(\frac{x}{2}+y\right)=\frac{\frac{3}{2}-\cos x}{2\sin\frac{x}{2}}$$

We will consider this equation as an equation in y. It is solvable if and only if the right member is, in absolute value, less than or equal to 1, that is, if

$$\left|\frac{\frac{3}{2}-\cos x}{2\sin\frac{x}{2}}\right| \leqslant 1$$

Since $3/2-\cos x>0$, this inequality can be rewritten

$$\frac{3}{2}-\cos x \leqslant 2\left|\sin\frac{x}{2}\right|$$

But $\cos x=1-2\sin^2(x/2)$ and, denoting $|\sin(x/2)|$ in terms of t, we have the inequality $4t^2-4t+1\leqslant 0$, which is only satisfied for $t=1/2$.

Thus, the right member of this equation in y does not exceed unity (in absolute value) only when $|\sin(x/2)|=1/2$, that is when $\sin(x/2)=\pm 1/2$. It has no solution for other values of x. Let us consider separately the cases $\sin(x/2)=1/2$ and $\sin(x/2)=-1/2$.

If $\sin(x/2)=1/2$, then $x=(-1)^k\pi/3+2k\pi$, where k is any integer. Then for arbitrary k,

$$\cos x=\cos\left[(-1)^k\frac{\pi}{3}+2k\pi\right]=\frac{1}{2}$$

and the equation at hand in y assumes the form $\sin(x/2+y)=1$, whence $x/2+y=\pi/2+2n\pi$ and therefore

$$y=2n\pi+\frac{\pi}{2}-\frac{x}{2}=2n\pi+\frac{\pi}{2}-(-1)^k\frac{\pi}{6}-k\pi$$

We have thus obtained one set (group) of solutions of the original equation

$$x=(-1)^k\frac{\pi}{3}+2k\pi,\quad y=\frac{\pi}{2}+(-1)^{k+1}\frac{\pi}{6}+(2n-k)\pi,$$

$$n,\ k=0,\ \pm 1,\ \pm 2,\ \ldots$$

In similar manner we consider the case of $\sin(x/2)=-1/2$, which leads us to the second group of solutions:

$$x=(-1)^{k+1}\frac{\pi}{3}+2k\pi,\quad y=-\frac{\pi}{2}+(-1)^k\frac{\pi}{6}+(2n-k)\pi,$$

$$n,\ k=0,\ \pm 1,\ \pm 2,\ \ldots$$

Note that the values of y would have been somewhat easier to find if we noted that the equation $|\sin(x/2)|=1/2$ is equivalent to $\cos x=1-2\sin^2(x/2)=1/2$, so that $x=\pm\pi/3+2k\pi$.

Second solution. This equation admits of a different solution based on a suitable grouping. Taking advantage of trigonometric formulas, reduce the equation to

$$4\cos^2\frac{x+y}{2}-4\cos\frac{x-y}{2}\cos\frac{x+y}{2}+1=0$$

It is easy to complete the square of a difference in the first two terms in the left member. The equation will then appear as

$$\left(2\cos\frac{x+y}{2}-\cos\frac{x-y}{2}\right)^2+\sin^2\frac{x-y}{2}=0$$

It is quite clear that the last equation is equivalent to the system

$$2\cos\frac{x+y}{2}=\cos\frac{x-y}{2}$$

$$\sin\frac{x-y}{2}=0$$

From the second equation of this system we have $x-y=2k\pi$, or $y=$ $=x-2k\pi$, where k is any integer. Substituting the value of y into the first equation, we get

$$2\cos(x-k\pi)=\cos k\pi$$

Since $\cos(x-k\pi)=(-1)^k\cos x$ and $\cos k\pi=(-1)^k$, it follows that $\cos x=1/2$, that is, $x=\pm\pi/3+2n\pi$, where n is any integer.

Thus, the solution of the original equation is given by the following pairs:

$$x=\pm\frac{\pi}{3}+2n\pi,\quad y=\pm\frac{\pi}{3}+2(n-k)\pi,\quad n,\ k=0,\ \pm1,\ \pm2,\ \ldots$$

(in the formulas, take both upper signs or both lower signs).

Note that the formulas obtained in this solution are quite different from those of the first solution, yet they describe one and the same set of pairs (x, y). True, it is not so easy to see this.

Also, the second solution which looks so simple is based on an aptly found grouping, which in all cases is somewhat artificial and carries with it an element of conjecture. The first solution, on the contrary, follows a very natural idea and merely involves manipulational difficulties.

The next problem reduces (depending on the approach chosen) either to the solution of a trigonometric inequality and then to a system of algebraic inequalities, or to the proof of a trigonometric inequality. Ingenuity is the keynote in both instances.

10. *Find all number pairs x, y that satisfy the equation*

$$\left(\sin^2 x+\frac{1}{\sin^2 x}\right)^2+\left(\cos^2 x+\frac{1}{\cos^2 x}\right)^2=12+\frac{1}{2}\sin y$$

First solution. The idea behind this solution is the same as in the first solution of the preceding problem. Let us try to express y in terms of x. We have an equation in y:

$$\sin y=2\left(\sin^2 x+\frac{1}{\sin^2 x}\right)^2+2\left(\cos^2 x+\frac{1}{\cos^2 x}\right)^2-24$$

To solve this equation it is necessary that the right member not exceed unity in absolute value.

We transform the right member to

$$2\left(\sin^4 x+\frac{1}{\sin^4 x}\right)+4+2\left(\cos^4 x+\frac{1}{\cos^4 x}\right)+4-24$$
$$=2(\sin^4 x+\cos^4 x)\left(1+\frac{1}{\sin^4 x\cos^4 x}\right)-16$$
$$=2(1-2\sin^2 x\cos^2 x)\left(1+\frac{1}{\sin^4 x\cos^4 x}\right)-16$$
$$=2\left(1-\frac{1}{2}\sin^2 2x\right)\left(1+\frac{16}{\sin^4 2x}\right)-16$$

Denoting $\sin^2 2x$ by z for brevity, we get a double inequality that ensures the solvability of the original equation:

$$-1\leqslant 2\left(1-\frac{z}{2}\right)\left(1+\frac{16}{z^2}\right)-16\leqslant 1$$

After the necessary manipulations we get the following system of two inequalities:

$$z^3+15z^2+16z-32\geqslant 0$$
$$z^3+13z^2+16z-32\leqslant 0$$

Although the first inequality is cubic, the solution is readily obtained by grouping the left member:

$$z^3-1+15z^2-15+16z-16=(z-1)(z^2+16z+32)$$

It is reduced to the form

$$(z-1)[z-(8-4\sqrt{2})][z-(8+4\sqrt{2})]\geqslant 0$$

Solving this inequality (say by the method of intervals), we obtain

$$1\leqslant z\leqslant 8-4\sqrt{2},\quad z\geqslant 8+4\sqrt{2}$$

However, $z=\sin^2 2x\leqslant 1$, so that only $z=1$ is the solution that interests us. Direct substitution shows that this solution also satisfies the second inequality, which means it is a solution of the system.

Thus, the equation in y which we wish to solve has a solution only when $z=\sin^2 2x=1$. Putting $z=1$ into the equation for $\sin y$ and using the transformed right member

$$\sin y=2\left(1-\frac{z}{2}\right)\left(1+\frac{16}{z^2}\right)-16$$

we get $\sin y=1$, and so the original equation is thus reduced to the system of equations

$$\sin^2 2x=1$$
$$\sin y=1$$

whose solutions, as can readily be seen, are pairs of the form

$$x=\frac{\pi}{4}+\frac{k\pi}{2},\quad y=\frac{\pi}{2}+2n\pi,\ k,\ n=0,\ \pm 1,\ \pm 2,\ \ldots$$

There is a still shorter way of solving the above system of algebraic inequalities, though it is less natural. Namely, since $z=\sin^2 2x$, then $0\leqslant z\leqslant 1$, and, hence,

$$z^3+15z^2+16z-32\leqslant 1+15+16-32=0$$

or the left member of the first inequality is nonpositive and obviously equal to zero only when $z=1$. Thus, when $0\leqslant z\leqslant 1$, this inequality is satisfied only for $z=1$. The second inequality is valid for $z=1$. It is of course more difficult to discern such a solution.

Note also that the success of the foregoing solution of the cubic inequality, which involves factoring out $z-1$, is ensured by the apt substitution $z=\sin^2 2x$. If in place of z we had introduced a new variable by the formula $t=\sin^2 x\cos^2 x$, the resulting inequality would have been more complicated: the left member would contain the factor $t-1/4$, which would have been more difficult to guess.

Second solution. Transforming the left member of the given equation as in the first solution, we get

$$(2-\sin^2 2x)\left(1+\frac{16}{\sin^4 2x}\right)=16+\sin y$$

But $2-\sin^2 2x\geqslant 1$, $1+\frac{16}{\sin^4 2x}\geqslant 1+16=17$ and both inequalities become equalities for $\sin^2 2x=1$. Therefore the left member of our equation is at least 17 and the right member is clearly not greater than 17. Thus, the original equation is equivalent to the system

$$\sin^2 2x=1$$
$$\sin y=1$$

which we obtained in the first solution.

Though the second solution is shorter, the first one is more natural, for we immediately comprehend the objective of each argument. The second solution makes use of a nonstandard procedure, a comparison of the values of the left and right members of the equation.

In the next equation in two unknowns it is clear that we cannot express one of the unknowns in terms of the other, but have to resort to quite different reasoning, again involving inequalities.

11. *Find all number pairs x, y that satisfy the equation*

$$\tan^4 x + \tan^4 y + 2\cot^2 x \cot^2 y = 3 + \sin^2(x+y)$$

Consider the left member. Taking advantage of the inequality $a^4 + b^4 \geqslant 2a^2b^2$, which evidently is true for arbitrary values of a and b, we get the inequality

$$\tan^4 x + \tan^4 y \geqslant 2\tan^2 x \tan^2 y$$

equality occurring only when $\tan^2 x = \tan^2 y$. Furthermore,

$$\tan^2 x \tan^2 y + \cot^2 x \cot^2 y \geqslant 2$$

as the sum of positive reciprocals, equality occurring only for $\tan^2 x \tan^2 y = 1$. Thus, the left member of our equation is greater than or equal to 4, the number 4 occurring only when the following equations are valid simultaneously:

$$\tan^2 x = \tan^2 y$$
$$\tan^2 x \tan^2 y = 1$$

On the other hand, $\sin^2(x+y) \leqslant 1$, and, hence, the right member is less than or equal to 4.

Thus, the original equation is satisfied if and only if both members are equal to 4, which occurs when x and y are solutions of the system

$$\tan^2 x = \tan^2 y$$
$$\tan^2 x \tan^2 y = 1$$
$$\sin^2(x+y) = 1$$

From the first two equations we have $\tan^2 x = \tan^2 y = 1$, that is, $\tan x = \pm 1$, $\tan y = \pm 1$, and all four combinations of signs are possible. Then the angles x and y can be written in the form

$$x = \frac{\pi}{4}(2k+1), \quad y = \frac{\pi}{4}(2n+1) \quad (k,\ n \text{ arbitrary integers})$$

Of these angles we can choose those which satisfy the third equation of the system. To do this, we substitute them into the third equation:

$$\sin^2(x+y) = \sin^2\frac{\pi}{2}(k+n+1) = \begin{cases} 1 & \text{if } k+n+1 \text{ is odd,} \\ 0 & \text{if } k+n+1 \text{ is even} \end{cases}$$

From this it follows that, of the solutions of the first two equations, we have to take the pairs k, n such that the sum $k+n$ is even, $k+n = 2m$, i.e., $n = 2m - k$, where m is any integer.

And so the solutions of the given equation are pairs x, y of the form

$$x = \frac{\pi}{4}(2k+1), \quad y = \frac{\pi}{4}(4m-2k+1) \quad (m,\ k \text{ any integers})$$

It is best to regard this solution as a rough solution and recast it in a logically more clear-cut final version. This can be done, say, as follows. Reduce the left member to the form

$$\begin{aligned}\tan^4 x+\tan^4 y+\cot^2 x\cot^2 y\\ &=(\tan^2 x-\tan^2 y)^2+2(\tan^2 x\tan^2 y+\cot^2 x\cot^2 y)\\ &=(\tan^2 x-\tan^2 y)^2+2(\tan x\tan y-\cot x\cot y)^2+4\end{aligned}$$

Then the equation becomes

$$(\tan^2 x-\tan^2 y)^2+2(\tan x\tan y-\cot x\cot y)^2+1=\sin^2(x+y)$$

The left member of this equation is greater than or equal to unity, the right member is less than or equal to unity. Therefore the equation is satisfied if and only if both members are equal to unity, that is, when x and y are solutions of the system

$$\begin{aligned}\tan^2 x-\tan^2 y&=0\\ \tan x\tan y-\cot x\cot y&=0\\ \sin^2(x+y)&=1\end{aligned}$$

This system is solved in the same way as in the rough solution.

Exercises

Solve the following equations, inequalities and systems:

1. $2^{|x|}=\sin x^2$. 2. $3^{|\sin\sqrt{x}\,|}=|\cos x|$.
3. $x^2=-\cos x$. 4. $3x^2=1-2\cos x$.
5. $2\cos(x/3)=2^x+2^{-x}$.
6. Show that the equation $2\sin^2\frac{x}{2}\sin^2\frac{x}{6}=x^2+\frac{1}{x^2}$ does not have a solution.
7. $8-x\cdot 2^x+2^{3-x}-x=0$.
8. $\log_2^2 x+(x-1)\log_2 x=6-2x$.
9. $x\cdot 2^x=x(3-x)+2(2^x-1)$.
10. $x^x=10^{x-x^2}$, $x>0$.
11. $x^2+(x+1)\sin\frac{\pi x}{6}=\frac{3+x}{2}$, $-2\leqslant x\leqslant 0$.
12. $x=\sin\pi\frac{x+1}{3}\sin\pi\frac{1-x}{3}$, $0\leqslant x\leqslant 1$.
13. $5x^2+5y^2+8xy+2x-2y+2=0$.
14. $x^2+4x\cos xy+4=0$.
15. $\frac{|\cot xy|}{\cos^2 xy}=\log_{1/3}(9y^2-18y+10)+2$.
16. $\log_2\left(\cos^2 xy+\frac{1}{\cos^2 xy}\right)=\frac{1}{y^2-2y+2}$.
17. $\tan^2\pi(x+y)+\cot^2\pi(x+y)=\sqrt{\frac{2x}{x^2+1}}+1$.

18. $\log_3|\pi x|+\log_{|\pi x|}3=\dfrac{2}{\sin^2(x+y)-2\sin(x+y)+2}$.

19. $\tan^3 x+2\tan x(\sin y+\cos y)+2=0$.

20. $2\sqrt{2}(\sin x+\cos x)\cos y=3+\cos 2y$.

21. $\cos x-y^2-\sqrt{y-x^2-1}\geqslant 0$.

22. $-x-y^2-\sqrt{x-y^2-1}\geqslant -1$.

23. $\cos x=-1/3$, $\tan x=2\sqrt{2}$.

24. $\sin x=-2/5$, $\cot x=-3/\sqrt{5}$.

25. $2(x^4-2x^2+3)(y^4-3y^2+4)=7$.

Find the real solutions of the following systems:

26. $x+y+z=2$, $2xy-z^2=4$.

27. $x^2+y^2+z^2=a^2$, $x^3+y^3+z^3=a^3$.

28. For which values of the parameter a does the system of equations

$$x^2+y^2=z$$
$$x+y+z=a$$

have a unique solution?

29. Find all the values of the parameters a and b for which the equation $(x^2+5)/2=x-2\cos(ax+b)$ has at least one solution.

4.3 Problems which are standard in aspect but are solved by nonstandard methods

The methods used in Sec. 4.2 for solving problems which are nonstandard in aspect (their employment there is absolutely necessary) can successfully be utilized in the solution of the most ordinary problems solved by the usual methods. Quite naturally, as a rule these methods offer a shorter and more elegant solution.

For instance, after Problems 5 and 6 that we discussed in Sec. 4.2, it is possible to solve the equation $\sin^3 x+\cos^3 x=1$ *orally*, whereas solving it in the ordinary way requires rather lengthy computations. Here are some more examples.

1. *Solve the inequality*

$$\sqrt{\sin x}+\sqrt{\cos x}>1$$

The domain of this inequality consists of the values of x such that simultaneously we have $\sin x\geqslant 0$ and $\cos x\geqslant 0$. Since, besides, $\sin x\leqslant 1$ and $\cos x\leqslant 1$, then, by the property of powers,

$$\sqrt{\sin x}\geqslant \sin^2 x,\quad \sqrt{\cos x}\geqslant \cos^2 x$$

Combining these inequalities, we get

$$\sqrt{\sin x}+\sqrt{\cos x}\geqslant 1$$

equality occurring only when we simultaneously have

$$\sqrt{\sin x}=\sin^2 x \quad \text{and} \quad \sqrt{\cos x}=\cos^2 x$$

Arguing as in Sec. 4.2, we readily see that these equations are simultaneously satisfied only when $x=2k\pi$ and when $x=\pi/2+2k\pi$, and thus the original inequality is valid for all values of x in the domain, with the exception of the values just indicated; that is, for all x such that simultaneously

$$\sin x > 0 \quad \text{and} \quad \cos x > 0$$

The common solution of these two inequalities and, hence, of the original inequality is

$$2k\pi < x < \frac{\pi}{2} + 2k\pi, \quad k=0, \ \pm 1, \ \pm 2, \ \ldots$$

By way of comparison, we give the ordinary (standard) solution of the given inequality. Its domain is defined by the inequalities $\sin x \geqslant 0$ and $\cos x \geqslant 0$. Both members of the original inequality are positive and therefore after squaring we get an inequality equivalent in the domain of the variable to the original one:

$$\sin x + \cos x + 2\sqrt{\sin x \cos x} > 1$$

Since $2\sin x \cos x = (\sin x + \cos x)^2 - 1$, then by replacing $\sin x + \cos x$ by u for brevity we get

$$\sqrt{2u^2-2} > 1-u$$

Since the left member of this inequality is nonnegative, it is automatically satisfied for $u > 1$, that is, all values $u > 1$ are solutions.

Now considering the values $u \leqslant 1$, we again have an inequality with nonnegative members. Squaring, we get (after some manipulating) the equivalent inequality $u^2+2u-3>0$, whose solutions are $u<-3$ and $u>1$. We consider the case $u \leqslant 1$ and so we need only retain the inequality $u<-3$.

And so the inequality $\sqrt{2u^2-2} > 1-u$ has the solution $u>1$ (from the first case) and $u<-3$. But $u=\sin x+\cos x=\sqrt{2}\,\sin(x+\pi/4)$ cannot be less than -3 and so it remains to solve the inequality

$$\sin x + \cos x > 1$$

In the domain of the original inequality, $\sin x + \cos x \geqslant 0$; therefore, after squaring both members of the inequality we get the equivalent inequality $\sin x \cos x > 0$, which is valid for all x in the domain, except those for which $\sin x \cos x = 0$, that is for values of x in the intervals

$$2k\pi < x < \frac{\pi}{2} + 2k\pi \quad k=0, \ \pm 1, \ \pm 2, \ \ldots$$

It is clear then that the first solution is much shorter, to say nothing of the fact that it is more universal in the sense that it is, for instance, inessential that both roots in the inequality be square roots; they could even be of different degrees. The "standard" solution in this case would have come up against insuperable difficulties.

The usefulness of nonstandard methods in the problems given so far lay solely in their brevity; we could easily have dispensed with them entirely.

The unusual aspect of the equations in the preceding section is actually a hint that tells us to seek unusual methods. A more treacherous problem is one in which the form is standard, ordinary, yet the problem is not amenable to ordinary methods. When dealing with such problems, we are never sure whether our approach is faulty or whether they indeed require certain nonstandard reasoning.

We will now investigate a number of solutions pertaining to the next two problems and show how multifarious certain nonstandard methods can be and what diversified ideas can underlie them.

2. *How many roots has the equation* $\sin x+2\sin 2x=3+\sin 3x$ *on the interval* $0\leqslant x\leqslant\pi$?

First solution. This appears to be a very standard type of equation and the natural thing from the very start is to attempt to solve it by orthodox methods, for example, by reducing it to one function, say to $\sin x$.

Using the formulas of the sine of a double and triple angle, we get the equation

$$4\sin x\cos x=3+3\sin x-4\sin^3 x-\sin x$$

An unpleasant pitfall already awaits us: in order to express $\cos x$ in terms of $\sin x$ without irrationalities, we have to square the equation and take the risk of introducing extraneous roots. As a result we get the equation

$$16\sin^2 x\,(1-\sin^2 x)=(3+2\sin x-4\sin^3 x)^2$$

which, after replacing $\sin x$ by y and some simple manipulations, reduces to the form

$$16y^6-24y^3-12y^2+12y+9=0$$

The resulting sextic equation suggests that we will have no end of trouble. True, we might still attempt to seek a suitable grouping of terms; combining the first, second and last terms, we represent the equation in the form

$$(4y^3-3)^2+12y\,(1-y)=0$$

But $y=\sin x$ and $0\leqslant x\leqslant\pi$ by hypothesis, so that we only need the roots y such that $0\leqslant y\leqslant 1$. However, for these y the second term in the left member is nonnegative, while the first term is nonnegative for any y, and therefore the equation is satisfied only for those values of y for which, simultaneously, $4y^3-3=0$ and $12y(1-y)=0$. Since obviously no such y exist, the equation has no solutions on the interval $0\leqslant y\leqslant 1$ under consideration. From this we conclude that neither does the original equation on the interval $0\leqslant x\leqslant \pi$.

This approach, which at first appeared to be quite standard, was successful solely because of an apt grouping. We will see below that this was probably the least suitable approach of all.

Second solution. This solution is also based on an apt grouping of terms that results after a few trigonometric transformations.

Rewrite the given equation as

$$\sin 3x-\sin x-2\sin 2x+3=0$$

and transform the left member; it is

$$\begin{aligned}&2\sin x\cos 2x-4\sin x\cos x+3\\&=\sin x(2\cos 2x-4\cos x)+3\\&=\sin x(4\cos^2 x-4\cos x-2)+3\\&=\sin x[(2\cos x-1)^2-3]+3\end{aligned}$$

We thus get the equation

$$\sin x(2\cos x-1)^2+3(1-\sin x)=0$$

We need solutions only on the interval $0\leqslant x\leqslant \pi$, and the inequality $\sin x\geqslant 0$ holds true on this interval. Therefore both terms in the resulting equation are nonnegative and so it is equivalent to the system of equations

$$\begin{aligned}\sin x(2\cos x-1)^2&=0\\3(1-\sin x)&=0\end{aligned}$$

From the second equation of the system we have $\sin x=1$. But then $\cos x=0$ and $\sin x(2\cos x-1)^2=1\neq 0$. Therefore not a single solution of the second equation is a solution of the first, and, hence, the system (and so also the original equation) does not have any solutions.

Third solution. Write the equation in the form

$$\sin x-\sin 3x+2\sin 2x=3$$

Using the formulas of the difference of sines and the sine of a double angle, we get the equation $\sin x(-4\cos^2 x+4\cos x+2)=3$. Since $\sin x\neq 0$ for the roots of this equation, it is equivalent to

$$-4\cos^2 x+4\cos x+2=\frac{3}{\sin x}$$

Regarding the trinomial $y=-4\cos^2 x+4\cos x+2$ as quadratic in $\cos x$, it is easy to see that its maximum value is equal to 3 and is attained when $\cos x=1/2$. On the other hand, we have, on the interval $0\leqslant x\leqslant\pi$, the inequality $0\leqslant\sin x\leqslant 1$, so that $3/\sin x\geqslant 3$, equality being attained only when $\sin x=1$. This means that our equation is only satisfied when $\cos x=1/2$ and $\sin x=1$ simultaneously. But this is clearly impossible and so the equation at hand does not have any solutions.

It will be noted that in all the previous solutions an essential element was the fact that $\sin x\geqslant 0$, since the roots were only sought in the interval $0\leqslant x\leqslant\pi$. However, this condition is given solely to simplify the problem, for the solution can be obtained without this restriction. Namely, we will prove that the given equation has no roots such that $\sin x<0$.

Indeed, if $\sin x<0$, then

$$\sin x+2\sin 2x<0+2=2$$

and

$$3+\sin 3x\geqslant 3+(-1)=2$$

That is, the left member of the original equation for all values of x *is strictly less than* 2, whereas the right member *is greater than or equal to* 2 for all values of x. This is made evident also by means of simple identity transformations: represent the equation in the form

$$\sin x=2(1-\sin 2x)+1+\sin 3x$$

The right member is nonnegative, whence it follows that $\sin x\geqslant 0$.

Thus, the given equation has no roots x for which $\sin x<0$; earlier it was proved that there are no roots x for which $\sin x\geqslant 0$. Hence, the original equation does not have any roots at all.

Fourth solution. This solution is the shortest one of all and is carried through irrespective of the restrictions imposed on x. Rewriting the equation in the form

$$\sin x-\sin 3x+2\sin 2x=3$$

or, what is the same thing,

$$-2\sin x\cos 2x+2\sin 2x=3$$

we can write the following chain:

$$\begin{aligned}|-2\sin x\cos 2x+2\sin 2x|&\leqslant|-2\sin x\cos 2x|\\&+|2\sin 2x|=2|\sin x||\cos 2x|+2|\sin 2x|\\&\leqslant 2(|\cos 2x|+|\sin 2x|)\end{aligned}$$

Since $|\cos 2x|+|\sin 2x|\leqslant\sqrt{2}$ for arbitrary x (this can be proved in a very simple fashion by removing the absolute-value signs or, still more simply, by squaring), the left member of the last equation does not exceed $2\sqrt{2}$ in absolute value and, hence, cannot be equal to 3.

In the next problem, which outwardly is a very ordinary one, standard procedures fail. For instance, in an attempt to obtain an equation in $\sin x$, we arrive at an equation of degree seven.

3. *Solve the equation*

$$\sin^2 x+\frac{1}{4}\sin^2 3x=\sin x\sin^2 3x$$

First solution. Write the equation

$$\sin^2 x-\sin x\sin^2 3x+\frac{1}{4}\sin^2 3x=0$$

and, considering the left member as a quadratic trinomial in $\sin x$, isolate the perfect square to get

$$\left(\sin x-\frac{1}{2}\sin^2 3x\right)^2+\frac{1}{4}\sin^2 3x\,(1-\sin^2 3x)=0$$

or, what is the same thing,

$$\left(\sin x-\frac{1}{2}\sin^2 3x\right)^2+\frac{1}{16}\sin^2 6x=0$$

This equation is clearly equivalent to the system of equations

$$2\sin x=\sin^2 3x$$
$$\sin 6x=0$$

The solution of the second equation is the set of values $x=k\pi/6$, where k is any integer. Of this set, we choose those values that satisfy the first equation. To find the corresponding values of k, substitute $x=k\pi/6$ into the first equation to get

$$\sin\frac{k\pi}{6}=\frac{1}{2}\sin^2\frac{k\pi}{2}=\begin{cases}0 & \text{if } k \text{ is even,}\\ \frac{1}{2} & \text{if } k \text{ is odd}\end{cases}$$

Let us now consider separately the cases when k is even and when it is odd. If k is even, we must have $\sin k\pi/6=0$; this occurs when $k\pi/6=m\pi$, where m is any integer. Thus, of all *even* values of k only the numbers $k=6m$, where m is any integer, are suitable. If k is odd, we have to have $\sin k\pi/6=\frac{1}{2}$; this is valid for $k\pi/6=(-1)^m\pi/6+$ $+m\pi$, where m is any integer. Of the *odd* values of k, we can only use the numbers $k=(-1)^m+6m$, where m is any integer.

Thus, the solutions of the original equation are the values of the following sets:

$$x=m\pi,\quad x=(-1)^m\pi/6+m\pi\ (m \text{ any integer})$$

A very simple solution of this system is this. The second equation separates into two: $\sin 3x=0$ and $\cos 3x=0$. If $\sin 3x=0$, then from the first equation we get $\sin x=0$, that is, $x=k\pi$. These values of x are solutions of the system. But if $\cos 3x=0$, then $\sin^2 3x=1$, and from the first equation we have $\sin x=1/2$. The solutions of this equation $x=(-1)^k\pi/6+k\pi$ are clearly solutions of the system. These two sets constitute the solution set of the original equation.

Comparing this reasoning with the preceding arguments, we note that the very unnatural separation of the simple equation $\sin 6x=0$ into two proved more effective than the natural approach involving its solution and the subsequent selection of roots.

Second solution. First of all, note that if $\sin^2 3x=0$, then from the equation it follows that $\sin x=0$ and it is easy to verify that all roots of the equation $\sin x=0$, that is, $x=k\pi$, where k is any integer, satisfy the equation. But if $\sin^2 3x\neq 0$, then the left member of our equation is positive and, hence, $\sin x>0$.

Now transpose everything to the left side of our equation and complete the square in the sum of squares, and do it in two ways: to obtain the square of a sum and the square of a difference. Then consider the equation in two forms:

$$\left(\sin x+\frac{1}{2}\sin 3x\right)^2+\sin x\,(-\sin 3x-\sin^2 3x)=0$$

and

$$\left(\sin x-\frac{1}{2}\sin 3x\right)^2+\sin x\,(\sin 3x-\sin^2 3x)=0$$

We consider the two cases separately.

(a) $\sin 3x>0$. Then $\sin 3x\geqslant \sin^2 3x$ and, hence, both terms in the second form of the equation are nonnegative, that is, x must satisfy the system

$$\sin x-\frac{1}{2}\sin 3x=0$$

$$\sin x\,(\sin 3x-\sin^2 3x)=0$$

But $\sin x\neq 0$, so from the second equation of the system we have $\sin 3x-\sin^2 3x=0$. By Condition (a), $\sin 3x\neq 0$, whence we therefore have $\sin 3x=1$; and so from the first equation, $\sin x=1/2$. Thus, we have to solve the system

$$\sin x=\frac{1}{2}$$

$$\sin 3x=1$$

But if $\sin x=1/2$, then $\sin 3x=3\sin x-4\sin^3 x=1$, that is, any solution of the first equation is a solution of the second equation, and the system reduces to the first equation. Thus, its solutions are the angles x in the set

$$x=(-1)^k\frac{\pi}{6}+k\pi, \text{ where } k \text{ is any integer.}$$

(b) $\sin 3x<0$. This case is handled in the same way as the preceding case, but here we make use of the first form of our equation. This case does not yield any new solutions. We can therefore write down the solution of our equation as

$$x=k\pi, \quad x=(-1)^k\frac{\pi}{6}+k\pi, \text{ where } k \text{ is any integer.}$$

This same solution can be carried out in more compact form if we make use of absolute values. Our equation can then be represented as

$$\left(\sin x-\frac{1}{2}|\sin 3x|\right)^2+\sin x\,(|\sin 3x|-\sin^2 3x)=0$$

But $\sin^2 3x\leqslant|\sin 3x|$ and therefore both terms are nonnegative so that we have the system

$$\sin x-\frac{1}{2}|\sin 3x|=0$$

$$\sin x\,(|\sin 3x|-\sin^2 3x)=0$$

The rest of the solution is similar to the preceding case.

Third solution. Noting first of all that the given equation implies $\sin x\geqslant 0$ and is satisfied for $\sin x=0$, or for $x=k\pi$, we will henceforth consider that $\sin x>0$.

Suppose that our equation is satisfied by some number x and we rewrite the equation as

$$\sin^2 x=\sin^2 3x\left(\sin x-\frac{1}{4}\right)$$

Since $\sin^2 x\neq 0$, it follows that $\sin^2 3x\neq 0$ and, hence, $\sin x-1/4>0$. But then, multiplying the inequality $\sin^2 3x\leqslant 1$ by the positive expression $\sin x-1/4$, we get the inequality

$$\sin^2 3x\left(\sin x-\frac{1}{4}\right)\leqslant\sin x-\frac{1}{4}$$

equality being attained only when $\sin^2 3x=1$. Hence, the inequality

$$\sin^2 x\leqslant\sin x-\frac{1}{4}$$

should also hold true or, what is the same, the inequality $(\sin x-1/2)^2\leqslant 0$, which is valid only when $\sin x=1/2$.

Thus, if some number x satisfies our equation, then it satisfies the system

$$\sin^2 3x = 1$$
$$\sin x = \frac{1}{2}$$

The converse is obvious: if x satisfies this system, then, substituting $\sin x = 1/2$ and $\sin^2 3x = 1$ into the original equation, we see that it too is satisfied. Thus, the given equation is equivalent to the system obtained and this system is solved in the same manner as the preceding ones.

This same mode of solution could be worked out somewhat differently. Namely, noting that $\sin x - 1/4 > 0$ and adding $-\sin x + 1/4$ to both members of the equation $\sin^2 x = \sin^2 3x (\sin x - 1/4)$, we obtain the equation

$$\left(\sin x - \frac{1}{2}\right)^2 = (\sin^2 3x - 1)\left(\sin x - \frac{1}{4}\right)$$

The left member is nonnegative, the right is nonpositive, and so the equation is satisfied if and only if both members are zero. This yields the already familiar system.

Fourth solution. Divide the equation through by its right member, but, so as not to lose any roots, first be sure that it is nonzero. Considering the case $\sin x = 0$, we get the roots of the original equation $x = k\pi$, where k is any integer, and for $\sin 3x = 0$ we again obtain $\sin x = 0$ from the equation so that we will have the very same roots as those just found.

Noting, as before, that $\sin x \geqslant 0$, we will henceforth consider that $\sin x > 0$. We can now divide:

$$\frac{\sin x}{\sin^2 3x} + \frac{1}{4 \sin x} = 1$$

Both terms in the left member are positive and so we can apply the inequality between the arithmetic mean and the geometric mean:

$$\frac{\sin x}{\sin^2 3x} + \frac{1}{4 \sin x} \geqslant 2 \sqrt{\frac{1}{4 \sin^2 3x}} = \frac{1}{|\sin 3x|} \geqslant 1$$

In the first weak inequality, equality is attained only when $\frac{\sin x}{\sin^2 3x} = \frac{1}{4 \sin x}$, in the second, when $|\sin 3x| = 1$. Therefore the left member of the equation is equal to unity if and only if both weak inequalities turn into equations, that is, if the following system is valid:

$$4 \sin^2 x = \sin^2 3x$$
$$|\sin 3x| = 1$$

This system is then easily solved.

Fifth solution. Noting that $\sin x \geqslant 0$, we write the following chain of inequalities

$$\sin^2 x + \frac{1}{4}\sin^2 3x \geqslant |\sin x|\,|\sin 3x| \geqslant \sin x \sin^2 3x$$

(Here we make use of the inequality $a^2+b^2 \geqslant 2|a||b|$ for $a=\sin x$, $b=\frac{1}{2}\sin 3x$ and of the fact that $|\sin x|=\sin x$ and $|\sin 3x| \geqslant \sin^2 3x$.) Therefore, our equation is satisfied if and only if both weak inequalities become equations, which occurs when we have the following system:

$$\sin x = \frac{1}{2}|\sin 3x|$$
$$\sin x |\sin 3x| = \sin x \sin^2 3x$$

The second equation is satisfied in the following four cases:

$$\sin x = 0, \quad \sin 3x = 0, \quad \sin 3x = \pm 1$$

In the first case we have the values $x=k\pi$, which are also solutions of the first equation of the system, that is, of the original equation as well. The second case, upon substitution into the first equation, leads to the same solutions. The third and fourth cases yield the solution $x=(-1)^k(\pi/6)+k\pi$, $k=0, \pm 1, \pm 2, \ldots$.

Exercises

Solve the following equations and inequalities.

1. $\sqrt{\dfrac{x^2-x-2}{\sin x}} + \sqrt{\dfrac{\sin x}{x^2-x-2}} = \dfrac{3}{2}$.
2. $\sqrt{\dfrac{x^2-2x-2}{x^2+4x+2}} + \sqrt{\dfrac{x^2+4x+2}{x^2-2x-2}} = 2$.
3. $\sqrt[4]{17+x} + \sqrt[4]{17-x} = 2$. 4. $\sqrt{1+x} + \sqrt{1-x} \geqslant 1$
5. $\sqrt{2x-1} + \sqrt{3x-2} > \sqrt{4x-3} + \sqrt{5x-4}$.
6 $\sin^4 x + \cos^{15} x = 1$. 7. $\sqrt{\sin 2x} + \sqrt{\cos 2x} = 1$.
8. $\sqrt{\sin^3 x} + \sqrt{\cos^3 x} = \sqrt{2}$.
9. $\sqrt{2+\cos^2 2x} = \sin 3x - \cos 3x$.
10. $\sqrt{5+\sin^2 3x} = \sin x + 2\cos x$.

4.4 Problems involving logical difficulties

Very considerable difficulties of a logical nature are ordinarily caused by equations, inequalities or systems containing *parameters*, in which it is required to find the values of the parameters for which certain supplementary requirements are fulfilled (say, the equation has a unique solution or, contrariwise, is satisfied by all admissible values of x, or every solution of one system of equations is a solution of another system, or every solution of one inequality is a solution of another, and the like).

This type of problem is probably the most difficult, for it requires a high degree of logical culture. The student must at every step clearly realize what has been done and what still remains to be done, and what the results obtained signify.

1. *For what values of a does the equation* $1+\sin^2 ax=\cos x$ *have a unique solution?*

It is clear that $\sin^2 ax$ cannot, for arbitrary values of a, be expressed in terms of $\sin x$ and $\cos x$. For this reason, the equation at hand cannot be solved by ordinary methods; a new idea for the solution is needed. The underlying idea will be similar to those employed in Sec. 4.2.

Due to the fact that we have the inequality $\cos x \leqslant 1 \leqslant 1+\sin^2 ax$, the original equation is valid if and only if one of the following systems of equations is fulfilled:

$$\begin{aligned} 1+\sin^2 ax &= 1 \\ \cos x &= 1 \end{aligned} \quad \text{or} \quad \begin{aligned} \sin ax &= 0 \\ \cos x &= 1 \end{aligned}$$

We thus have to solve the last system and investigate for which values of a it has a unique solution. Since the original equation is equivalent to this system, the values of a thus found will be the required values.

Here is where the most serious logical complications begin. It is precisely at this point that we see which student understands the problem and which merely performs the manipulations without realizing what he is doing and why it is necessary.

Here is an instance of one student's solution of the system:

$$\text{“}ax=\pi k, \quad x=2\pi n, \quad 2a\pi n=\pi k, \quad a=\frac{k}{2n}\text{”}$$

That and nothing else! Not a single word, merely the equation $a=k/(2n)$ was underlined and this was apparently taken to mean the answer. This is no solution of course.

We now give a real solution, which repeats the manipulations of the preceding "solution" but is supported by arguments that were lacking there.

The solutions of the first equation of the latter system are

$$ax=k\pi, \quad k=0, \pm 1, \pm 2, \ldots$$

The solutions of the second equation are also obvious:

$$x=2n\pi, \quad n=0, \pm 1, \pm 2, \ldots$$

We need x such that satisfy simultaneously both equations, that is, we have to find the numbers k and n for which we obtain one and the same value of x for both sets. Thus, we have to solve one equation in two unknowns n and k (and also with the parameter a):

$$2an\pi = k\pi \tag{1}$$

It is obvious that for *any* a the number pair $n=0$, $k=0$ is a solution of this equation. To it corresponds the root $x=0$. Thus, for *arbitrary* a the original equation has the solution $x=0$. If $n \neq 0$, then equation (1) can be rewritten

$$a = k/(2n) \tag{2}$$

Now let us recall our basic problem: not to *solve* the equation but merely to *determine for which values of a it has a unique solution.* Now any pair of numbers k and n which satisfies Condition (2) will yield a solution of the original equation $x=2\pi n=\pi k/a$. Since for arbitrary a we have already found one root of the original equation ($x=0$), we must now seek values of a for which no integers k and n exist such that relation (2) is valid. Clearly, if a is irrational, then no such k and n exist. The first result is obtained: *if a is an irrational number, then the given equation has a unique solution.*

Is the problem solved? Of course not, since we have not yet investigated the rational values of a. However, if a is rational, that is $a=p/q$, then it can be written in the form $a=(2p)/(2q)$ and in equation (2) we get the solution $k=2p$, $n=q$. Hence, in this case, with the exception of $x=0$, there will at least be one solution (actually there will even be infinitely many). To summarize then: *for a rational, the original equation has more than one solution.* The problem is solved.

All these steps are needed so as to solve the problem for ourselves and obtain the answer. This might be called the rough solution. We now show what the final version might look like.

Obviously, $x=0$ is a root of the equation for *any* a. We will demonstrate that for a irrational there are no other solutions and for a rational, there are. Indeed, first suppose that a is *irrational*. From the inequalities $\cos x \leqslant 1 \leqslant 1+\sin^2 ax$ it follows that x is a solution if and only if the following system is satisfied:

$$\begin{array}{r} 1+\sin^2 ax = 1 \\ \cos x = 1 \end{array} \quad \text{that is} \quad \begin{array}{r} \sin ax = 0 \\ \cos x = 1 \end{array}$$

If $x \neq 0$ is a solution of the last system, then, firstly, $ax=\pi k$, k an integer, and, secondly, $x=2\pi n$, n an integer, $n \neq 0$. But then $2a\pi n=\pi k$, whence $a=k/(2n)$, that is to say, a is a rational number, which runs counter to the assumption. Now let a be *rational*, $a=p/q$. Then $x=2\pi q$ will clearly be a solution and, moreover, one different from zero. Thus, the given equation has a unique solution if and only if the number a is irrational.

This problem can also be solved graphically. We assume that $a \neq 0$, since for $a=0$ the equation clearly has an infinitude of solutions. Let us rewrite our equation thus:

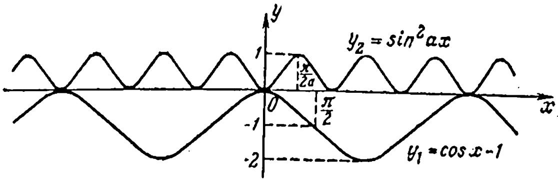
$$\sin^2 ax = \cos x - 1$$

and denote

$$y_1 = \cos x - 1, \quad y_2 = \sin^2 ax = \frac{1-\cos 2ax}{2}$$

We draw both graphs on one drawing (Fig. 168). The original equation clearly has a solution if and only if the graphs of the functions y_1 and y_2 have a point in common.

Fig. 168



It is evident from the drawing that for arbitrary a there is a point of intersection of the graphs: $x=0$. The drawing also shows that subsequent intersection of graphs is only possible at points where both curves touch the x-axis, that is, at points where, simultaneously, $\sin^2 ax=0$ and $\cos x=1$. But $\sin^2 ax=0$ for $ax=n\pi$, where $n=0, \pm 1, \pm 2, \ldots$ and $\cos x=1$ for $x=2k\pi$, where $k=0, \pm 1, \pm 2, \ldots$. Therefore the point $x=0$ will be the sole meeting point of the graphs only when $2ak\pi \neq n\pi$ for no nonzero integers n and k. In other words, we have shown that there will be a unique solution only when $a \neq n/(2k)$, where n and k are nonzero integers. Then, as before, it is demonstrated that this occurs only when a is an irrational number.

In solving problems involving parameters, one often reasons as follows. Let a parameter a be some fixed number *that satisfies the condition of the problem*; such values of a will be called *suitable*. We then derive consequences from the statement of the problem and assumptions concerning a. We thus obtain certain conditions which the suitable values of the parameter must satisfy. The values of the parameter which do not satisfy these consequences are automatically classed as unsuitable, and we have only to consider the values of the parameter that satisfy the consequences obtained. In particular, if these consequences are satisfied solely by certain concrete values, then the problem reduces to a verification of these values.

2. *Find all the values of a for which the system*

$$\begin{gathered} 2^{|x|} + |x| = y + x^2 + a \\ x^2 + y^2 = 1 \end{gathered}$$

has only one solution (a, x, y are real numbers).

In accordance with the above, we first assume that a is some suitable number, that is, a number that satisfies the statement of the problem. In other words, for this value of a the given system of equations has exactly one solution; denote it by (x_0, y_0). But it is easy to notice that both equations of the system remain unchanged upon replacing x by $-x$, which means that the pair $(-x_0, y_0)$ is also a solution of the system for the value under consideration. The original assumption, however, was that the system has a *unique* solution with respect to a. There is only one way out, (x_0, y_0) and $(-x_0, y_0)$ are one and the same pair. This then simply means that $x_0=-x_0$, that is, $x_0=0$. So far this reasoning does not yield any information about y_0. But if we put the solution $(0, y_0)$ into the original system, we get the equations

$$1=y_0+a, \quad y_0^2=1$$

whence it follows that y_0 is equal either to 1 or to -1; accordingly, a is equal either to 0 or to 2.

We have thus demonstrated that if a is a suitable number, then either $a=0$ or $a=2$. It must be stressed that the foregoing reasoning has in no way proved that the numbers 0 and 2 are suitable. Quite the contrary, that is precisely what we must now find out.

We first consider the value $a=0$. In this case we have the system

$$\begin{aligned} 2^{|x|}+|x|&=y+x^2 \\ x^2+y^2&=1 \end{aligned} \tag{3}$$

If we can prove that this system has a unique solution, then this will signify that the value $a=0$ satisfies the condition of the problem. Note that the value $a=0$ was obtained above when we substituted the pair (0, 1) into the original system. It is easy to verify that this pair does indeed satisfy the system (3) and, thus, for $a=0$ the original system already has one solution. Now let us find out whether (3) has any other solutions.

This system is not solvable by ordinary procedures. We will have to reason in a special way. From the second equation of the system it follows that $|x|\leqslant 1$, $|y|\leqslant 1$, whence $x^2\leqslant |x|$ and $y\leqslant 1$. Besides, $2^{|x|}\geqslant 1$, since $|x|\geqslant 0$. From all these inequalities we get

$$2^{|x|}+|x|\geqslant 1+x^2\geqslant y+x^2$$

and, hence, the first equation is satisfied only when equality occurs in both weak inequalities; that is, when

$$2^{|x|}=1, \quad |x|=x^2, \quad y=1$$

and this is true only for $x=0$, $y=1$. Thus, for $a=0$ the given system has the unique solution of (0, 1).

We now consider the value $a=2$. In this case we have the system

$$2^{|x|}+|x|=y+x^2+2$$
$$x^2+y^2=1$$

As before, we note that the pair $(0, -1)$ is a solution and again we have to find out whether there are any other solutions. But by substituting $x=1$, $y=0$, we are assured that the pair $(1, 0)$ is also a solution of the system and, hence, for $a=2$ the system has more than one solution.

To summarize then, the given system has a unique solution for $a=0$ alone.

The foregoing solution requires a remark or two, not of a mathematical but rather of a psychological nature. As so often happens, the solution is easy to understand, but how is it found? There is of course no cutand-dry answer to that.

In our solution there are three possible guesses.

Firstly, we noticed that the system does not change upon replacement of x by $-x$. This was an essential hint. Anyone with some idea about the evenness and oddness of a function and with some experience in handling functions would realize this.

Secondly, we started out by working system (3) by nonstandard procedures using inequalities. This conjecture is somewhat more complicated, but the examples of earlier sections showed us that it is often necessary to employ inequalities in equation solving.

Finally, we realized that for $a=2$ the original system has yet another solution: $x=1$, $y=0$. We therefore tried simply to pick a solution, and it worked out. This approach proved successful merely due to the existence of "good" integral solutions. In certain cases, such a choice is the only possible route for solving a problem.

3. *Find all the values of a and b for which the system*

$$xyz+z=a$$
$$xyz^2+z=b$$
$$x^2+y^2+z^2=4$$

has only one solution (a, b, x, y, z are real numbers).

Let (a, b) be a suitable pair of values of the parameters and (x_0, y_0, z_0) the corresponding unique solution. It is readily seen that the system remains unchanged if, simultaneously, we replace x by $-x$ and y by $-y$. This implies that the triple $(-x_0, -y_0, z_0)$ is also a solution of the system and, as in the preceding problem, we conclude that $x_0=y_0=0$. Substituting the triple $(0, 0, z_0)$ into the system, we get $z_0=a$, $z_0=b$, $z_0^2=4$, whence $z_0=\pm 2$ and $a=b=\pm 2$.

Thus, if the pair (a, b) is suitable, then either $a=b=2$ or $a=b=-2$.

Again, as in the preceding problem, we have to establish whether these pairs of values of the parameters are suitable or not.

For $a=b=2$ we have the system

$$\begin{aligned} xyz+z &= 2 \\ xyz^2+z &= 2 \\ x^2+y^2+z^2 &= 4 \end{aligned}$$

one of the solutions of which, as can readily be verified, is $x=0$, $y=0$, $z=2$. From the second and first equations it follows that $xy(z^2-z)=0$. If $x=0$, then from the second and third equations we get $z=2$ and $y=0$. We already know this solution. The same solution is obtained if $y=0$.

We will now assume that $z^2-z=0$, i.e., $z=0$ or $z=1$. However, when $z=0$ we see that the first two equations are contradictory, and for $z=1$ we get the system

$$\begin{aligned} xy &= 1 \\ x^2+y^2 &= 3 \end{aligned}$$

which, as it is easy to see, has four real solutions. Thus, for $a=b=2$ the original system has five solutions, and therefore the pair $a=b=2$ is not a suitable one.

Now let $a=b=-2$. We have the system

$$\begin{aligned} xyz+z &= -2 \\ xyz^2+z &= -2 \\ x^2+y^2+z^2 &= 4 \end{aligned}$$

one of the solutions of which, as we can readily see, is $x=0$, $y=0$, $z=-2$. Reasoning as before, we see that the system does not have any other real solutions and so for $a=b=-2$ the original system has a unique solution, which means this pair of values of the parameters is suitable.

Hence the condition of the problem is satisfied only by the values $a=b=-2$.

4. *Find all the values of a for which the system*

$$\begin{aligned} (x^2+1)^a+(b^2+1)^y &= 2 \\ a+bxy+x^2y &= 1 \end{aligned}$$

has at least one solution for any value of b (a, b, x, y are real numbers).

Let a be a suitable value of the parameter, that is, a value for which the given system has at least one solution for *any* value of b. We choose some value of b; this can be done in arbitrary fashion, but we will choose b so that the system takes on the simplest possible aspect. Clearly, the best to choose is $b=0$. Then the system looks like this

$$\begin{aligned} (x^2+1)^a &= 1 \\ a+x^2y &= 1 \end{aligned}$$

and since a is a suitable value, the system has at least one solution, which we denote by (x_0, y_0).

In this solution, x_0 is either zero or nonzero. If $x_0=0$, then from the second equation we get $a=1$, and if $x_0\neq 0$, then $x_0^2+1\neq 1$, and from the first equation we get $a=0$.

Thus, if a is a suitable number, then either $a=0$ or $a=1$. Now we have to determine whether these values are indeed suitable or not.

When $a=0$ the system is of the form

$$(b^2+1)^y=1$$
$$bxy+x^2y=1$$

We now have to find out whether this system has any solutions for arbitrary values of b. For $b\neq 0$ it follows from the first equation that $y=0$, and then the second equation is inconsistent. Hence, the value $a=0$ is not a suitable value.

Let $a=1$, then the system is

$$x^2+(b^2+1)^y=1$$
$$bxy+x^2y=0$$

Clearly, $x=y=0$ for any b is a solution and so $a=1$ is a suitable value.

Thus, the condition of the problem is satisfied by the unique value $a=1$.

5. *Find all the numbers a for each of which any root of the equation*

$$\sin 3x = a\sin x + (4-2|a|)\sin^2 x \tag{4}$$

is a root of the equation

$$\sin 3x + \cos 2x = 1 + 2\sin x \cos 2x \tag{5}$$

and, contrariwise, any root of the latter equation is a root of the former.

The problem can more briefly be stated thus: *for which values of a are the equations* (4) *and* (5) *equivalent*? There are fundamentally two ways of determining the equivalence of two equations: the first is to obtain each equation from the other by means of certain manipulations, the second, in accord with the definition of equivalence, is to prove that every root of one equation is a root of the other, and vice versa.

In our example, the first approach is apparently inapplicable and we have to take advantage of the second approach. Here, too, however, things are not so simple. It is hard to reason about the coincidence of the roots of two equations which are so unlike. The only thing that can save us is a knowledge of all these roots or the roots of at least one of the equations.*

---

* More pictorial examples of such a situation can be given: the equations $x^2=2x^3-1$ and $x2^x=2$ (in the domain of real numbers) are equivalent since $x=1$ is the only root, but try to convince yourself that they are equivalent without first solving both equations.

In our case, equation (5) has a simple solution and so the problem readily reduces to the following one: *for which values of a does equation* (4) *have exactly the same roots as* (5)?

To simplify computations, denote $\sin x$ by y. Then (5) becomes

$$2y^2 - y = 0 \tag{6}$$

This equation has the roots $y_1=0$, $y_2=1/2$. Similarly, upon replacing $\sin 3x = 3y - 4y^3$, equation (4) becomes

$$[4y^2 + (4-2|a|)y + a - 3]\,y = 0 \tag{7}$$

Many students replaced $\sin x$ by y and this "helped" them to make two serious mistakes. Thus, many decided immediately that the required values of a do not exist since equation (6) is quadratic and equation (7) is cubic and, hence, they are not equivalent because they have different numbers of roots. This argument contains two mistakes at once. Firstly, a quadratic and a cubic equation can be equivalent (for instance, the equations $x^2=0$ and $x^3=0$ both have the *unique* root $x=0$) and, secondly, as we will see for ourselves below, (4) and (5) can be equivalent even if (6) and (7) are not.

Therein lies the second mistake. At first glance it would appear quite obvious that our problem was reduced to the following: for which values of a does equation (7) have only the roots 0 and 1/2. But actually, if we recall that $y = \sin x$, we can indicate yet another possibility for the value of a to be suitable: if (7) has the roots 0, 1/2 and its third root y_3 is greater than unity in absolute value, then (4) and (5) are equivalent because the corresponding value $\sin x = y_3$ will not give equation (4) any additional solutions. Naturally, the equations (4) and (5) are equivalent when the third root of (7) is equal to 0 or to 1/2.

Our problem is now clear: we have to *find values of a such that* (7) *has the roots* 0, 1/2 *and its third root is either* 0 *or* 1/2 *or exceeds* 1 *in absolute value*.

It is evident at once that 0 is a root of equation (7) so that we will henceforth consider the equation

$$4y^2 + (4-2|a|)y + a - 3 = 0 \tag{8}$$

One of the roots of this equation must be 1/2. Substituting $y=1/2$ into it, we find that 1/2 is a root when $a=|a|$, or $a \geqslant 0$. By Viète's theorem, the second root is equal to $(a-3)/2$ and according to what has just been said, the value of a will be suitable in the following three cases:

$$(1)\ \frac{a-3}{2}=0, \quad (2)\ \frac{a-3}{2}=\frac{1}{2}, \quad (3)\ \left|\frac{a-3}{2}\right|>1$$

(also bear in mind that $a \geqslant 0$).

We then have the answer:

$$a=3,\ a=4,\ 0\leqslant a<1,\ a>5$$

6. *Find all the numbers a such that for every root of the equation*

$$2\sin^7 x-(1-a)\sin^3 x+(2a^3-2a-1)\sin x=0 \tag{9}$$

is a root of the equation

$$2\sin^6 x+\cos 2x=1+a-2a^3+a\cos^2 x \tag{10}$$

and, contrariwise, every root of the second equation is a root of the first equation.

Here, both the given equations are complicated and so we cannot proceed as in the preceding problem. But we can note that equation (9) has solutions of the form $x=k\pi$, where k is any integer, and perhaps some other solutions. This remark will enable us to solve the problem.

Let a be a suitable value of the parameter. Then the values $x=k\pi$—roots of (9)—are roots of (10), and this immediately yields the equation $a^3=a$ (since $\sin^6 k\pi=0$, $\cos 2k\pi=\cos^2 k\pi=1$). Therefore, the suitable values are to be chosen from among only three numbers: 0, 1 and -1. The task now is to verify all three values.

Let $a=0$, then the equations will assume the form

$$\sin x(\sin^2 x-1)(2\sin^4 x+2\sin^2 x+1)=0$$
$$\sin^2 x(\sin^2 x-1)(\sin^2 x+1)=0$$

Since $1+\sin^2 x>0$ and $2\sin^4 x+2\sin^2 x+1>0$, these equations are equivalent.

Let $a=1$. Then the equations can be rewritten as

$$\sin x(2\sin^6 x-1)=0 \quad \text{and} \quad \sin^2 x(2\sin^4 x-1)=0$$

Since the first equation has the solution $\sin x=\sqrt[6]{1/2}$, which does not satisfy the second equation, these equations are not equivalent.

Let $a=-1$. We then have the equations

$$\sin x(2\sin^6 x-2\sin^2 x-1)=0 \quad \text{and} \quad \sin^2 x(2\sin^4 x-3)=0$$

Since $2\sin^4 x-3<0$ and $2\sin^6 x-2\sin^2 x-1=2\sin^2 x(\sin^4 x-1)-1<0$, it is clear that the equations are equivalent. Thus, the condition of the problem is satisfied only by $a=0$ and $a=-1$. In this problem too, many students replaced $\sin^2 x$ by y and could not figure out what to do with the value $a=-1$, since in the inequalities that have to be proved in this case, essential use is made of the fact that $0\leqslant y\leqslant 1$.

7. *Find all number pairs a, b for which every number pair x, y ($x\neq\pi/2+k\pi$, $y\neq\pi/2+n\pi$; $k, n=0, \pm1, \pm2, \ldots$) that satisfies the*

equation $x+y=a$ also satisfies the equation

$$\tan x+\tan y+\tan x\tan y=b \qquad (11)$$

Let a and b be a suitable pair of values of the parameters. We take a number pair $x=0$, $y=a$ which clearly satisfies the equation $x+y=a$. If $a\neq\pi/2+n\pi$, then this pair satisfies the restrictions imposed on x and y in the statement of the problem and, for this reason, by virtue of (11), the equation

$$\tan a=b$$

must be valid. Let us now take the number pair $x=\pi/4$, $y=a-\pi/4$, which also satisfies the equation $x+y=a$. If $a\neq 3\pi/4+k\pi$, then this pair likewise satisfies the restrictions on x and y and therefore (since a and b are assumed to be suitable) the following equation must hold:

$$1+2\tan\left(a-\frac{\pi}{4}\right)=b \qquad (12)$$

Since $b=\tan a$, then a thus satisfies the equation

$$1+2\tan\left(a-\frac{\pi}{4}\right)=\tan a$$

which can readily be reduced to the quadratic equation $\tan^2 a-2\times$ $\times\tan a+1=0$. Hence, $\tan a=1$ and the suitable pairs a, b must be sought among the infinitude of pairs of the form

$$a=\frac{\pi}{4}+m\pi, \quad b=1, \; m=0, \pm 1, \pm 2, \ldots$$

Let us determine which of these pairs are actually suitable. Let $x+y=\pi/4+m\pi$ for some integer m, and $x\neq\pi/2+k\pi$, $y\neq\pi/2+n\pi$, k, $n=0, \pm 1, \pm 2, \ldots$. Then $y=\pi/4+m\pi-x$ and, hence,

$$\begin{aligned}\tan x+\tan y+\tan x\tan y &=\tan x+\tan\left(\frac{\pi}{4}+m\pi-x\right)+\tan x\tan\left(\frac{\pi}{4}+m\pi-x\right)\\ &=\tan x+\tan\left(\frac{\pi}{4}-x\right)+\tan x\tan\left(\frac{\pi}{4}-x\right)\\ &=\tan x+\frac{1-\tan x}{1+\tan x}+\tan x\,\frac{1-\tan x}{1+\tan x}\end{aligned}$$

The last expression is equal to unity. Thus all pairs $a=\pi/4+m\pi$, $m=0, \pm 1, \pm 2, \ldots$, $b=1$ are suitable.

However, the solution is not yet complete since in the course of our discussion we excluded the values $a=\pi/2+n\pi$ and $a=3\pi/4+k\pi$, k, $n=0, \pm 1, \pm 2, \ldots$. It remains to consider these values as well.

Let $a=\pi/2+n\pi$, n an integer. In this case, obviously, $a\neq 3\pi/4+$ $+k\pi$ and therefore (12) must be valid, from which it follows, for the values of a under consideration, that $b=3$. We will attempt to find suitable pairs among

$$a=\frac{\pi}{2}+n\pi, \quad b=3, \quad n=0, \pm 1, \pm 2, \ldots$$

The pair $x=-\pi/4$, $y=3\pi/4+n\pi$ satisfies the equation $x+y=a$; on the other hand,

$$\tan\left(-\frac{\pi}{4}\right)+\tan\left(\frac{3\pi}{4}+n\pi\right) + \tan\left(-\frac{\pi}{4}\right)\tan\left(\frac{3\pi}{4}+n\pi\right)=-1\neq 3$$

and therefore there are no suitable pairs among the number pairs a, b under consideration.

Now let $a=3\pi/4+k\pi$, k an integer. Since in this case $a\neq\pi/2+n\pi$, the equation $\tan a=b$ must be true, i.e., $b=-1$. We wish to find suitable pairs among the pairs

$$a=\frac{3\pi}{4}+k\pi, \quad b=-1, \quad k=0, \pm 1, \pm 2, \ldots$$

The pair $x=3\pi/8$, $y=3\pi/8+k\pi$ satisfies the equation $x+y=a$, on the other hand,

$$\tan\frac{3\pi}{8}+\tan\left(\frac{3\pi}{8}+k\pi\right)+\tan\frac{3\pi}{8}\tan\left(\frac{3\pi}{8}+k\pi\right) = 2\tan\frac{3\pi}{8}+\tan^2\frac{3\pi}{8}>0$$

(because the angle $3\pi/8$ lies in the first quadrant) and so the left member of (11) is different from -1 and thus there are no suitable pairs among the pairs a, b under consideration.

The final answer is this: the condition of the problem is satisfied by an infinity of pairs

$$a=\frac{\pi}{4}+m\pi, \quad b=1, \quad m=0, \pm 1, \pm 2, \ldots$$

We conclude this example with a final remark: at the examination the students gave many different solutions, some of which were shorter, but in all cases certain facts were assumed obvious which could be regorously substantiated only by means of extremely lengthy and subtle reasoning. Unfortunately, the reasoning was absent. What is more, the students did not even realize that such reasoning is necessary for the solution to be considered exhaustive.

8. *Find all the values of a for which any value of x that satisfies the inequality*

$$ax^2+(1-a^2)x-a>0$$

does not exceed two in absolute value.

In its logical form, the statement of this problem is quite analogous to that of the preceding problem. Namely, it is required to find the values of the parameter a for which from the inequality $ax^2+(1-a^2)x-a>0$ follows the inequality $-2\leqslant x\leqslant 2$. However, the method of solution applied in the preceding problem is not suitable here mainly because we have to do with inequalities and not with equations.

In this problem we have to determine for which values of a all the solutions of the given inequality lie in the interval $-2\leqslant x\leqslant 2$. If $a\neq 0$, then the given inequality is quadratic, and we first consider this general case. Thus, let $a\neq 0$. We know that the solutions of a quadratic inequality, if they exist, form on the number line either a finite interval, two infinite intervals, or the entire set of real numbers; and this depends on the signs of the discriminant and the leading coefficient. We therefore compute the discriminant of the quadratic trinomial in the left member of the inequality:

$$D=(1-a^2)^2+4a^2=a^4+2a^2+1=(a^2+1)^2$$

Thus, $D>0$ for any a, so that the roots of the trinomial are real and distinct and easily found: $x_1=a$, $x_2=-1/a$.

Now, depending on the sign of the number a, the solutions of the given inequality form an interval between roots (for $a<0$ when the parabola is concave downwards) or two infinite intervals (for $a>0$).

By hypothesis, we need values of a for which all solutions of the inequality lie in the interval $-2\leqslant x\leqslant 2$. Therefore the values $a>0$ are not suitable: two infinite intervals cannot fit into a finite interval. It remains only to consider the values $a<0$. In this case, $x_1<0<x_2$ and the solution of the given inequality is the interval $a<x<-1/a$.

We want the entire interval $a<x<-1/a$ to lie in the interval $-2\leqslant x\leqslant 2$, and this occurs obviously if and only if the endpoints of the first interval lie on the interval $-2\leqslant x\leqslant 2$ (coincidence of endpoints is admissible), that is, if the inequalities

$$-2\leqslant a<-\frac{1}{a}\leqslant 2$$

hold true. From the inequality $-1/a\leqslant 2$, taking into account that $a<0$, we get $a\leqslant -1/2$ and, hence, $-2\leqslant a\leqslant -1/2$.

Thus, any solution of the original inequality does not exceed two in absolute value when $-2\leqslant a\leqslant -1/2$, but we obtained this on the assumption that $a\neq 0$. To complete the solution we have yet to consider this special case. For $a=0$, the original inequality assumes the form $x>0$ and not all its solutions fail to exceed two in absolute value so that the value $a=0$ is not a suitable value. To summarize, then, the inequality obtained above

$$-2\leqslant a\leqslant -\frac{1}{2}$$

is the final answer.

9. *Find all the values of a for which, for all x not exceeding unity in absolute value, we find the inequality*

$$\frac{ax-a(1-a)}{a^2-ax-1}>0$$

to be valid.

First replace the given inequality by the equivalent but more customary quadratic inequality

$$(ax+1-a^2)\,[ax-a(1-a)]<0$$

We have been a bit hasty in calling this inequality quadratic, since we have not yet checked to see if, when the brackets are removed, the coefficient of x^2 is different from zero. This coefficient is equal to a^2 and is zero when $a=0$, but for $a=0$ the given inequality takes the form $0<0$, which is to say that it does not hold for any x. Therefore, $a=0$ is not a suitable value and we discard it, considering henceforth $a\neq 0$ everywhere.

We will solve the resulting quadratic inequality by following the same ideas as in the preceding problem. We see at once that the roots of the quadratic trinomial are real and so the discriminant need not be computed. Besides, the leading coefficient a^2 is positive and, hence, the solutions of the quadratic inequality form an interval between its roots $x_1=(a^2-1)/a$, $x_2=1-a$, if these roots are distinct. But if the roots coincide, the quadratic inequality is not satisfied for a single value of x and therefore the corresponding values do not interest us.

We thus need the values of a for which the entire interval $-1\leqslant x\leqslant 1$ lies between the numbers $a-1/a$ and $1-a$. But in order to write this geometric condition in the language of inequalities, we have to know which of the two numbers is greater. This clearly depends on the number a, and so we consider two cases.

(a) $(a^2-1)/a<1-a$. As in the preceding problem, for the interval $-1\leqslant x\leqslant 1$ to lie entirely within the interval $(a^2-1)/a<x<1-a$, it is necessary that the endpoints (-1 and 1) lie in the interval; thus, the following inequalities must be valid:

$$\frac{a^2-1}{a}<-1<1<1-a$$

(Coincidences of extreme points, that is, the equations $a-1/a=-1$ and $1-a=1$ are not admissible, since, for instance, when $a-1/a=-1$, the number -1 in the interval $-1\leqslant x\leqslant 1$ is exterior to the interval $-1<x<1-a$.)

From the inequality $1<1-a$ it follows that $a<0$, and then from the inequality $(a^2-1)/a<-1$ we get $a^2-1>-a$ or $a^2+a-1>0$. The solutions of this inequality are the values of $a<(-1-\sqrt{5})/2$ and $a>(-1+\sqrt{5})/2$. Since $a<0$, we only leave the values $a<(-1-\sqrt{5})/2$.

Now, from the resulting values of a we have to choose those which satisfy Condition (a), i.e., the inequality $a-1/a<1-a$. But this condition is automatically satisfied for $a<(-1-\sqrt{5})/2$. Indeed, the indicated values are obtained as solutions of the inequalities $a-1/a<-1<1<1-a$.

(b) $1-a<(a^2-1)/a$. In this case we have to solve the inequalities

$$1-a<-1<1<\frac{a^2-1}{a}$$

From $1-a<-1$ we have $a>2$ and then from $(a^2-1)/a>1$ follows $(a^2-1)>a$, or $a^2-a-1>0$. The solutions of this inequality are the values $a<(1-\sqrt{5})/2$ and $a>(1+\sqrt{5})/2$. Since $a>2$, we only leave the values $a>2$. As in Case (a), these values automatically satisfy Condition (b). Thus, the condition of the problem is satisfied by the values $a<(-1-\sqrt{5})/2$ and $a>2$.

Exercises

1. Prove that if the equations

$$a\sin x+b\cos x+c=0 \quad \text{and} \quad 2a\tan x+b\cot x+2c=0$$

both have no solutions, then $a=b=0$, $c\neq 0$.

2. For which values of the parameter a are the systems of equations

$$\begin{array}{c} \sin(x+y)=0 \\ x^2+y^2=a \end{array} \quad \text{and} \quad \begin{array}{c} x+y=0 \\ x^2+y^2=a \end{array}$$

equivalent?

3. For which values of the parameter a do the equations

$$x^2+x+a=0 \quad \text{and} \quad x^2+ax+1=0$$

(a) have a common root? (b) For which values are they equivalent? (a, x are real numbers).

4. Find all the numbers a for each of which any root of the equation

$$a\cos 2x+|a|\cos 4x+\cos 6x=1$$

is a root of the equation

$$\sin x\cos 2x=\sin 2x\cos 3x-\frac{1}{2}\sin 5x$$

and, conversely, every root of the latter equation is a root of the former.

5. Find all numbers a for each of which every root of

$$4\cos^2 x-\cos 3x=a\cos x-|a-4|(1+\cos 2x)$$

is a root of

$$2\cos x\cos 2x=1+\cos 2x+\cos 3x$$

and, conversely, every root of the latter equation is a root of the former.

6. Find all the values of a and b for which the system

$$\left|\frac{x^y-1}{x^y+1}\right|=a$$
$$x^2+y^2=b$$

has only one solution (a, b, x, y are real numbers, $x>0$).

7. Find all the values of a for which the system

$$2^{bx}+(a+1)by^2=a^2$$
$$(a-1)x^3+y^3=1$$

has at least one solution for any value of b (a, b, x, y real).

8. Find all the values of a for which the system

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

$$x^3+ax^2y+xy^2=1$$

has at least one solution and every solution satisfies the equation $x+y=0$ (a, x, y real).

9. Given the inequality $ax+k^2>0$. For which values of a

(a) is the inequality valid for all values of x and k?
(b) are there values of x and k for which the inequality is valid?
(c) is there a value of x such that the inequality is valid for any value of k?
(d) is there a value of k for which the inequality is valid for every value of x?
(e) is there a value of k such that the inequality is valid for all values of x?
(f) is there a value of x for which the inequality is valid for every value of k?

4.5 Problems involving the location of roots of a quadratic trinomial

The quadratic trinomial can with full justification be called the principal function of all school mathematics. If one dismisses the very simple linear function, then this is the only function for which the school course of mathematics requires rigorous proof of all the properties needed in the theory and in problem solving. Every student should have a good working knowledge of all the necessary properties of the quadratic trinomial.

The number of problems solved with the aid of the properties of the quadratic trinomial is very great and the problems themselves are highly diversified. Besides problems whose solutions are obtained directly from familiar theorems (these include the solution of quadratic equations and inequalities, finding the conditions of the existence of real roots, determining the signs of roots, finding the largest and smallest values of a quadratic trinomial), there are many problems where the usual group of elementary theorems no longer suffices.

It is quite impossible to describe in full all possible types of problems involving quadratic trinomials. We here investigate a few that deal with the location of roots. There is one problem, among the standard school problems listed above, that belongs to this type: to determine the signs of the roots actually means to determine the location of the roots relative to the point 0. It will be recalled that this problem is solved with the aid of Viète's theorem. But what do we do if it is required to locate the roots with respect to some other point? Can it be that the point 0 is so much "better" than the other points that for 0 the problem is solved immediately while for other points it is a hard nut to crack? Not in the least. The only thing that distinguishes the point 0 from other points is that we have a theorem at hand to handle it (Viète's theorem), whereas for other points we have to think up new similar theorems.

But then another complication arises. If we were to think up theorems to solve root-location problems with the same simplicity as determining the signs of roots, then the number of such theorems would be tremendous and there would be no hope of remembering them. Indeed, it is quite natural (and this actually occurs in problems) to be interested in the location of the roots in a certain specified interval $c<x<d$ or in the infinite interval $x<c$ or in the infinite interval $x>d$. With respect to each one of these intervals, we can ask the following questions. Under what condition: Do both roots occur in the interval? Does no root lie in the interval? Is there exactly one root? Is there at least one root? Using these questions alone, we could make up 12 different problems. And what if, besides, we considered intervals of the type $c\leqslant x<d$ or $x\geqslant d$? And what if we take into account that the properties of a quadratic trinomial are essentially dependent on the sign of the leading coefficient (the coefficient of x^2)? It is quite obvious that the number of required theorems would be practically unsurveyable. Memorizing theorems is clearly not the way out. Then the only way is to learn to think up a treorem every time for every concrete problem, and then of course to remember the most important ones.

For devising such theorems, the student should have an excellent knowledge of the properties of a quadratic trinomial and be able freely to handle such expressions and to reason simultaneously in two spheres: the algebraic and the geometric. What this means is that for any property stated verbally or algebraically, the student should be able to interpret it geometrically in the form of a graph. Contrariwise, the student must be able to state any graphical property verbally and describe it in formal algebraic terms. For instance, the leading coefficient is less than zero; this means that the parabola is concave downwards, the trinomial has no real roots and hence the parabola does not cross or touch the x-axis (axis of abscissas); the graph of the quadratic ax^2+bx+c lies entirely above the x-axis and so $a>0$ and $b^2-4ac<0$. This last geometric statement can be formulated in at least three different ways: the inequality $ax^2+bx+c>0$ is valid for arbitrary values of x; the inequality $ax^2+bx+c\leqslant 0$ has no solutions; the trinomial does not have any real roots and the leading coefficient is positive.

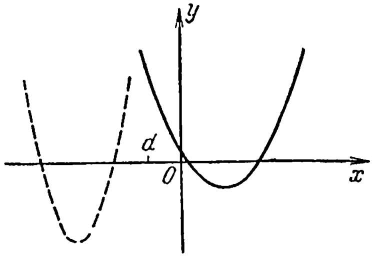
The solution of many of the problems that follow actually amounts to an expansion of this "translation glossary" from the language of algebra to that of geometry and vice versa. Before taking up specific problems let us examine the procedure of a solution in several cases of a theoretical nature.

Let $f(x)=ax^2+bx+c$. *All reasoning will be conducted on the assumption that* $a>0$. Accordingly, when solving concrete problems, we will always restate them so as to be able to utilize the properties of a trinomial with positive leading coefficient. We denote the roots by x_1 and x_2 and the discriminant by D.

1. *Under what conditions are both roots of this trinomial greater than a certain specified number d?*

In order to formulate these conditions, let us begin with a graph of the trinomial that satisfies this condition (Fig. 169). Firstly, it cuts

Fig. 169



the x-axis or touches it; hence, $D \geqslant 0$; secondly, the value $f(d)$ at point d is positive. But this is not sufficient: the trinomial whose graph is shown dashed in Fig. 169 also has these properties yet does not satisfy the condition of the problem. In order to distinguish our trinomials from others of the same type it is clearly sufficient to require that the vertex of the parabola (more precisely, its abscissa) lie to the right of point d; that is, $-b/(2a) > d$. We have thus found the required condition: both roots exceed d if and only if $D \geqslant 0$, $f(d) > 0$ and $-b/(2a) > d$.

The arguments given above are of course not at all rigorous and are to be regarded simply as a rough solution, an exploratory search. We now undertake a rigorous proof, which turns out to be rather simple.

Let both roots exceed d. Then $D \geqslant 0$ since the roots are real, the abscissa of the vertex $-b/(2a)$ exceeds d because it lies between the roots and, finally, $f(d) > 0$ because d is exterior to the interval between the roots. Conversely, let the three indicated conditions hold. Then both roots are real; the condition $f(d) > 0$ means that the point d is exterior to the interval between the roots, and the third condition ensures that d is less than the least root, otherwise d would be greater than the largest root and, hence, greater than the half sum of the roots, which is equal to $-b/(2a)$.

In the examples that follow we will confine ourselves to rougn solutions and leave the rigorous proof to the reader as a useful and even necessary exercise.

2. *Under what conditions do the roots of the trinomial lie on either side of the number d?*

We get the answer immediately if the problem is rephrased thus: Under what conditions does the number d lie between the roots of the given trinomial? Now this statement, as we know, is equivalent to the fact that $f(d)<0$. (Recall that the leading coefficient is taken to be positive!)

3. *Under what conditions does exactly one root of the trinomial lie in the interval $d<x<e$?*

Here too the answer is sufficiently obvious. It is true if and only if the trinomial has values of distinct signs at the points d and e. So as to avoid considering separately the cases when $f(d)>0$, $f(e)<0$ and $f(d)<0$, $f(e)>0$, it is advisable to write these conditions in a more compact form: $f(d)\,f(e)<0$.

The foregoing three examples make the general approach to problems of this nature quite clear. In most problems, however, the question is not often stated so directly as in the examples we have just analyzed. A problem often has to be reformulated in order to arrive at a desirable statement. We now examine a number of specific problems.

Let us first recall the last problem of Sec. 4.4. It was required to find out under what condition the interval $-1\leqslant x\leqslant 1$ lies between the roots of the trinomial $(ax+1-a^2)(ax-a+a^2)$. But this assertion clearly signifies that the numbers -1 and 1 lie between the roots of the trinomial, which occurs if and only if $f(-1)<0$ and $f(1)<0$; that is,

$$(-a+1-a^2)(-2a+a^2)<0, \quad (a+1-a^2)a^2<0$$

This system of inequalities is easily solved by the method of intervals.

4. *Find out for which values of m the quadratic x^2+mx+m^2+6m is negative for all values of x that satisfy the condition $1<x<2$.*

For this quadratic to be negative for all values of x in the interval $1<x<2$, it is necessary, firstly, that it have a positive discriminant and, secondly, that the interval $1<x<2$ lie between its roots, coincidence of the endpoints with the roots not being excluded. Reasoning as in the preceding problems, we get the following system of inequalities:

$$m^2-4(m^2+6m)>0$$

$$\frac{-m-\sqrt{-3m^2-24m}}{2}\leqslant 1<2\leqslant\frac{-m+\sqrt{-3m^2-24m}}{2}$$

There is nothing fundamentally complicated in the solution of this system, but computationally it can cause a lot of trouble. So let us apply the suggestions mentioned earlier. The necessary requirement is fulfilled if the numbers $f(1)$ and $f(2)$ are negative; i.e., if the system of inequalities

$$m^2+7m+1\leqslant 0$$
$$m^2+8m+4\leqslant 0$$

is valid and the solution is readily found on the number line:

$$-\frac{7+3\sqrt{5}}{2}\leqslant m\leqslant -4+2\sqrt{3}$$

5. *Find the values of p for which the equation* $1+p\sin x=p^2-\sin^2 x$ *has a solution.*

Denoting $\sin x$ by y, rewrite the equation in the form

$$y^2+py+1-p^2=0$$

Here is where many students made a grave mistake, reasoning something like this: "Since y is $\sin x$, then $-1\leqslant y\leqslant 1$ and hence we have to find values of p such that the roots of the foregoing quadratic equation lie in the interval $-1\leqslant y\leqslant 1$." Actually there is no necessity whatsoever in requiring that both roots lie on this interval; it is sufficient for at least one of them to be located there. If for some p, one root y_1 does not exceed unity in absolute value, then the equation $\sin x=y_1$ has a solution, and so also does the original equation. Such a number p is a suitable value.

To summarize: we have to find values of p such that at least one of the roots of (5) lies in the interval $-1\leqslant y\leqslant 1$. This problem can be reduced to solving inequalities: firstly, the discriminant must be nonnegative, $5p^2-4\geqslant 0$; secondly, at least one of the double inequalities

$$-1\leqslant\frac{-p-\sqrt{5p^2-4}}{2}\leqslant 1,\quad -1\leqslant\frac{-p+\sqrt{5p^2-4}}{2}\leqslant 1$$

must be valid. Computationally, however, the solution will be extremely awkward. We will choose a different approach.

To facilitate subsequent reasoning, we will first consider the case when both roots of the quadratic $y^2+py+1-p^2$ are different from -1 and 1. Then the required condition holds in two cases: when exactly one root lies in the interval $-1<y<1$ and when both roots lie in that interval. According to Problem 3, the former evidently occurs if and only if the product of the values of the quadratic under consideration is negative at the points -1 and 1; that is, for values of p that satisfy the inequality

$$(p^2-p-2)(p^2+p-2)<0$$

the solution of which is easily found by the method of intervals:

$$-2<p<-1,\quad 1<p<2$$

Now let us write down the conditions under which the second case holds. First of all, the quadratic must be positive for $y=-1$ and for $y=1$. But this condition does not yet determine the case under consideration; it is also true of quadratics which have no roots at all and of these whose two roots lie to the right of 1 or to the left of -1. To eliminate quadratics with no roots, we have to demand besides that the discriminant be greater than zero; and to eliminate the remaining "extra" quadratics, we must impose the condition that the abscissa of the vertex of the parabola, that is, $y=-p/2$, lie within the interval $-1<y<1$.

Thus, the set of conditions under which the case at hand holds true is this: (a) the values of the quadratic for $y=-1$ and $y=1$ are positive, (b) the discriminant is positive, (c) the vertex of the parabola lies in the interval $-1<y<1$. This yields the following system of inequalities:

$$\begin{gathered} 2-p-p^2>0 \\ 2+p-p^2>0 \\ 5p^2-4\geqslant 0 \\ -1<-\frac{p}{2}<1 \end{gathered}$$

The solutions of this system are found with ease:

$$-1<p\leqslant-\frac{2}{\sqrt{5}}, \quad \frac{2}{\sqrt{5}}\leqslant p<1$$

Finally, we consider the temporarily dismissed case when the quadratic has the roots $y=-1$ and $y=1$. Clearly, $y=-1$ is a root when $p=1$ and $p=-2$, and $y=1$ is a root when $p=-1$ and $p=2$. These values of p are likewise suitable values.

Collecting together all the values of p thus found, we get the final answer: the original equation has a solution when

$$-2\leqslant p\leqslant-\frac{2}{\sqrt{5}}, \quad \frac{2}{\sqrt{5}}\leqslant p\leqslant 2$$

6. *Indicate in the coordinate plane all the points whose coordinates (x, y) are such that the expression*

$$\left(2\cos t+\frac{1}{2}\cos x\cos y\right)\cos x\cos y+1+\cos x-\cos y+\cos 2t$$

is positive for any value of t and depict the region formed by these points.

If t runs through all the real numbers (or at least any interval of length 2π), then $z=\cos t$ runs through the interval $-1\leqslant z\leqslant 1$. Therefore the given problem may be rephrased as follows: For which values of x and y is the inequality

$$2z^2+2z\cos x\cos y+\frac{1}{2}\cos^2 x\cos^2 y+\cos x-\cos y>0$$

valid for arbitrary values of z lying in the interval $-1\leqslant z\leqslant 1$?

Under what condition is the quadratic positive over the entire interval $-1\leqslant z\leqslant 1$? First of all, it will be true if the discriminant is less than zero. Therefore, the pairs (x, y) for which

$$D=\cos^2 x\cos^2 y-2\left(\frac{1}{2}\cos^2 x\cos^2 y+\cos x-\cos y\right)$$
$$=-2(\cos x-\cos y)<0$$

or $\cos x - \cos y>0$ will be suitable values.

But if the discriminant is nonnegative, then the suitable pairs (x, y), as will readily be seen, are determined by the following conditions: the values of the quadratic at the points -1 and 1 are positive and the abscissa of the vertex of the parabola is exterior to the interval $-1\leqslant z\leqslant 1$. But in the case of our quadratic, the abscissa of the vertex is equal to $-1/2\cos x\cos y$, which is clearly less than unity in absolute value and, hence, the abscissa of the vertex does indeed lie between -1 and 1. Therefore, the case of $D\geqslant 0$ does not yield any new solutions.

The concluding part of the solution—representation of the solutions of the inequality $\cos x - \cos y>0$ in the plane—has been done in Problem 30 of Sec. 1.13.

This problem may be solved somewhat differently by representing the inequality in the form

$$(2z+\cos x\cos y)^2+2(\cos x-\cos y)>0$$

This approach was taken by many students at the examination, but there was a logical subtlety in this instance which very few students could handle. Most wrote that the last inequality is valid for arbitrary z if and only if $\cos x- \cos y>0$, and proof was offered that otherwise, if we take z so that the first term vanishes, we violate the inequality.

This argument would be correct if the problem called for *arbitrary* z. But we have $z=\cos t$ which assumes a value only between -1 and 1. Therefore, when attempting to make the first term vanish, one has also to justify the existence of the number z *in the interval* $-1\leqslant z\leqslant 1$, for which number this parenthesis is equal to zero. Lack of this proof is a definite logical drawback, although it does not lead to any actual mistake. In reality, such a number z always exists; this is due to the peculiarities of the specific example under consideration. It follows from the fact that the expression $-1/2\cos x\cos y$ is always less than 1 in absolute value.

It will be noted that the proof of this fact in the second solution corresponds to the investigation of the position of the abscissa of the vertex in the first solution.

7. *Solve the equation*

$$9^{-|x-2|}-4\cdot 3^{-|x-2|}-a=0$$

for every real number a.

Denoting $3^{-|x-2|}$ by y and noting that $0<3^{-|x-2|}\leqslant 1$ for every x, we get the equation

$$y^2-4y-a=0$$

for which we have to find the roots lying in the interval $0<y\leqslant 1$. The abscissa of the vertex of the graph of the trinomial $f(y)=y^2-4y-a$ is equal to 2 so that if the trinomial has roots, than the greater root exceeds 2 and does not interest us. It therefore remains to write down the condition under which there is exactly one root in the interval $0<y\leqslant 1$.

First of all, $y=1$ is a root when $a=-3$. Furthermore, since $f(0)=-a$, $f(1)=-a-3<f(0)$, then, by Problem 3, there is exactly one root in the interval $0<y<1$ if and only if $f(0)>0$ and $f(1)<0$, which occurs when $-3<a<0$.

Thus, only for the values $-3\leqslant a<0$ does the equation have exactly one root in the interval $0<y<1$. This is the smallest root $y=2-\sqrt{4+a}$. Now solving the equation $3^{-|x-2|}=2-\sqrt{4+a}$, which for the values of a thus found has a solution (this follows from the fact that the value of y thus found lies between 0 and 1), we get

$$|x-2|=-\log_3(2-\sqrt{4+a}),\quad x_{1,2}=2\pm\log_3(2-\sqrt{4+a})$$

8. *Indicate all the values of a for which the equation*

$$\log_{\sqrt{x}} a^2\left|\log_{a^2}\frac{x}{2}\right|+\log_a x=\log_{a^2}4\cdot\log_{\sqrt{x}} a$$

has a solution and find all the corresponding solutions.

Denoting $\log_a x$ by y and $\log_a 2$ by b and using the formulas for transforming logarithms, we get the equation

$$\frac{4}{y}\left|\frac{1}{2}(y-b)\right|+y=\frac{2b}{y}$$

or, expanding the domain of the variable by adjoining zero, the equation

$$y^2+2|y-b|-2b=0$$

We first seek the roots of this equation that satisfy the condition $y>b$, and, naturally, are nonzero. Then our equation turns into the quadratic

$$y^2+2y-4b=0$$

We want the nonzero roots of this equation that exceed b. For such roots to exist, it is first of all necessary that the discriminant $D=1+4b$ be nonnegative. Therefore, we will henceforth assume that $b\geqslant -1/4$.

The abscissa of the vertex of the graph of $y^2+2y-4b$ is equal to -1 and so the smaller root is less than $-1<b$ since $b\geqslant -1/4$. Hence, we will only be interested in the larger root and whether it will be located

to the right of b. From graphical considerations, it is quite obvious that the larger root esceeds b if and only if the quadratic is negative at point b; that is, $b^2-2b<0$. This inequality holds true for $0<b<2$ and for these values of b we have the root

$$y=-1+\sqrt{1+4b}$$

which is clearly nonzero and, hence, satisfies all the required conditions.

Now let us consider the values $y\leqslant b$. In this case our equation assumes the form $y^2-2y=0$ and its roots are $y_1=2$ and $y_2=0$. Since we only need nonzero roots not exceeding b, it follows that y_2 can be discarded and $y_1=2$ is only suitable for $b\geqslant 2$.

Combining the results of these two cases, we get the (nonzero!) solutions of the equation in y: for $b\leqslant 0$ there is no solution; for $0<b<2$ we have $y=-1+\sqrt{1+4b}$; for $b\geqslant 2$ we have $y=2$.

It remains to pass to the unknown x and the parameter a. The transition to $x=a^y$ does not involve any difficulties, but the transition to a involves the solution of three inequalities:

$$\log_a 2\leqslant 0,\quad 0<\log_a 2<2,\quad \log_a 2\geqslant 2$$

They are all easily solved: the first inequality is valid for $0<a<1$, the second, for $a>\sqrt{2}$, and the third for $1<a\leqslant\sqrt{2}$.

To summarize, then: for $0<a<1$ there is no solution; $x=a^{-1+\sqrt{4+\log_a 2}}$ for $a>\sqrt{2}$; $x=a^2$ for $1<a\leqslant\sqrt{2}$.

Exercises

1. Find all the values of the parameter d for which both roots of the quadratic equation $x^2-6dx+(2-2d+9d^2)=0$ are real and exceed 3.

2. Find all the values of the parameter a for which both roots of the quadratic equation $x^2-ax+2=0$ are real and lie between 0 and 3 (extreme values excluded).

3. For which values of a is one of the roots of the polynomial $(a^2+a+1)x^2+(a-1)x+a^2$ greater than 3 and the other less than 3?

4. Find all the values of a for which the roots of the equation $x^2+x+a=0$ exceed a.

5. For which values of a do the roots of the polynomial $2x^2-2(2a+1)x+a(a-1)$ satisfy the inequalities $x_1<a<x_2$?

6. For which values of a does the equation

$$(1+a)\left(\frac{x^2}{x^2+1}\right)^2-3a\frac{x^2}{x^2+1}+4a=0$$

have real roots?

7. Find all values of m for which the inequality $mx^2-4x+3m+1>0$ is valid for all $x>0$.

8. Find all the values of a for which from the inequality $x^2-a(1+a^2)x+a^4<0$ follows the inequality $x^2+4x+3>0$.

9. For which values of y is the following statement true: "There exists at least one value of x for which the inequality $2\log_{0.5}y^2-3+2x\log_{0.5}y^2-x^2>0$ is valid?"

10. For which values of y is the following statement true: "For any x the inequality

$$x^2\left(2-\log_2\frac{y}{y+1}\right)+2x\left(1+\log_2\frac{y}{y+1}\right)-2\left(1+\log_2\frac{y}{y+1}\right)>0$$

holds true?"

11. Find all the values of a for which from the inequality $ax^2-x+1-a<0$ follows the inequality $0<x<1$.

12. Find all the values of a for which from the inequality $0\leqslant x\leqslant 1$ follows the inequality $(a^2+a-2)x^2-(a+5)x-2\leqslant 0$.

13. Find all the values of a for which, for all x not exceeding unity in absolute value, the inequality $2x^2-4a^2x-a^2+1>0$ holds true.

Solve the following equations and inequalities:

14. $x+\sqrt{x}=a$.

15. $\left(\frac{1+x}{\sqrt{x}}\right)^2+2a\frac{1+x}{\sqrt{x}}+1=0$.

16. $4^x-4m2^x+2m+2=0$.

17. $4^{\sin x}+m2^{\sin x}+m^2-1=0$, where $-1<m<1$.

18. $(\log_{10}\sin x)^2-2a\log_{10}\sin x-a^2+2=0$.

19. $(\log_a\sin x)^2+\log_a\sin x-a=0$, $a>0$, $a\neq 1$.

20. $\sqrt{|x|+1}-\sqrt{|x|}=a$.

21. $\sqrt{a(2^x-2)+1}=1-2^x$.

22. $2|\log_{10}(ax)|\cdot\log_x 10=(4\log_{10}a-3)\log_{x^2}10-\frac{1}{2}\log_{10}x$.

23. $\log_{\sqrt{x}}a\cdot\left|\log_a\frac{x}{2}\right|=\log_{a^2}2\cdot\log_{\sqrt{x}}a^2-\log_a\sqrt{x}$.

24. $\log_{100}x^2=\log_{\sqrt{x}}10\left(\log_{10}10a-\left|\log_{10}\frac{x}{a}\right|\right)$.

25. $\log_{\sqrt{x}}a^2\cdot\left|\log_{a^2}\frac{x}{2}\right|+\log_a x=\log_{a^2}4\cdot\log_{\sqrt{x}}a$.

26. $\log_a x\,|\log_x a-\log_2 a|=\log_x(ax)-\log_2 a^2+\log_a x\cdot\log_2 a$.

Represent, in the plane, points for which the following conditions hold true.

27. For every value of t

$$\cos 2(t+x)+2\sin(t+x)\cos y-\frac{1}{2}(\cos y-1)^2-\sin x<\frac{1}{2}.$$

28. For every value of t

$$\sin^2(t+x)+\sin(t+y)+\sin(t+2x-y)+\frac{1}{4}>0.$$

29. For every value of t

$$\left(2\cos t+\frac{1}{2}\cos x\cos y\right)\cos x\cos y+1+\cos x-\cos y+\cos 2t>0.$$

30. For at least one value of t

$$\cos(t+3x+y)-\cos(t+x-y)-\sin^2(t+2x)>\frac{1}{4}$$

31. For at least one value of t

$$\sin^2 t\cos^2 x+\cos^2 t\sin^2 x+\frac{1}{2}\sin 2x\sin 2t+2(\cos 2x+\cos y)<0.$$

ANSWERS TO EXERCISES*

Sec. 1.1

1. (c) An equation of the form $ax^2+bx+c=0$, where x is the unknown and a, b, c are given numbers, $a \neq 0$, (d) see Sec. 1.4, (e) see Sec. 1.5, (f) see Sec. 1.7. **2.** (a) Axiom, (b) theorem, (c) definition, (d) theorem. **3.** For $a > 0$, $b > 0$. **4.** The converse theorem: if a quadratic equation has two distinct real roots, then the discriminant is positive, is true. The inverse theorem: if the discriminant of a quadratic equation is nonpositive (that is, is negative or zero), then the equation cannot have two distinct real roots, is true. The contrapositive theorem: if a quadratic equation does not have two distinct real roots, then the discriminant is nonpositive, is true. (Prove all these theorems!). **5.** Prove by contradiction. **6.** $(a_1-\alpha)\ldots(a_n-\alpha)=0$. **7.** $|ab|+|bc|+|ca|=0$. **8.** Either $a<0$, $b>0$, or a and b are numbers of the same sign, and $a>b$. This follows from the fact that the function $y=1/x$ is negative and decreases on the set of negative numbers and is positive and decreases on the set of positive numbers **9.** If arbitrary x_1 and x_2 are such that $-1 \leqslant x_1 < x_2 \leqslant 1$, then $3x_1-x_1^2 < 3x_2-x_2^2$. **10.** Necessary condition.

Sec. 1.2

1. The number $\overline{a_n\ldots a_3a_2a_1a_0}$ is divisible by 11 if and only if the number $|a_0-a_1+a_2-a_3+\ldots+(-1)^n a_n|$ is. **2.** Prove by contradiction. Let $N=m^2$; consider two cases: when m is divisible by 3 and when it is not. **3.** Prove that $(p-1)(p+1)$ is divisible by 3 and by 8. **4.** $p=3$. One out of three consecutive numbers p, $p+1$, $p+2$ must be divisible by 3. Therefore, either $p=3$ or (for $p>3$) the number $p+1$ is divisible by 3. But then $p+4=(p+1)+3$ cannot be prime. **5.** Assure yourself that the last digit of the square of any integer may be only one of the digits 0, 1, 4, 5, 6, 9. **6.** 494. **7.** There are ten such numbers: 17, 32, 47, 62, 77, 92, 107, 122, 137, 152. **8.** For n representable in the form $5s-3$, where s is a positive integer. If $3n+4=5k$, then $3(n+3)=5(k+1)$, and so $n+3$ must be divisible by 5. **9.** Not for any. If we had the equations $2n+3=kp$ and $5n+7=kq$, $k>1$, then we would have $1=(5p-2q)k$. **10.** 3762. Assure yourself by simple enumeration that there is a unique representation $13=2^2+3^2$ and only two representations $85=2^2+9^2=6^2+7^2$. From the third condition it follows that the first digit of the desired number cannot be less than the last, and so we have to consider four numbers: 3292, 3922, 3672, 3762. **11.** 857. Note that $\overline{abc}\,1=10\cdot\overline{abc}+1$ and $2\,\overline{abc}=\overline{abc}+2000$.

---

* In the answers to trigonometric equations and inequalities, k, l, m, n denote arbitrary integers unless otherwise stated.

12. 34 056, 34 452, 34 956. Use the criteria of divisibility by 4 and by 9. **13.** $n \geqslant 2$. Factor n^4+4. **14.** Show that among the numbers k, m, n there are either two even numbers or two odd numbers. Then utilize the identity $k^3+m^3+n^3=(k+m+n)^3-3(k+m)(m+n)(n+k)$. **15.** $\underbrace{\overline{1\ldots 1}}_{2n \text{ times}}-\underbrace{\overline{2\ldots 2}}_{n \text{ times}}=$ $=9^{-1}\cdot\underbrace{\overline{9\ldots 9}}_{2n \text{ times}}-2\cdot 9^{-1}\cdot\underbrace{\overline{9\ldots 9}}_{n \text{ times}}=9^{-1}(10^{2n}-1)-2\cdot 9^{-1}\cdot(10^n-1)=[(10^n-1)/3]^2=$ $=(\underbrace{3\ldots 3}_{n \text{ times}})^2$. **16.** $x_1=3$, $y_1=1$; $x_2=-3$, $y_2=-1$. Regarding the equation as a quadratic in x, conclude that its discriminant $25y^2+56$ must be the square of an integer, that is, $56=(k+5y)(k-5y)$. Then run through the possible reprèsentations of the number 56 as a product of two integers. **17.** $x_1=2$, $y_1=2$; $x_2=2$, $y_2=-2$; $x_3=-2$, $y_3=2$; $x_4=-2$, $y_4=-2$. Express x^2 in terms of y^2 and conclude that y can only assume integral values for which $3<y^2\leqslant 12$. **18.** For example, the numbers $(a+b)/2$ and $a+2^{-1/2}(b-a)$. **19.** No, they cannot. **20.** Show that from the rational nature of the number $\tan 5°$ it would follow that $\tan 30°$ is a rational number.

Sec. 1.4

1. Yes. For $a=0$ the absolute value of a is simultaneously equal to a and $-a$ **2.** (a) If $x\geqslant 0$, then $-x\leqslant 0$, and then $|-x|=-(-x)=x=|x|$; if $x<0$, then $-x>0$ and then $|-x|=-x=|x|$. (b) If $x\geqslant 0$, then $|x|=x$ and, hence, $x\leqslant|x|$; if $x<0$, then $x<-x$, i.e., $x<|x|$, whence $x\leqslant|x|$. **4.** $a_1=\ldots=a_n=0$. **5.** For the sake of definiteness let $a<b$; then $|b-a|=b-a$. Three cases are possible: $0\leqslant a<b$, $a<0\leqslant b$, $a<b<0$. In the first case, the distance is equal to $b-a=|b-a|$, in the second, $b+|a|=b-a=|b-a|$, in the third, $|a|-|b|=-a-(-b)=b-a=|b-a|$. **6.** (a) $a-b<x<a+b$, (b) $x\leqslant a-b$, $x\geqslant a+b$. **7.** $x_1=3/2$, $x_2=7/6$. **8.** $x=-1$. **9.** No solution. **10.** x any number. **11.** No solution. **12.** $7/6\leqslant x\leqslant 3/2$. **13.** $x_1=0$, $x_2=-1$. Since $|x|+x^3>0$ for $x>0$, then there are no positive roots; for $x\leqslant 0$ we have the equation $-x+x^3=0$ for which nonpositive roots are sought. **14.** $x=-1$. It is clear that the roots satisfy the inequality $x+1\leqslant 0$; rewrite the equation in the form $(x+1)(1-|x-1|)=0$ and choose the root for which the condition $x\leqslant -1$ is valid. **15.** No solution. Assure yourself that $-x^2+2x-3<0$ for all x; obtain the equation $x^2-2x+2=0$, which has no real roots. **16.** $x=-1$, $x\geqslant 0$. Rewrite the inequality in the form $|1+x|^2\geqslant|1+x||1-x|$; $x=-1$ is a soluiton and if $x\neq -1$, then it is possible to cancel out $|x+1|>0$. **17.** $1\leqslant x\leqslant 3$, $x=4$. **18.** $x<-3$, $-3<x<-2$, $x>0$. **19.** $x\leqslant 3/2$. **20.** $0\leqslant x\leqslant 2$. Since $x^2+x+1>0$ for all x, then x^2+x+1 may be cancelled out of both members of the inequality. **21.** $x_1=2$, $x_2=5$. **22.** $x_1=1/\sqrt{2}$, $x_{2,3}=-1\pm(1/\sqrt{2})$. **23.** $x=-2$. **24.** $x_{1,2}=\pm\log_3 2$. Note that the equation remains unaltered when x is replaced by $-x$, so that it suffices to find only nonnegative roots (the nonpositive roots are then obtained by reversing the sign). **25.** $x=-1$, $0\leqslant x\leqslant 1$. **26.** $x<1/3$, $x>3$. **27.** $4/3<x<4$. **28.** $2<x<5$. **29.** $x\leqslant -2-\sqrt{2}$, $x\geqslant 1+\sqrt{3}$. **30.** $1-\sqrt{17}\leqslant x\leqslant\sqrt{5}-1$. **31.** $x_1=2$, $y_1=1$; $x_2=0$, $y_2=-3$; $x_3=-6$, $y_3=9$. **32.** $x_{1,2}=\pm(1-\sqrt{1-4a})/2$ if $a<0$; $x=0$ for $a=0$; for $a>0$ there is no solution. Rewrite the equation in the form $|x|^2+|x|+a=0$. **33.** $x_{1,2}=\pm\log_{12}(1+\sqrt{1-a})$ for $a<1$; $x=0$ for $a=1$; for $a>1$ there is no solution. **34.** $x_{1,2}=2\pm\log_3(2-\sqrt{4+a})$ for $-3\leqslant a<0$; there is no solution for the other values of a. **35.** Use the inequality between the arithmetic mean and the geometric mean **36.** $\sqrt{|x|}$, x^2, $|x|$, x, $|x|^3\sqrt{y}$, x^3y^2. **37.** For $a\geqslant 0$. **38.** $-2a$. **39.** $2/(2-x)$. Note that $x\pm 2\sqrt{x-1}=(\sqrt{x-1}\pm 1)^2$. **40.** $|\cos\alpha-\cos\beta|$.

Sec. 1.5.

1. (a) On a circle of radius 20 with centre at the origin, (b) on a circle oj radius 5 with centre at the point (2, 0), (c) on a circle of radius 15 with centre at the point (—1, 0). **2.** Within a circle of radius 1 with centre at the origin **3.** On a circle of radius 2 with centre at the origin and exterior to this circle. **4.** Inside an annulus bounded by circles with centres at the origin and radii 1 and 2 and on the first circle. **5.** Within a circle of radius 2 with centre at the origin. **6.** On a circle of radius 3 with centre at the point (—1, 0). **7.** Inside a circle of radius 1 with centre at the point (0, 1). **8.** On a circle of radius 7 with centre at the point (—1, 2). **9.** Exterior to a circle of radius 9/2 with centre at the point (—1/2, 1/2). **10.** Inside an annulus bounded by circles with centres at the point (0, —1) and radii 2 and 3 and on both circles. **11** On a straight line perpendicular to the line segment connecting the points (0, 0) and (0, 1/3) and passing through the midpoint of the line segment. **12.** The point (0, $\sqrt{3}/3$). **13.** The point (0, —2). **14.** No such points. **15.** On a circle of radius 1 with centre at the origin **16.** On a circle of radius $\sqrt{3/2}$ with centre at the origin. **17.** There are two such complex numbers: $z_1=6+17i$ and $z_2=6+8i$. **18.** $\frac{z_1+z_2}{2}$ **19.** $z_1+z_2-z_3$ $z_1+z_3-z_2$, $z_2+z_3-z_1$. **20.** $\cos 150°+i\sin 150°$. **21.** $2\cos 20°(\cos 20°+i\sin 20°)$. **22.** $\cos(\pi-\alpha)+i\sin(\pi-\alpha)$. **23.** $\cos\left(\frac{3\pi}{2}+\alpha\right)+i\sin\left(\frac{3\pi}{2}+\alpha\right)$. **24.** $\frac{1}{\cos\alpha}\left[\cos\left(\frac{3\pi}{2}+\alpha\right)+i\sin\left(\frac{3\pi}{2}+\alpha\right)\right]$ if $0<\alpha<\frac{\pi}{2}$, $\frac{1}{|\cos\alpha|}\left[\cos\left(\frac{\pi}{2}+\alpha\right)+i\sin\left(\frac{\pi}{2}+\alpha\right)\right]$ if $\frac{\pi}{2}<\alpha<\pi$. **25.** On a ray issuing from the origin at an angle of $\pi/4$ to the positive x-axis **26.** On a ray issuing from the origin at an angle of $7\pi/6$ to the positive x-axis. **27.** The portion of the plane containing the second and third quadrants including the y-axis minus the origin, and a portion of the first quadrant located between rays emerging from the origin at angles of $\pi/3$ and $\pi/2$. **28.** A line segment of the x-axis between the points (—1, 0) and (0, 0) without the endpoints. **29.** The point (0, 2) **31.** $z=12+16i$. The point representing this number is the point of tangency of a ray issuing from the origin and lying in the first quadrant to a circle of radius $\sqrt{15}$ with centre at the point (0, 25). **32.** $\varphi/2$ if $0\leqslant\varphi<\pi/3$ or $\pi<\varphi<5\pi/3$; $\varphi/2+\pi$ if $\pi/3<\varphi<\pi$ or $5\pi/3<\varphi<2\pi$; if $\varphi=\pi/3$ or $\varphi=\pi$ or $\varphi=5\pi/3$, then $z_1=0$. **33.** Generally speaking, no; for example, $z_1=i$, $z_2=2i$. **34.** Generally speaking, no; for example $z_1=1+i$, $z_2=2-i$. **35.** Note that if the imaginary parts of the numbers z_1, z_2 are equal to zero, i.e., if these numbers are real, then the assertion is not true. **37.** All real numbers. **38.** All pure imaginaries. **39.** All complex numbers with real parts equal to 1. **40.** 0. **41.** $0, -1, \frac{1}{2}-i\frac{\sqrt{3}}{2}, \frac{1}{2}+i\frac{\sqrt{3}}{2}$. **42.** $0, i$. **43.** For $x=1$, $y=-4$ and for $x=-1$, $y=-4$. **44.** (a) Exterior to a circle of radius 10 with centre at the point (—1, 0), (b) inside a circle of radius 5 with centre at the origin. **45.** If $a=1$, then $z=-1-i$; if $1<a<\sqrt{2}$, then $z_{1,2}=\frac{-a^2\pm a\sqrt{2-a^2}}{a^2-1}-i$; if $a=\sqrt{2}$, then $z=-2-i$; if $a>\sqrt{2}$, then the equation has no solution **46.** If $0\leqslant a\leqslant 1/2$, then the equation does not have a solution; if $a>1/2$, then $z=\frac{4a+\sqrt{4a^2+3}}{16^2-4}+\frac{i}{4}$. **47.** $z_1=1$, $w_1=-1$; $z_2=-i$ $w_2=1$.

Sec. 1.6.

1. $N^{-\frac{\alpha\beta\gamma}{s}}$. 2. 0. 3. 0. 4. $-1/4$. 5. 1. 6. 2. 7. $\log_b N$. 8. 0. 9. 0. 10. 0 because $2^{\log_3 5}=(5^{\log_5 2})^{\log_3 5}=5^{\log_3 2}$. 11. $7\frac{1}{196}$. 12. $\frac{16}{\sqrt[3]{5}}$. 13. $\log_3 2<\log_2 3$. 14. $\log_4 5>\log_6 5$. 15. $\log_2 3<\log_3 11$ 16. If $a>1$, then $\log_2 a>\log_3 a$; if $a=1$, then $\log_2 a=\log_3 a=0$; if $0<a<1$, then $\log_2 a<\log_3 a$. 17. If $a>1$, then $\log_a 2<\log_a 3$; if $0<a<1$, then $\log_a 2>\log_a 3$. 18. $\sqrt[3]{0.01}<\sqrt[5]{0.001}$. 20. $\frac{1+ab}{a(8-5b)}$. 21. $3(1-a-b)$. 22. $a>0$, $a\neq 1$, $N>0$, $N\neq 1$; $b>0$, $b\neq 1/a$. 24. 1. 25. For $a=1$ the equation is obvious; if $a\neq 1$, then pass to logarithms to the base a. 27. Since $0<\log_3 2<1$ and $1/2<1$, it follows that the inequality is valid. 28. First prove that $\log_b a\log_c b\log_d c=\log_d a$, $b\neq 1$, $c\neq 1$, $d\neq 1$, $a>0$, $b>0$, $c>0$, $d>0$. 29. For $a>1$ and $b>1$ or for $0<a<1$ and $0<b<1$. 30. Prove by contradiction. 31. $\frac{1-\sqrt{5}}{2}\leqslant x<0$, $x\geqslant\frac{1+\sqrt{5}}{2}$. 32. $x_1=1/8$, $x_2=2$. 33. $x=2$. 34. $x_1=10$, $x_2=\sqrt[9]{10}$. 35. $x_1=16$, $x_2=1/2$. 36. $x_1=1/4$, $x_2=\frac{1}{\sqrt[4]{2}}$. 37. $x_1=9$, $x_2=1/9$. 38. $x_1=a^{1/\sqrt{2}}$, $x_2=a^{-1/\sqrt{2}}$. 39. $x_1=a^{\frac{3+\sqrt{5}}{2}}$, $x_2=a^{\frac{3-\sqrt{5}}{2}}$. 40. $x_1=9$, $x_2=1/9$. 41. $x=\sqrt{2}$, $y=1$. 42. $x=a^{1-4a}$, $y=a^{2a}$. 43. $x_1=2$, $y_1=4$, $x_2=1/2$, $y_2=1/4$. 44. $x=9$, $y=1/3$. 45. $x=a^{4m-6n}$, $y=a^{6m-12n}$. 46. $x=2/3$, $y=27/8$, $z=32/3$. 47. $x=1$, $y=2$. 48. $x_1=4$, $y_1=32$; $x_2=-1$, $y_2=1$. 49. $0<x<3^{\frac{1}{1-\log_2 3}}$. 50. $0<x<\sqrt[10]{10}$. 51. $a\geqslant 1/16$.

Sec. 1.7.

1. $x=3$. 3. $3-2\sqrt{2}$, 1, $3+2\sqrt{2}$. 4. $\sqrt{\frac{\sqrt{5}+1}{2}}$, $\sqrt{\frac{\sqrt{5}-1}{5}}$. 5. $A=2$, $B=32$. 6. 25 stones. 7. 27, 18 and 12 years old. 8. If $b_1>0$, then arbitrary values of $q<1$ and arbitrary values of $q>3$ are admissible; if $b_1<0$, then arbitrary values of q in the interval $1<q<3$ are admissible. 9. 25 harvesting combines. 10. 192 litres. 11. 18 litres. 13. $x=y=z=2$. 15. Note that the numbers $\frac{1}{a_1}$, $\frac{1}{a_2}$, ..., $\frac{1}{a_n}$ constitute a geometric progression and compute the sum. $P=(S/T)^{n/2}$. 16. $x_1=2$, $x_2=-2$. 17. $x_1=\sqrt{2}$, $x_2=-\sqrt{2}$.

Sec. 1.8.

1. Multiply the inequality by 2 and pass to the obvious inequality $(a-b)^2+(b-c)^2+(c-a)^2\geqslant 0$. 3. Rewrite the inequality in the form $(a-1)^2+(b-1)^2+(c-1)^2\geqslant 0$. 7. Rewrite the inequality in the form $(x+y+1)^2+2(y+1)^2\geqslant 0$. 8. Rewrite the inequality in the form $(1-\sin^2 x)(5-\sin^2 x)\geqslant 0$. 9. For $x\leqslant 0$, all terms are nonnegative; for $0<x<1$ we have $x^2>x^5$, $1>x$; for $x\geqslant 1$ we have $x^8\geqslant x^5$, $x^2\geqslant x$. 10. Since $a/c+b/c=1$, then $0<a/c<1$,

$0 < b/c < 1$, and so $(a/c)^{2/3} > (a/c)$, $(b/c)^{2/3} > (b/c)$. **11.** Extract the nth root of both members of the inequality. **12.** Assure yourself that in the left member of the inequality the sum of the numbers equidistant from the ends is not less than $2(n+1)^{-1}$. **13.** Represent $(n!)^2 = (1\cdot n)\cdot[2\cdot(n-1)]\ldots[(n-1)\cdot 2]\cdot(n\cdot 1)$ and prove that for arbitrary k, $1 < k < n$, the inequality $k(n-k+1) > n$ is valid. **14.** Apply the method of mathematical induction. **15.** Apply the method of mathematical induction. **19.** If p is the semiperimeter of a triangle with sides a, b and c, and S is the area, then, by Hero's formula and the inequality between the geometric mean and the arithmetic mean, we get $\sqrt{S} \leqslant p/2$. **20.** Introduce an auxiliary angle. **21.** The least value 2/3 is attained when $x=0$, the greatest value is absent. **22.** The least value 0 is attained when $x=0$, the greatest value 1/2 is attained when $x=-1$ and when $x=1$. To determine the least value, represent the function in the form $y=(x^2+x^{-2})^{-1}$. **23.** The greatest value 1 is attained when $x=1$; the least value -1 is attained when $x=-1$. **24.** Apply the inequality between the arithmetic mean and the geometric mean and then utilize the inequality $\sin x+\cos x \geqslant -\sqrt{2}$. Equality is attained when $x=(5\pi/4)+2k\pi$. **25.** Express $\cot x$ in terms of $\cot(x/2)$.

Sec. 1.9

1. $x=2$. **2.** $x=(14+2\sqrt{15})/3$. **3.** $x=45-16\sqrt{7}$. **4.** No solution. **5.** $x=2$. **6.** $x=8$. **7.** $x=5$. **8.** $x_1=0$, $x_2=4$. **9.** $x_1=1$, $x_2=3/2$, $x_3=2$. **10.** $x_1=190/63$, $x_2=2185/728$. **11.** $x=\sqrt[5]{a^4}/(1-\sqrt[5]{a^4})$ for $0 < a < 1$; for the remaining a, there are no solutions. **12.** $x=-2$. **13.** $x_1=0$, $x_2=2$. **14.** $x=\log_{2/7} 3$. **15.** $x_{1,2}=1 \pm \sqrt{1+\log_{2+\sqrt{3}} 10}$. **16.** $x=\log_2 3$. **17.** $x=4\log_3 2$. **18.** $x_1=\log_3 28-3$, $x_2=\log_3 10$. **19.** $x=2$. **20.** $x_1=1/8$, $x_2=2$. **21.** $x=[\log_2(1+\sqrt{41})-1]^3/6$. **22.** $x_1=10$, $x_2=\sqrt[9]{10}$. **23.** $x_1=9$, $x_2=1/9$. **24.** $x_1=2$, $x_2=8$. **25.** $x_1=0$, $x_2=1$. **26.** No solution. **27.** $x_1=100$, $x_2=1/100$. **28.** $x_1=9$, $x_2=1/9$. **29.** $x_1=1$, $x_2=2$, $x_3=1/\sqrt[4]{8}$. **30.** $x_1=\sqrt[4]{3}$, $x_2=1/\sqrt[4]{3}$. **31.** $x=2^{-\log_3 a}$. **32.** No solution. **33.** No solution. **34.** $x=5/2$. **35.** $x=-25$. **36.** $x=5$. **37.** $x=8$. **38.** $x=1/9$. **39.** $x=3$. **40.** $x=5$. **41.** $x=-4$. **42.** $x=2$. **43.** $x=1$. **44.** $0 \leqslant x \leqslant 1$, $x=4$. **45.** $x=4$. **46.** $x_1=2k\pi/5$, $x_2=(6k\pm 2)\pi/15$. **47.** $x_1=(4k-1)\pi/4$, $x_2=(2k+1)\pi/2$. **48.** $x_1=k\pi$, $x_2=[(-1)^{k+1}\pi/6]+k\pi$. **49.** $x=[(-1)^k\pi/6]+k\pi$. **50.** $x=(8k+1)\pi/4$. **51.** $x=(8k+1)\pi/4$. **52.** $x=(6k+1)\pi/3$. **53.** $x=(1/2)\arccos(1/7)+2k\pi$. **54.** $x=(16k+1)\pi/8$. **55.** $x_1=(20k+1)\pi/5$, $x_2=(20k+9)\pi/5$. **56.** $x=\frac{1}{2}\arctan 4+2k\pi$. **57.** $x=(4k+1)\pi/2$. **58.** $x_1=(4k+1)\pi/4$, $x_2=\arcsin[(\sqrt{5}-1)/2]+(2k+1)\pi$. **59.** $x_1=2$, $x_2=-5$, $x_3=(2\pi/3)+4k\pi$, $x_4=(-2\pi/3)+4k\pi$ $(k \neq 0)$, $x_5=2\times\arccos[(2\sqrt{3}-3)/2]+4k\pi$, $x_6=-2\arccos[(2\sqrt{3}-3)/2]+4k\pi$ $(k \neq 0)$. **60.** $x_1=(2k+1)\pi/6$, $x_2=(4k+1)\pi/4$. **61.** $x=k\pi/3$. **62.** $x_1=(2k+1)\pi/4$, $x_{2,3}=(6k\pm 1)\pi/6$. **63.** $x=-\arcsin[(\sqrt{5}-1)/2]+(2k+1)\pi$. **64.** $x=\pm\arccos(2-2\sqrt{2})+2k\pi$. **65.** $x_1=k\pi$, $x_2=(12k+1)\pi/6$. **66.** $x=2$. **67.** $x=2$.

Sec. 1.10

1. $1 < x < 3$. **2.** $x > 5/2$. **3.** $x > \log_{\tan\frac{\pi}{8}} \frac{3+\sqrt{13}}{2}$. **4.** $0 < x < a$, $1 < x < 1/a$. **5.** $-1 < x < 0$, $0 < x < 1$. **6.** $2k\pi < x < (\pi/4)+2k\pi$. **7.** $0 < x < \frac{1}{2}$,

$x > 32$. **8.** $1/16 < x < 1/8$, $8 < x < 16$. **9.** $-1 < x < -2/\sqrt{5}$, $2/\sqrt{5} < x < 1$. **10.** $\log_2 5 - 2 < x < \log_2 3$. **11.** $2^{-1/\sqrt{2}} \leqslant x < 1$. **12.** $0 \leqslant x < \log_3^2 2$, $x > 3/2$. **13.** $-2 < x < -1$, $-1 < x < 0$, $0 < x < 1$, $x > 2$. **14.** $0 < x < \pi/24$, $5\pi/24 < x < \pi/4$. **15.** $1 < x < 2$. **16.** $(\sqrt{61} - 9)/2 < x < 2$. **17.** $2k\pi - (\pi/4) + \arcsin(3/4) < x < (3\pi/4) - \arcsin(3/4) + 2k\pi$. **18.** $x < -3$, $-2 < x < -1$. **19.** $x < 1$, $(3/2) < x < 2$, $x > 3$. **20.** $-\sqrt{(4-\pi)/\pi} \leqslant x \leqslant \sqrt{(4-\pi)/\pi}$. **21.** $(\pi/6) + k\pi < x < (5\pi/6) + k\pi$. **22.** $(\pi/4) + k\pi < x < (3\pi/4) + k\pi$. **23.** $0 < x < \pi/2$, $2\arctan\left[(1+\sqrt{5})/2\right] < x < \pi$. **24.** $x < \log_2 3$. **25.** $-1 \leqslant x \leqslant 1$. **26.** $0 < x < 1$, $x \geqslant 2$. **27.** $1/2 < x < 1$. **28.** $x < -7$, $5 - < x \leqslant -2$, $x \geqslant 4$. **29.** $x \leqslant -2/3$, $1/2 \leqslant x \leqslant 2$. **30.** $2k\pi - (\pi/4) + \arcsin(2\sqrt{2}/3) < x < 2k\pi + (3\pi/4) - \arcsin(2\sqrt{2}/3)$, $2k\pi - \pi < x < 2k\pi - (\pi/2)$. **31.** $-1 < x < (1-\sqrt{5})/2$, $(1+\sqrt{5})/2 < x < 2$. **32.** $-1 \leqslant x < 1 - (\sqrt{31}/8)$. **33.** $(3-a)/(2-a) < x \leqslant 2$ if $0 < a < 1$; $2 \leqslant x < (3-a)/(2-a)$ if $1 < a < 2$; $x \geqslant 2$ if $a = 2$; $x < (3-a)/(2-a)$, $x \geqslant 2$ if $a > 2$. **34.** $-1 < x < 0$, $2 < x < 3$. **35.** $1/2 < x < 1$. **36.** $-5/2 < x < -2$, $-3/2 < x < 3/2$, $3/2 < x < 8/3$. **37.** $-1 - \sqrt{23/2} \leqslant x \leqslant \sqrt{23/2} - 1$ if $a > 1$; $-\sqrt{24} < x \leqslant 1 - \sqrt{23/2}$, $\sqrt{23/2} - 1 \leqslant x < \sqrt{24}$ if $0 < a < 1$. **38.** $(14 + \sqrt{7})/2 \leqslant x \leqslant 9$. **39.** $(\pi/4) + k\pi < x < (\pi/3) + k\pi$. **40.** $1 < x < 3/2$, $2 < x < 5/2$, $x > 3$. **41.** $-(\pi/2) + k\pi < x < -\arctan 2 + k\pi$, $-(\pi/4) + k\pi < x < (\pi/4) + k\pi$. **42.** $1/4 + k \leqslant x < 1/2 + k$, $1/2 + k < x \leqslant 1 + k$. **43.** $2k\pi < x < (\pi/6) + 2k\pi$, $(\pi/6) + 2k\pi < x < (2k+1)\pi$. **44.** $x \geqslant 1$. **45.** $1 - \sqrt{6} < x \leqslant 2 - \sqrt{10}$, $1 + \sqrt{6} < x \leqslant 2 + \sqrt{10}$. **46.** $(\pi/4) + k\pi < x \leqslant (\pi/3) + k\pi$. **47.** $x < -1$, $x > 1$. **48.** $0 < x < 1$. **49.** $-1/2 \leqslant x < -1/4$, $\frac{3}{4} < x \leqslant 1$. **50.** $0 \leqslant x \leqslant 81$, $x \geqslant 1296$. **51.** $\arctan 5 + 2k\pi < x < \pi + 2k\pi$. **52.** $x = k\pi/3$. **53.** $-6 \leqslant x < 0$, $3 < x \leqslant 4$. **54.** $x \leqslant -2 - \sqrt{2}$, $1 + \sqrt{3} \leqslant x$. **55.** $x < \log_5 3$. **56.** $-(3+\sqrt{5})/2 < x \leqslant 1$. **57.** $3 < x < \pi$, $\pi < x < 3\pi/2$, $3\pi/2 < x < 5$. **58.** $0 < x < 2$, $x > 4$. **59.** $-(\pi/4) + k\pi < x \leqslant k\pi$, $(\pi/4) + k\pi < x < (\pi/2) + k\pi$. **60.** $-(7\pi/6) + 2k\pi \leqslant x \leqslant (\pi/6) + 2k\pi$.

Sec. 1.11

1. $x = -4$, $y = 6$. **2.** $x = 7$, $y = 5$. **3.** $x_1 = 1$, $y_1 = 1$; $x_2 = 16/81$, $y_2 = 4/9$. **4.** $x = 2/3$, $y = 27/8$, $z = 32/3$. **5.** $x_1 = 1$, $y_1 = 6$; $x_2 = 2$, $y_2 = 7$; $x_3 = 3$, $y_3 = 8$. **6.** $x_1 = 4$, $y_1 = 32$; $x_2 = -1$, $y_2 = 1$ **7.** $x = 8$, $y = 2$. **8.** $x_1 = 0$, $y_1 = 2\log_2 3 - 2$; $x_2 = 3$, $y_2 = -1$. **9.** $x = 5$, $y = 0$. **10.** $x_1 = \sqrt{3}$, $y_1 = 4$; $x_2 = -\sqrt{3}$, $y_2 = 4$. **11.** $x = 1$, $y = 0$. **12.** $x = 20$, $y = 16$. **13.** No. **14.** $-2 < k < 4$. **15.** $m = 0$. **16.** $\alpha \neq (\pi/2) + k\pi$.

Sec. 1.12.

1. 20 km/hr, 25 km/hr, 15 km/hr. **2.** 27, 18 and 12 years old. **3.** 30 km. **4.** 1/6. **5.** 19π cm/sec and 27π cm/sec. **6.** No. **7.** 3 hours, 6 hours, 2 hours. **8.** 5 grams and 20 grams. **9.** $5m(t-t_2) - 5m^{-1}(t-t_1)$. **10.** The area of the forest is 40 km². Obtain the equation $AC = 5 + 1/4\, BC^2 + 1/16\ AB^2$ from the statement of the problem; besides, for any three points A, B and C the inequality $AC \leqslant AB + BC$ is valid, whence $5 + 1/4\ BC^2 + 1/16\, AB^2 \leqslant AB + BC$ or $(1/2\, BC - 1)^2 + (1/4\, AB - 2)^2 \leqslant 0$, which is possible only for $AB = 8$ and $BC = 2$. **11.** The number of marks 2, 3, 4 and 5 are equal, respectively, to 11, 7, 10, and 2. **12.** Velocities are: motorcycle, 40 km/hr, Moskvich car, 60 km/hr, and Volga car 80 km/hr. **13.** The water is delivered twice as fast. **14.** 1:3. **15.** 20 km/hr and 80 km/hr. **16.** No. **17.** No. **18.** The rate of the cyclist is 20 km/hr, that of the truck, 40 km/hr, of the Volga car 80 km/hr. The distance from A to D is 60 km. **19.** No. **20.** 60 cubic metres. **21.** 2 minutes. **22.** 0.6 km/min. **23.** 12/7 days. **24.** 12 hours. **25.** Four boxes of the third type and 25 boxes of the second type. **26.** 12 sheets. **27.** The first pipe will fill the pool in 2 hours, the second in 4 hours. **28.** 4 hours. **29.** 16 hours and 45 minutes. **30.** 18 hours.

Sec. 1.13.

1. $-1/2 \leqslant x \leqslant 1/2$. **2.** The function is not defined for a single value of x. **3.** $x=0$. **4.** $-(\pi/4)+k\pi \leqslant x \leqslant (\pi/4)+k\pi$. **5.** $x<0$, $0<x\leqslant 2$, $x\geqslant 3$. **6.** $x=k$. **16.** Note that $y=-2$ for all $x\neq 3$. **24.** The graph consists of points on the x-axis with abscissas $x=(\pi/2)+2k\pi$. **25.** Note that $y=1$. **26.** Note that $y=-\cos 2x$. **27.** Introduce an auxiliary angle. **32.** Take advantage of the method of adding graphs. **36.** Note that $y=\sqrt{1-x^2}$; see Sec. 2.5. **40.** The graph of the function y_1 consists of two branches, while the graph of the function y_2 consists of only one. **41.** The graph of the function y_1 is the bisector of the first quadrantal angle (minus the origin). **42.** The function y_1 is not defined for $x=k\pi/2$. **43.** Pieces of the parabola $y=(x-2)^2$ corresponding to $x\leqslant 1$ and $x\geqslant 3$, and pieces of the parabola $y=-x^2+4x-2$ corresponding to $x\leqslant 1$ and $x\geqslant 3$. **44.** The bisector of the first quadrantal angle and the third quadrantal angle (including the negative x-axis and negative y-axis). **45.** The interior (including the boundary) of the square with vertices at the points (2, 0), (1, −1), (2, −2), (3, −1). **46.** The half-plane lying below the straight line $y=x-2$ and the half-plane above the straight line $y=x+2$ (the lines themselves excluded). **47.** The interior (minus the boundary) of a rectangle with vertices (1, 2), (−1, 2), (−1, −2), (−1, −2), (1, −2). **48.** Pieces of a sine curve $y=\sin x$ corresponding to $2k\pi\leqslant x\leqslant(2k+1)\pi$ and pieces symmetric to them about the x-axis. **49.** The straight lines $y=2x+(2k+1)\pi$ and $y=-2x+(2k+1)\pi$. **50.** The portion of the plane to the right of the curve $x=\sin|y|$ (including the curve). **51.** The strip $0<y<1$ minus vertical line segments drawn at points $x=k\pi/2$. **52.** The curvilinear angle bounded from above by the straight line $y=x$ and from below by the curve $y=-2^x$ (minus these curves themselves), from which the positive x-axis is deleted. **53.** $y=1+(x-1)^{-1}$, $x>0$, $y>0$. **54.** $y=\sqrt{2}\,(x-x^{-1})$, $x>0$, $y>0$. **55.** $x=-y^2+(y/4)$, $y>x>0$.

Sec. 2.1

1. (e) See Sec. 2.4. **2.** (a) Theorem, b) definition, (c) theorem, (d) theorem. **3.** The converse theorem: "If $\cos\varphi\leqslant 0$, then the angle φ has its terminal side in the second quadrant" is not true. The inverse theorem: "If the angle φ does not have its terminal side in the second quadrant, then $\cos\varphi>0$" is not true. The contrapositive theorem: "If $\cos\varphi>0$, then the angle φ does not have its terminal side in the second quadrant" is true. **5.** (a) $\sin 1^\circ<\sin 1$, (b) $\tan 1>\tan 2$. **6.** (a) $\alpha=(-1)^k\beta+k\pi$ or, what is the same thing, $\beta=(-1)^n\alpha+n\pi$. In other words, if α and β are certain concrete angles such that $\sin\alpha=\sin\beta$, then there exists an integer k such that the angles α and β are connected by the relation $\alpha=(-1)^k\beta+k\pi$, (b) $\alpha=\pm\beta+2k\pi$, (c) $\alpha=\beta+k\pi$, $\alpha\neq(\pi/2)+n\pi$, $\beta\neq(\pi/2)+m\pi$, (d) $\alpha=(\pi/2)\pm\beta+2k\pi$.

7. $\cos(\alpha/2)=-\frac{1}{2}\sqrt{2+2\sqrt{1-\sin^2\alpha}}=-\frac{1}{2}\left(\sqrt{1+\sin\alpha}+\sqrt{1-\sin\alpha}\right)$,

$\sin(\alpha/2)=\pm\frac{1}{2}\sqrt{2-2\sqrt{1-\sin^2\alpha}}$, the plus sign being taken if $270^\circ\leqslant\alpha\leqslant 360^\circ$, the minus sign, if $360^\circ\leqslant\alpha\leqslant 450^\circ$; for all α in the interval $270^\circ\leqslant\alpha\leqslant 450^\circ$ we can write $\sin(\alpha/2)=\frac{1}{2}\left(\sqrt{1-\sin\alpha}-\sqrt{1+\sin\alpha}\right)$. **8.** 1. **9.** −1/8. Multiply and divide the expression at hand by $\sin(\pi/7)$.

Sec. 2.2

1. $x=18°$. Find the root of the equation $\cos 4x=\sin x$ lying between $0°$ and $90°$. **2.** Assure yourself that $y=\sin^2\alpha$ for arbitrary x. **3.** $(1+\cos^4 x)^{-1/2}$. **4.** $(a+b\sin^2 x)^{-1}$. **5.** $\tan^{3/2} x$, $x=(\pi/3)+k\pi$. **6.** $\sin a/(\cos a\cos x-1)$. **9.** $\sqrt{2}\cos x\cos^{-1}(x/2)\sin[(\pi/4)+(x/2)]$. **10.** 0. **11.** $a+b=2ab$. **12.** $(p^2-q^2)^2=-pq$. **13.** $y=a$, $z=c$. **14.** Multiply both parts of the equation being proved by $\sin(\pi/7)$. **15.** Use the equation $142°30'=90°+45°+(1/2)\cdot 15°$. **16.** $\alpha+\beta\neq(\pi/2)+k\pi$, $\alpha\neq(\pi/2)+k\pi$, $\beta\neq\pi+2k\pi$. **18.** $(\sqrt{7}-2)/3$. **19.** Establish that $0°<\alpha<30°$, $0°<\beta<30°$, i.e. $0°<\alpha+2\beta<90°$ and then compute $\sin(\alpha+2\beta)$. **20.** Transform the expression $[(a/b)(A/B)+1]/[(a/b)+(A/\beta)]$; the condition $aB+bA\neq 0$ signifies that $\sin[2x-(\alpha+\beta)]\neq 0$. **21.** q. **22.** $(\sin 2\alpha+\sin 2\beta+\sin 2\gamma)/4$. **23.** 2. **24.** 3. **25.** Replacing the cotangents in the sum $\cot(\alpha/2)+\cot(\beta/2)+\cot(\gamma/2)$ in terms of sines and cosines and reducing to a common denominator, transform the numerator to a product using the equation $\alpha+\beta+\gamma=\pi$.

Sec. 2.3

1. (a) $\alpha=(-1)^n\beta+n\pi$, where n is an integer or, what is the same thing, $\beta=(-1)^k\alpha+k\pi$, where k is an integer. In other words, if α and β are certain concrete angles such that $\sin\alpha=\sin\beta$, then there exists an integer n such that the angles α and β are connected by the relation $\alpha=(-1)^n\beta+n\pi$. (b) $\alpha=\pm\beta+2n\pi$. (c) $\alpha=\beta+n\pi$ and α, $\beta\neq(\pi/2)+k\pi$. (d) $\alpha=(\pi/2)\pm\beta+2n\pi$. **2.** If $|c|>\sqrt{a^2+b^2}$, then there is no solution; if $c=\sqrt{a^2+b^2}$, then $x=(\pi/2)-\varphi+2k\pi$; if $c=-\sqrt{a^2+b^2}$, then $x=-(\pi/2)-\varphi+2k\pi$; if $|c|<\sqrt{a^2+b^2}$, then $x=(-1)^k\arcsin\left[c/\sqrt{a^2+b^2}\right]-\varphi+k\pi$. Here, φ is an auxiliary angle defined from the conditions $\sin\varphi=b/\sqrt{a^2+b^2}$, $\cos\varphi=a/\sqrt{a^2+b^2}$. **3.** $x_1=(\pi/4)+k\pi$, $x_2=(5\pi/12)+k\pi$. **4.** $x_1=(\pi/30)+(2k\pi/5)$, $x_2=-(\pi/2)+2n\pi$. **5.** $x_1=(\pi/12)+(k\pi/7)$, $x_2=(\pi/4)+n\pi$. **6.** $x=\arccos(4/5)+2k\pi$. Express $\sin 3x$ in terms of $\sin x$. **7.** $x=1/4\arccos(3/5)+(k\pi/2)$. Express $\sin^2 2x$ in terms of $\cos 4x$. **8.** $z_1=-(\pi/3)+2k\pi$, $z_2=(\pi/9)+(2n\pi/3)$. **9.** $x_1=(\pi/2)+k\pi$, $x_2=\pm\arcsin(1/\sqrt{5})+n\pi$. Pass to the equation $\cos(\pi\cos^2 x)=\cos(\pi\sin 2x)$, whence either $2\sin 2x-\cos 2x=1-4x$ or $2\sin 2x+\cos 2x=4k-1$, where k is an integer. In both cases, the only value of k for which solutions exist is $k=0$. **10.** $x=\pm(\pi/3)+2k\pi$. **11.** $x_1=(2k+1)\pi$, $x_2=(-1)^n(\pi/3)+n\pi$. **12.** $x_1=(\pi/4)+k\pi$; $x_2=-\arctan 3+n\pi$. Prove that $\cos x\neq 0$ and divide all the terms of the equation by $\cos^2 x$. **13.** $x_1-\arctan(3/2)+k\pi$, $x_2=-\arctan(1/2)+n\pi$. Pass to the function $\tan x$. **14.** $x_1=k\pi$, $x_2=(-1)^n(\pi/6)+n\pi$. Pass to the function $\sin x$. **15.** $x_1=2(2k+1)\pi$, $x_2=(-1)^n(2\pi/3)+4n\pi$. **16.** $x_1=3k\pi$, $x_2=\pm(\pi/4)+3n\pi$. Denote $x/3$ by y and express $\tan 3y$ in terms of $\tan y$. **17.** $x=(-1)^n(\pi/20)-(6/5)+(n\pi/5)$. Denote $5x+6$ by y. **18.** $x_1=\dfrac{1-2n}{2n+3}$, $x_2=\dfrac{7-6k}{6k+5}$, $x_3=\dfrac{(m-1)\pi-\operatorname{arccot}2\sqrt{3}}{(m+1)\pi+\operatorname{arccot}2\sqrt{3}}$. Denote $\dfrac{\pi-\pi x}{1+x}$ by y. **19.** $x=\pm(\pi/3)+k\pi$. Pass to the function $\cos 2x$. **20.** $t_1=k\pi$, $t_2=\pm(\pi/6)+n\pi$. Pass to the function $\cos 2t$. **21.** $x=\pm(2\pi/3)+2k\pi$. Assure yourself that $\sin^4(x/2)-\cos^4(x/2)=-\cos x$. **22.** $x=(\pi/8)+(k\pi/4)$. Assure yourself that $\sin^8 x+\cos^8 x=2+12\cos^2 2x+2\cos^4 2x$ and then pass to $\cos 4x$. **23.** $x=\pm(\pi/6)+(k\pi/2)$. Use the formula for a sum of cubes. **24.** $x_1=2(2k+1)\pi$, $x_2=(2n+1)\pi$, $x_3=(2\pi/5)+(4m\pi/5)$. Get rid of the squares of cosines and transform the resulting sum into a product. **25.** $y_1=2k\pi$, $y_2=(3\pi/8)+(n\pi/2)$, $y_3=-(\pi/2)+2m\pi$. **26.** $x_1=k\pi/4$, $x_2=\pm 1/4\arccos\left[(1+\sqrt{5})/4\right]+(n\pi/2)$, $x_3=\pm 1/4\arccos\left[(1-\sqrt{5})/4\right]+(m\pi/2)$. **27.** $x=2\beta\pm(2\pi/3)+2k\pi$ Transform the product of cosines into a sum. 28. $x=2k\pi/[n(n+1)]$, where k is an arbitrary integer. Transform the products of sines to differences of cosines. 29. $x_1=(2k+1)\pi$, $x_2=(\pi/2)+n\pi$, $x_3=2m\pi/5$. 30. $x_1=n\pi/3$, $x_2=2m\pi/9$. 31. $x_1=k\pi$, $x_2=\pm\arctan\sqrt{5/7}+n\pi$. 32. $x_1=(\pi/4)+k\pi$, $x_2=(\pi/2)+2n\pi$, $x_3=(2m+1)\pi$. Employ the formulas $\cos 2x=(\cos x+\sin x)(\cos x-\sin x)$ and $1-\sin 2x=(\cos x-\sin x)^2$. 33. $x_1=2k\pi$, $x_2=(\pi/2)+n\pi$, $x_3=-(\pi/4)+m\pi$. Transform the left member of the equation to the form $2\cos x(\sin x+\cos x)^2$. 34. $x_1=\pm(\pi/4)+k\pi$, $x_2=-(\pi/6)+(n\pi/2)$. Denote $\tan x-\cot x$ by y; note that $\tan x-\cot x=-2\cot 2x$. 35. $x_1=\pm(\pi/6)+2k\pi$, $x_2=(-1)^{n+1}\arcsin(1/4)+n\pi$. Represent the radicand as a perfect square. 36. $-(\pi/3)+k\pi\leqslant x\leqslant-(\pi/6)+k\pi$, $(\pi/6)+k\pi\leqslant x\leqslant(\pi/3)+k\pi$. 37. $x=\log_2 p$, where $p=1, 2, \ldots$. 38. $x_1=\log_4 p$, $x_2=\log_4(4p-1)-1$, $x_3=\log_4(4p-3)-3/2$, where $p=1, 2, \ldots$. 39. $x_1=\pm(1/2)\arccos(1/3)+k\pi$. $x_2=\pm(1/2)\arccos(-1/4)+k\pi$. Rewrite the equation in the form $5(\tan x+\cot 3x)=\tan 2x-\tan x$ and pass to sines and cosines. 40. $x=(\pi/4)+2k\pi$. Rewrite the equation in the form $4(\sin x+\cos x)+(\sin x+\sin 3x)+(\cos x-\cos 3x)=2\sqrt{2}(2+\sin 2x)$. 41. $x=13\pi/4$. 42. $x_1=\pi/6$, $x_2=3\pi/10$, $x_3=7\pi/6$, $x_4=13\pi/10$. 43. $x=(2p+1)\pi/18$, where p is any integer except numbers of the form $9m+4$, $m=0, \pm1, \pm2, \ldots$. 44. $x_1=\pi/3$, $x_2=5\pi/3$. 45. $x_1=5\pi/6$, $x_2=13\pi/6$. 46. $x=(\pi/2)+2k\pi$. 47. No solution. 48. No solution. 49. $x=(2k+1)\pi$. 50. $x=-(\pi/6)+k\pi$. 51. $x=4$. 52. No solution. 53. $x=k\pi$. Use the formula for the difference of tangents and obtain the equation $\cos x\cos 2x=-1$. 54. Prove that the equation $\sin 2x=\sin 3x=1$ is not possible for a single value of x. 55. $x=\pm(1/2)\arccos[(a^2-2)/2]+k\pi$ for $|a|\leqslant 2$. 56. $a=1$, $x=k$. Reduce the equation to the form $\cos^2\pi(a+x)=1/(2a-a^2)$. 57. $x_1=7\pi/6$, $x_2=11\pi/6$ for $a=-1/4$; $x_1=\pi+\arcsin\left(1+\sqrt{1+4a}\right)/2$, $x_2=2\pi-\arcsin\left(1+\sqrt{1+4a}\right)/2$, $x_3=\pi+\arcsin\left(1-\sqrt{1+4a}\right)/2$, $x_4=2\pi-\arcsin\left(1-\sqrt{1+4a}\right)/2$ for $-1/4<a<0$; $x_1=0$, $x_2=\pi$, $x_3=3\pi/2$ for $a=0$; $x_1=\arcsin\left(\sqrt{1+4a}-1\right)/2$, $x_2=\pi-\arcsin\left(\sqrt{1+4a}-1\right)/2$ for $0<a<2$; $x=\pi/2$ for $a=2$; no solutions for $a<-1/4$ and for $a>2$. Write out the conditions for which the equation $z^2+z-a=0$ has real roots, the smaller root being at least -1, and the larger root not exceeding 1. 58. $x=k\pi$ for arbitrary $m>0$; besides, $x=\pm\frac{1}{2}\arccos\left(\sqrt{1+2m}-1\right)/2+n\pi$ for $0<m\leqslant 4$. Pass to the functions of the angle $2x$. 59. $x=(-1)^k\arcsin[b/(b-1)]+k\pi$ for $b<1/2$, $b\neq-1$, $b\neq 1/3$, $b\neq 0$. Reduce the equation to the form $\sin x=b/(b-1)$ and take into account the restrictions $\tan x\neq 0$, $\cos x\neq 0$, $2\cos 2x\neq 1$, $\tan^2 x\neq 1/3$ 60. $x=\pm\arcsin\sqrt{2/(1+a^2)}+k\pi$ for $|a|>1$, $|a|\neq\sqrt{3}$. Reduce the equation to the form $\sin^2 x=2/(1+a^2)$ and take into account the restrictions $\tan^2 x\neq 1$, $\cos 2x\neq 0$, $\cos x\neq 0$.

Sec. 2.4

1. $x_1=(7\pi/36)+k\pi$, $y_1=(5\pi/36)+k\pi$; $x_2=-(11\pi/36)+k\pi$, $y_2=-(13\pi/36)+k\pi$. 2. $x=(\pi/2)+k\pi$, $y=(\pi/6)-k\pi$. 3. $x=-\frac{\pi}{6}+\frac{(-1)^k}{2}\arcsin\frac{2-3\sqrt{3}}{6}+\frac{k\pi}{2}$, $y=\frac{\pi}{6}+\frac{(-1)^k}{2}\arcsin\frac{2-3\sqrt{3}}{2}+\frac{k\pi}{2}$. 4. $x_1=(\pi/4)+k\pi$, $y_1=\arctan 2+n\pi$, $z_1=(3\pi/4)-\arctan 2-(k+n)\pi$; $x_2=-(\pi/4)+k\pi$, $y_2=\arctan 2+n\pi$, $z_2=(5\pi/4)+\arctan 2-(k+n)\pi$. 5. $x=(-1)^k\arcsin\frac{2-\sqrt{2}}{2}+k\pi$, $y=\pm\arccos(2-\sqrt{2})+2n\pi$. 6. $x=\frac{1}{2}\left[(-1)^k\arcsin\frac{2}{5}+(-1)^n\arcsin\frac{4}{5}+(k+n)\pi\right]$, $y=\frac{1}{2}\left[(-1)^k\arcsin\frac{2}{5}-(-1)^n\arcsin\frac{4}{5}+\right.$

$+(k-n)\pi\Big]$. 7. $x=2$, $y=1$. 8. $x_1=\pi/3$, $y_1=\pi/6$, $x_2=0$, $y_2=\pi/2$. 9. $x_1=\pi$, $y_1=\pi$ for arbitrary $a>0$; if $1\leqslant a\leqslant 2$, then the solutions are also $x_2=\arcsin\sqrt{(4-a^2)/3}$, $y_2=\arcsin\sqrt{(4-a^2)/3a^2}$; $x_3=\pi-\arcsin\sqrt{(4-a^2)/3}$, $y_3=\pi-\arcsin\sqrt{(4-a^2)/3a^2}$; $x_4=\pi+\arcsin\sqrt{(4-a^2)/3}$, $y_4=$ $=\arcsin\sqrt{(4-a^2)/(3a^2)}$; $x_5=2\pi-\arcsin\sqrt{(4-a^2)/3}$, $y_5=\pi-$ $-\arcsin\sqrt{(4-a^2)/(3a^2)}$ 10. $x_1=\arctan\sqrt{a}+k\pi$, $y_1=\pm\arccos\dfrac{2\sqrt{a}}{b(1-a)}+$ $+2n\pi$; $x_2=-\arctan\sqrt{a}+k\pi$, $y_2=\pm\arccos\dfrac{2\sqrt{a}}{b(a-1)}$ if $a\geqslant 0$, $a\neq 1$, $2\sqrt{a}\leqslant|b(1-a)|$.

Sec. 2.5

1. $-\pi/3$, $\pi/2$. 2. $5\pi/6$, π, $\pi/2$. 3. $-\pi/6$, $-\pi/4$. 4. $2\pi/3$, $3\pi/4$, $\pi/2$. 5. $5-2\pi$, $4\pi-10$, $2\pi-6$, $4\pi-10$. 6. $-\arcsin\dfrac{4\sqrt{15}-3}{20}$, $\arctan\dfrac{1}{13}$, $\pi+\arctan\dfrac{1}{8}$. 13. $x\geqslant 0$. 14. $x>0$. 15. No solution. 16. $x=1$. 17. $x=0$. 18. $x=\sqrt{3}/2$.

Sec. 3.1.

1. (e) If two planes intersect, then four dihedral angles are formed at the straight line that is the line of their intersection (two pairs of vertical dihedral angles); any one of them can be taken as the angle between the planes. (f) Note two possibilities: when one of the radii of the circular sector, the revolution of which generates a spherical sector, coincides with the axis of revolution (in this case, the spherical sector is a convex figure, the base of the spherical sector is a segmental surface), and when the axis of revolution does not intersect the arc of this circular sector (in this case, the spherical sector is a nonconvex figure, its base is a zone of a sphere). 2. (a) Theorem, (b) definition, (c) axiom, (d) theorem. 3. Converse theorem: "If a straight line L is parallel to a plane π, then L is parallel to the straight line l" is not true. The inverse theorem: "If a straight line L is not parallel to a straight line l, then L is not parallel to the plane π" is not true. The contrapositive theorem: "If a straight line L is not parallel to a plane π, then L is not parallel to the straight line l" is true. 4. This follows from the converse of the Pythagorean theorem. 5. (a) Impossible. For example, we inscribe a triangle ABC in a circle and consider the inscribed polygons for which the chords AB and BC are always sides (i.e., all vertices, except B, are chosen on the arc AC). (b) Impossible. For example, let us take the sequence of inscribed triangles $A_nB_nC_n$ for which the arcs AB and BC are equal to $2\pi/n$ (what is the limit of the sequence of perimeters of these triangles?). 6. Compare the circumference with the perimeter of an inscribed regular hexagon and with the perimeter of a circumscribed square. 7. Not valid. The proof is incorrect and is admissible only if AB and AB_2 and BC and CB_2 do not constitute a single straight line; then tne triangles are indeed equal. But a configuration is also possible for which AB and AB_2 (or BC and CB_2) lie on one straight line. 8. Two distinct configurations are possible: either all three straight lines are parallel or all three intersect in a single point.

Sec. 3.2

1. (a) Draw the straight lines L^* and L^{**} parallel to L and distant a from it, and the straight lines l^* and l^{**} parallel to l and distant a from it. Consider the bisectors of the angles, which form at the points of intersection of L with l^* and l^{**}

and also l with L^* and L^{**}. (b) Join the arbitrary point M of the desired locus to points A and B; complete the triangle AMB to a parallelogram with diagonal AB and assure yourself that $MO^2=2a-AB^2=\text{const}$; hence, for $2a \geqslant AB^2$ the desired locus is a circle with centre at the point O (for $2a=AB^2$ it degenerates into a point). There are no such points M for $2a<AB^2$. **2.** A circle with diameter AB minus the points A and B. Draw a common tangent at point M and use the theorem on the equality of line segments of tangents drawn from a single point to the same cicrle. **3.** The entire plane. Show that from any point (as centre) in the plane it is always possible to draw a circle intersecting all three straight lines. **4.** A circle constructed on the line segment AB as on a diameter. **5.** A circle constructed on the line segment OA (where O is the centre of the given circle) as a diameter. **6.** A circle concentric with the given one and having radius r cosec $(\alpha/2)$, where r is the radius of the given circle. **7.** A circle with centre at the midpoint of the line segment OA (where O is the centre of the given circle) and radius half the radius of the given circle. **8.** A circle with centre O at the midpoint of the line segment AB and radius AB. Prove that $MO=AB$. **9.** (a) A straight line perpendicular to the plane of the triangle ABC and passing through the centre of the circle circumscribed about the triangle. (b) No such points. **10.** Four straight lines resulting from the intersection of the bisecting planes of dihedral angles between the planes Π and π with two planes parallel to Π and distant a from it. **11.** Four straight lines perpendicular to the plane π and passing through the centres of an inscribed circle and excribed circle of a triangle formed by the lines of intersection of the three given planes with the π plane. **12.** A sphere constructed on the line segment AB as a diameter. **13.** A great circle cut out of a sphere, which is constructed on the line segment AP (where P is the foot of a perpendicular dropped from point A on straight line l) as a diameter, by the plane drawn through AP perpendicular to the straight line l. **14.** All points lying inside the given dihedral angle. Show that if a plane is drawn through any interior point M perpendicular to the edge of the dihedral angle, then on the sides of the resulting plane angle it is always possible to choose points A and B so that M is the midpoint of the line segment AB. **15.** All points of plane Π drawn through the straight line L parallel to l, with the exception of points of the line L itself, and all points of the plane π drawn through the straight line l parallel to L except the points of the line l itself. **16.** Consider separately the cases when $l>a/\sqrt{2}$, $l=a/\sqrt{2}$, $l<a/\sqrt{2}$. **17.** A circle. Drawing a plane through the straight line AB and the point M of tangency, compute the distance KM, where K is the point of intersection of the straight line AB and the π plane. **18.** If $l>a$, then the desired locus consists of all points of a circle of radius $\frac{1}{2}\sqrt{l^2-a^2}$ with centre at the centre of a cube, which circle lies in the horizontal plane passing through the centre of the cube. If $l=a$, then the desired locus consists of a single point, the centre of the cube. If $l<a$, then not a single point of space has the required property. **19.** Note that $\sqrt[4]{a^4+b^4}=\sqrt{c\sqrt{(ac/b)^2+(bc/a)^2}}$, where $c=\sqrt{ab}$. **20.** Take advantage of the fact that if O is the centre of a circle circumscribed about the triangle ABC and OM is a perpendicular dropped from point O to the side $BC=a$, then $BM=MC=a/2$. It is thus easy to construct the triangle OBM and then the vertex C. **21.** If the straight lines AB and l intersect at point C, then take advantage of the equation $AC\cdot BC=CD^2$, where D is the point of tangency of the desired circle and line l. **22.** The centre of the desired circle is equidistant from the centre O of the given circle and from the point lying on a perpendicular erected to the straight line l at the point A, and at a distance from A equal to the radius of the given circle. **23.** From the ratio between the areas, express AC in terms of AB and construct the line segment AC on the basis of this formula. **24.** Set up an equation for the line segment BD and express it in terms of the side AB and the altitude h of triangle ABC dropped on the side AB. Construct line segment h and then construct BD by the formula thus obtained. **25.** From the equality of the volumes of the pyramid and prism express a leg of the desired isosceles right triangle in terms of h and a side of the base of the pyramid and perform the construction by this formula.

Sec. 3.3

1. $\sqrt{(bc-l^2)(b+c)^2/bc}$. 2. $\frac{1}{2}h \operatorname{cosec}\alpha \sqrt{-\cos 2\alpha}$. 3. $\sin A:\sin B:\sin C = \sqrt{5}:2\sqrt{2}:3$. 4. 1/16. 5. $\arctan[24va^{-3}n^{-1}\sin^2(\pi/n)\sec(\pi/n)]$. 6. b/a. 7. $1/6\,\pi a^3 \sin 2\alpha \sec^2 2\alpha$. 8. $2\arctan\frac{1}{4}$. 9. $\frac{1}{3}\pi bc(b+c)\sin 2\cos(\alpha/2)$. Note that the axis of revolution is perpendicular to the bisector of the angle A. 10. $R^2 \arcsin[r/(R-r)]$. 11. $\frac{1}{2}a\sec(\alpha/2)$. 12. $\arccos(-\cos^2\alpha)$. 13. $[bc/(b+c)]^2$. 14. $2m\cos\alpha\cot(\alpha/2)$; $\frac{1}{4}\,m\operatorname{cosec}^2(\alpha/2)$. 15. $R(\pi/2)+\alpha-\beta]$, $R[(\pi/2)-\alpha+\beta]$, $R[-(\pi/2)+\alpha+\beta]$, $R[(3\pi/2)-\alpha-\beta]$, where $\alpha=\arcsin(18/\sqrt{445})$, $\beta=\arcsin(21/\sqrt{445})$. 16. S/l. 17. $2/3\,r$. 18. $\frac{2p(p+q)R^2\sin 2\varphi\cos^2\varphi}{p^2+q^2+2pq\cos 2\varphi}$. 19. $\tan^2(\varphi/2)$. 20. $\sqrt{4/3(a^2-a+1)}$. 21. $3+2\sqrt{3}$. 22. $2ab/(a+b)$. 23. $8\sin(\alpha/2)\sin(\beta/2)\sin(\gamma/2)$. 24. $2\arcsin\frac{h_a h_b}{l(h_a+h_b)}$. 25. 9/4. 26. 12/5. 27. $1/2\, b\,|\cos(3\alpha/2)|\sec(\alpha/2)\operatorname{cosec}\alpha$. Note that for $\alpha<\pi/3$ the centre of the circumscribed circle lies closer to the base of the triangle than the centre of the inscribed circle; for $\alpha>\pi/3$ it is farther away. 28. $(a-r)/(a+r)$. Draw the straight line $APOQ$ through the point A and the centre O of the circle and note that the desired quantity is equal to the ratio of the areas of the triangles BPC and BQC. Then compute the ratio of the altitudes of these triangles dropped on side BC 29. $2\sin^2\alpha/[\alpha(1+\sin 2\alpha+\sin^2\alpha)]$, where α is the measure of the angle AOB in radians. 30. $2\arcsin\left[(1-m)/\sqrt{m^2+2m}\right]$ for $1/4<m<l$. 31. $-2S\cos 2\alpha\cos^2\alpha$. 32. $1/4\,a^2\,|\cos 2\alpha|\tan\alpha(1+2\cos\alpha)^{-1}$.

33. $a\sqrt{1+4\dfrac{\cos^2\dfrac{\alpha-\beta}{2}\sin^2\dfrac{\alpha+\beta}{2}}{\cos^4\dfrac{\alpha+\beta}{2}}}$. 34. $\dfrac{a^2+b^2-2ab\cos\alpha}{2b\sin\alpha}$, $\dfrac{|a^2-b^2|}{b}$.

35. 2 metres and 2 metres. 36. 2 cm, 1/2 cm, and 5/2 cm. 37. 3 and 4. 38. $2(1-\alpha)$. 39. $(1-\alpha)/\beta$. 40. Only an equilateral triangle. 41. No. 42. Not in all cases. Yes, if the base of the pyramid is a square. 43. $\cos\alpha=2/3$. Show that the angle at the vertex A of the triangle must be acute and the point H of intersection of the altitudes is diametrically opposite the point at which the inscribed circle is tangent to the base of the triangle. Further note that $\angle HBC=90°-\alpha$ and obtain the equation $\cot\alpha=2\tan(\alpha/2)$. 44. 4. 45. $\sin^2[(\pi/4)-(\alpha/4)]\operatorname{cosec}^2[(\pi/4)+(3\alpha/4)]$. 46. $1/4\,h\sin^2(\beta-\alpha)\operatorname{cosec}^2\alpha\times\operatorname{cosec}^2\beta\,[a-1/2\,h\sin(\alpha+\beta)\operatorname{cosec}\alpha\operatorname{cosec}\beta]$. 47. $(\pi/4)\pm\arccos\left[(1+2\sqrt{2})/4\right]$. 48. $\arccos\left(1/\sqrt[4]{2}\right)$. 49. $AB=BC=2$, $CD=1$, $AD=\sqrt{3}$, $S=3\sqrt{3}/2$. 50. $6\sqrt{7}$.

Sec. 3.4

1. (a) $\arccos(\sqrt{3}/3)$, (b) 60°, (c) 90°, (d) $a/\sqrt{2}$ (e) $a/\sqrt{3}$. 2. 90°, $\arccos\frac{1}{3}$, $a/\sqrt{2}$. 3. Yes. However, the proof is not correct, since it is not shown that the constructed plane π is not dependent on the choice of point A on the straight line L. It must also be proved that if we take another point on the line L and carry out the same constructions, then the resulting plane will coincide with the π plane. 4. Only provided $L\perp l$. 5. No. This straight line only exists if all three given skew lines are parallel to one plane. 6. Yes. Draw a plane π through one of the straight lines l_1 and a point A on another

straight line l_2; prove that point A may be chosen so that the π plane and the third straight line l_3 are not parallel. Draw line L through A and the point of intersection of line l_3 with the π plane; prove that A may be chosen so that the lines L and l_1 are not parallel. **7.** No. **8.** $1/3\, a\sqrt{2}$ and $2/3\, a\sqrt{2}$. Note that the common perpendicular is equal to the altitude of the triangular pyramid cut out of the cube by a plane passing through the endpoints of three edges converging to one vertex. **10.** Yes. **11.** From 0 to π, irrespective of the size of the given dihedral angle. **12.** A parallelogram. **14.** $\arcsin(\sin\varphi \operatorname{cosec}\alpha)$. **15.** $\arctan(\sqrt{2}/3)$ and $\arctan(5\sqrt{2}/7)$. **16.** $2\arcsin(b/\sqrt{4b^2-a^2})$. **17.** $\arcsin\sqrt{\sin^2\alpha+\sin^2\beta}$. **18.** $\arccos(\cos\alpha\sin\beta)$. **19.** $2\arccos\frac{3}{4}$ and $\arccos\frac{3}{4}$. **20.** $\operatorname{cosec}\alpha\ \sqrt{1+2\cos\alpha\cos\beta\cos\gamma-\cos^2\alpha-\cos^2\beta-\cos^2\gamma}$. **21.** $2/3\ R\sqrt{3+6\cos\alpha}$. **22.** Both angles are equal to $\arcsin[\sin\alpha/\sqrt{2}]$. **23.** $\arcsin(\sin\beta\sec\alpha)$. **24.** b/c. **25.** $BD=\frac{1}{5}\sqrt{9a^2+4b^2+25c^2-12ab\cos\alpha}$, $BC=\frac{1}{5}\sqrt{9a^2+9b^2+25c^2+18ab\cos\alpha}$. **26.** $\arccos\frac{c^2-d^2}{\sqrt{(c^2+d^2)^2-4(a^2-b^2)^2}}$. **27.** $\arcsin[(2/\sqrt{3})\sin(\alpha/2)]$. **28.** $\tan\varphi=\sqrt{7}$, $a\sqrt{3/7}$. **29.** $9a^3\cos^2(\alpha/2)\times\cot(\alpha/2)[3-4\sin^2(\alpha/2)]$. **30.** $1/8\,p^3\sqrt{3}\cot^2\alpha\,(4+\tan^2\alpha)^{3/2}$. **31.** $2\sin(\varphi/2)\sqrt{\sin(\varphi/2)\sin(3\varphi/2)}$. **32.** $\sin\varphi=2\sqrt{2/3}$. **33.** $\arccos(\cot^2\gamma)$. **34.** 45° or 60°. Through one of the skew lines draw a plane parallel to the second straight line and project this line on the constructed plane. Two cases are possible, depending on whether the line segment joining the projections of points A and B on this plane subtends an acute or an obtuse angle **35.** $\sqrt{9m^2\operatorname{cosec}^2\alpha-l^2}+\sqrt{9m^2\operatorname{cosec}^2\beta-l^2}$ if the angles of the triangle ABM at the side AB are both acute; $\left|\sqrt{9m^2\operatorname{cosec}^2\alpha-l^2}-\sqrt{9m^2\operatorname{cosec}^2\beta-l^2}\right|$ if one of these angles is obtuse. **36.** $2\arcsin[1/2\tan(\alpha/2)]$.

Sec. 3.5

4. Utilize the similarity of the triangles obtained in the construction so as to be able to replace the ratios in the left member of the equation being proved with new ratios in which the denominators will contain one and the same side of the triangle ABC. **13.** Use the formula for the sine of a triple angle. **16.** 4:9. **21.** $3/4\,a^2$; $\arccos\sqrt{1/3}$. **29.** $1/2\,b\tan(\alpha/2)[3-4\sin^2(\alpha/2)]^{-1/2}$. **30.** $\frac{2\sqrt{3}h\tan[(\alpha-\pi/6]}{9\tan^2[(\alpha-\pi)/6]-3}$. **31.** $3\sqrt[3]{\frac{48V\sin(\alpha/2)}{\sqrt{3-4\sin^2(\alpha/2)}}}$.

Sec. 3.6.

1. Take two straight lines l_1 and l_2 of the three given ones, construct parallel planes π_1 and π_2 containing l_1 and l_2, respectively. Let A and B be the points of intersection of the third line l_3 with the planes π_1 and π_2. Through point A in plane π_1 draw a straight line k_1 parallel to l_2 and through point B in plane π_2 a straight line k_2 parallel to l_1. Then build equal parallelograms: in the π_1 plane with vertex A and diagonals lying on lines l_1 and k_1; in the π_2 plane with vertex B and diagonals lying on the lines l_2 and k_2. **2.** No. **3.** No. **4.** Yes. The cutting plane has to be taken parallel to the lines of intersection lying opposite the faces of the given tetrahedral angle. **5.** No. **8.** A square. **9.** A regular hexagon. **10.** Construct a section of the cube by a plane passing through the points A, B, C, D dividing the corresponding edges of the cube in the ratio 1:3 (Fig. 170). Assure yourself that $ABCD$ is a square with side $3\sqrt{2}/4\approx 1.06$ (if an edge of the cube is taken to be 1). For the section of the desired hole you can take a square with side unity lying strictly within the square $ABCD$. **11.** $1/3\,a(3+2\sqrt{3})$. **12.** $2a\sqrt{2-\sqrt{3}}$. **13.** $a(3-\sqrt{3})$. **14.** $1/2\,a(\sqrt{6}-\sqrt{2})$. **15.** Yes. **16.** $4\pi/27$. **17.** $\arctan\left[\sqrt{2}\cot(\alpha/2)\right]$. **18.** $a/\sqrt{2}$. **19.** $2Rl^{-1}\sqrt{l^2-R^2}\times$ $\times\sin(\pi/n)$; $2\arcsin[Rl^{-1}\sin(\pi/n)]$. **20.** $2/3\,\pi h^3$. **21.** $a\sqrt{3/2}$. **22.** $l\cos^2(\alpha/2)\times$ $\times\sin(\alpha/2)$. **23.** $a^3b^3c^3(ab+bc+ca)^{-3}$. **24.** $1/12\,\pi a^3[1+2\cot^2(\alpha/2)]$. **25.** $1/2\,c\sin B\times$ $\times\operatorname{cosec} A$, $1/2\,c\sin A\operatorname{cosec} B$, $1/2\,c\sin A\sin B\operatorname{cosec}^2(A+B)$. Express the sides

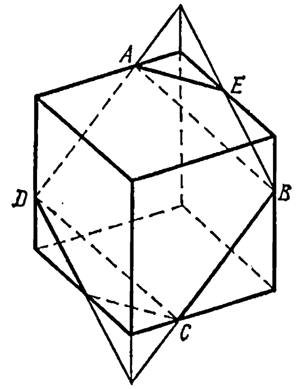


Fig. 170

of the triangle in terms of the radii of the spheres. **26.** $1/3\,\pi b^3\cos\beta\cos(\alpha/2)\times$ $\times\sin^2(\alpha/2)$. **27.** $\arctan\left[\sin\alpha/\sqrt{\cos 2\alpha}\right]$. If $2\alpha>\pi/2$, then the cones do not have common tangent planes (not passing through a common generatrix). **28.** $2l\sec(\alpha/2)\times\sec\beta\sqrt{\sin(\alpha-\beta)\sin\beta}$. **29.** $(2+\sqrt{3})/4$.

Sec. 3.7

1. $a^2\sin 2\alpha\cos\alpha\operatorname{cosec}^2 3\alpha$. **2.** $1/4\,a^2\sqrt{4\cos^2\alpha-1}$. **3.** $a^2\cos^3\alpha$. **4.** 1:15. **5.** $a^3\sqrt{3/2}$. **6.** $\sqrt{2}+1$. **7.** $3/4\sqrt{3}h^2(l^2-h^2)(h-a)^{-2}$. **8.** $(3\sqrt{3}+1)/26$. **9.** $3/\sqrt{14}$. **10.** $\sqrt{3}V^{2/3}\cos^{-1/3}\alpha\sin^{-2/3}\alpha$. **11.** $a(1+2\cos\varphi)^{-1}$. **12.** $d^2\sqrt{2}\operatorname{cosec}\varphi$. **13.** $3+2\sqrt{2}$. **14.** $\sqrt[4]{3}\,S^{1/2}\cos^{1/2}\alpha\sin\alpha$. **15.** 13:23. **16.** 1:47. **17.** $3a$. **18.** $3\sqrt{3}\,a^{2/4}$. **19.** $1/4\,(2a+b)^2$. **20.** $3a^2/2\sqrt{2}$. **21.** $a^2\left(1+\sqrt{4+2\sqrt{2}}\right)$. **22.** $23V/24$. **23.** $2/5\,a$. **24.** $12h^3/247$. **25.** $25a\sqrt{4h^2+3a^2}/64$. **26.** $5S/4$. **27.** $25S/16$. **28.** 100/69. **29.** $4\sqrt{6}h^2/9$. **30.** $\arccos\frac{3}{4}$. **31.** $(a/3)\sqrt{15+9\sqrt{3}}$, $(a/12)\sqrt{15+9\sqrt{3}}$. **32.** $3\sqrt{3}/32$. **33.** $\frac{3d^2}{4}$. **34.** $3\sqrt{3}/4$ square decimetres. **35.** $3\sqrt{7}/16$ square decimetres

Sec. 3.8

1. $\pi h\sqrt{2R(2R-h)}$. **2.** $a\sqrt{6}/8$. **3.** $4\pi rR^2(2R-r)^{-1}$. **4.** $\arccos\sqrt{\frac{4V-2\pi R^3}{3V+2\pi R^3}}$. **5.** 4:21. **6.** $2\pi/3$. **7.** $10\pi h^3/9$. **8.** 16:9. **9.** $8\sqrt{3a^3}/27$. **10.** $6\sqrt{3}/\pi$. **11.** $b\cos(\alpha/2)\tan(\alpha/4)$. **12.** $\sqrt{3}\,r^3\cot^3(\alpha/2)\tan\alpha$. **13.** $\frac{1}{2}\,l\sin^2\alpha$.

14. $2r\operatorname{cosec} 2\varphi(1+\sin\varphi)$. **15.** $a\sqrt{6}/2$. **16.** $1/3(1+2\sqrt{2})^3\pi R^3$. **17.** The condition of the problem is satisfied by two spheres with radii $r[1+2\tan^2(\alpha/2)\pm\tan(\alpha/2)\sqrt{3+4\tan^2(\alpha/2)}]$. **18.** $S_{lat}=1/3\pi l^2$, $R=3l/4\sqrt{2}$. **19.** $\sqrt{2}\pi a^3/36$.

20. $V=a^2h/2$, where for $a/2\leqslant r<a/\sqrt{2}$, $h=\dfrac{r\pm\sqrt{r^2-a^2/4}}{1-2r^2/a^2}$; for $r=a/\sqrt{2}$, $h=a/4$; for $r>a/\sqrt{2}$, $h=\dfrac{r-\sqrt{r^2-a^2/4}}{1-2r^2/a^2}$. **21.** $\dfrac{1}{4}a(\sqrt{3}-1)^2$. **22.** $b\sqrt{4a^2-b^2}/(4a+10b)$. **23.** $1/18\pi a^3\sqrt{3}$. **24.** $r(1+\sqrt{3}+2/3\sqrt{6})$. **25.** $1/3\,r(6+\sqrt{3}+\sqrt{27+12\sqrt{3}})$. **26.** $1/2\,r(2+\sqrt{6})$. **27.** 2. **28.** $\dfrac{9a^2h^2}{a^2+12h^2}\arctan\dfrac{\sqrt{a^2+12h^2}}{a\sqrt{3}}$. **29.** $\dfrac{a(2b-a)\sqrt{3}}{2\sqrt{3b^2-a^2}}$. **30.** $5a^3\sqrt{2}/96$.

Sec. 4.2

1. There is no solution. **2.** $x=0$. **3.** There is no solution. **4.** There is no solution. The right member is nonnegative for $\cos x\leqslant 1/2$; but all such x satisfy the inequality $|x|\geqslant\pi/3$, i.e., $x^2\geqslant\pi^2/9>1$ so that $3x^2>3\geqslant 1-2\cos x$. **5.** $x=0$. **7.** $x=2$. Reduce the equation to the form $x2^x=8$ and by trial and error find the root $x=2$; there are no other roots because for $x\geqslant 0$ the left member of this equation increases monotonically, and for $x<0$ it is nonpositive. **8.** $x_1=1/4$, $x_2=2$. Reduce the equation to the form $(\log_2 x+2)(\log_2 x+x-3)=0$. The equation $\log_2 x=3-x$ has a unique root, $x=2$; for $x>2$ the left member is greater than 1, the right member is less than 1; for (admissible) $x<2$, vice versa. **9.** $x_1=0$, $x_2=2$. Reduce the equation to the form $(x-2)2^x=(x-2)(1-x)$. **10.** $x=1$. For $x>1$ we have $x^2>1$; on the other hand, for these x we have $x-x^2<0$, i.e., $10^{x-x^2}<1$. For $0<x<1$, contrariwise, $x^x<1$ and $10^{x-x^2}>1$. **11.** $x_1=-1$, $x_2=1$. Reduce the equation to the form $2(x+1)\sin(\pi x/6)=(x+1)(3-2x)$. If $-2\leqslant x\leqslant 2$, then $-\pi/3\leqslant\pi x/6\leqslant\pi/3$ and so the left member of the equation $2\sin(\pi x/6)=3-2x$ increases, the right member decreases so that $x=1$ is the only root of the last equation. **12.** $x=1/2$. Reduce the equation to the form $4x-1=2\cos(2\pi x/3)$; its root is $x=1/2$. There are no other roots because for $0\leqslant x<1/2$ we have $4x-1<2\cos(2\pi x/3)$ and for $1/2<x\leqslant 1$ we have $4x-1>2\cos(2\pi x/3)$. **13.** $x=-1$, $y=1$. **14.** $x_1=2$, $y_1=(\pi/2)+k\pi$; $x_2=-2$, $y_2=k\pi$. **15.** $x=(2k+1)\pi/4$, $y=1$. **16.** $x=k\pi$, $y=1$. **17.** $x=1$ $y=(2k-3)/4$. **18.** $x=\pm 3/\pi$, $y=[(4k+1)\pi/2]\mp(3/\pi)$. **19.** $x_1=-\arctan\sqrt{2}+k\pi$, $y_1=(\pi/4)+2n\pi$; $x_2=\arctan\sqrt{2}+k\pi$, $y_2=(5\pi/4)+2n\pi$. Reduce the equation to the form $[\tan x+(\sin y+\cos y)]^2+(1-\sin 2y)=0$. **20.** $x_1=(\pi/4)+2k\pi$, $y_1=(2n+1)\pi$; $x_2=(5\pi/4)+2k\pi$, $y_2=2n\pi$. Using the substitution $\cos y=z$, reduce the equation to the form $z^2-2z\sin[x+(\pi/4)]+1=0$ or $[z-\sin[x+(\pi/4)]]^2+[1-\sin^2[x+(\pi/4)]]=0$. **21.** $x=0$, $y=1$. If the inequality is true, then $y\geqslant x^2+1\geqslant 1$; on the other hand, $\cos x\geqslant y^2+\sqrt{y-x^2-1}\geqslant y^2$, that is, $\cos x\geqslant 1$. Consequently, all the inequalities written down become equalities: $y=x^2+1=1$, whence $x=0$, $y=1$. A check shows that this number pair does indeed satisfy the inequality. **22.** $x=1$, $y=0$. **23.** $x=\arccos\dfrac{1}{3}+(2k+1)\pi$. **24.** No solution. **25.** $x_{1,2}=\pm 1$, $y_{1,2}=\pm\sqrt{3/2}$; $x_{3,4}=\pm 1$, $y_{3,4}=\mp\sqrt{3/2}$. **26.** $x=2$, $y=2$, $z=-2$. **27.** $x_1=a$, $y_1=0$, $z_1=0$; $x_2=0$, $y_2=a$, $z_2=0$; $x_3=0$, $y_3=0$, $z_3=a$. **28.** $a=-1/2$. **29.** a any number, $b=(2k+1)\pi-a$.

Sec. 4.3.

1. No solution. Utilize the fact that the sum of two positive reciprocal expressions cannot be less than 2. **2.** $x=-2/3$. Take advantage of the fact that the sum of two positive reciprocal expressions is equal to 2 only when they are both equal to 1. **3.** No solution. One of the terms is greater than 2, depending on the sign of x. **4.** $-1 \leqslant x \leqslant 1$. **5.** $4/5 \leqslant x < 1$. For $x < 1$ (and $x \geqslant 4/5$ from the condition making all roots meaningful) every term in the left member is greater than every term in the right member; the situation is reversed when $x > 1$. **6.** $x_1 = (\pi/2) + k\pi$, $x_2 = 2k\pi$. **7.** $x_1 = (\pi/4) + k\pi$, $x_2 = k\pi$. **8.** No solution. Since $\sqrt{\sin^3 x} \leqslant \sin x$ and $\sqrt{\cos^3 x} \leqslant \cos x$, it follows that the left member does not exceed $\sin x + \cos x \leqslant \sqrt{2}$; however, $\sin x + \cos x = \sqrt{2}$ only if $\sin x = \cos x = \sqrt{2}/2$, but for such x we have the strict inequality $\sqrt{\sin^3 x} < \sin x$. **9.** $x = (5\pi/4) + 2m\pi$. **10.** No solution.

Sec. 4.4

1. If the first equation has no solution, then $c^2 > a^2 + b^2$; in particular, $c \neq 0$. The second equation can be represented in the form $(2a \tan^2 x + 2c \tan x + b) \times \cot x = 0$. If $a \neq 0$, then the discriminant of the trinomial $2ay^2 + 2cy + b$ is equal to $c^2 - 2ab > a^2 + b^2 - 2ab \geqslant 0$, which means the second equation has two real and distinct roots; then at least one of them is nonzero and therefore $\cot x$ is meaningful. If $a = 0$, then $\tan x = -b/2c$; for $b \neq 0$ we have $\tan x \neq 0$, i.e., the second equation has a solution. **2.** $a < \pi^2/2$. **3.** (a) $a_1 = 1$, $a_2 = -2$, (b) $a = 1$. **4.** $a < -1$, $a = 0$. **5.** $a = 4$, $a > 5$. **6.** $a = 0$, $0 < b \leqslant 1$. **7.** $a = -1$. **8.** $a = -1$, $a = 1$. **9.** (a) Not for any a, (b) for any a, (c) $a \neq 0$, (d) for any a, (e) $a = 0$, (f) $a \neq 0$.

Sec. 4.5.

1. $d \geqslant 11/9$. **2.** $2\sqrt{2} \leqslant a < 11/9$. **3.** Not for any values. **4.** $a < -2$. **5.** $a < -3$, $a > 0$. **6.** $-1/2 \leqslant a < 0$. **7.** $m > 1$. **8.** $a \leqslant -3$, $a \geqslant -1$. **9.** $y < -2\sqrt{2}$, $-1/\sqrt{2} < y < 0$, $0 < y < 1/\sqrt{2}$, $y > 2\sqrt{2}$. **10.** $0 < y < 1$. **11.** $1/2 \leqslant a \leqslant 1$. **12.** $-3 \leqslant a \leqslant 3$. **13.** There are no such values. **14.** $x = (2a + 1 - \sqrt{1 + 4a})/2$ for $a \geqslant 0$; no solution for $a < 0$. **15.** $x_{1,2} = \left(\sqrt{a^2 - 1} - a \pm \sqrt{2a^2 - 5 - 2a\sqrt{a^2 - 1}}\right)^2$ for $a < -5/4$; $x = 1$ for $a = -5/4$; no solution for $a > -5/4$. Find the roots of the quadratic $f(y) = y^2 + 2ay + 1$ that satisfy the condition $y \geqslant 2$. If $D = a^2 - 1 < 0$, then there are no real roots at all, and so it must be true that $a^2 - 1 \geqslant 0$. If $f(2) = 4a + 5 < 0$, then the greater root $y = -a + \sqrt{a^2 - 1}$ will exceed 2. If $f(2) \geqslant 0$ and $-a \geqslant 2$, then both roots are suitable, but this case does not occur for any a whatsoever. **16.** $x = \log_2 \left(2m + \sqrt{4m^2 - 2m - 2}\right)$ for $m < -1$; $x = 1$ for $m = 1$; $x_{1,2} = \log_2 \left(2m \pm \sqrt{4m^2 - 2m - 2}\right)$ for $m > 1$; there are no solutions for the other values of m. Find the positive roots of the quadratic $y^2 - 4my + 2m + 2$. The discriminant $D = 4m^2 - 2m - 2$ is nonnegative for $m \leqslant -1/2$, $m \geqslant 1$. For $m \geqslant 1$ we have $y_{1,2} = 2m + \sqrt{4m^2 - 2m - 2}$ for $-1 \leqslant m \leqslant -1/2$ there are no positive roots and for $m < -1$ the greater root $y = 2m + \sqrt{4m^2 - 2m - 2}$ is positive. **17.** $x = (-1)^k \arcsin \log_2 \left[(-m + \sqrt{4 - 3m^3})/2\right] + k\pi$. Find the roots of the quadratic $f(y) = y^2 + my + m^2 - 1$ such that $1/2 \leqslant y \leqslant 2$. By virtue of $-1 < m < 1$ the discriminant $D = 4 - 3m^2 > 0$ and the constant term is negative, so we can only be interested in the greater root. It does not exceed 2 if and only if $f(2) \geqslant 0$, which occurs for arbitrary m. **18.** $x_1 = (-1)^k \times \arcsin 10^{a + \sqrt{2a^2 - 2}} + k\pi$, $x_2 = (-1)^k \arcsin 10^{a - \sqrt{2a^2 - 2}} + k\pi$ for $-\sqrt{2} \leqslant a \leqslant 1$; $x = x_2$ for $a < -\sqrt{2}$, $a = -1$, $b \geqslant \sqrt{2}$, there are no solutions for the remaining values of x. **19.** $x = (-1)^k \arcsin a^{(\sqrt{1+4a} - 1)/2} + k\pi$ for $0 < a < 1$, $x = (-1)^k \arcsin a^{-(\sqrt{1+4a} + 1)/2} + k\pi$ for $a > 1$. **20.** $x_{1,2} = \pm (1 - a^2)^2/4a^2$ for

$0 < a < 1$, $x=0$ for $a=1$, and there are no solutions for the remaining values of a. **21.** $x=\log_2 a$ for $0 < a \leqslant 1$, there are no solutions for the remaining values of a. **22.** $x=1/10$ for $a=10$; $x_1=1/10$, $x_2=10^{2-\sqrt{1+8\log_{10} a}}$ for $10 < a < 1000$; $x_1=1/10$, $x_2=1/1000$ for $a \geqslant 1000$; there are no solutions for the remaining values of a. **23.** $x=a^4$ for $1 < a \leqslant \sqrt[4]{2}$; $x=a^{\sqrt{8\log_a 2+4}-2}$ for $a > \sqrt[4]{2}$; there are no solutions for the remaining values of a. **24.** $x_1=10^{1-\sqrt{3}}$, $x_2=10^{-1+\sqrt{4\log_{10}a+3}}$ for $10^{1-\sqrt{3}} \leqslant a \leqslant 10^{1+\sqrt{3}}$; $x_{1,2}=10^{1\pm\sqrt{3}}$ for $a > 10^{1+\sqrt{3}}$; there are no solutions for the remaining values of a. **25.** $x=a^2$ for $1 < a \leqslant \sqrt{2}$, $x=a^{-1+\sqrt{4+\log_a 2}}$ for $a > \sqrt{2}$; no solutions for the remaining values of a. **26.** $x_1=a^{1/\left(\log_2 a+\sqrt{\log_2^2 a-2\log_2 a}\right)}$, $x_2=a^{1/(2\log_a 2-2)}$ for $0 < a < 1$; $x=x_2$ for $1 < a < 2$ and $2 < a < 4$; $x=x_1$ for $a \geqslant 4$. **27.** It reduces to the inequality $\sin x-\cos y > 0$. **28.** It reduces to the inequality $-1/2 < \cos(x-y) < 1/2$. **29.** It reduces to the inequality $\cos x-\cos y > 0$. **30.** It reduces to the inequality $|\sin(x+y)| > 1/2$. **31.** It reduces to the inequality $\cos 2x+\cos y < 0$.

INDEX

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:
Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR

Printed in the Union of Soviet Socialist Republics