CHAPTER III

ANALYTIC GEOMETRY

§1. Introduction

In the first half of the 17th century a completely new branch of mathematics arose, the so-called analytic geometry, establishing a connection between curves in a plane and algebraic equations in two unknowns.

A quite rare event thereby happened in mathematics: In one or two decades there appeared a great, entirely new branch of mathematics based on a very simple concept, which until then had not received proper attention. The appearance of analytic geometry in the first half of the 17th century was not accidental. The transition in Europe to the new capitalistic methods of manufacture required the advance of a whole series of sciences. A short time before, contemporary mechanics was being created by Galileo and other scientists, experimental data were being accumulated in all regions of natural science, the means of observation were being perfected, and instead of absolete scholastic theories new ones were being created. In astronomy, among the foremost scientists the teachings of Copernicus had finally triumphed. The rapid development of long-range navigation insistently called for knowledge of astronomy and the elements of mechanics.

The art of warfare also required mechanics. Ellipses and parabolas, whose geometric properties as conic sections were already well known in detail to the ancient Greeks almost 2000 years earlier, ceased to be only part of geometry, as they were to the Greeks. After Kepler had discovered that the planets revolve around the sun in ellipses, and Galileo that a stone thrown into the air traces out a parabola, it was necessary to calculate these ellipses and to find the parabolas along which bullets fly from a gun; it was necessary to discover the law by which the atmospheric pressure, discovered by Pascal, decreases with the height; it was necessary actually to calculate the volumes of various bodies, and so forth.

All these questions almost simultaneously called to life three entirely new mathematical sciences: analytic geometry, differential calculus, and integral calculus, including the solution of the simplest differential equations.

These three new fields qualitatively changed the face of the whole of mathematics. They made it possible to solve problems never even dreamed of before.

In the first half of the 17th century, i.e., at the beginning of the 1600's, a group of the most outstanding mathematicians was already close to the idea of analytic geometry, but there were two of them, in particular, who understood clearly the possibility of creating a new branch of mathematics. These were Pierre Fermat, a counsellor of the parliament of the French city of Toulouse and a world-famous mathematician, and the famous French philosopher René Descartes. Descartes is credited with being the chief creator of analytic geometry. He was the one who, as a philosopher, raised the question of its complete generality. Descartes published the great philosophical treatise "Discourse on the method of rightly conducting the reason and seeking the truth in the sciences, with applications: dioptrics, meteorology and geometry."

The last part of this work, entitled "Geometry" and published in 1637, contains a sufficiently complete, although somewhat confusing, presentation of the mathematical theory that since then has been called analytic geometry.

§2. Descartes' Two Fundamental Concepts

Descartes wished to create a method that could equally well be applied to the solution of all problems of geometry, that is, which would provide a general method for their solution. Descartes' theory is based on two concepts: the concept of coordinates and the concept of representing by the coordinate method any algebraic equation with two unknowns in the form of a curve in the plane.

The concept of coordinates. By the *coordinates of a point* in the plane Descartes means the abscissa and ordinate of this point, i.e., the numerical values x and y of its distances (with corresponding signs) to two mutually perpendicular straight lines (coordinate axes) chosen in this plane (see Chapter II). The point of intersection of the coordinate axes, i.e., the point having coordinates (0, 0) is called the *origin*.

With the introduction of coordinates Descartes constructed, so to speak, an "arithmetization" of the plane. Instead of determining any point geometrically, it is sufficient to give a pair of numbers x, y and conversely (figure 1).

The notion of comparison of equations with two unknowns with curves in the plane. Descartes' second concept is the following. Up to the time of Descartes, where an algebraic equation in two unknowns $F(x, y) = 0$ was given, it was said that the problem was indeterminate, since from the equation it was impossible to determine these unknowns; any value could be assigned to one of them, for example to x, and substituted in the equation; the result was an equation with only one unknown y, for which, in general, the equation could be solved. Then this *arbitrarily chosen* x together with the *so-obtained* y would satisfy the given equation. Consequently, such an "indeterminate" equation was not considered interesting.

Descartes looked at the matter differently. He proposed that in an equation with two unknowns x be regarded as the abscissa of a point and the corresponding y as its ordinate. Then if we vary the unknown x, to every value of x the corresponding y is computed from the equation, so that we obtain, in general, a set of points which form a curve (figure 2).*

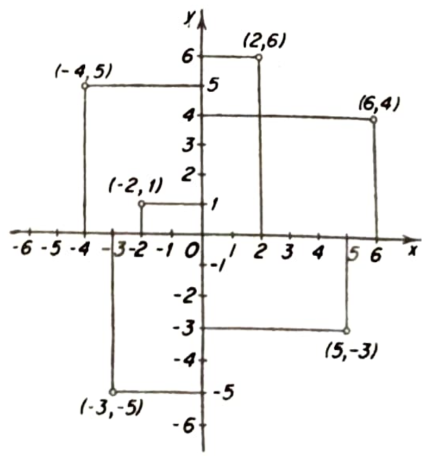


FIG. 1.

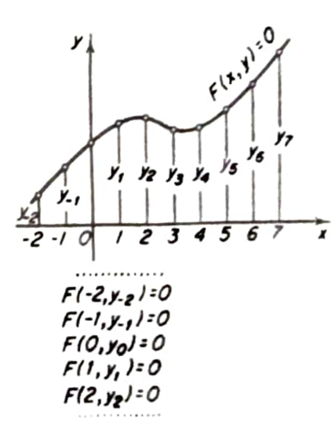


FIG. 2.

* Sometimes, the equation is not satisfied by any point (x, y) with real coordinates, sometimes by one or a few such points. In this case we say that the curve is imaginary or reduces to points (see §7).

Thus, to each algebraic equation with two variables, $F(x, y) = 0$, corresponds a completely determined curve of the plane, namely a curve representing the totality of all those points of the plane whose coordinates satisfy the equation $F(x, y) = 0$.

This observation of Descartes *opened up an entire new science.*

The basic problems solved by analytic geometry and the definition of analytic geometry. Analytic geometry provides the possibility: (1) of solving construction problems by computation (see for example, the division of a segment in a given ratio, see §3); (2) of finding the equation of curves defined by a geometric property (for example, of an ellipse defined by the condition that the sum of distances to two given points is constant, see §7); (3) of proving new geometric theorems algebraically (see, for example, the derivation of Newton's theory of diameters, §6); (4) conversely, of representing an algebraic equation geometrically, to clarify its algebraic properties (see, for example, the solution of third- and fourth-degree equations from the intersection of a parabola with a circle, §5).

Thus, analytic geometry is that part of mathematics which, applying the coordinate method, investigates geometric objects by algebraic means.

§3. Elementary Problems

The coordinates of a point that divide a segment in a given ratio. Given the coordinates (x_1, y_1) and (x_2, y_2) of two points M_1 and M_2, let us find the coordinates (x, y) of the point M dividing the segment M_1M_2 in the ratio m to n (figure 3). From the similarity of the shaded triangles we obtain:

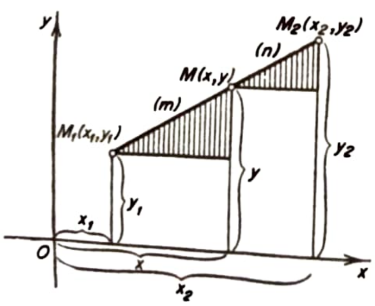


FIG. 3.

$$\frac{x - x_1}{x_2 - x} = \frac{m}{n}, \text{ from which}$$

$$x = \frac{nx_1 + mx_2}{m + n},$$

$$\frac{y - y_1}{y_2 - y} = \frac{m}{n}, \text{ from which}$$

$$y = \frac{ny_1 + my_2}{m + n}.$$

Distance between two points. Let us find the distance between the points M_1 and M_2, whose coordinates are (x_1, y_1) and (x_2, y_2) respectively. From the shaded right triangle (figure 4), we obtain by the theorem of Pythagoras

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

The area of a triangle. Let us find the area S of the triangle $M_1M_2M_3$ (figure 5) if the coordinates of its vertices are respectively (x_1, y_1), (x_2, y_2), (x_3, y_3). Considering the area of the triangle as the sum of the areas of trapezoids with bases y_1, y_3 and y_3, y_2 minus the area of the trapezoid with bases y_1, y_2 and writing the product $-(y_1 + y_2)(x_2 - x_1)$ in the form $(y_1 + y_2)(x_1 - x_2)$, we obtain

$$S = \tfrac{1}{2}[(y_1 + y_2)(x_1 - x_2) + (y_2 + y_3)(x_2 - x_3) + (y_3 + y_1)(x_3 - x_1)].$$

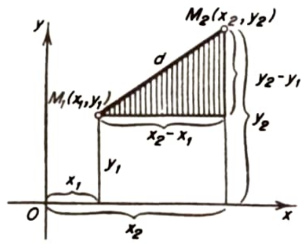


FIG. 4.

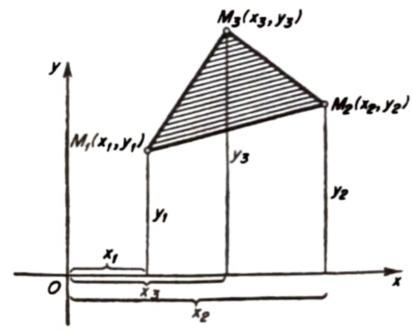


FIG. 5.

In these problems it only remains to verify that the derived formulas remain valid without any change in those cases when one or more coordinates or their differences are negative. Such verification easily follows.

Determination of the points of intersection of two curves. Relying on the second fundamental idea that the equation $F(x, y) = 0$ represents a curve, it is particularly simple to find the points of intersection of two curves. In order to find the coordinates of the points of intersection of two curves, it is obviously necessary to solve simultaneously the equations that represent them. The pair of numbers x, y obtained from the ordinary solution of these two equations will determine the point whose coordinates satisfy both of the equations, i.e., the point that lies on the first as well as on the second curve, and this is the point of their intersection.

The solution of geometric problems by the tools of analytic geometry, as we see, is very convenient for practical purposes, especially because every solution is at once obtained in the convenient form of numbers. *Such a geometry, such a science, was exactly what was lacking at that time.*

§4. Discussion of Curves Represented by First- and Second-Degree Equations

First degree equation. Making use of his second idea, Descartes first of all examined what curves correspond to an equation of the first-degree,

$$Ax + By + C = 0, \tag{1}$$

i.e., to an equation where A, B, C are numerical coefficients with A and B not both zero. It turned out that in the plane a straight line always corresponds to such an equation.

We shall prove that equation (1) always represents a straight line, and conversely, that to every line in the plane there corresponds a completely determined equation of the form (1). In fact, let us suppose, for example, that $B \neq 0$; then equation (1) can be solved for y

$$y = kx + l,$$

where $k = -\frac{A}{B}$; $l = -\frac{C}{B}$.

We examine first the equation $y = kx$. It obviously represents a straight line passing through the origin and making an angle ϕ with the x-axis whose tangent $\tan \phi$ is k (figure 6). Indeed, the equation can be

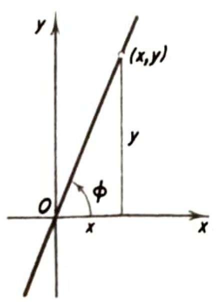


FIG. 6.

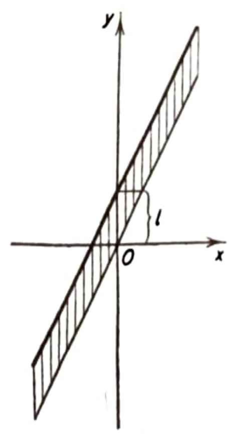


FIG. 7

written as $y/x = k$, so that the coordinates of every point (x, y) on the straight line satisfy the equation, and the coordinates of no point $(x, \bar{y})$ not lying on the straight line satisfy the equation, since for such a point y/x will be either greater than or smaller than k. In addition, if $\tan \phi > 0$, then for this line either both x and y are positive or both negative, and if $\tan \phi < 0$ their signs are opposite.

Thus, the equation $y = kx$ represents a straight line passing through the origin O, and consequently the equation $y = kx + l$ also represents a line, namely the one which is obtained from the previous line by the parallel translation such that the ordinate of each of its points is increased by l (figure 7).

The earlier derived formulas of the coordinates of a point dividing a segment in a given ratio, the distance between two given points, and the area of a triangle as well as the information about the equation of a straight line already enable us to solve a large number of problems.

The equation of a straight line passing through one or two given points. Let M_1 be the point with coordinates x_1, y_1 and let k be a given number. The equation $y = kx + l$ represents a straight line making with the Ox-axis an angle whose tangent is equal to k and intersecting the Oy-axis at a distance l from O. Let us choose l such that this line goes through the point (x_1, y_1). For this, the coordinates of the point M_1 must satisfy the equation, i.e., we must have $y_1 = kx_1 + l$, from which $l = y_1 - kx_1$.

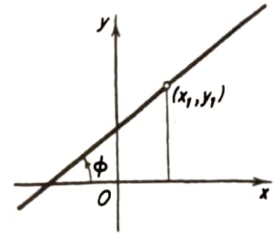


FIG. 8.

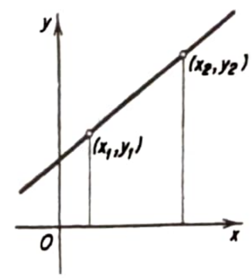


FIG. 9.

Substituting this value for l, we obtain the equation of the line that passes through the given point (x_1, y_1) and makes with the Ox-axis an angle whose tangent is equal to k (figure 8). This equation is $y = kx + y_1 - kx_1$ or

$$y - y_1 = k(x - x_1).$$

Example. Let the angle between the line and the Ox-axis be equal to 45°, and let the point M have coordinates (3, 7); then the equation of the corresponding line (since $\tan 45° = 1$) will be: $y - 7 = 1 \cdot (x - 3)$ or $x - y + 4 = 0$.

If we require that the line passing through the point (x_1, y_1) also go through the point (x_2, y_2), it follows that the condition $y_2 - y_1 = k(x_2 - x_1)$ must be imposed on k. Finding k from this and substituting it in the previous equation, we obtain the equation of the line passing through two given points (figure 9):

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}.$$

Descartes' result concerning second-degree equations. Descartes also investigated the question as to what kinds of curves in the plane are represented by the second-degree equation with two variables whose general form is

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

and showed that such an equation, generally speaking, represents an ellipse, a hyperbola, or a parabola; i.e., curves very well known to the mathematicians of antiquity.

These are Descartes' most important achievements. However, his book was far from being restricted to these topics; he also investigated the equations of a number of interesting geometric loci, examined certain theorems on transformation of algebraic equations, mentioned without proof his famous *law of signs* for the number of positive roots of an equation whose roots are all real (see Chapter IV, §4) and, finally, presented a remarkable method for determining the real roots of third- and forth-degree equations from the intersection of the parabola $y = x^2$ with circles.

§5. Descartes' Method of Solving Third- and Fourth-Degree Algebraic Equations

Transformation of third- and fourth-degree equations to an equation of the fourth-degree not involving the x^3-term. We will show that the solution of an arbitrary third- or fourth-degree equation can be reduced to the solution of an equation of the form

$$x^4 + px^2 + qx + r = 0. \tag{2}$$

Let the given third-degree equation be $z^3 + az^2 + bz + c = 0$. Substituting $z = x - a/3$, we obtain

$$(x - a/3)^3 + a(x - a/3)^2 + b(x - a/3) + c = 0.$$

The x^2-terms in the expansion of the parentheses will cancel out, so that we get an equation of the form $x^3 + px + q = 0$. Multiplying this equation by x, we bring it to the form (2) with $r = 0$, which also admits a root $x_4 = 0$.

An equation of the fourth-degree $z^4 + az^3 + bz^2 + cz + d = 0$ can be reduced to the form (2) by the substitution $z = x - a/4$. Hence, the solution of all third- and fourth-degree equations can be reduced to the solution of an equation of the form (2).

The solution of third- and fourth-degree equations by the intersection of a circle with the parabola $y = x^2$. Let us first derive the equation of a circle with center (a, b) and radius R. If (x, y) is any of its points, then the square of its distance to the point (a, b) is equal to $(x - a)^2 + (y - b)^2$ (see §3). Thus, the equation of the circle in question is

$$(x - a)^2 + (y - b)^2 = R^2.$$

Now we try to find the points of intersection of this circle with the parabola $y = x^2$. In order to do this, by virtue of what was said in §3, it is necessary to solve simultaneously the equation of this circle

$$x^2 + y^2 - 2ax - 2by + a^2 + b^2 - R^2 = 0$$

and the equation of the parabola

$$y = x^2.$$

Substituting y from the second equation into the first, we obtain a fourth-degree equation in x:

$$x^2 + x^4 - 2ax - 2bx^2 + a^2 + b^2 - R^2 = 0$$

or

$$x^4 + (1 - 2b)\,x^2 - 2ax + a^2 + b^2 - R^2 = 0.$$

If we choose a, b and R^2 such that

$$1 - 2b = p,\ -2a = q,\ a^2 + b^2 - R^2 = r,$$

then exactly equation (2) is obtained. For this purpose we have to take

$$a = -\frac{q}{2},\ b = \frac{1 - p}{2},\ R^2 = \frac{q^2}{4} + \frac{(1 - p)^2}{4} - r. \tag{3}$$

In the last formula (3), generally speaking, R^2 may turn out to be negative. However, in the case when equation (2) has even one real root x_1, the following equality holds

$$x_1^4 + (1 - 2b)\, x_1^2 - 2ax_1 + a^2 + b^2 - R^2 = 0. \tag{4}$$

Denoting x_1^2 by y_1, equation (4) can be rewritten as

$$x_1^2 + y_1^2 - 2ax_1 - 2by_1 + a^2 + b^2 - R^2 = 0$$

or as

$$(x_1 - a)^2 + (y_1 - b)^2 = R^2.$$

Hence, in the case when equation (2) has a real root, the number $R^2 = [(1 - p)^2 + q^2]/4 - r$ is *positive*, the equation

$$(x - a)^2 + (y - b)^2 = R^2$$

is the equation of a circle, and all real roots of equation (2) are the abscissas of points of intersection of the parabola $y = x^2$ with this circle. (In case $r = 0$, $R^2 = a^2 + b^2$, this circle passes through the origin.)

Thus, if the coefficients p, q, r of equation (2) are given, and it is necessary to find a, b and R^2 by formulas (3), then if $R^2 < 0$, equation (2) is known to have no real roots. But, if $R^2 \geqslant 0$ then the abscissas of the points of intersection of the circle with center (a, b) and radius R with the parabola $y = x^2$ (drawn once and for all) give all the real roots of equation (2); and also in case $R^2 < 0$, the resulting circle cannot intersect the parabola and equation (2) does not have real roots.

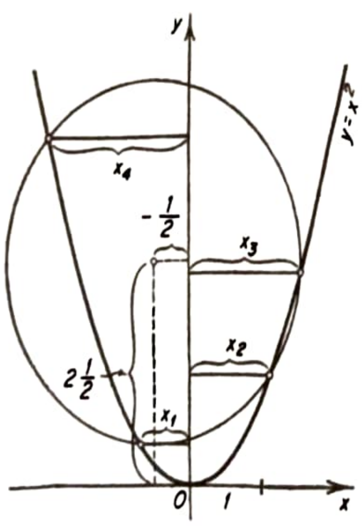


FIG. 10.

Example. Let the given fourth-degree equation be:

$$x^4 - 4x^2 + x + \frac{5}{2} = 0.$$

Then we have

$$a = -\frac{1}{2}, b = \frac{5}{2} = 2\frac{1}{2},$$

$$R = \sqrt{\frac{1}{4} + \frac{25}{4} - \frac{5}{2}} = 2.$$

Figure 10 shows the corresponding circle and the roots x_1, x_2, x_3, x_4 of the given equation.

The 1st, 2nd, 3rd and 4th sections above contain, in an abbreviated and somewhat more modern form, the essential content of Descartes' book.

From Descartes' time up to the present, analytic geometry has undergone an immense development that has been very fruitful for many different parts of mathematics. We will attempt in the following sections of this chapter to trace the most important stages of this development.

First of all, it is necessary to say that the inventors of the infinitesimal analysis were already in possession of Descartes' method. Whether it was a question of tangents or normals (perpendiculars to the tangents at the point of tangency) to curves, or of maxima or minima of functions considered geometrically, or of the radius of curvature of a curve at a given point, etc., the equation of the curve was considered first, by the method of Descartes, and then the equations of the normal, the tangent, and so forth, were found. Thus infinitesimal analysis, namely the differential and integral calculus, would have been inconceivable without the preliminary development of analytic geometry.

§6. Newton's General Theory of Diameters

The first mathematician to take a further great step forward in analytic geometry itself was Newton. In 1704 he examined the theory of third-order curves, i.e., curves which are represented by third-degree algebraic equations in two unknowns. At the same time he found, among other things, an elegant general theorem about "diameters," which correspond to secants in a given direction. He proved the following.

Let an nth-order curve be given, i.e., a curve which is represented by an nth-degree algebraic equation in two unknowns; then an arbitrary straight line intersecting it has in general n common points with it. Let M be the point of the secant that is the "center of gravity" of these points of its intersection with the given nth-order curve, i.e., the center of gravity of a set of n equal point masses situated at these points. It turns out that if we

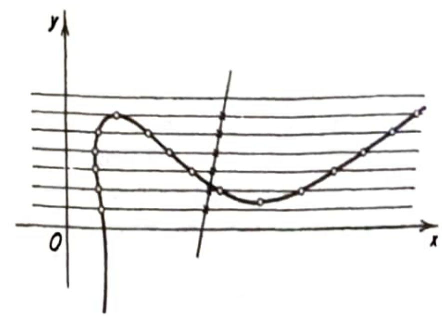


FIG. 11.

take all possible sets of mutually parallel secants and for each of them consider these centers of mass M, then for any given set of parallel secants all the points M lie on a straight line. Newton called this line the "diameter" of the nth-order curve corresponding to the given direction of the secants. Since the proof of this theorem is quite easy with the help of analytic geometry, we will give it here.

Let an nth-order curve be given and some set of mutually parallel secants of the curve. Choose the coordinate axes so that these secants are parallel to the Ox-axis (figure 11). Then their equations will have the form $y = l$, where the constant l will be different for different secants. Let $F(x, y) = 0$ be the equation that represents the nth-order curve with respect to these coordinate axes. It is easy to show that under a transformation from one rectangular coordinate system to another, although the equation of the curve changes, its order does not change (this will be shown in §8). Therefore $F(x, y)$ will also be an nth-degree polynomial. To determine the abscissas of the points of intersection of the curve with the secant $y = l$, it is necessary to solve the simultaneous equations $F(x, y) = 0$ and $y = l$; as a result, in general, an nth-degree equation in x is obtained

$$F(x, l) = 0, \tag{5}$$

from which we find the abscissas $x_1, x_2, \cdots, x_n$. The abscissa x_c of the center of gravity of the n points of intersection is equal, by the very definition of center of gravity, to

$$x_c = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

But, as is known from the theory of algebraic equations, the sum $x_1 + x_2 + \cdots + x_n$ of the roots of an equation is equal to the coefficient of the $(n - 1)$th power of the unknown x, taken with the opposite sign, divided by the coefficient of the nth power of x. But because the sum of the exponents of x and y in every term of $F(x, y)$ is equal to or less than n, the term in x^n does not contain y at all but has the form Ax^n, where A is a constant; and if the terms in x^{n-1} contain y, they do so to no higher than the first power; i.e., they have the form $x^{n-1}(By + C)$. Consequently, the coefficient of x^n is A and that of x^{n-1} is $Bl + C$, and we have for any given l

$$x_c = -\frac{Bl + C}{nA}.$$

But the secant is parallel to the Ox-axis so that for all of its points $y = l$, and hence the ordinate y of the center of gravity of the points of its

intersection with the given nth-order curve is also equal to l; thus finally we obtain $nAx_c + By_c + C = 0$, i.e., the coordinates x_c, y_c of the centers of gravity for all these secants satisfy a first-degree equation, and consequently lie on a straight line.

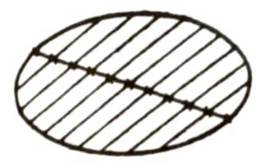


FIG. 12.

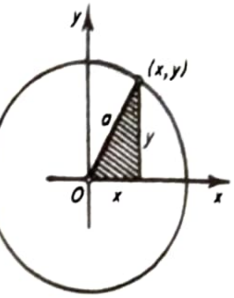


FIG. 13.

The case when $F(x, y)$ does not contain x^n can be investigated analogously.

In case the curve is of the 2nd order ($n = 2$) the center of gravity of two points is simply the midpoint between them, so that the locus of midpoints of parallel chords of a second-order curve is a straight line (figure 12), a result that for the ellipse, as well as for the hyperbola and the parabola, was already well known to the ancients. But this was proved by them, even though only for these partial cases, with quite difficult geometric arguments, and here a new general theorem, unknown to the ancients, is proved in an entirely simple way.

Such examples reveal the power of analytic geometry.

§7. Ellipse, Hyperbola, and Parabola

In this and the following sections, we consider second-order curves. Before investigating the general second-degree equation, it is useful to examine some of its simplest forms.

The equation of a circle with center at the origin. First of all, we consider the equation

$$x^2 + y^2 = a^2.$$

It evidently represents a circle with center at the origin and radius a, as follows from the theorem of Pythagoras applied to the shaded right triangle (figure 13), since whatever point (x, y) of this circle is taken, its x and y coordinates satisfy this equation, and conversely, if the coordinates x, y of a point satisfy the equation, then the point belongs to the circle; i.e., the circle is the set of all those points of the plane that satisfy the equation.

The equation of an ellipse and its focal property. Let two points F_1 and F_2 be given, the distance between which is equal to $2c$. We will find the equation of the locus of all points M of the plane; the sum of whose distances to the points F_1 and F_2 is equal to a constant $2a$ (where, of course, a is greater than c). Such a curve is called an *ellipse* and the points F_1 and F_2 are its *foci*.

Let us choose a rectangular coordinate system such that the points F_1 and F_2 lie on the Ox-axis and the origin is halfway between them. Then the coordinates of the points F_1 and F_2 will be $(c, 0)$ and $(-c, 0)$. Let us take an arbitrary point M with coordinates (x, y), belonging to the locus in question, and let us write that the sum of its distances to the points F_1 and F_2 is equal to $2a$,

$$\sqrt{(x-c)^2+(y-0)^2}+\sqrt{(x+c)^2+(y-0)^2}=2a. \tag{6}$$

This equation is satisfied by the coordinates (x, y) of any point of the locus under consideration. Obviously the converse is also true, namely that any point whose coordinates satisfy equation (6) belongs to this locus. Equation (6) is therefore the equation of the locus. It remains to simplify it.

Raising both sides to the second power, we obtain

$$x^2-2cx+c^2+y^2+2\sqrt{(x^2-2cx+c^2+y^2)(x^2+2cx+c^2+y^2)}$$
$$+x^2+2cx+c^2+y^2=4a^2,$$

or after simplification

$$x^2+y^2+c^2-2a^2=-\sqrt{(x^2+y^2+c^2)^2-4c^2x^2}.$$

Squaring again both sides, we obtain

$$(x^2+y^2+c^2)^2-4a^2(x^2+y^2+c^2)+4a^4=(x^2+y^2+c^2)^2-4c^2x^2$$

or after simplification

$$(a^2-c^2)x^2+a^2y^2=(a^2-c^2)a^2.$$

Let us set $a^2-c^2=b^2$ (as may be done since $a>c$); then we obtain $b^2x^2+a^2y^2=a^2b^2$, and dividing by a^2b^2 we have

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \tag{7}$$

The coordinates (x, y) of any point M of the locus thus satisfy equation (7).

It can be shown on the other hand that if the coordinates of a point satisfy equation (7) then they also satisfy equation (6). Consequently, equation (7) is the equation of this locus, i.e., the equation of the ellipse (figure 14).

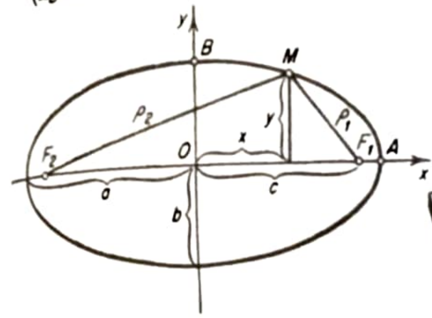


FIG. 14.

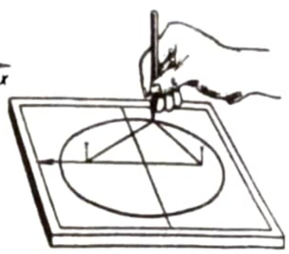


FIG. 15.

This argument is a classical example of finding the equation of a curve given by some of its geometrical properties.

The well-known method of tracing an ellipse by means of a thread (figure 15) is based on the property of the ellipse that the sum of the distances of any of its points to two given points is a constant.

Remark. In order to determine an ellipse, we could have taken, instead of the focal property considered here, any other geometric property characteristic of it, for example, that the ellipse is the result of a "uniform contraction" of a circle toward one of its diameters or any other property.

Substituting $y = 0$ in equation (7) of the ellipse, we obtain $x = \pm a$, i.e., a is the length of the segment OA (see figure 14), which is called the *major semiaxis* of the ellipse. Analogously, substituting $x = 0$, we obtain $y = \pm b$, i.e., b is the length of the segment OB, which is called the *minor semiaxis* of the ellipse.

The number $\epsilon = c/a$ is called the *eccentricity* of the ellipse, so that, since $c=\sqrt{a^2-b^2}<a$, the eccentricity of an ellipse is less than 1. In the case of a circle, $c = 0$ and consequently $\epsilon = 0$; both foci are at one point, the center of the circle (since $OF_1 = OF_2 = 0$), but the previous method of drawing the curve with a thread is still valid.

Laws of planetary motion. In studying Tycho Brahe's long-continued observations on the motion of the planet Mars, Kepler discovered that the planets revolve around the Sun in ellipses such that the Sun occupies one focus of the ellipse (the other focus remains unoccupied and plays no part in the motion of a planet around the Sun) (figure 16). Kepler also observed that the focal radius ρ in equal times sweeps out sectors of equal area,* and Newton showed that the necessity of such a motion follows mathematically from the law of inertia (proportionality of acceleration to force) and the law of universal gravitation.

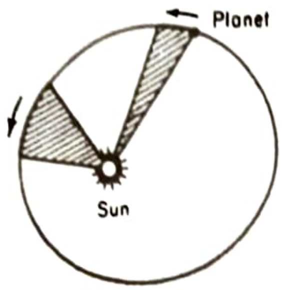


FIG. 16.

The ellipse of inertia. As an example of the application of the ellipse in a technical problem, we consider the so-called ellipse of inertia of a plate.

Let the plate be of uniform thickness and homogeneous material, for example a zinc plate of arbitrary shape. We rotate it around an axis in its plane. A body in rectilinear motion has, as is well known, an inertia with respect to this rectilinear motion that is proportional to its mass (independently of the shape of the body and the distribution of the mass). Similarly, a body rotating around an axis, for instance a flywheel, has inertia with respect to this rotation. But in the case of rotation, the inertia is not only proportional to the mass of the rotating body but

FIG. 17a.

FIG. 17b.

* The eccentricities of planetary orbits are not very large, so that the orbits of planets are almost circles.

also depends on the distribution of the mass of the body with respect to the axis of rotation, since the inertia with respect to rotation is greater if the mass is farther from the axis. For example, it is very easy to bring a stick at once into fast rotation around its longitudinal axis (figure 17a). But if we try to bring it at once to fast rotation around an axis perpendicular to its length, even if the axis passes through its midpoint, we will find that unless this stick is very light, we must exert considerable effort (figure 17b).

It is possible to show that the inertia of a body with respect to rotation about an axis, the so-called *moment of inertia* of the body relative to the axis, is equal to $\Sigma r_i^2 m_i$ (where by $\Sigma r_i^2 m_i$ we mean the sum $r_1^2 m_1 + r_2^2 m_2 + \cdots + r_n^2 m_n$ and think of the body as decomposed into very small elements, with m_i as the mass of the ith element and r_i the distance of the ith element from the axis of rotation, the summation being taken over all elements of the body).

Let us return to our plate. Let O (figure 18) be a point of this plate. We consider the moments of inertia J_u of the plate relative to an axis

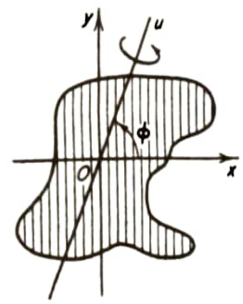


FIG. 18.

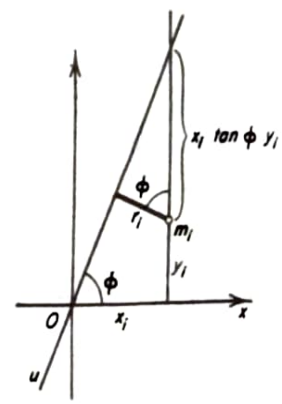


FIG. 19.

of rotation u passing through O and lying in the plane of the plate. For this purpose we take the point O as the origin of a Cartesian coordinate system and choose arbitrary axes Ox and Oy in the plane of the plate; then we will characterize the axis of rotation u by the angle ϕ which it makes with the Ox-axis. It is easy to see (figure 19) that

$$r_i = |(x_i \tan\phi - y_i)\cos\phi| = |x_i \sin\phi - y_i \cos\phi|.$$

Hence

$$\Sigma r_i^2 m_i = \Sigma (x_i^2 \sin^2\phi - 2x_i y_i \sin\phi\cos\phi + y_i^2 \cos^2\phi)\, m_i$$
$$= \sin^2\phi\, \Sigma x_i^2 m_i - 2\sin\phi\cos\phi\, \Sigma x_i y_i m_i + \cos^2\phi\, \Sigma y_i^2 m_i .$$

The quantities $\sin^2\phi$, $2\sin\phi\cos\phi$, and $\cos^2\phi$ are taken outside the summation sign, since they are constant for a given axis u. We now write

$$\Sigma x_i^2 m_i = A, \quad -\Sigma x_i y_i m_i = B, \quad \Sigma y_i^2 m_i = C.$$

The quantities A, B, and C do not depend on the choice of the axis u, but only on the shape of the plate, the distribution of its mass, and the fixed choice of the coordinate axes Ox and Oy. Consequently,

$$J_u = A\sin^2\phi + 2B\sin\phi\cos\phi + C\cos^2\phi.$$

We consider all possible axes u in the plane of the plate passing through the point O and lay off on each of these axes from the point O a length equal to ρ, the inverse of the square root of the moment of inertia J_u of the plate relative to that axis, i.e., $\rho = 1/\sqrt{J_u}$. Then we obtain

$$\frac{1}{\rho^2} = A\sin^2\phi + 2B\sin\phi\cos\phi + C\cos^2\phi.$$

But

$$x = \rho\cos\phi, \quad y = \rho\sin\phi,$$

so that the equation of this locus has the following form:

$$Cx^2 + 2Bxy + Ay^2 = 1.$$

A second-order curve is obtained that is evidently finite and closed; i.e., it is an ellipse (figure 20), since all other second-order curves, as we will later show, are either infinite or reduce to one point.

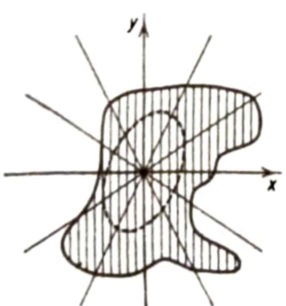


FIG. 20.

The following remarkable result is obtained: Whatever may be the form and size of a plate and the distribution of its mass, the magnitude of its moment of inertia (more precisely, of the quantity ρ inversely proportional to the square root of the moment of inertia) with respect to the various axes lying in the plane of the plate and passing through the given point O, is characterized by a certain ellipse. This ellipse is called the ellipse of inertia of the plate relative to the point O. If the

point O is the center of gravity of the plate, then the ellipse is called its central ellipse of inertia.

The ellipse of inertia plays a great role in mechanics; in particular, it has an important application in the strength of materials. In the theory of strength of materials, it is proved that the resistance to bending of a beam with given cross section is proportional to the moment of inertia of its cross section relative to the axis through the center of gravity of the cross section and perpendicular to the direction of the bending force. Let us clarify this by an example. We assume that a bridge across a stream consists of a board that sags under the weight of a pedestrian passing over it. If the same board (no thicker than before) is placed "on its edge," it scarcely bends at all, i.e., a board placed on its edge is, so to speak, stronger. This follows from the fact that the moment of inertia of the cross section of the board (it has the shape of an elongated rectangle that we may think of as evenly covered with mass) is greater relative to the axis perpendicular to its long side than relative to the axis parallel to its long side. If we set the board not exactly flat nor on edge but obliquely, or even if we do not take a board at all but a rod of arbitrary cross section, for example a rail, the resistance to bending will still be proportional to the moment of inertia of its cross section relative to the corresponding axis. The resistance of a beam to bending is therefore characterized by the ellipse of inertia of its cross section.

For an ordinary rectangular beam this ellipse will have the form shown in figure 21. The rigidity of such a beam under a load in the direction of the Oz-axis is proportional to bh^3.

Steel beams often have a ʃ -shaped cross section; for such beams the cross section and the ellipse of inertia are represented in figure 22. The

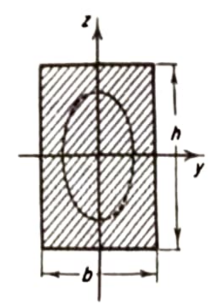


FIG. 21.

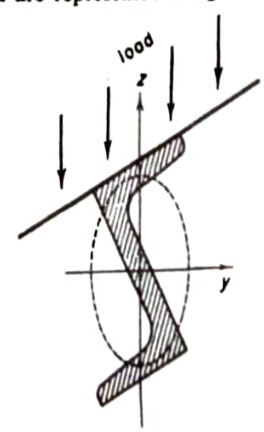


FIG. 22.

greatest resistance to bending is in the z direction. When they are used, for example as roof rafters under a load of snow and their own weights, they work directly against bending in a direction close to this most advantageous direction.

The hyperbola and its focal property. Now we consider the equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

representing a curve which is called a *hyperbola*. If we denote by c a number such that $c^2 = a^2 + b^2$, then it is possible to show that a hyperbola is the locus of all points the difference of whose distances to the points F_1 and F_2 on the Ox-axis with abscissas c and $-c$ is a constant: $\rho_2 - \rho_1 = 2a$ (figure 23). The points F_1 and F_2 are called the *foci*.

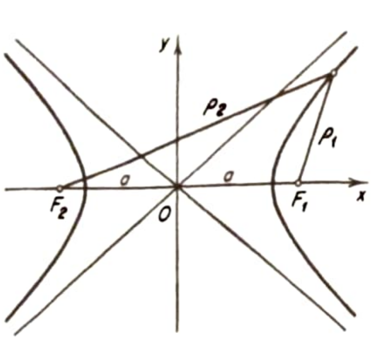


FIG. 23.

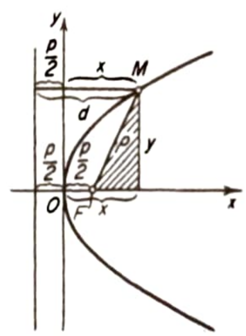


FIG. 24.

The parabola and its directrix. Finally, we consider the equation

$$y^2 = 2px$$

and call the corresponding curve a *parabola*. The point F lying on the Ox-axis with abscissa $p/2$ is called the *focus* of the parabola, and the straight line $y = -p/2$, parallel to the Oy-axis, is its *directrix*. Let M be any point of the parabola (figure 24), ρ the length of its focal radius MF, and d the length of the perpendicular dropped from it to the directrix. Let us compute ρ and d for the point M. From the shaded triangle we obtain $\rho^2 = (x - p/2)^2 + y^2$. As long as the point M lies on the parabola, we have $y^2 = 2px$, hence

$$\rho^2 = \left(x - \frac{p}{2}\right)^2 + 2px = \left(x + \frac{p}{2}\right)^2.$$

But directly from the figure it is clear that $d = x + p/2$. Therefore $\rho^2 = d^2$, i.e., $\rho = d$. The inverse argument shows that if for a given point we have $\rho = d$, then the point lies on the parabola. Thus a parabola is the locus of points equidistant from a given point F (called the focus) and a given straight line d (called the directrix).

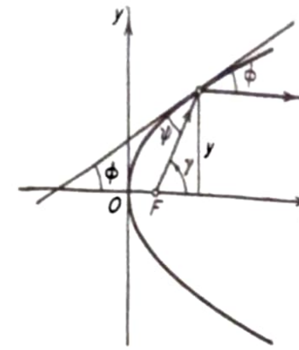


FIG. 25.

The property of the tangent to a parabola. Let us examine an important property of the tangent to a parabola and its application in optics. Since for a parabola $y^2 = 2px$ we have $2y\,dy = 2p\,dx$, it follows that the derivative, or the slope of the tangent, is equal to $dy/dx = \tan\phi = p/y$ (figure 25).

On the other hand, it follows directly from the figure that

$$\tan\gamma = \frac{y}{x - p/2}.$$

But

$$\tan 2\phi = \frac{2\,p/y}{1 - p^2/y^2} = \frac{2py}{y^2 - p^2} = \frac{2py}{2px - p^2} = \frac{y}{x - p/2},$$

i.e., $\gamma = 2\phi$, and since $\gamma = \phi + \psi$, therefore $\psi = \phi$. Consequently, by virtue of the law (angle of incidence is equal to angle of reflection) a beam of light, starting from the focus F and reflected by an element of the parabola (whose direction coincides with the direction of the tangent) is reflected parallel to the Ox-axis, i.e., parallel to the axis of symmetry of the parabola.

On this property of the parabola is based the construction of reflecting telescopes, as invented by Newton. If we manufacture a concave mirror whose surface is a so-called paraboloid of revolution, i.e., a surface obtained by the rotation of a parabola around its axis of symmetry, then all the light rays originating from any point of a heavenly body lying strictly in the direction of the "axis" of the mirror are collected by the mirror (figure 26) at one point, namely its focus. The rays originating from some other point of the heavenly body, being not exactly parallel to the axis of the mirror, are collected almost at one point in the neighborhood of the focus. Thus, in the so-called focal plane through the focus of the mirror and perpendicular to its axis, the inverse image of the star is obtained; the farther away this image is from the focus, the more diffuse it will be, since it is only the rays exactly parallel to the axis of the mirror that are collected by the mirror at one point. The image so obtained can be viewed in a special microscope, the so-called

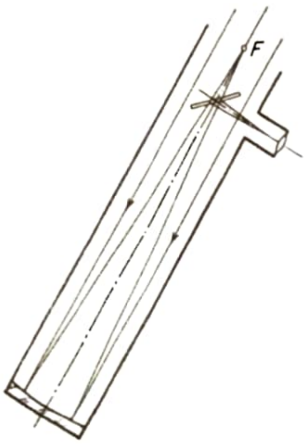


FIG. 26.

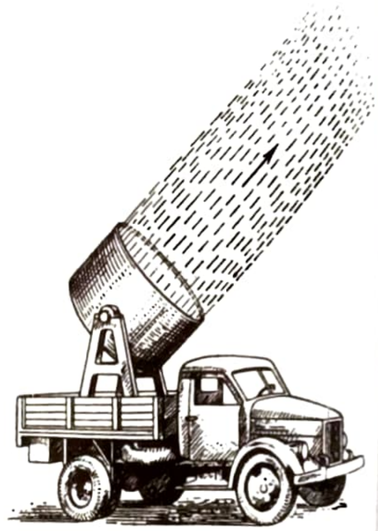


FIG. 27.

eye piece of the telescope, either directly or, in order not to cut off the light from the star with one's own head, after reflection in a small plane mirror, attached to the telescope near the focus (somewhat nearer than the focus to the concave mirror) at an angle of 45°.

The searchlight (figure 27) is based on the same property of the parabola. In it, conversely, a strong source of light is placed at the focus of a paraboloidal mirror, so that its rays are reflected from the mirror in a beam parallel to its axis. Automobile headlights are similarly constructed (figure 28).

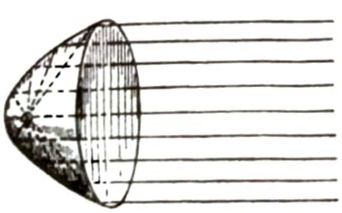


FIG. 28.

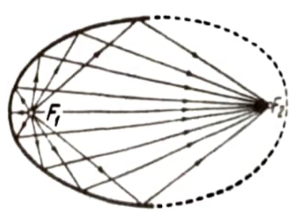


FIG. 29.

In the case of an ellipse, as it is easy to show, the rays issuing from one of its foci F_1 and reflected by the ellipse are collected at the other focus F_2 (figure 29), and in the hyperbola the rays originating from one of its foci F_1 are reflected by it as if they originated from the other focus F_2 (figure 30).

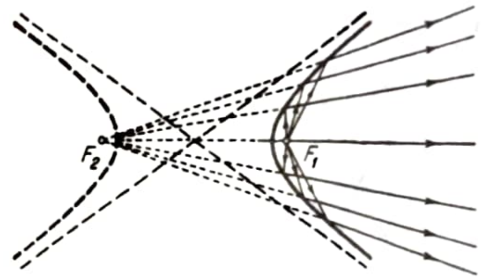


FIG. 30.

The directrices of the ellipse and the hyperbola. Like the parabola, the ellipse and the hyperbola have directrices, in this case two apiece. If we consider a focus and the directrix "on the same side with it," then for all points M of the ellipse we have $\rho/d = \epsilon$, where the constant ϵ is the eccentricity, which for an ellipse is always smaller than 1; and for all points of the corresponding branch of the hyperbola, we also have $\rho/d = \epsilon$, where ϵ is again the eccentricity, which for a hyperbola is always greater than 1.

Thus the ellipse, the parabola and one branch of the hyperbola are the loci of all those points in the plane for which the ratio of their distance ρ from the focus to their distance d from the directrix is constant (figures 31 and 32). For the ellipse this constant is smaller than unity, for the parabola it is equal to unity, and for the hyperbola it is greater than unity. In

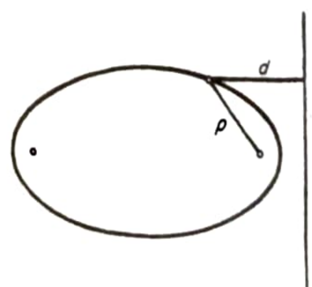


FIG. 31.

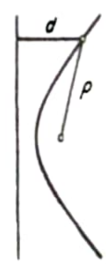


FIG. 32.

this sense the parabola is the "limiting" or "transition" case from the ellipse to the hyperbola.

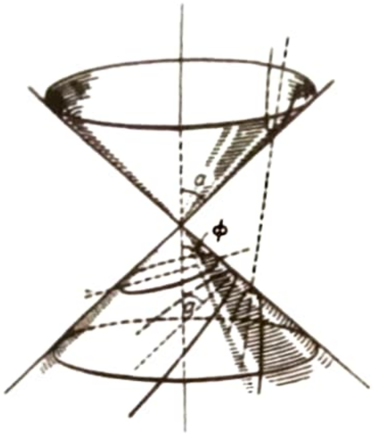


FIG. 33.

Conic sections. The ancient Greeks had already investigated in detail the curves obtained by intersecting a straight circular cone by a plane. If the intersecting plane makes with the axis of the cone an angle ϕ of 90°, i.e., is perpendicular to it, then the section obtained is a circle. It is easy to show that if the angle ϕ is smaller than 90°, but greater than the angle α which the generators of the cone make with its axis, then an ellipse is obtained. If ϕ is equal to α, a parabola results and if ϕ is smaller than α, then we obtain a hyperbola as the section (figure 33).

The parabola as the graph of quadratic proportion and the hyperbola as the graph of inverse proportion. We recall that the graph of quadratic proportion

$$y = kx^2$$

is a parabola (figure 34) and that the graph of inverse proportion

$$y = \frac{k}{x} \quad \text{or} \quad xy = k$$

is a hyperbola (figure 35), as we will easily prove later. A hyperbola was defined earlier as the curve represented by the equation

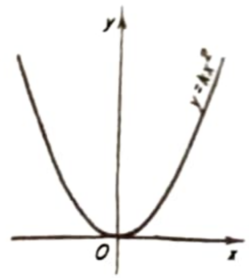


FIG. 34.

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

In the special case $a = b$ the so-called *rectangular hyperbola* plays the same role among hyperbolas as the circle plays among ellipses. In this case, if we rotate the coordinate axes by 45° (figure 36) the equation in the new coordinates (x', y') will have the form

$$x'y' = k.$$

We have now considered three important second-order curves: the ellipse, the hyperbola, and the parabola, and for their definitions we have taken the so-called canonical equations

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \quad \text{and} \quad y^2 = 2px,$$

by which they are represented.

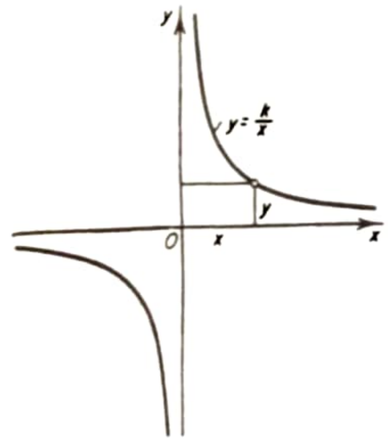


FIG. 35.

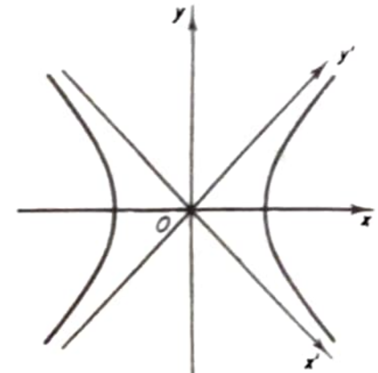


FIG. 36.

We now pass to the study of the general second-degree equation in two unknowns, namely to the question what kinds of curves are represented by this equation.

§8. The Reduction of the General Second-Degree Equation to Canonical Form

The first consistent presentation of analytic geometry by Euler. A significant step in the development of analytic geometry was the appearence in 1748 of the book "Introduction to analysis" in the second volume of which, among other things related to the theory of functions and other branches of analysis, for the first time a presentation was given of analytic geometry in the plane with a detailed investigation of second-order curves, very close to the one given in contemporary textbooks of analytic geometry, and also with an investigation of higher order curves. This was the first exposition of analytic geometry in the contemporary sense of the word.

The notion of reducing an equation to canonical form. A second-degree equation*

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0$$

contains six terms, not three or only two as in the above canonical equations of the ellipse, hyperbola, and parabola. This is not because such an equation represents a more complicated curve but because the system of coordinates is possibly not suited to it. It turns out that if we select a suitable Cartesian coordinate system, then a second-degree equation with two variables always can be reduced to one of the following canonical forms:

1. $\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0.$ Ellipse
2. $\frac{x^2}{a^2} + \frac{y^2}{b^2} + 1 = 0.$ Imaginary ellipse
3. $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0.$ Point (a pair of imaginary lines intersecting in a real point)
4. $\frac{x^2}{a^2} - \frac{y^2}{b^2} - 1 = 0.$ Hyperbola
5. $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0.$ A pair of intersecting lines
6. $y^2 - 2px = 0.$ Parabola
7. $x^2 - a^2 = 0.$ A pair of parallel lines
8. $x^2 + a^2 = 0.$ A pair of imaginary parallel lines
9. $x^2 = 0.$ A pair of coincident straight lines

where a, b, p, are not equal to zero.

Equations 1, 4, and 6 of the enumerated canonical forms are already well known to us; these are the canonical equations of the ellipse, hyperbola, and parabola. Two of them are not satisfied by any points, namely equations 2 and 8. Indeed, the square of a real number is always positive or zero, so that on the left-hand side of equation 2 the sum of the terms $x^2/a^2 + y^2/b^2$ is never negative, and since the term $+1$ also appears, the

* The coefficients of xy, x, y will be denoted not by B, D, E but by $2B$, $2D$, $2E$ for simplicity of subsequent formulas.

left-hand side cannot be equal to zero; analogously in equation 8, the number x^2 is not negative, and a^2 is positive. From these considerations, it follows that only $(x = 0, y = 0)$ satisfies equation 3, i.e., one point, the origin. Equation 5 can be written as $(x/a - y/b)(x/a + y/b) = 0$, from which we see that it is satisfied by those points and only those points for which one of the first-degree expressions $x/a - y/b$ or $x/a + y/b$ is equal to zero; so the curve it represents is this pair of intersecting lines. Equation 7 analogously gives $(x - a)(x + a) = 0$; i.e., the corresponding curve is a pair of parallel lines $x = a$ and $x = -a$. Finally, curve 9 is a special limiting case of curve 7, when $a = 0$; i.e., it is a pair of coincident lines.

Formulas of coordinate transformations. In order to obtain the indicated important result about the possible types of second-order curves, it is necessary first to deduce the formulas by which the rectangular coordinates of points vary under a change of the coordinate system.

Let x, y be the coordinates of a point M relative to the axes Oxy. Let us translate these axes parallel to themselves to the position $O'x'y'$ and let the coordinates of the new origin O' relative to the old axes

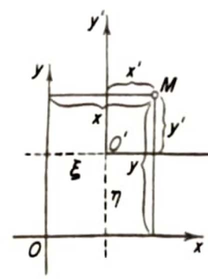


FIG. 37.

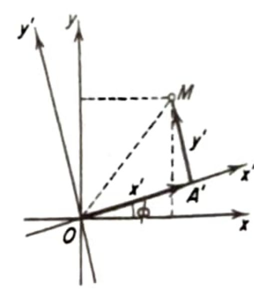


FIG. 38.

be ξ and η. It is evident (figure 37) that the new coordinates x', y' of the point M are connected with its old coordinates x, y by the formulas

$$x = x' + \xi,$$

$$y = y' + \eta,$$

which are the formulas of the so-called parallel translation of axes. If we rotate the original axes Oxy about the origin counterclockwise by an angle ϕ then, as is easy to see (figure 38), if we project the polygonal line $OA'M$ composed of the new coordinate segments x', y' on the Ox-axis and the Oy-axis, respectively, we obtain

$$x = x' \cos\phi - y' \sin\phi,$$
$$y = x' \sin\phi + y' \cos\phi,$$

which are the formulas for transformation of coordinates under rotation of a rectangular coordinate system.

If we are given an equation $F(x, y) = 0$ of a curve relative to the axes Oxy and we wish to write the transformed equation of the same curve, i.e., relative to the new axes $O'x'y'$, then we must replace x and y in the equation $F(x, y) = 0$ by their expressions in terms of x' and y', given by the formulas of the transformation. For example, under parallel translation of the axes, we obtain the transformed equation

$$F(x' + \xi, y' + \eta) = 0,$$

and under rotation of the axes the equation

$$F(x' \cos\phi - y' \sin\phi, x' \sin\phi + y' \cos\phi) = 0.$$

We note that under a transformation to new axes the degree of an equation does not change. Indeed, the degree cannot increase, since the transformation formulas are of the first-degree. But the degree cannot decrease either, since then the inverse coordinate transformation would increase it (and it is also of the first degree).

The reduction of a general second-degree equation to one of the 9 canonical forms. We now show that given any second-degree equation in two unknowns we can always first rotate the axes and then translate them parallel to themselves in such a way that the transformed equation for the final axes will have one of the forms 1, 2, ···, 9.

Indeed, let the given second-degree equation have the form

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0. \tag{8}$$

Let us rotate the axes through some angle ϕ, which we select in the following way. Replacing x and y in equation (8) by their expressions in terms of the new coordinates (according to the formulas for rotation), we find, after collecting similar terms, that the coefficient $2B'$ in the transformed equation

$$A'x'^2 + 2B'x'y' + C'y'^2 + 2D'x' + 2E'y' + F' = 0$$

is equal to

$$\begin{aligned} 2B' &= -2A \sin\phi \cos\phi + 2B(\cos^2\phi - \sin^2\phi) + 2C \sin\phi \cos\phi \\ &= 2B \cos 2\phi - (A - C) \sin 2\phi. \end{aligned}$$

Setting it equal to zero, we obtain $2B \cos 2\phi = (A - C) \sin 2\phi$, from which

$$\cot 2\phi = \frac{A - C}{2B}.$$

Since the cotangent varies from $-\infty$ to $+\infty$, we can always find an angle ϕ for which this equality is satisfied. By rotating the axes through this angle, we find that for the rotated axes $Ox'y'$ the equation of our curve, represented for the initial axes by equation (8), has the form

$$A'x'^2 + C'y'^2 + 2D'x' + 2E'y' + F = 0, \tag{9}$$

i.e., that it does not contain the term with the product of the coordinates (F remains unchanged, since the formulas of rotation do not contain constant terms).

Now we translate the already rotated axes $Ox'y'$ parallel to themselves to the position $O''x''y''$, and let the coordinates of the new origin O'' relative to the axes $Ox'y'$ be ξ', η'. The equation of our curve for these final axes will be

$$A'(x'' + \xi')^2 + C'(y'' + \eta')^2 + 2D'(x'' + \xi') + 2E'(y'' + \eta') + F = 0. \tag{10}$$

We now show that we can always select ξ' and η' (i.e., we can translate the axes $Ox'y'$ parallel to themselves) in such a way that the final equation for the axes $O''x''y''$ has one of the canonical forms 1, 2, ···, 9.

Removing all parentheses in equation (10) and collecting similar terms, we obtain

$$A'x''^2 + C'y''^2 + 2(A'\xi' + D')x'' + 2(C'\eta' + E')y'' + F' = 0, \tag{10'}$$

where we have denoted by F' the sum of all constant terms; its value does not interest us at the moment.

We consider three possible cases.

I. A' and C' both not equal to zero. In this case, taking $\xi' = -D'/A'$, $\eta' = -E'/C'$, we eliminate the terms with the first powers of x'' and y'' and obtain an equation of the form

$$A'x''^2 + C'y''^2 + F' = 0. \tag{I}$$

II. $A' \neq 0, C' = 0$, but $E' \neq 0$. Letting $\xi' = -D'/A', \eta' = 0$, i.e., $y'' = y'$, we obtain the equation

$$A'x''^2 + 2E'y' + F' = 0,$$

or

$$A'x''^2 + 2E'\left(y' + \frac{F'}{2E'}\right) = 0.$$

Then making a parallel translation along the Oy'-axis by an amount $\eta'' = -F'/2E'$, we find that $y' = y'' - F'/2E'$, i.e., $y' + F'/2E' = y''$ so that we obtain the equation

$$A'x''^2 + 2E'y'' = 0. \tag{II}$$

If we have $A' = 0$, $C' \neq 0$, $D' \neq 0$, we can simply interchange the roles of x and y and obtain the same result.

III. $A' \neq 0$, $C' = 0$, $E' = 0$. Taking again $\xi' = -D'/A'$, $\eta' = 0$, we obtain the equation

$$A'x''^2 + F' = 0. \tag{III}$$

If we have $A' = 0$, $C' \neq 0$, $D' = 0$, we can again interchange the roles of x and y.

We have now considered all the possibilities, in view of the fact that A' and C' cannot simultaneously be zero, since then the degree of the equation would be reduced, and we have seen that under our coordinate transformations this degree does not change.

Thus, with the appropriate choice of rectangular coordinates every second-degree equation can be brought to one of the three so-called "reduced" equations (I), (II), (III).

Let the equation have the form (I) (in this case A' and C' are not equal to zero). If $F' \neq 0$, then writing equation (I) as

$$\frac{x''^2}{-F'/A'} + \frac{y''^2}{-F'/C'} - 1 = 0,$$

we arrive, depending on the signs of A', C', F', at one of the equations 1, 2, or 4. If the denominator of x''^2 is negative and that of y''^2 is positive, then we must also interchange the axes $O''x''$ and $O''y''$.

If $F' = 0$, then equation (I) can be written in the form

$$\frac{x''^2}{1/A'} + \frac{y''^2}{1/C'} = 0,$$

and we arrive at equations 3 or 5.

If the equation has the form (II) (in this case A' and E' are not both zero), then we can write it as

$$x''^2 + \frac{2E'}{A'}y'' = 0,$$

and denoting $-E'/A'$ by p and interchanging the names of the axes $O''x''$ and $O''y''$ we obtain equation 6.

Finally, if we have an equation of form (III) (where $A' \neq 0$), it can be rewritten as $x''^2 + F'/A' = 0$ and one of the equations 7, 8, or 9 is obtained.

This important theorem on the possibility of reducing every 2nd-degree equation to one of the 9 canonical forms was already examined in detail by Euler. The arguments in Euler's book differ only in form from the ones just given.

§9. The Representation of Forces, Velocities, and Accelerations by Triples of Numbers; Theory of Vectors

Following Euler an important step was taken by Lagrange. In his "Analytic mechanics," published in 1788, Lagrange arithmetized forces, velocities and accelerations in the same way as Descartes arithmetized points. This idea that Lagrange developed in his book subsequently took the form of the so-called theory of vectors and proved to be an important help in physics, mechanics, and technology.

Rectangular coordinates in space. We remark, first of all, that neither Descartes nor Newton developed analytic geometry in space. This was done later on, in the first half of the 18th century, by Laguerre and Clairaut. In order to specify a point M in space they selected three mutually perpendicular axes Ox, Oy, and Oz and considered (figure 39) the numerical values of the distances of the point M from the planes Oyz, Oxz, and Oxy, taken with the corresponding signs, the so-called *abscissa* x, *ordinate* y, and *altitude* z of the point M.

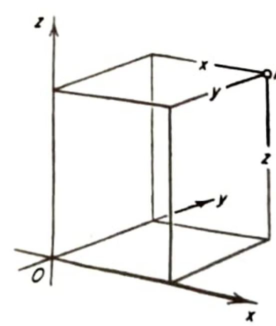


FIG. 39.

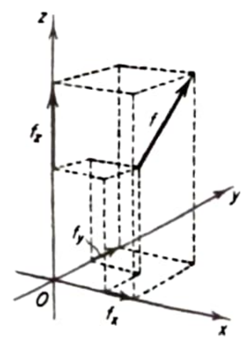


FIG. 40.

Arithmetization of forces, velocities, and accelerations, introduced by Lagrange. We consider (figure 40) a force f which can be represented in conventional units by a segment with an arrow, having a specific length and direction. Lagrange points out that this force f can be decomposed into three components f_x, f_y, and f_z in the direction of the corresponding axes Ox, Oy, and Oz; these components, as directed segments on the axes, can be given simply by numbers, positive or negative depending on whether the component is directed in the positive or the opposite direction of the axis. Thus, we can consider, for example, the force (2, 3, 4) or the force (1, —2, 5), etc. In the composition of forces according to the parallelogram law, as can easily be shown (it will be shown later), their corresponding components have to be added. For example, the sum of the given forces is the force

$$(2 + 1, 3 - 2, 4 + 5) = (3, 1, 9).$$

The same can be done for velocities and accelerations. In every problem of mechanics, all the equations connecting forces, velocities, and accelerations can also be written as equations connecting their components, i.e., connecting simply numbers; then the mechanical equation will necessarily be written in the form of three equations: first for the x's, the second for the y's, and the third for the z's.

But it was only after a hundred years from the time of Lagrange that mathematicians and physicists, particularly under the influence of the developing theory of electricity, began on a wide scale to consider the general theory of such segments, having a definite length and direction. Such segments were called vectors.

The theory of vectors has a great significance in mechanics, physics, and technology, and its algebraic side, the so-called algebra of vectors (in contrast to vector analysis) appears at once as an essential constituent part of analytic geometry.

Algebra of vectors. Any directed segment (whether it represents a force, a velocity, an acceleration, or some other entity) i.e., a segment having a given length and a definite direction, is called a *vector*. Two vectors are said to be *equal*, if they have the same length and the same direction; i.e., in the very concept of "vector" only its length and its direction are taken into account. Vectors can be added. Let the vectors $\mathbf{a}, \mathbf{b}, \cdots, \mathbf{d}$ be given. We lay out the vector $\mathbf{a}$ from some point, then from its end point we draw the vector $\mathbf{b}$, etc. We obtain a so-called vector polygon $\mathbf{ab} \cdots \mathbf{d}$ (figure 41). The vector $\mathbf{m}$ whose initial point coincides with the initial point of the first vector $\mathbf{a}$ of this polygon, and whose

end point coincides with the end point of the last vector $\mathbf{d}$, is called the *sum* of these vectors

$$\mathbf{m} = \mathbf{a} + \mathbf{b} + \cdots + \mathbf{d}. \tag{11}$$

It is easy to show that the vector $\mathbf{m}$ does not depend on the order in which the summands $\mathbf{a}, \mathbf{b}, \cdots, \mathbf{d}$ are taken.

FIG. 41.

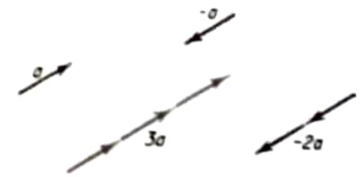


FIG. 42.

The vector equal in length to the vector $\mathbf{a}$ but opposite in direction is called its *inverse* vector and is denoted by $-\mathbf{a}$.

Subtraction of the vector $\mathbf{a}$ is defined as addition of its inverse vector.

In vector calculus ordinary real numbers are customarily called scalars. Let a vector $\mathbf{a}$ (figure 42) and a scalar λ be given, then by the product of the vector $\mathbf{a}$ with the scalar (number) λ, i.e., $\lambda\mathbf{a}$, is meant the vector whose length is equal to the product of the length $|\mathbf{a}|$ of the vector $\mathbf{a}$

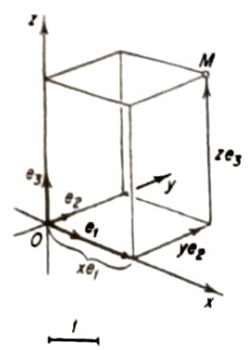


FIG. 43.

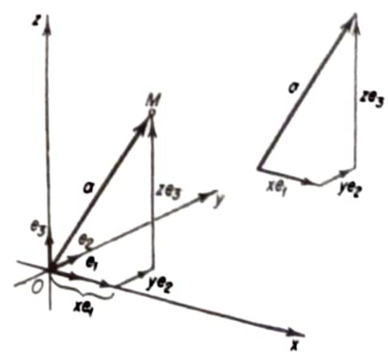


FIG. 44.

and the absolute value $|\lambda|$ of the number λ, and whose direction is the same as that of $\mathbf{a}$ if $\lambda > 0$ and the opposite if $\lambda < 0$.

Let us consider a system of rectangular Cartesian coordinates $Oxyz$ and the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ having length equal to unity and directions coinciding with the positive directions of the axes Ox, Oy, Oz, respectively. It is obvious that any given point M (figure 43) of space can be reached from the origin O by traversing a certain number of "times" (an integral, fractional or irrational, positive or negative "number of times") the vector $\mathbf{e}_1$, then so many "times" the vector $\mathbf{e}_2$, and finally so many "times" the vector $\mathbf{e}_3$. It is clear that the numbers x, y, z showing how many "times" it is necessary to traverse the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, are simply the Cartesian coordinates of the point M.

Let a vector $\mathbf{a}$ be given; if we cause a point to move from the initial point of $\mathbf{a}$ to its end point and decompose this motion into motions parallel to the axes Ox, Oy, and Oz, and if it is hereby necessary to shift the point through a distance $x\mathbf{e}_1$ parallel to the Ox-axis, through $y\mathbf{e}_2$ parallel to the Oy-axis and through $z\mathbf{e}_3$ parallel to the Oz-axis, then

$$\mathbf{a} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3. \tag{12}$$

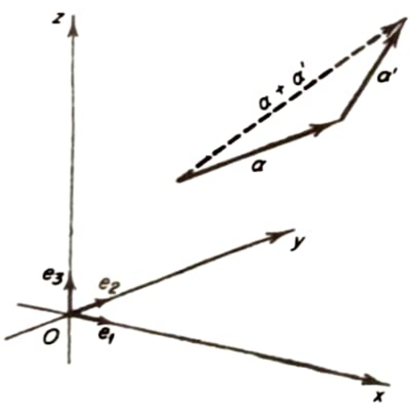


FIG. 45.

The numbers x, y, z are called the *coordinates* of the vector $\mathbf{a}$. These are obviously just the coordinates of the end point M of this vector, if its initial point lies at the origin O of the coordinate system (figure 44). From this it follows at once that in adding vectors their corresponding coordinates are to be added, and in subtraction they are to be subtracted. If the first vector "carries" us along the Ox-axis by a distance $x\mathbf{e}_1$, and the second by $x'\mathbf{e}_1$, then clearly their sum "carries" us along the Ox-axis by a distance $(x + x')\mathbf{e}_1$, etc. (figure 45). It also follows at once that in multiplication of a vector by a number, its coordinates are multiplied by the number.

Scalar product and its properties. If we are given two vectors $\mathbf{a}$ and $\mathbf{b}$, then the number equal to the product of their lengths by the cosine of the angle between them $|\mathbf{a}|\,|\mathbf{b}|\cos\phi$ is called their *scalar product* and is denoted by $\mathbf{ab}$ or $(\mathbf{ab})$. Let x, y, z be the coordinates of the vector $\mathbf{a}$

and $\bar{x}$, $\bar{y}$, $\bar{z}$ the coordinates of the vector $\mathbf{b}$; then the scalar product is equal to

$$\mathbf{ab} = x\bar{x} + y\bar{y} + z\bar{z}, \tag{13}$$

i.e., to the sum of the products of their corresponding coordinates.

This important result can be proved as follows. First we make the following remarks:

(1) If we multiply one of the vectors of a scalar product, for example $\mathbf{a}$, by a number λ, this is obviously the same as multiplying their scalar product by the same number, i.e.,

$$(\lambda\mathbf{a})\mathbf{b} = \lambda(\mathbf{ab}).$$

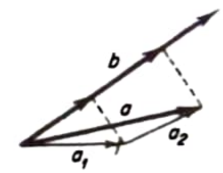


FIG. 46.

(2) The scalar product is distributive, i.e., if $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$, then $\mathbf{ab} = \mathbf{a}_1\mathbf{b} + \mathbf{a}_2\mathbf{b}$.

In fact, the left-hand side of this equality is equal to the product of the length of the vector $\mathbf{b}$ by the numerical value of the projection of the vector $\mathbf{a}$ on the axis of the vector $\mathbf{b}$ (figure 46), and the right-hand side is equal to the product of the length of $\mathbf{b}$ by the sum of the numerical values of the projections of the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ on the axis of $\mathbf{b}$. But proj $\mathbf{a}$ = proj $\mathbf{a}_1$ + proj $\mathbf{a}_2$, which proves the equality.

Now we consider two vectors $\mathbf{a}$ and $\mathbf{b}$ whose decompositions in terms of the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ are $\mathbf{a} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$, $\mathbf{b} = \bar{x}\mathbf{e}_1 + \bar{y}\mathbf{e}_2 + \bar{z}\mathbf{e}_3$, so that

$$\mathbf{ab} = (x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3)(\bar{x}\mathbf{e}_1 + \bar{y}\mathbf{e}_2 + \bar{z}\mathbf{e}_3).$$

By the distributivity (2) of the scalar product, the sums of vectors in parentheses can be multiplied as polynomials, and by (1) the scalar factors in each of the terms can be taken outside the parentheses, so that

$$\begin{aligned}\mathbf{ab} = x\bar{x}\mathbf{e}_1\mathbf{e}_1 + x\bar{y}\mathbf{e}_1\mathbf{e}_2 + x\bar{z}\mathbf{e}_1\mathbf{e}_3 + y\bar{x}\mathbf{e}_2\mathbf{e}_1 + y\bar{y}\mathbf{e}_2\mathbf{e}_2 + y\bar{z}\mathbf{e}_2\mathbf{e}_3 \\ + z\bar{x}\mathbf{e}_3\mathbf{e}_1 + z\bar{y}\mathbf{e}_3\mathbf{e}_2 + z\bar{z}\mathbf{e}_3\mathbf{e}_3.\end{aligned}$$

But

$$|\mathbf{e}_1| = |\mathbf{e}_2| = |\mathbf{e}_3| = 1, \quad \cos 0^\circ = 1 \text{ and } \cos 90^\circ = 0.$$

Consequently,

$$\begin{aligned}\mathbf{e}_1\mathbf{e}_1 &= 1, & \mathbf{e}_1\mathbf{e}_2 &= 0, & \mathbf{e}_1\mathbf{e}_3 &= 0,\\ \mathbf{e}_2\mathbf{e}_1 &= 0, & \mathbf{e}_2\mathbf{e}_2 &= 1, & \mathbf{e}_2\mathbf{e}_3 &= 0,\\ \mathbf{e}_3\mathbf{e}_1 &= 0, & \mathbf{e}_3\mathbf{e}_2 &= 0, & \mathbf{e}_3\mathbf{e}_3 &= 1.\end{aligned}$$

Thus,

$$\mathbf{ab} = x\bar{x} + y\bar{y} + z\bar{z}. \tag{14}$$

We remark, in particular, that if the vectors **a** and **b** are mutually perpendicular, then $\phi = 90°$ and $\cos\phi = 0$. Therefore the equality

$$x\bar{x} + y\bar{y} + z\bar{z} = 0 \tag{15}$$

serves as an easily verifiable condition of perpendicularity of the vectors **a** and **b**.

Angle between two directions. Let us consider a direction characterized by its angles α, β, γ with the coordinate axes. We draw the line

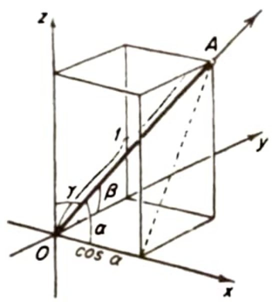


FIG. 47.

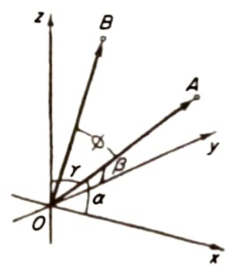


FIG. 48.

in this direction through the origin of the coordinate system and mark off on it from the origin a segment OA of unit length (figure 47). In this case the coordinates of the point A, i.e., the coordinates of the vector $\overrightarrow{OA}$ are exactly $\cos\alpha$, $\cos\beta$, and $\cos\gamma$. If we have a second direction given by the angles $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$, then the analogous vector $\overrightarrow{OB}$ for this second direction has coordinates $\cos\bar{\alpha}$, $\cos\bar{\beta}$, $\cos\bar{\gamma}$ (figure 48). Let ϕ be the angle between these vectors; then their scalar product is equal to $1 \cdot 1 \cos\phi$ from which we find

$$\cos\phi = \cos\alpha\cos\bar{\alpha} + \cos\beta\cos\bar{\beta} + \cos\gamma\cos\bar{\gamma}. \tag{16}$$

This is the very important formula for the cosine of the angle between two directions.

§10. Analytic Geometry in Space; Equations of a Surface in Space and Equations of a Curve

If an equation $z = f(x, y)$ is given and if x and y are regarded as the abscissa and ordinate and z the altitude of a point, then this equation

itself represents some surface P, which can be obtained by erecting perpendiculars of length z at the points (x, y) of the Oxy-plane. The locus of the end points of these perpendiculars gives the surface P represented by this equation. If the equation connecting x, y, and z is not already solved with respect to z, then it can be solved for z and after that we can construct the surface P. In general, in analytic geometry the totality of all those points of space whose coordinates x, y, z satisfy a given equation (figure 49) in three variables x, y, z is said to be the surface represented by the equation.

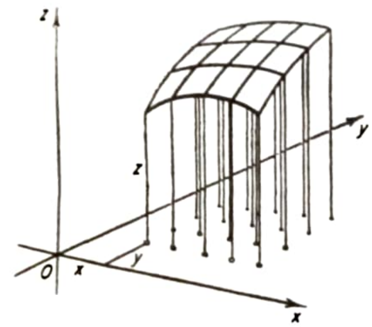


FIG. 49.

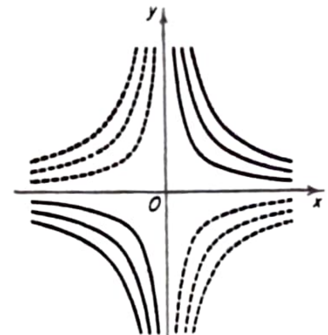


FIG. 50.

A function of two variables $f(x, y)$, as was pointed out already in Chapter II, can represent not only a surface P, but also its system of level curves, i.e., curves in the Oxy-plane on each of which the function $f(x, y)$ has a constant value. This system of curves is clearly nothing else than the topographical map of the surface P on the Oxy-plane.

Example. The equation $xy = z$ gives, for instance, the level curves: $\cdots, xy = -3, xy = -2, xy = -1, xy = 0, xy = 1; xy = 2, xy = 3, \cdots$. All of them are hyperbolas (figure 50) except $xy = 0$, which represents the two coordinate axes. What is obtained is clearly a saddlelike surface (figure 51) (the so-called hyperbolic paraboloid).

In order to define a curve in space, we can give the equations of any two surfaces P and Q which intersect along the curve. For example, the system

$$xy = z,$$
$$x^2 + y^2 = 1$$

gives a space curve (figure 52). The equation $xy = z$ determines the earlier hyperbolic paraboloid, and the equation $x^2 + y^2 = 1$ determines a circular cylinder of unit radius, whose axis is the Oz-axis. The system of equations consequently defines the curve of intersection of the paraboloid with the cylinder, which is represented in figure 52.

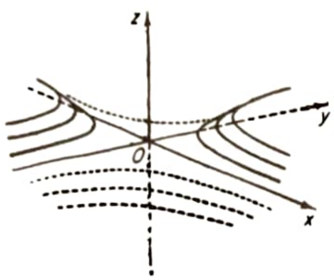


FIG. 51.

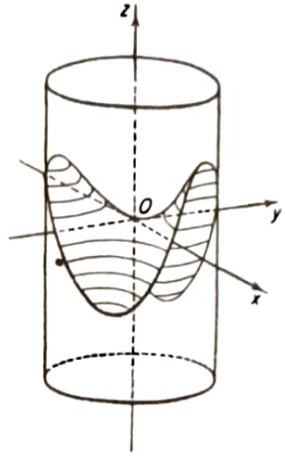


FIG. 52.

If in this system one of the unknowns, say x, is chosen arbitrarily, and then the system is solved with respect to y and z, we will obtain the coordinates x, y, z of the various points of the curve.

Equation of a plane and equations of a straight line. It can be shown that every equation of the first degree with three variables

$$Ax + By + Cz + D = 0$$

represents a plane, and conversely. By what has already been said, it is clear that a line can be given by a system of two such equations:

$$\begin{aligned} A_1x + B_1y + C_1z + D_1 = 0, \\ A_2x + B_2y + C_2z + D_2 = 0, \end{aligned}$$

i.e., as the curve of intersection of two planes.

The general second-degree equation in three variables and its 17 canonical forms. A second-degree equation in three variables

$$\begin{aligned} A_1x^2 + A_2y^2 + A_3z^2 + 2B_1yz + 2B_2xz + 2B_3xy \\ + 2C_1x + 2C_2y + 2C_3z + D = 0, \quad (17) \end{aligned}$$

contains 10 terms. Analogously to what was done earlier for an equation with two variables, it can be shown that by a suitable rotation of the given coordinate system about the origin, equation (17) can be reduced to the form

$$A_1'x'^2 + A_2'y'^2 + A_3'z'^2 + 2C_1'x' + 2C_2'y' + 2C_3'z' + D = 0, \quad (18)$$

i.e., so as to eliminate the terms with products of the variables. However, the proof here of the possibility of such a simplification of the equation is considerably more difficult than in the case of the plane. The difficulty of the proof arises from the fact that in the plane a rotation about a point is given by one angle ϕ, which we selected suitably, while in space the rotation of a body about a fixed point is given by three independent angles (Euler angles) ϕ, θ, ψ and in a quite complicated way. So the equation must be cleared of the terms with products of variables in a roundabout way (see Chapter XVI on the theory of reduction by orthogonal transformations of a quadratic form to a sum of squares). Then, as in the case of the plane, a parallel translation of the axes is made and the equation is simplified, after which equation (18) finally assumes one of the following canonical forms:

1. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} - 1 = 0$  Ellipsoid

2. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} + 1 = 0$  Imaginary ellipsoid

3. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} - 1 = 0$  Hyperboloid of one sheet

4. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} + 1 = 0$  Hyperboloid of two sheets

5. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 0$  Second-order cone

6. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 0$  Imaginary second-order cone

7. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - 2cz = 0$  Elliptic paraboloid

8. $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - 2cz = 0$  Hyperbolic paraboloid

9. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - 1 = 0$  Elliptic cylinder

10. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + 1 = 0$  Imaginary elliptic cylinder

11. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 0$  A pair of intersecting imaginary planes

12. $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - 1 = 0$  Hyperbolic cylinder

13. $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 0$  A pair of intersecting planes

14. $y^2 - 2px = 0$  Parabolic cylinder

15. $x^2 - a^2 = 0$  A pair of parallel planes

16. $x^2 + a^2 = 0$  A pair of imaginary parallel planes

17. $x^2 = 0$  A pair of coincident planes.

The last nine canonical equations 9-17 do not contain terms in z and represent exactly the canonical equations of second-order curves in the Oxy-plane. In space these equations represent cylinders, whose directrices are the corresponding second-order curves in the Oxy-plane and whose generators are parallel to the Oz-axis. Indeed, if one of these equations is satisfied by a point with coordinates $(x_1, y_1, 0)$, then it will also be satisfied by any point with coordinates (x_1, y_1, z) whatever z may be, since there are in any case no terms with z in the equation.

Among the equations 1-8 as can easily be seen, equation 2 is not satisfied by any point with real x, y, z and equation 6 is satisfied only by one such point (0, 0, 0), i.e., the origin. It remains, therefore, to study only the six equations 1, 3, 4, 5, 7, 8.

Ellipsoid. Let us compare the surfaces represented by the equations $x^2/a^2 + y^2/b^2 + z^2/c^2 - 1 = 0$ and $x^2 + y^2 + z^2 - 1 = 0$. The second of these is obviously the equation of a sphere C with center at the origin

and with unit radius, since $x^2 + y^2 + z^2$ is the square of the distance from the point (x, y, z) to the origin O. If (x, y, z) is a point lying on the sphere, i.e., satisfying the second equation, then (ax, by, cz) is a point whose coordinates satisfy the first equation. The surface represented by the first equation is thus obtained from the sphere C if all abscissas x of points of the sphere are replaced by ax, y by by, and z by cz, i.e., if the sphere C is uniformly stretched from the Oyz-, Oxz-, and Oxy-planes with coefficients of expansion a, b and c, respectively. This surface is called an ellipsoid (figure 53).

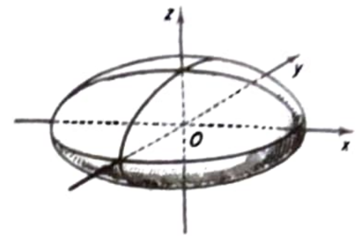


FIG. 53.

Hyperboloids and the second-order cone. Let us consider equations 3, 4, and 5, i.e., the equation of the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = \delta, \tag{19}$$

where $\delta = 1, -1$ or 0. Let us compare it with the equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} - \frac{z^2}{c^2} = \delta, \tag{20}$$

in which the denominator of y^2 is also a^2 and not b^2, as in equation (19). As before, we observe that surface (19) is obtained from surface (20) by expansion from the Oxz-plane with coefficient b/a.

Let us now see what surface is represented by (20). We take a plane $z = h$ perpendicular to the Oz-axis and examine its intersection with the surface (20). Substituting $z = h$ in equation (20), we obtain

$$x^2 + y^2 = a^2\left(\delta + \frac{h^2}{c^2}\right).$$

If $\delta + h^2/c^2$ is positive, then this equation together with $z = h$ gives a circle, lying in the plane $z = h$ with center on the Oz-axis. If $\delta + h^2/c^2$ is negative, which can be the case only with $\delta = -1$ and h^2 small, then the plane $z = h$ does not intersect surface (20) at all, since the sum of squares $x^2 + y^2$ cannot be a negative number.

The whole surface (20) thus consists of circles lying in planes perpendicular to the Oz-axis and having their centers on the Oz-axis. But in this case the surface (20) is a surface of revolution about the Oz-axis.

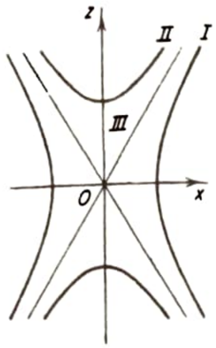


FIG. 54.

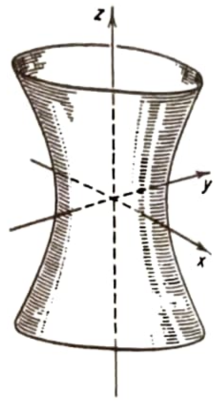


FIG. 55.

If we intersect it with a plane passing through the Oz-axis, we obtain its "meridian," i.e., a curve, lying in a plane passing through the axis, by the revolution of which the surface is generated.

If we intersect the surface (20) with the coordinate plane Oxz, i.e., the plane $y = 0$ (figure 54), by substituting $y = 0$ in equation (20), we obtain the equation of the meridian $x^2/a^2 - z^2/c^2 = \delta$. In case $\delta = 1$

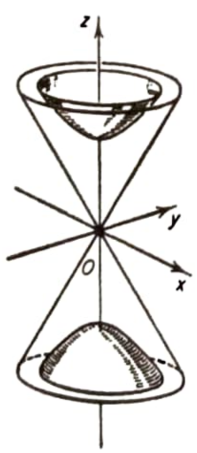


FIG. 56.

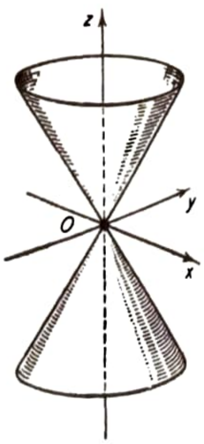


FIG. 57.

this is the hyperbola I, for $\delta = -1$, it is the hyperbola II, and for $\delta = 0$, the pair of intersecting lines III. By revolution around the z-axis these produce, respectively, a so-called hyperboloid of revolution of one sheet (figure 55), a hyperboloid of revolution of two sheets (figure 56) and a straight circular cone (figure 57).

The general hyperboloid of one sheet, hyperboloid of two sheets, and second-order cone 3, 4, and 5 are obtained from these surfaces of revolution by an expansion from the Oxz-plane with coefficient b/a.

Paraboloids. Only equations 7 and 8 remain. Let us compare the first of these $x^2/a^2 + y^2/b^2 = 2cz$ with the equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} = 2cz,$$

which we investigate in the same way as before. It represents a surface obtained by revolving the parabola $x^2 = 2a^2cz$ about the Oz-axis,

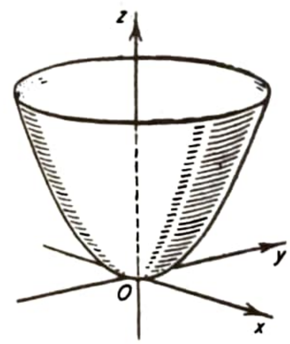


FIG. 58.

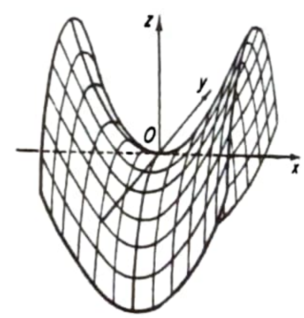


FIG. 59.

namely the so-called paraboloid of revolution (figure 58) discussed earlier in connection with parabolic mirrors. The general elliptic paraboloid 7 is obtained from the paraboloid of revolution by an expansion from the Oxz-plane.

The surface 8 has to be studied in a different way, namely by examining its intersections with planes $z = h$, which are hyperbolas. The contour map of the surface 8 is represented in figure 50; in a different position of the coordinate axes we considered this surface in figure 51. It is saddle-shaped, as illustrated in figure 59 and is called a hyperbolic paraboloid. Its intersections with planes parallel to the Oxz-plane turn out to be identical parabolas. The same result is obtained by intersections with planes parallel to the Oyz-plane.

Rectilinear generators of a hyperboloid of one sheet. It is a very curious and not at all obvious fact that the hyperboloid of one sheet and the hyperbolic paraboloid can be obtained, just like the cone and the cylinder, by the motion of a straight line. In case of the hyperboloid, it is sufficient to prove this fact for a hyperboloid of revolution of one sheet $x^2/a^2 + y^2/b^2 - z^2/c^2 = 1$, since the general hyperboloid of one sheet is obtained by a uniform expansion from the Oxz-plane and under such an expansion any straight line will go into a straight line. Let us

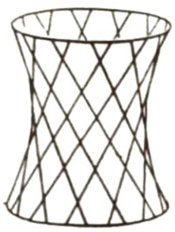


FIG. 60.

FIG. 61.

FIG. 62.

intersect the hyperboloid of revolution with the plane $y = a$ parallel to the Oxz-plane. Substituting $y = a$ we obtain

$$\frac{x^2}{a^2} + \frac{a^2}{a^2} - \frac{z^2}{c^2} = 1 \quad \text{or} \quad \frac{x^2}{a^2} - \frac{z^2}{c^2} = 0.$$

But this equation together with $y = a$ gives in the plane $y = a$ a pair of intersecting lines: $x/a - z/c = 0$ and $x/a + z/c = 0$.

Thus we have already discovered that there is a pair of intersecting lines lying on the hyperboloid. If now we revolve the hyperboloid about the Oz-axis, then each of these lines obviously traces out the entire hyperboloid (figure 60).

It is easy to show that: (1) two arbitrary straight lines of one and the same family of lines so obtained do not lie in the same plane (i.e., they are skew lines), (2) any line of one of these families intersects all the lines of the other family (except its opposite, which is parallel to it), and (3) three lines of one and the same family are not parallel to any one and the same plane.

With two matches and a needle it is easy to obtain a representation of the hyperboloid of revolution of one sheet. Let us puncture one of the matches through its middle by the needle, and on the sharp end point of the needle we pin the other match parallel to the first match. If we then revolve the whole apparatus about the first match as an axis, the second match will trace the surface of a cylinder (figure 61). But if the second match is not parallel to the first match, then during a revolution it will trace the surface of a hyperboloid of revolution of one sheet, as can easily be visualized if the rotation is rapid (figure 62).

Summary of the investigation of the second-degree equation. Although the general second-degree equation with three variables can represent essentially 17 different surfaces, it is not difficult to remember them. The last nine are cylinders over the nine possible second-order curves, while the first eight are divided into four pairs: two ellipsoids (real and imaginary), two hyperboloids (of one sheet and two sheets), two second-order cones (real and imaginary), and two paraboloids (elliptic and hyperbolic). All these surfaces play an essential role in mechanics, physics, and technology (ellipsoid of inertia, ellipsoid of elasticity, hyperboloid in the Lorentz transformation in physics, paraboloid of revolution for parabolic mirrors, etc.).

§11. Affine and Orthogonal Transformations

The next important step in the development of analytic geometry was the introduction into it, and into geometry in general, of the theory of transformations. Here it will be necessary to explain the matter in some detail.

"Contraction" of the plane toward a line. Let us consider one of the simplest transformations of the plane, namely uniform "contraction" toward a line with coefficient k. In the plane let there be given a line a and a positive coefficient k, for example, $k = 2/3$. All points of the line a are fixed, and every point M not lying on this line is sent into the point M' such that M' lies on the same side of the line as M on the perpendicular from M to a at a distance from a equal to 2/3 of the distance from M to a. If the coefficient k, as here, is smaller than unity, then we have a proper contraction of the

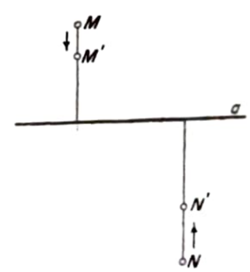


FIG. 63.

plane to the line; but if k is greater than unity, we have an expansion of the plane from the line, but for convenience we will in this and other cases talk about "contraction," except that the word "contraction" will be put in quotation marks.

The point or figure to be transformed is called the preimage and the one into which it is sent is its image. The point M', for example, is the image of the point M (figure 63).

We show that under a uniform "contraction" of a plane to a line, any line of the plane is transformed into a line. For let the plane be "contracted" to a line a lying in it with coefficient of "contraction" k. Let b be any line of the plane, O the point in which it intersects the line a, B another arbitrary point of b, and BA the perpendicular to the line a from the point B (figure 64). In the "contraction" the point B goes to

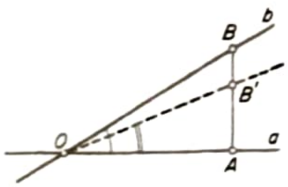


FIG. 64.

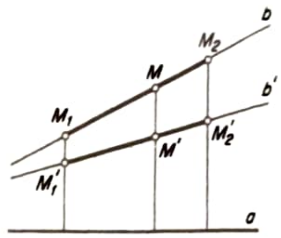


FIG. 65.

the point B' on this perpendicular such that $B'A = k \cdot BA$. Therefore, the tangent of the angle $B'OA$ will be equal to $AB'/OA = k \cdot AB/OA$, i.e., will be equal to k times the tangent of the angle which the line b makes with line a, i.e., for all points B' into which different points of the line b are transformed, it will be one and the same. All points B' consequently lie on one and the same line, passing through the point O and making with line a an angle with this tangent.

Under "contraction" parallel lines remain parallel. Indeed, if the tangents of the angles which lines b and c make with line a are the same, then the tangents of those angles which the images b' and c' make with a differ from them only by a factor k, i.e., they are still equal to each other, which means that the lines b' and c' are also parallel to each other.

Any rectilinear segment of the plane under "contraction" to a line is contracted (or expanded) uniformly (although to various degrees for segments of various directions). When we speak here of "uniform" contraction, we mean that the midpoint of the segment remains the midpoint, the third remains the third, etc., i.e., the segment shrinks uniformly along its full length. Indeed, in whatever ratio the point M

divides the segment M_1M_2, its image M' will divide $M_1'M_2'$, in the same ratio, since parallel lines (in this case perpendiculars to the line a) cut lines intersecting them (in this case b and b') in proportional parts (figure 65).

The ellipse as the result of "contraction" of a circle. We consider a circle with center at the origin and radius a. By the theorem of Pythagoras its equation is $x^2 + \bar{y}^2 = a^2$, where we have written $\bar{y}$ instead of y, since y will be needed later. Let us see what this circle is contracted into if we "contract" the plane to the Ox-axis with coefficient b/a (figure 66). After this "contraction" the x-values of all points remain the same, but the $\bar{y}$-values become equal to $y = \bar{y}(b/a)$, i.e., $\bar{y} = (a/b)\,y$. Substituting for $\bar{y}$ in the above equation of the circle, we will have:

$$x^2 + \frac{a^2}{b^2}y^2 = a^2 \quad \text{or} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

as the equation, in the same coordinate system, of the curve obtained from the given circle by contraction to the Ox-axis. As we see, we obtain an ellipse. Thus we have proved that an ellipse is the result of a "contraction" of a circle.

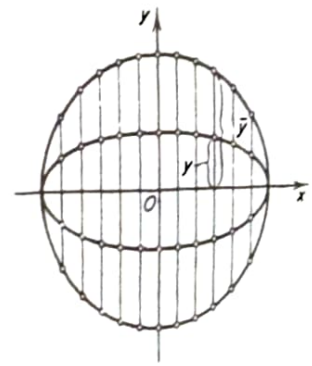


FIG. 66.

From the fact that an ellipse is a "contraction" of a circle, many properties of the ellipse follow directly. For example, the aforementioned property of diameters, namely that if parallel secants of an ellipse are given, then their midpoints lie on a straight line (see figure 12), can be shown in the following way. We perform the inverse expansion of the ellipse into the circle. Under this expansion parallel chords of the ellipse go into parallel chords of the circle, and their midpoints into the midpoints of these chords. But the midpoints of parallel chords of a circle lie on a diameter, i.e., on a straight line, and so that the midpoints of parallel chords of the ellipse also lie on a straight line. Namely, they lie on that line which is obtained from the diameter of the circle under the "contraction" which sends the circle into the ellipse.

Here is another application of the theory of "contraction." Since any vertical strip of the circle under its contraction to the Ox-axis does not change its width and its length is multiplied by b/a, the area of this strip after contraction is equal to its initial area multiplied by b/a, and since the area of the circle is equal to πa^2, the area of the corresponding ellipse is equal to $\pi a^2 (b/a) = \pi ab$.

Example of the solution of a more complicated problem. Let an ellipse be given and let it be required to find the triangle with smallest area circumscribed to this ellipse. We first solve the problem for a circle. We show that in the case of a circle, this is an equilateral triangle. Indeed, let the circumscribed triangle be nonequilateral; i.e., the smallest of its angles (denoted by B) is less than 60°, and the largest $C > 60°$. If then, without varying the angle A, we move side BC into the position B_0C_0 (figure 67) by shifting the vertex B toward A until one of the angles B_0 or C_0 becomes equal to 60°, we obtain a circumscribed triangle AB_0C_0

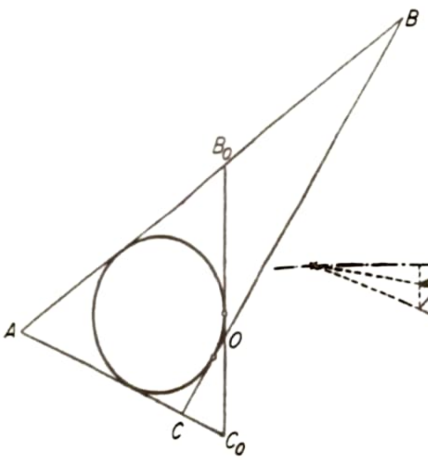


FIG. 67.

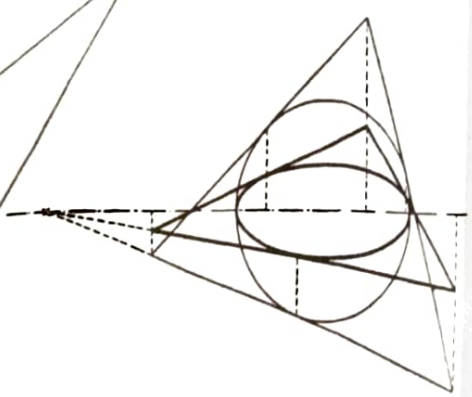


FIG. 68.

with smaller area, since here* $OC < OB$, $OC_0 \leqslant OB_0$ and therefore the discarded area OBB_0 is greater than the added one OCC_0. If the triangle so obtained is not equilateral, then by repeating the above procedure we reduce its area still further and arrive at an equilateral triangle. Hence, any nonequilateral triangle circumscribed to a given circle has a greater area than an equilateral one.

We now return to the ellipse. Let us make an expansion of it from the major axis, thereby converting it back into the circle from which it was obtained by "contraction." Under this expansion (figure 68): (1) all

* As can be easily shown.

triangles circumscribed to the ellipse are transformed into triangles circumscribing the resultant circle; (2) the areas of all figures, and in particular of these triangles are increased in one and the same ratio. From this we see that the triangles circumscribing the given ellipse with the smallest area will be those that are converted into equilateral triangles circumscribing the circle. There are infinitely many such triangles; each of them has its center of gravity at the center of the ellipse and the points of tangency are in the middle of its sides. Any of these triangles can easily be constructed (figure 68), starting from the aforementioned circle.

"Contractions" of the plane to a line are only a particular case of more general, so-called *affine* transformations of the plane.

General affine transformations. A pair of vectors e_1, e_2 starting from a common origin O and not lying on the same line will be called a coordinate "frame" of the plane. The coordinates of a point M of the plane relative to this frame Oe_1e_2 will then be numbers x, y such that in order to reach the point M from the origin O it is necessary to lay off from the point O x-times the vector e_1 and then y-times the vector e_2. This is a general Cartesian coordinate system of the plane. Analogously, a general Cartesian coordinate system can be introduced in space. The ordinary, so-called rectangular Cartesian coordinate system that we have made use of up to now corresponds to the particular case when the coordinate vectors e_1, e_2 are mutually perpendicular and their lengths are equal to the unit of measurement.

A general affine transformation of the plane is one under which a given net of equal parallelograms is transformed into another arbitrary net of equal parallelograms. More precisely, it is a transformation of the plane under which a given coordinate frame Oe_1e_2 is transformed into a certain other frame (generally speaking, with another "metric," i.e., with different lengths for the vectors e_1' and e_2' and a different angle between them) and an arbitrary point M is sent into the point M' having the same coordinates relative to the new frame as M had relative to the old (figure 69).

"Contraction" to the Ox-axis with coefficient k is a special case in which the rectangular frame $Oe_1'e_2'$ passes into the frame $Oe_1'ke_2'$.

It can easily be shown that under an affine transformation every straight line is sent into a straight line, parallel lines are mapped into parallel lines, and if a point divides a segment in a given ratio, then its image divides the image of this segment in the same ratio. Moreover, we can prove the remarkable theorem that any affine transformation of the plane can be obtained by performing a certain rigid motion of the plane onto itself, and then, in general, two "contractions" with different coefficients k_1 and k_2 to two mutually perpendicular lines.

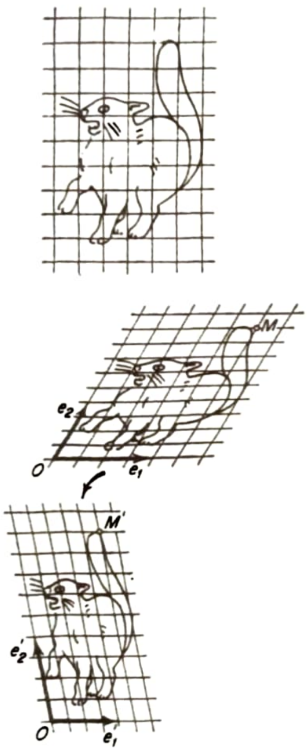


FIG. 69.

For the proof of this assertion, we consider all radii of some circle of the plane (figure 70). Let radius OA be the one which, after the transformation, turns out to be the shortest, and let it be mapped into $O'A'$. The perpendicular AB to OA is then transformed into $A'B'$, which must be perpendicular to $O'A'$, since if the perpendicular $O'C'$ were different from $O'A'$, then it would be the image of the oblique OC, and the image $O'D'$ of the radius OD would be a part of the perpendicular $O'C'$, i.e., shorter than the oblique $O'A'$, contrary to assumption.

The mutually perpendicular lines OA and AB are therefore mapped into mutually perpendicular lines $O'A'$ and $A'B'$. Consequently, the square net constructed on OA and AB is transformed into a net of equal rectangles (figure 71) and uniform "contractions" take place along the straight lines of this square net.

In a completely analogous way a general affine transformation of space can be defined as one under which a space coordinate frame $Oe_1e_2e_3$ is transformed into some other frame $O'e_1'e_2'e_3'$, generally speaking, with a different "metric," i.e., with unit segments of different lengths and with different angles between them, and a point M is sent into point M' having the same coordinates relative to the new frame as those of the point M relative to the old frame.

All the properties enumerated here also hold for affine transformations of space, except that in the last theorem there will be a rigid motion of space and then three "contractions" to three mutually perpendicular planes with certain coefficients k_1, k_2, k_3.

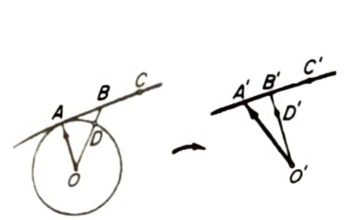


FIG. 70.

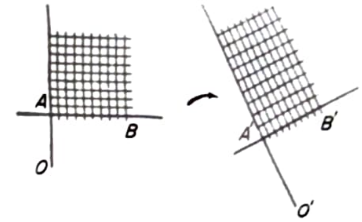


FIG. 71.

Applications of affine transformations. The most important applications of affine transformations are:

1. In the first place there is the application in geometry to solving problems concerning affine properties of figures, i.e., properties that are preserved under affine transformations. The theorem about the diameters of an ellipse and the problem of circumscribed triangles were examples. To solve such problems we make an affine transformation of the figure to some simpler one, for which we prove the desired property and then return to the original figure.

2. Second, there is the application in analytic geometry to the classification of second-order curves and surfaces. The main point is, as can be shown, that different ellipses are related to one another in the sense that one can be obtained from another by an affine transformation (the Latin word *affinis* means "related"). Also all hyperbolas are affine to one another, and so are all parabolas. But we cannot convert an ellipse into a parabola, or a hyperbola into a parabola, by an affine transformation, i.e., they are not affinely related to one another. It is natural to divide up all second-order curves into affine classes of curves, affinely related to one another. It turns out that the reduction of an equation to canonical form gives exactly this classification; i.e., there are nine affine classes of second-order curves. (We will not go into detail why imaginary ellipses and pairs of imaginary parallel lines belong to different affine classes. Properly speaking, neither in one case nor in the other are there any curves on the plane at all. The question here is really about algebraic properties of the equation itself.)

Similarly, the classification of second-order surfaces according to their canonical equations into 17 forms is the same as the affine classification.

Let us give a simple example of the application of the affine classification of second-order surfaces. We show that if we arbitrarily select in space three lines a, b, c such that (1) any two of them do not lie in the same plane (i.e., they are skew to each other) and (2) they are not all parallel to one and the same plane, then the set of all straight lines d of space, each of which simultaneously intersects all three given lines a, b, c (figure 72) constitutes the entire surface of a hyperboloid of one sheet.

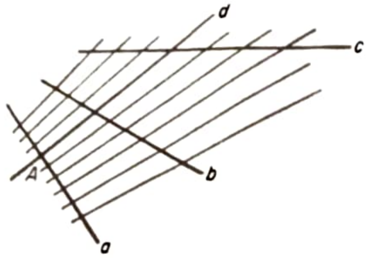


FIG. 72.

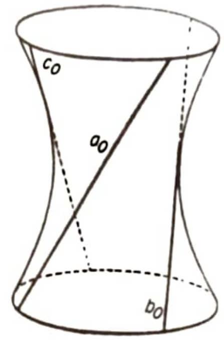


FIG. 73.

Let us explain more fully the set of lines d we are discussing here. Through an arbitrary point A of line a, we can pass a plane P containing the line b and a plane Q containing the line c. These planes P and Q intersect in a unique line d, which passes through the point A of line a and intersects lines b and c. Drawing all such lines d through arbitrary points of line a, we obtain the set of all those lines d of space each of which intersects all three given lines $a, b,$ and c. This collection of lines determines a surface. We note that any given hyperboloid of one sheet can be obtained in this way, since we only need to take for the lines $a, b,$ and c three distinct straight lines a_0, b_0, c_0 of one family (figure 73) and for the lines d all the straight lines of the other family. Conversely, let there be given three arbitrary pairwise skew lines of space a, b, c, not all parallel to one and the same plane. Then, as can be shown, these lines always form the three edges (without common points) of some parallelepiped (figure 74). After constructing such parallelepipeds for the given lines a, b, c and for three lines a_0, b_0, c_0 of one and the same family of an arbitrary hyperboloid of one sheet, we make an affine transformation of space that sends the parallelepiped a_0, b_0, c_0 into the parallelepiped a, b, c; obviously, this transformation maps the hyperboloid onto the surface in question. But according to the affine classification of second-order surfaces, the affine image of a hyperboloid of one sheet is again a hyperboloid of one sheet.

3. Third, there is the application to the theory of continuous transformations of continuous media, for example, in the theory of elasticity,

in the theory of electric or magnetic fields, etc. Very small elements of the given continuous medium transform "almost" affinely. We say, "in the small the transformation is linear" (we call a first-degree expression linear, and in the following section we see that in analytic geometry the formulas of affine transformations are of the first degree). This is evident in figure 75. On the lines of the large square net, their distortion or

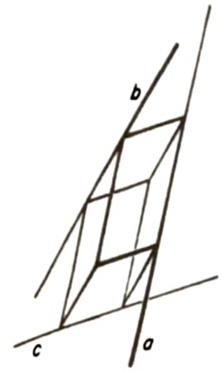


FIG. 74.

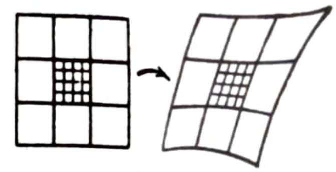


FIG. 75.

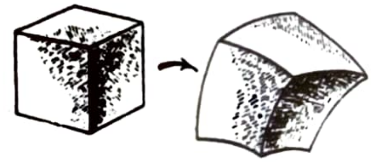


FIG. 76.

"fanning out" is clearly noticeable. But for a small piece of the very dense square net, all this shows itself very little, and the square net transforms "almost" into a net of equal parallelograms. A similar picture is obtained in space also (figure 76). By the fact that any affine transformation of space reduces to a motion and three mutually perpendicular "contractions," it follows that an element of a body under an elastic deformation first moves as a rigid body and then undergoes three mutually perpendicular "contractions."

Formulas of affine transformations. If the frame Oe_1e_2 is affinely transformed and $O'e_1'e_2'$ is its image, while the coordinates of the new origin O' relative to the old frame are ξ, η and the coordinates of the vectors e_1' and e_2' relative to the old frame are a_1, a_2 and b_1, b_2, respectively, then the formulas of the affine transformation, as can be easily seen from figure 77, are

$$x' = a_1x + b_1y + \xi,$$
$$y' = a_2x + b_2y + \eta$$

in the sense that if x, y are the coordinates of any point M relative to the old frame Oe_1e_2, then x', y' given by these formulas, are the coordinates relative to the same frame of the image M' of this point.

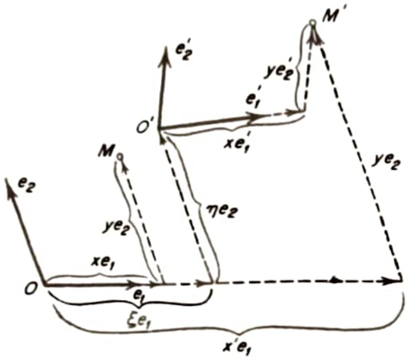


FIG. 77.

Indeed, let Oe_1e_2 be a frame before transformation, and $O'e'_1e'_2$ its image, while M is an arbitrary point of the plane and M' is its image. Then by the very definition of an affine transformation, if the coordinates of the point M relative to the frame Oe_1e_2 are x, y, then the coordinates of its image M' relative to the image $O'e'_1e'_2$ of this frame are exactly the same x, y.

Now consider a vector m' joining the origin O of the old frame to the image M' of the point M. Then $m' = x'e_1 + y'e_2$. But this vector is equal to a certain vector sum

$$m' = \xi e_1 + \eta e_2 + xe'_1 + ye'_2,$$

and the vectors e'_1 and e'_2 are

$$e'_1 = a_1e_1 + a_2e_2, \quad e'_2 = b_1e_1 + b_2e_2$$

so that

$$m' = \xi e_1 + \eta e_2 + a_1xe_1 + a_2xe_2 + b_1ye_1 + b_2ye_2$$

or

$$m' = (a_1x + b_1y + \xi)e_1 + (a_2x + b_2y + \eta)e_2.$$

Comparing this expression with the first expression for m' we obtain

$$\left.\begin{aligned} x' &= a_1x + b_1y + \xi, \\ y' &= a_2x + b_2y + \eta. \end{aligned}\right\} \tag{21}$$

The determinant

$$\Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1,$$

as can be shown, is not zero and is equal to the ratio of area of the parallelogram constructed on the vectors of the new frame to the area

of the same parallelogram constructed on the vectors of the old frame.

Analogous formulas are obtained for space

$$\left.\begin{aligned} x' &= a_1x + b_1y + c_1z + \xi, \\ y' &= a_2x + b_2y + c_2z + \eta, \\ z' &= a_3x + b_3y + c_3z + \zeta, \end{aligned}\right\} \tag{22}$$

where (ξ, η, ζ) are the coordinates of the origin O' of the transformed frame $O'e'_1e'_2e'_3$ and (a_1, a_2, a_3), (b_1, b_2, b_3), (c_1, c_2, c_3) are the coordinates of its vectors e'_1, e'_2, e'_3 relative to the old frame $Oe_1e_2e_3$.

The determinant*

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1$$

is not zero and is equal to the ratio of the volume of the parallelepiped formed by the vectors of the new frame to the volume of the parallelepiped formed by the vectors of the old frame.

Orthogonal transformations. Rigid motions of the plane onto itself or such motions plus a reflection about a line lying in the plane, are called *orthogonal transformations of the plane*, and rigid motions of space, or such motions plus a reflection of the space about one of its planes, are called *orthogonal transformations of space*. It is clear that orthogonal transformations are affine transformations under which the "metric" of the frame does not change, since it only undergoes a rigid motion, or else such a motion plus a reflection.

We will investigate orthogonal transformations by means of rectangular coordinates, i.e., when the vectors of the original frame are mutually perpendicular and have lengths equal to the unit of measurement. After an orthogonal transformation the vectors of the frame remain mutually perpendicular, i.e., their scalar product remains equal to zero and their lengths remain equal to unity. Therefore (see formula (14), this chapter) in the case of the plane, we have

$$a_1b_1 + a_2b_2 = 0, \quad a_1^2 + a_2^2 = 1, \quad b_1^2 + b_2^2 = 1, \tag{21'}$$

and in the case of space

$$\begin{aligned} &a_1b_1 + a_2b_2 + a_3b_3 = 0, \quad a_1^2 + a_2^2 + a_3^2 = 1, \\ &a_1c_1 + a_2c_2 + a_3c_3 = 0, \quad b_1^2 + b_2^2 + b_3^2 = 1, \\ &b_1c_1 + b_2c_2 + b_3c_3 = 0, \quad c_1^2 + c_2^2 + c_3^2 = 1. \end{aligned} \tag{22'}$$

* On determinants see Chapter XVI.

Hence, if the initial frame is taken to be rectangular, then formulas (21) give an orthogonal transformation if and only if the conditions (21′) of orthogonality are fulfilled, and formulas (22) give an orthogonal transformation of space if the conditions (22′) of orthogonality are satisfied. It can be shown that if $\Delta > 0$, we have a rigid motion, and if $\Delta < 0$ a rigid motion plus a reflection.

§12. Theory of Invariants

The concept of invariant.* Invariants of a second-degree equation with two variables. In the second half of the last century still another important new concept was introduced, that of invariant.

Consider, for example, a second-degree polynomial in two variables

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F. \tag{23}$$

If we regard x, y as rectangular coordinates and make a transformation to new rectangular axes, then after replacing x, y in (23) by their expressions in terms of the new coordinates x', y', removing parentheses, and reducing similar terms, we obtain a new transformed polynomial with different coefficients

$$A'x'^2 + 2B'x'y' + C'y'^2 + 2D'x' + 2E'y' + F'. \tag{24}$$

It turns out that there exist expressions formed from the coefficients which under this transformation do not change their numerical value, although the coefficients themselves change. Such an expression in A', B', C', D', E', F' has exactly the same numerical value as when it is formed with the A, B, C, D, E, F.

Expressions of this kind are called *invariants* of the polynomial (23) with respect to the group of orthogonal transformations (i.e., relative to transformations from one set of rectangular coordinates x, y to any other rectangular coordinates x', y').

Invariants of this sort, as it turns out, are

$$I_1 = A + C,$$

$$I_2 = \begin{vmatrix} A & B \\ B & C \end{vmatrix} = AC - B^2,$$

$$I_3 = \begin{vmatrix} A & B & D \\ B & C & E \\ D & E & F \end{vmatrix} = ACF + 2BDE - AE^2 - CD^2 - FB^2,$$

* *Invarians* in Latin means "unchanged."

i.e.,

$$A + C = A' + C', \quad AC - B^2 = A'C' - B'^2,$$

$$ACF + 2BDE - AE^2 - CD^2 - FB^2$$
$$= A'C'F' + 2B'D'E' - A'E'^2 - C'D'^2 - F'B'^2.$$

It is possible to prove the important theorem that any orthogonal invariant of the polynomial (23) can be expressed in terms of these three basic invariants.

If we equate the polynomial (23) to zero, we obtain an equation of some second-order curve. Any quantity, connected with this curve but not with its location in the plane, will clearly not depend on what coordinates its equation is written in, and therefore, when expressed in terms of the coefficients, it will be an orthogonal invariant of the polynomial (23), and thus it will be expressible in terms of the three basic invariants. Moreover, since under multiplication of all six coefficients of the equation by any given number t (different from zero) the curve represented by the equation remains the same, an expression of any property of the curve in terms of the I_1, I_2, I_3 must certainly be such that if the A, B, C, D, E, F in it are multiplied by t, the number t cancels out. The expression in question must be, as they say, homogeneous of degree zero relative to A, B, C, D, E, F.

Let us verify this by an example. For instance, let the equation

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0$$

represent an ellipse. Since the equation completely determines this ellipse, we can calculate from it (i.e., from its coefficients) all the basic quantities connected with the ellipse. For example, we can calculate its semiaxes a and b, i.e., we can express the semiaxes in terms of the coefficients. The expressions for these semiaxes will be invariants and therefore, expressible in terms of I_1, I_2, I_3. By reduction of the equation to canonical form and some subsequent calculation, the following rather complicated expressions for the semiaxes are obtained in terms of I_1, I_2, I_3:

$$\sqrt{\frac{2\,|I_3|}{|I_2|\,|I_1 \pm \sqrt{I_1^2 - 4I_2}\,|}},$$

which are homogeneous relative to A, B, C, D, E, F.

From this it is clear that the invariants I_1, I_2, I_3, themselves, being homogeneous but not of degree zero, do not have straightforward geometric meanings; they are algebraic entities.

It can be shown that the expression

$$K_1 = \begin{vmatrix} A & D \\ D & F \end{vmatrix} + \begin{vmatrix} C & E \\ E & F \end{vmatrix} = AF - D^2 + CF - E^2$$

can be varied by parallel translation but not by pure rotation of the given rectangular axes, and it is therefore called a *semi-invariant*.

As an example of an application of invariants and semi-invariants, we give Table 1, which if we calculate I_1, I_2, I_3 and K_1, allows us to determine directly from its equation the affine class of a second-order curve.

In Table 1 the necessary and sufficient conditions are given that an

Table 1

Criterion of the class	Name	Reduction equation	Canonical equation
$I_2 > 0,\ I_1 I_3 < 0$	ellipse		$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$
$I_2 > 0,\ I_1 I_3 > 0$	imaginary ellipse		$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$
$I_2 > 0,\ I_3 = 0$	point	$\lambda_1 x^2 + \lambda_2 y^2 + \frac{I_3}{I_2} = 0$	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$
$I_2 < 0,\ I_3 \neq 0$	hyperbola		$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$
$I_2 < 0,\ I_3 = 0$	pair of intersecting lines		$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$
$I_2 = 0,\ I_3 \neq 0$	parabola	$I_1 x^2 + 2\sqrt{-\frac{I_3}{I_1}}\,y = 0$	$x^2 = 2py$
$I_1 = 0, I_3 = 0,\ K_1 < 0$	pair of parallel lines		$x^2 = a^2$
$I_2 = 0,\ I_3 = 0,\ K_1 > 0$	pair of imaginary parallel lines	$I_1 x^2 + \frac{K_1}{I_1} = 0$	$x^2 = -a^2$
$I_2 = 0,\ I_3 = 0,\ K_1 = 0$	pair of coincident lines		$x^2 = 0$

equation of a second-order curve be reducible to one or another of the nine canonical forms ($I_1 I_3$ designates the product of I_1 and I_3).

Consider, for example, the equation $x^2 - 6xy + 5y^2 - 2x + 4y + 3 = 0$. We have $A = 1$, $B = -3$, $C = 5$, $D = -1$, $E = 2$, $F = 3$, so that $I_1 = 6$, $I_2 = -4$, $I_3 = -9$. The conditions of the 4th line of the table are satisfied: $I_2 < 0$, $I_3 \neq 0$, i.e. this is a hyperbola. Its semiaxes are equal to

$$\sqrt{\frac{2 \cdot 9}{4 \cdot |6 \pm \sqrt{36 + 16}\,|}} \approx 0.57 \quad \text{and} \quad 1.93.$$

The coefficients of the reduced equation (I), (II) and (III) are given in terms of invariants and semi-invariants as follows:

$$\lambda_1 x''^2 + \lambda_2 y''^2 + \frac{I_3}{I_2} = 0, \tag{I}$$

$$I_1 x''^2 + 2\sqrt{-\frac{I_3}{I_1}}\,y'' = 0, \tag{II}$$

$$I_1 x''^2 + \frac{K_1}{I_1} = 0, \tag{III}$$

where λ and λ_2 are the roots of the so-called characteristic quadratic equation

$$\lambda^2 - I_1 \lambda + I_2 = 0.$$

Formulas (I-III) allow a quick calculation of the semiaxes a and b of an ellipse and a hyperbola, the parameter p of an ellipse and the distance $2a$ between parallel lines. The formulas for semiaxes were given earlier. The parameter p is equal to

$$p = \sqrt{-\frac{I_3}{I_1^3}} \quad \text{and the distance} \quad 2a = 2\sqrt{-\frac{K_1}{I_1^2}}.$$

A completely analogous theory of invariants and semi-invariants, with a corresponding table for the determination of the affine class and the formulas of the coefficients of reduced equations, can be given for second-order surfaces in three-dimensional space.

It should be pointed out that so far we have been discussing only those invariants that are considered in analytic geometry for curves and surfaces of the second order. The concept of invariant, however, has a far broader meaning.

By an invariant of some object under study, relative to certain of its transformations, we mean any quantity numerical, vectorial, etc. connected with this object that does not vary under these transformations. In the previous problem the object is a second-degree polynomial with two variables (i.e., more precisely, the set of its coefficients), and the transformations are those of the polynomial obtained by the transition from one rectangular coordinate system to another.

Another example: The object is a given mass of a given gas under a given temperature. The transformations are changes in volume or pressure of this mass of gas. The invariant, according to the Boyle-Mariotte law, is the product of the volume by the pressure. We can speak of lengths of segments in space or the size of angles as invariants of the group of motions of space, of ratios in which a point divides a segment, or of ratios of areas, as the invariants of the group of affine transformations of space, etc.

Various invariants are particularly important in physics.

§13. Projective Geometry

Perspective projections. Artists began long ago to study the laws of perspectivity. This was necessary because a human being sees objects in perspective projection on the retina of the eye, in such a way that the

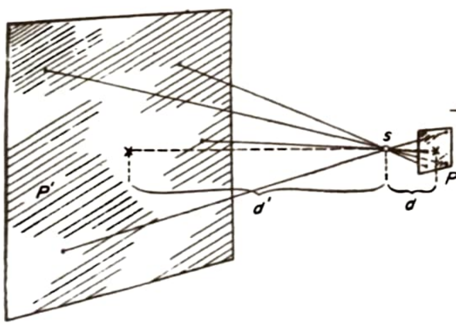


FIG. 78.

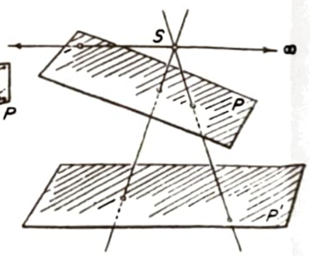


FIG. 79.

form and mutual location of objects are distorted in a characteristic manner. For example, telegraph poles in the distance look smaller and closer together, parallel tracks of a railway seem to converge, etc. We will not consider here space perspectivities, i.e., properties of perspective

projections of objects in space onto a plane but only the properties of perspective projections of a plane onto a plane.

Let us consider a photograph (for example, one frame of a moving picture film) P, a screen P', and between them a lens S (figure 78). Then, if the photograph is transparent and is illuminated from behind (if it is nontransparent, let it be illuminated from the front, i.e., from the side where the lens is), then the illuminated points of the photograph radiate beams of light, which are collected by the lens in such a way that they appear again on the screen P' in the form of points. We will assume that this projection takes place as if the points of the photograph P were projected on the screen P' on straight lines passing through the optical center S of the lens.

The situation will be a very simple one if the planes P and P' are parallel. In this case we will obviously obtain on the plane P' an undistorted image of everything that is on the plane P. This image will be smaller or larger than the original depending on whether the ratio $d':d$, where d and d' are the distances from the center of the lens to the planes P and P' respectively, is smaller or larger than 1.

The situation will be considerably more difficult if the planes P and P' are not parallel (figure 79). In this case, under projection through the point S not only the size of the figure changes but also its form is distorted. Parallel lines under such projection may become convergent, the ratio in which a point divides a segment may change, etc. In general, some of the relations that remain invariant under arbitrary affine transformation may change here.

This sort of projection takes place, for example, in aerial photography. The airplane oscillates in flight and therefore the photographic apparatus (figure 80a) rigidly attached to it is, in general, not oriented altogether vertically but at the moment of exposure is usually in an oblique position, i.e., we obtain a distorted image of the locality (which we assume to be plane).

How are we to correct this image? For this it is necessary to study the properties of projection of a plane P onto another plane Π (in general, the two planes are not parallel) by lines passing through a point S which is not on plane P nor on plane Π. Such projections are called *perspective projections*.

We will prove later the following important theorem.

Theorem. *If we have two perspective projections of a plane P on plane Π such that under both projections the points A, B, C, D of a quadruple of points of "general position" on the plane P (i.e., a quadruple in which no three of the points lie on one line), are projected into the same points A',*

B', C', D' respectively of plane II, *then all points of plane P are also projected under both projections into the same points of plane* II.

In other words, the result of a perspective projection is completely determined if it is known into which points this projection sends the points of an arbitrary quadruple of points of general position in the figure to be projected.

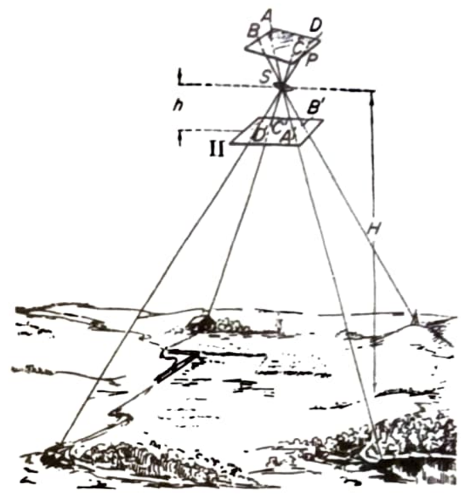
This is the so-called uniqueness theorem of the theory of projective transformations or the fundamental theorem of plane perspectivity.

Application of the fundamental theorem of plane perspectivity in aerial photography. Let us show how this theorem provides a suitable method for correcting this image in photography.

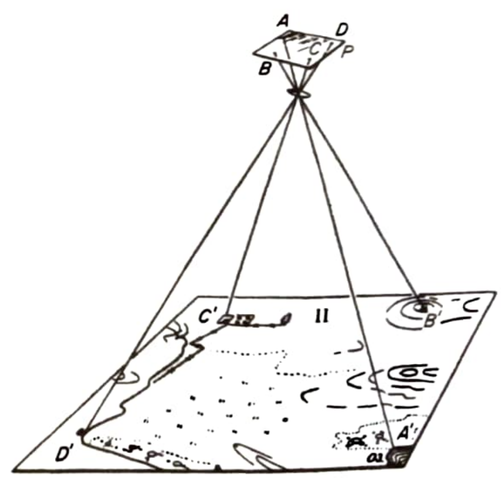
If at the moment of aerial exposure, we imagine a horizontal screen II placed at a distance h below the center S of the lens (figure 80a), then the projection onto this screen through the center S of the image recorded on the photographic plate P will obviously not be distorted but will be similar to the horizontal locality with a scale h:H, where H is the height of the airplane at the moment of exposure. In order to correct the image received on the photograph P so as to convert it into an undistorted image, we treat it as follows. The developed photograph P is placed in a projecting apparatus resting on a special tripod on which, by means of adjustable screws, the apparatus can be moved closer to the screen II or farther from it and can be rotated in every way.

To the screen II (figure 80b) we attach a topographical map of the locality made by measurements on the surface of the Earth (not a detailed map, since the details of interest to us are to be provided by the aerial photograph). On this map attached to the screen II we select four points A', B', C', D' that can be found easily on the photograph also (for example, an intersection of roads, a corner of a house, etc.), and at the corresponding points A, B, C, D of the picture P we pierce the film with a needle. We then place a projection lamp behind the plate P in such a position that the picture is projected onto the screen II through a lens S of the supporting apparatus. By using the adjustable screws we arrange that the light beams from the pinholes fall on the corresponding points A', B', C', D' of the map attached to the screen. After this has been done, we replace the topographical map by a plateholder with a photographic plate and then, without changing the settings of the screws, we photograph the image projected on the screen II of the picture P taken from the airplane.

By the theorem stated previously, we thereby obtain a true (i.e., similar to the locality) and not a distorted map of the photographed region.



(a)



(b)

FIG. 80.

We now pass to the presentation of the theory necessary for proving the fundamental theorem.

The projective plane. The totality of *all* lines and planes of space passing through a given point S of the space is called the projecting bundle of lines and planes with center S. If this bundle is intersected by a plane P, not passing through the center, then to every point of the plane P will correspond a line of the bundle intersecting the plane P in this point, and to each line of the plane P will correspond that plane of the bundle which intersects the plane P along this line. However, we do not in this way establish a one-to-one mapping from the set of lines and planes of the bundle on the set of points and lines of the plane P. As a matter of fact, the lines and planes of the bundle which are parallel to the plane P do not in this sense correspond to any points or lines of the plane P, since they do not intersect it. Nevertheless, we *agree* to say that these lines of the bundle intersect the plane P but in its ideal (or infinitely distant) points, lying in the corresponding directions, and that such a plane of the bundle intersects the plane P along an ideal (or infinitely distant) line. The plane P, complemented by these ideal points and ideal line, is called a *complemented* or *projective plane*. We will denote it by P^*. The sets of lines and planes of the bundle S are then mapped one-to-one onto the sets of points (real and ideal) and lines (real and ideal) of this projective plane P^*.

Hence, we *agree* to say that a point (real or ideal) lies on a line (real or ideal) of the projective plane P^* if the corresponding line of the bundle lies in the corresponding plane of the bundle. From this point of view, any two lines of the projective plane intersect (in a real or ideal point), since any two planes of the bundle intersect along some line of the bundle. It follows from this, among other things, that the ideal line consists simply of the set of all ideal points.

In essence, the complementation of the plane by its ideal elements means that we use this plane as a cross section to study the bundle of all lines and planes passing through one point.

Projective mappings; the fundamental theorem. By a *projective mapping* we understand such a mapping of a projective plane P^* onto some other projective plane $P^{*\prime}$ (which can also coincide with the plane P^*, in which case we speak of a projective transformation of the plane P^*), which, first of all, is pointwise a one-to-one mapping, and second, is such that collinear sets of points of the plane P^* go into collinear sets of points of the plane $P^{*\prime}$, and conversely. (Here, by points and lines we always understand real as well as ideal points and lines.)

It is clear that two arbitrary perspective projections of one and the same plane P^* onto a plane Π^* may be obtained from each other by projective transformations.

In fact, 1, their points (real or ideal) are in one-to-one correspondence with the points (real or ideal) of the protective plane P^* and consequently with each other, and 2, collinear points of the first projection correspond to collinear points of the plane P^* and consequently also of the second projection, and conversely. Therefore, the aforementioned theorem of the theory of perspectivities is a direct consequence of the following theorem about projective transformations: if under a projective transformation of the plane Π^*, four of its points A, B, C, D, forming a quadruple of general position remain fixed, then all of its points remain fixed.

Let us outline the idea of the proof of this theorem by means of the so-called Möbius net.

We note that (1) if under a projective transformation two points remain fixed, then the line that passes through them is mapped into itself, and (2) if two lines are mapped into themselves, then the point of their intersection remains fixed. Therefore, from the fact that the points A, B, C, D of the plane Π^* remain fixed, it follows in turn that also the points E, F, G, H, K, L, etc. remain fixed (figure 81). The con-

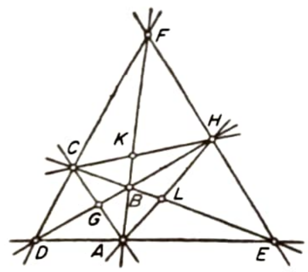


FIG. 81.

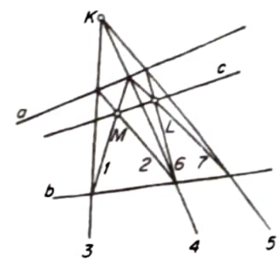


FIG. 82.

struction of such points can be continued by joining the points already obtained. This is the so-called Möbius net. By continuing its construction, we can find points as densely placed as we like. It can be shown that the set of these nodes everywhere densely covers the whole plane. Therefore, if we further assume the continuity of a projective transformation (which in fact, already follows from its definition, although the proof of this fact is not easy), the result is that if under a projective transformation of the plane Π^* the points A, B, C, D remain fixed, then all the points of the plane Π^* remain fixed.

Projective geometry. By *two-dimensional projective geometry* we mean the totality of theorems about those properties of figures in the projective plane, i.e., the ordinary plane complemented by ideal elements, which do not change under arbitrary projective transformations.

Here is an example of a problem of projective geometry. Given two lines a and b and a point M (figure 82), the problem is to construct the line c passing through the point M and through the point of intersection of lines a and b, not using this point of intersection (as may be necessary, if this point is very distant). If through the point M we draw the two secants 1 and 2 and then the lines 3 and 4 through the points of their intersection with lines a and b, we obtain the point K. Let us draw through it line 5 and secants 6 and 7; then it can be shown that the line c passing through point L of intersection of lines 6 and 7 and point M, is the desired line.

From the theory of conic sections, it follows (figure 83) that the ellipse, hyperbola and parabola are perspective projections of one another, and moreover all of them are perspective projections of the circle.

If we regard perspective projections as projective transformations of projective planes P^* and $P^{*\prime}$ one onto the other, then by superposing these planes we obtain the result that all ellipses, hyperbolas, and parabolas are projective transformations of the circle. The difference in them

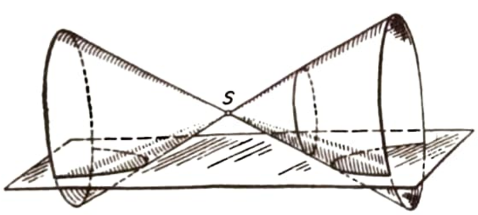


FIG. 83.

FIG. 84.

is that projective images of the circle under transformations in which a line not intersecting the circle is mapped into the infinitely distant line are ellipses; on the other hand, if a line tangent to the circle is mapped into the infinitely distant line, then a parabola is obtained, and if a secant, then a hyperbola (figure 84).

The notation of projective transformations in formulas. If on the plane P^* we take an ordinary Cartesian coordinate system, then, as can be shown, the formulas for projective transformations of the plane are as follows

$$x' = \frac{a_1x + b_1y + c_1}{a_3x + b_3y + c_3}, \quad y' = \frac{a_2x + b_2y + c_2}{a_3x + b_3y + c_3},$$

where the determinant

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \neq 0,$$

and conversely.

If for some point (x, y) the denominators are equal to zero, this means that its image (x', y') is an ideal (infinitely distant) point. The equation

$$a_3x + b_3y + c_3 = 0$$

represents the line which under the given projective transformation goes into the ideal (infinitely distant) line.

§14. Lorentz Transformations

The derivation of the formulas of the Lorentz transformation for motion on a straight line and in the plane from the condition of the constancy of the speed of light. At the very end of the 19th century a fundamental contradiction was discovered in physics. Michelson's well-known experiment, in which the speed of light (which is about 300,000 km/sec) was measured in the direction of motion of the Earth along its orbit around the Sun (the speed of the Earth is about 30 km/sec) and perpendicular to this direction, showed irrefutably that all moving bodies in nature, even if they are moving in a vacuum, are contracted in the direction of motion. The theory of this contraction was investigated in detail by the Dutch physicist, Lorentz. He showed that this contraction is greater as the speed of the moving body gets closer to the speed of light in a vacuum, and at a speed equal to the speed of light the contraction becomes infinite. Lorentz derived the formulas for this contraction. But shortly afterwards the physicist Einstein introduced into this problem a completely different point of view, to which Poincaré was already close. Einstein argued as follows. If we assume that for the propagation of light, as for ordinary motion of a material body, Galileo's law of composition of velocities is valid, then the speed of light is $c' = c + v$, where v is the speed of the observer moving toward the source of the light, and c is the speed of light for a stationary observer. From Michelson's experiment it follows that $c' = c$. The law $c' = c + v$ is based on the transformation

$$\begin{aligned} x' &= x + v_x t, \\ t' &= t, \end{aligned} \tag{25}$$

connecting the coordinate x of a point relative to a coordinate system I with its coordinate x' relative to a coordinate system II which has its axes parallel to the axes of system I and which moves parallel to the Ox-axis with velocity v_x relative to system I. Clearly, these are the formulas, as Einstein says, that must be changed.

It can be shown, as was recently done, for example, by A. D. Aleksandrov, that from the equality of the speed of light in both coordinate systems x, y, z, t and x', y', z', t' it already follows that the formulas of transformation from coordinates x, y, z, t to coordinates x', y', z', t' are linear and homogeneous, i.e., have the form

$$\begin{aligned} x' &= a_1x + b_1y + c_1z + d_1t, \\ y' &= a_2x + b_2y + c_2z + d_2t, \\ z' &= a_3x + b_3y + c_3z + d_3t, \\ t' &= a_4x + b_4y + c_4z + d_4t. \end{aligned} \tag{26}$$

From other considerations one can show that their determinant* is equal to unity.

If a point in system I moves rectilinearly and uniformly in an arbitrary given direction with the speed of light c, then $x = v_x t$, $y = v_y t$, $z = v_z t$ and $v_x^2 + v_y^2 + v_z^2 = c^2$, from which

$$x^2 + y^2 + z^2 - c^2t^2 = 0. \tag{27}$$

But according to Michelson's experiment this point in system II also necessarily moves with the same speed of light c, so that it is also necessary that

$$x'^2 + y'^2 + z'^2 - c^2t'^2 = 0.$$

Consequently the formulas (26) are not just arbitrary transformations which are linear, homogeneous, and with determinant equal to 1, but must at the same time satisfy the condition that if the coordinates x, y, z, t are such that

$$x^2 + y^2 + z^2 - c^2t^2 = 0,$$

then the transformed coordinates x', y', z', t' must also satisfy this equation. Such transformations (26) are called *Lorentz transformations*.

* See Chapter XVI.

Let us first consider the simplest case, when the point moves along the Ox-axis. In this case formulas (26) have the form

$$\begin{aligned} x' &= a_1x + d_1t, \\ t' &= a_2x + d_2t, \end{aligned} \tag{26'}$$

and equation (27)

$$x^2 - c^2t^2 = 0. \tag{27'}$$

Let us introduce the notation $ct = u$, when formulas (26′) and equation (27′) take the form

$$\begin{aligned} x' &= a_1x + \frac{d_1}{c}u, \\ u' &= a_2cx + \frac{d_2c}{c}u \end{aligned} \tag{26$_1$}$$

and

$$x^2 - u^2 = 0.$$

Let us find the explicit forms of formulas (26$_1$). Consider x and u as a Cartesian rectangular system in the plane, i.e., consider the problem geometrically; then we may regard formulas (26$_1$) as those of an affine transformation of the plane Oxu (whose determinant, as was shown is equal to 1). We will denote this transformation by L. If, as we assume, $x^2 - u^2 = 0$ implies $x'^2 + u'^2 = 0$, then this transformation translates the intersecting straight lines

$$x^2 - u^2 = 0$$

into themselves. The transformation L is therefore a combination of a contraction and expansion with identical coefficients τ along these lines.

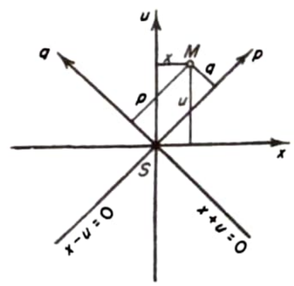


FIG. 85.

From figure 85 we obtain

$$x = \frac{p}{\sqrt{2}} - \frac{q}{\sqrt{2}},$$

$$u = \frac{p}{\sqrt{2}} + \frac{q}{\sqrt{2}}.$$

But after the transformation L, the numbers p and q will go into $p' = p/\tau$ and $q' = q\tau$, so that

$$x'\sqrt{2} = \frac{p}{\tau} - q\tau,$$

$$u'\sqrt{2} = \frac{p}{\tau} + q\tau.$$

Expressing p and q in terms of x and u from the first pair of equations, substituting into the second and simplifying, we obtain

$$x' = \frac{x - \dfrac{\tau^2 - 1}{\tau^2 + 1}ct}{\dfrac{2\tau}{\tau^2 + 1}}, \quad t' = \frac{t - \dfrac{1}{c}\dfrac{\tau^2 - 1}{\tau^2 + 1}x}{\dfrac{2\tau}{\tau^2 + 1}},$$

or, setting $(\tau^2 - 1)/(\tau^2 + 1)c = v$, we have

$$x' = \frac{x - vt}{\sqrt{1 - (v/c)^2}}, \quad t = \frac{t - (vx/c^2)}{\sqrt{1 - (v/c)^2}},$$

which are the famous *Lorentz formulas*.

In particular, if we take $x = 0$, i.e., if we consider the motion of the origin of coordinate system I, we obtain

$$x' = \frac{-vt}{\sqrt{1 - (v/c)^2}}, \quad t' = \frac{t}{\sqrt{1 - (v/c)^2}},$$

or $x' = -vt'$, from which obviously v is the speed of motion of coordinate system II relative to system I.

Suppose, for example, that we are given two points on the Ox-axis with coordinates x_1 and x_2 relative to system I, so that the distance between them relative to system I is $r = |x_1 - x_2|$. Let us see what the distance between them is for an observer attached to system II. We have

$$x_1' = \frac{x_1 - vt}{\sqrt{1 - (v/c)^2}}, \quad x_2' = \frac{x_2 - vt}{\sqrt{1 - (v/c)^2}},$$

from which

$$r' = |x_1' - x_2'| = \frac{|x_1 - x_2|}{\sqrt{1 - (v/c)^2}}.$$

The factor $\sqrt{1 - (v/c)^2}$ is exactly the coefficient of the Lorentz contraction. Since c is very large, this coefficient is very close to 1 for moderately large v, and therefore the contraction is not significant. But such elementary particles as electrons or positrons often move with velocities comparable to the speed of light, and therefore in studying their motion it is necessary to take this contraction into account, or, as they say, to consider the relativistic effect.

We pass now to the case next in complexity, namely, when the point moves in the Oxy-plane. For this case the transformations (26) will have the form

$$\begin{aligned} x' &= a_1x + b_1y + d_1t, \\ y' &= a_2x + b_2y + d_2t, \\ t' &= a_3x + b_3y + d_3t, \end{aligned} \tag{26''}$$

where

$$\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix} = 1,$$

and equation (27) will be

$$x^2 + y^2 - c^2t^2 = 0. \tag{27''}$$

These are the Lorentz formulas for motion in the Oxy-plane.

Again we put $ct = u$. Then transformations (26″) can be rewritten as

$$\begin{aligned} x' &= a_1x + b_1y + \frac{d_1}{c}u, \\ y' &= a_2x + b_2y + \frac{d_2}{c}u, \\ u' &= a_3cx + b_3cy + \frac{d_3c}{c}u, \end{aligned} \tag{26$_2$}$$

where the determinant will again be equal to one, and equation (27″) will assume the simpler form

$$x^2 + y^2 - u^2 = 0. \tag{27$_2$}$$

We will regard x, y, u as the Cartesian rectangular coordinates of a point in ordinary three-dimensional space and will consider formulas (26$_2$) as those of affine transformations of this space. Equation (27$_2$) represents a straight circular cone K with an angle of 90° at the vertex (figure 86).

From the point of view of this geometric interpretation (we call it geometric because here we regard $u = ct$ simply as a space coordinate) of a Lorentz transformation, the set of motions in the plane is identical with the set of all equi-affine (i.e., affine and volume-preserving) transformations of the space which map the cone K onto itself.

Let us consider some special Lorentz transformations.

1. It is clear that any simple rigid rotation about the axis of the cone K through an angle ω is an equi-affine transformation of space, mapping the cone K into itself, i.e., it is a special Lorentz transformation. We will denote it by ω.

2. Reflections of the space in an arbitrary plane π passing through the axis of the cone K are clearly also Lorentz transformations. We will denote them by π.

3. Finally, let us consider the following transformation (figure 87). Let v and w be any pair of opposite generators of the cone, and let P and Q be the planes tangent to the cone along these generators. These

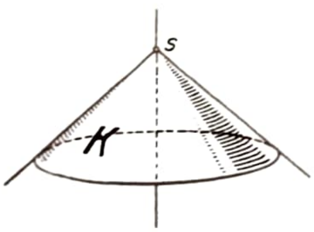


FIG. 86.

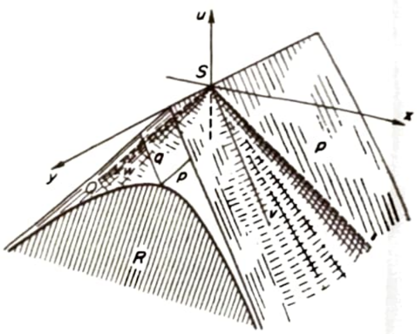


FIG. 87.

planes are mutually perpendicular. Let us make a contraction of the space to the plane P and an expansion of it with the same coefficient from the plane Q, or conversely. For example, we contract the space by a factor of three to the plane P and expand it also by a factor of three from the plane Q. Such a transformation of space is clearly also affine and preserves all volumes. We will denote it by L. We show that this transformation maps the cone K into itself. Because the cone K has the axis u as its axis of revolution, any figure can be rotated in such a way that the generators v and w lie, for example, in the plane Sxu. Therefore it is sufficient to carry out the proof for this case.

For the proof we intersect the cone K by an arbitrary plane R parallel to the Sxu plane. The equation of this plane is $y = b$, where b is a

constant. Substituting this value in the equation of the cone K, we obtain

$$x^2 - u^2 = -b^2.$$

This is the equation of a hyperbola for which the lines of intersection of the plane R with the planes P and Q are exactly the asymptotes. But since for a point of such a hyperbola it is characteristic that the product of distances p and q to the asymptotes, i.e., to the planes P and Q, is constant, under transformation L all points of this hyperbola remain on the same hyperbola, and the hyperbola is mapped onto itself. But the whole surface of the cone K consists of such hyperbolas, and therefore under the transformation L of the space the cone K is sent into itself. This transformation L is therefore also a Lorentz transformation.

Since under affine transformations straight lines go into straight lines, and intersecting lines go into intersecting lines, therefore a bundle S of straight lines under any Lorentz transformation is mapped one-to-one onto itself. Moreover, under affine transformations of space all planes go into planes, so that under these transformations of the bundle S onto itself a projective transformation of the bundle is obtained. If we intersect this bundle by a plane II perpendicular to the axis of the cone K, which as a whole is not altered by the given Lorentz transformation of space, and extend this plane to the projective plane II* and then trace the points of intersection of the lines of the bundle S with the plane II*, we have the result that the Lorentz transformations of the bundle will simultaneously produce projective transformations Λ of the plane II* and these latter will transform the circle α, in which the plane II* intersects the interior part of the cone K, into itself. To analyze the properties of Lorentz transformations, it is easier to examine these projective transformations Λ of the circle α into itself.

Projective transformations of a circle into itself. A point, a ray or half line issuing from it, and one of the half planes cut off by the entire line will be called a "frame" of the plane II* (not to be confused with a coordinate frame, §11). We show (figure 88) that if we take two arbitrary frames M and M' containing interior points of the circle α, then by means of the transformations L, ω, π we can send one of these frames into the other. For this it is sufficient to make the transformations $\Lambda = L_1 \cdot \omega \cdot L_2^{-1}$ (or else $\Lambda = L_1 \cdot \omega \cdot \pi \cdot L_2^{-1}$). The transformation L_1 sends the first frame M to the center O of the circle α, the transformation ω rotates it as necessary, and finally the transformation L_2^{-1} brings it into coincidence with the second frame M'.

Let us show, in addition, that there is only one transformation Λ which translates a given frame M into a given frame M'. In order to do this we observe first that if there were two transformations Λ_1 and Λ_2 sending frame M into frame M', then the transformation $\Lambda = \Lambda_1\Lambda_2^{-1}$ would not be the identity transformation Λ and would send frame M into itself. Therefore, it is sufficient to show that if a transformation Λ sends frame M into itself, then it is the identity, i.e., leaves all points of the plane of circle α fixed.

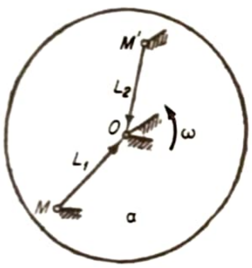


FIG. 88.

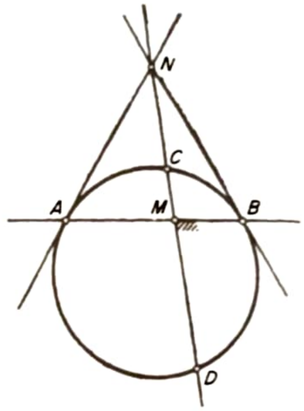


FIG. 89.

Let us show this. Suppose that the transformation Λ sends frame M into itself (figure 89). Then it maps the line AB of this frame into itself, but since it sends the circumference of the circle α into itself, it therefore leaves points A and B fixed, or else interchanges them. The latter, however, is impossible, since the half line of the frame is mapped into itself. Let us draw the tangents at points A and B to the circle α. They are mapped into themselves, since if such a tangent were mapped into a secant $A\bar{A}$, then the inverse transformation would send the different points A and $\bar{A}$ of the circle α into the one point A. But the Λ are projective transformations, and consequently one-to-one. Since under the transformations Λ these tangents go into themselves, therefore the point N of their intersection remains fixed, and consequently the line MN is mapped into itself. From the fact that the half line of the frame M is mapped into itself, we conclude as above that points C and D are not interchanged, but remain fixed. Hence, under the given projective transformation Λ of the projective plane Π^* four of its points A, B, C, D, no three of which lie on the same line, remain fixed. According to the uniqueness theorem of projective transformations this is the identity transformation.

Later, in §5 of Chapter XVII it will be shown that by using these properties of the Lorentz group, it is easy to construct a model of

Lobačevskiĭ's plane geometry, and if we consider Lorentz transformations for the general case of motion of a point in space, then we can do the same for Lobačevskiĭ's space geometry, and thereby prove its consistency.

We see that the theory of Lorentz transformations, projective geometry and the theory of perspectivity and non-Euclidean geometry are closely related to one another. It turns out that there is still another theory that is also closely related to them, namely, the so-called conformal transformations in the theory of functions of a complex variable, which solve such important problems of mathematical physics as the distribution of temperature in a heated plate, the flow of air around the wing of an airplane, the distribution of charge in a plane electrostatic field, the problems of elasticity in the plane, and many others.

Conclusion

Analytic geometry is an absolutely indispensable method for the investigation of other branches of mathematics, physics, and other natural sciences. Therefore it is studied not only at universities but in all technical higher institutions of learning, and also in some vocational schools. It is also a question of whether we should not include a fairly detailed treatment of the elements of analytic geometry in high school courses.

Various coordinates. The essential elements of the concept of analytic geometry, as we have seen, are the coordinate method and the investigation of equations connecting these coordinates. Besides Cartesian coordinates, other different ones can be considered. For example, in the plane, we can choose a point P (the so-called pole) and a ray originating from it (the polar axis) and determine the position of a point M by the length ρ of the polar radius from the pole to the point and the value ω of the angle made by this radius with the polar axis (figure 90).

In particular, the ellipse, hyperbola, or parabola, if for the pole we take

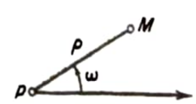


FIG. 90.

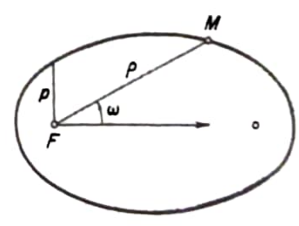


FIG. 91.

a focus, and for the polar axis the ray passing from the focus along the axis of symmetry to the side opposite the nearer vertex (figure 91), have one and the same equation

$$\rho = \frac{p}{1 - \epsilon \cos \omega},$$

where ϵ is the eccentricity of the curve, and p is its so-called parameter. This equation is of a great importance in astronomy. For it was with its help that the result was derived, from the law of inertia and the law of universal gravitation, that the planets revolve about the Sun in ellipses.

The geographical coordinates, latitude and longitude, by which the position of a point is given on a sphere, are well known.

Analogously, we can take a coordinate network on an arbitrary surface, as is done in differential geometry (see Chapter VII), etc.

Many-dimensional and infinite-dimensional analytic geometry; algebraic geometry. It would seem that in the 19th century analytic geometry underwent such an immense development, described earlier in a general way, and produced so many ideas, that it would have necessarily exhausted itself, but this is not so. In very recent times, two new, extensive branches of mathematics have been rapidly developed and have extended the concepts of analytic geometry, namely so-called *functional analysis* and general *algebraic geometry*. It is true that both of these only halfway represent a straightforward continuation of classical analytic geometry: Much of functional analysis is analysis, and in algebraic geometry there is more than a little of the theory of functions and of topology.

Let us explain what we mean. In the middle of the last century mathematicians had already begun to consider four-dimensional and general n-dimensional analytic geometry, i.e., to study those questions of algebra that are straightforward generalizations of algebraic questions of the kind involved in two- and three-dimensional analytic geometry, to the case when there are four or n unknowns. At the very end of the 19th century a series of outstanding analysts came to the idea that for the purposes of analysis and mathematical physics it is significant to consider infinite-dimensional analytic geometry.

At first glance it may seem that n-dimensional or even four-dimensional spaces seem like farfetched mathematical fictions, then the same can also be said about an infinite-dimensional space. But it is not really so. The arguments concerning an infinite-dimensional space are not at all difficult. They now constitute an important branch of mathematics, functional analysis (see Chapter XIX).

It is curious that infinite-dimensional analytic geometry has most important practical applications and plays a fundamental role in contemporary physics.

As to algebraic geometry, it is a more immediate continuation of ordinary analytic geometry, which is itself only a part of algebraic geometry. Algebraic geometry can be regarded as that part of mathematics which is occupied with curves, surfaces, and hypersurfaces, represented in Cartesian coordinates by algebraic equations of not only first and second degree, but also of higher degrees. It turns out that in these investigations it is advantageous to consider not only real but also complex coordinates, i.e., to consider everything in a so-called complex space. The most important results in this domain were obtained in the last century by Riemann. As a brilliant example of theorems about higher order curves, we point out a remarkably general result of I. G. Petrovskiĭ about the number of ovals into which an nth-order curve can be decomposed. Petrovskiĭ showed that if p is the number of such ovals which do not lie at all in other ovals, or lie in an even number of ovals, and m is the number of those ovals which lie in an odd number of ovals, and if we consider only curves whose component ovals neither intersect themselves nor each other (figure 92), then

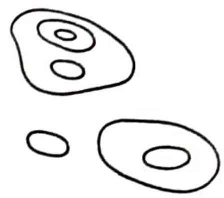


FIG. 92.

$$p - m \leqslant \frac{3n^2 - 6n}{8} + 1,$$

where n is the order of the curve, i.e., the degree of the equation by which the curve is represented.

This result is the more important as up to then almost nothing had been known about the general form of a higher order curve. It is no doubt one of the most important recent general theorems in analytic geometry.

Suggested Reading

A. A. Albert, *Solid analytic geometry*, McGraw-Hill, New York, 1949.

J. L. Coolidge, *A history of the conic sections and quadratic surfaces*, Oxford University Press, Oxford, 1945.

H. S. M. Coxeter, *The real projective plane*, 2nd ed., Cambridge University Press, New York, 1960.

L. P. Eisenhart, *Coordinate geometry*, Dover, New York, 1960.
A. Jaeger, *Introduction to analytic geometry and linear algebra*, Holt, Rinehart and Winston, New York, 1960.
D. J. Struik, *Lectures on analytic and projective geometry*, Addison-Wesley, Cambridge, Mass., 1953.