CHAPTER VII

CURVES AND SURFACES

§1. Topics and Methods in the Theory of Curves and Surfaces

In a school course, geometry involves only the simplest curves: straight lines, broken lines, and circumferences and arcs of circles; and as for surfaces, merely planes, surfaces of polyhedra, spheres, cones, and cylinders. In more extended courses other curves are considered, chiefly the conic sections: ellipses, parabolas, and hyperbolas. But the study of an arbitrary curve or surface is completely alien to elementary geometry. At first sight it is even unclear how any general properties could be selected for investigation when we are speaking of arbitrary curves and surfaces. Yet such an investigation is completely natural and necessary.

In every kind of practical activity and experience of nature, we constantly encounter curves and surfaces of widely different forms. The path of a planet in space, of a ship at sea, or of a projectile in the air, the track of a chisel on metal, of a wheel on the road, of a pen on the tape of a recording device, the shape of a camshaft governing the valves of a motor, the contours of an artistic design, the form of a dangling rope, the shape of a spiral spring coiled for some specific purpose, such examples are endless. The surfaces of various objects, thin shells, cisterns, the framework of an airplane, casings, sheetlike materials, provide an endless diversity of surfaces. Methods for the processing of products, the optical properties of various objects, the streamlining of bodies, the rigidity or deformability of thin shells, these and many other features depend to a great extent on the geometric form of the surfaces of objects.

Of course, the gouge left by a chisel on metal is not a mathematical curve. A cistern, even with thin walls, is not a mathematical surface. But to a first approximation, which is sufficient for the study of many questions, actual objects may be represented mathematically by curves and surfaces.

In introducing the concept of a mathematical curve, we disregard all the reasons why we cannot decrease the thickness without limit. By means of this abstract concept, we succeed in representing those (completely concrete) properties of an object that are preserved when its thickness and breadth are decreased in comparison with its length.

Similarly, if we disregard the limitations on our ability to decrease the thickness of a shell or to determine precisely the actual boundaries of a given object, we are led to the concept of a mathematical surface. We will not give a rigorous description of these well-known concepts but will only remark that the exact mathematical definitions are not simple and belong to topology.

Finally, an important source of interest in various curves and surfaces has been the development of mathematical analysis. It is sufficient to remember, for example, that a curve is the geometric representation of a function, which is the most important concept of analysis. Moreover, every one is familiar with graphs quite apart from any study of analysis.

In elementary geometry as created by the ancient Greeks, there was nothing about arbitrary curves or surfaces, but even in elementary analytic geometry we are accustomed to say "every curve is represented by an equation" or "every equation in the two variables x and y represents a curve in the coordinate plane." Similarly the coordinates of surfaces are given by the equations $z = f(x, y)$ or $F(x, y, z) = 0$, and in general the coordinate method, by establishing a close connection between elementary geometry and analysis, enables us to define many different curves and surfaces.

But analytic geometry, being restricted to the methods of algebra and elementary geometry, goes no further than the investigation of certain specific types of figures. The study of arbitrary curves and surfaces represents a new branch of mathematics, known as *differential geometry*.

It must be admitted at once that differential geometry imposes on its curves and surfaces certain conditions arising from the methods of analysis. However, this is not an essential limitation on the diversity of the allowable curves and surfaces, since in the great majority of cases they are capable of representing actual objects with the necessary degree of precision. The name "differential geometry" itself gives an indication of the methods of the theory; its basic tool is the differential calculus and it primarily investigates the "differential" properties of the curves

and surfaces, i.e., their properties "at a point."* Thus, the direction of a curve at a point is determined by its tangent at that point and the amount by which it twists is described by its curvature (the exact definition of this term will be given below). Differential geometry investigates the properties of small segments of curves and surfaces and only in its later developments does it proceed to the study of their properties "in the large," i.e., in their entire extent.

The development of differential geometry is inseparably connected with the development of analysis. The basic operations of analysis, namely differentiation and integration, have a direct geometric meaning. As was mentioned in Chapter II, differentiating a function $f(x)$ corresponds to drawing a tangent to the curve

$$y = f(x).$$

The slope of the tangent line (i.e., the trigonometric tangent of the angle it makes with the axis Ox) is precisely the derivative $f'(x)$ of the function $f(x)$ at the corresponding point (figure 1), and the area "under the curve"

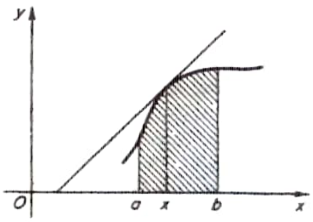

FIG. 1.

$$y = f(x)$$

is precisely the integral $\int_a^b f(x)\, dx$ of this function, evaluated between the corresponding limits. Just as in analysis we investigate arbitrary functions, so in differential geometry we examine arbitrary curves and surfaces. In analysis, the first object of study is the general course of a curve on a plane, its rise and fall, its greater or smaller curvature, the direction of its convexity, its points of inflection, and so forth. The close connection between analysis and the curves is indicated by the name of the first textbook in analysis, by the French mathematician l'Hôpital in 1695: "Infinitesimal analysis applied to the study of curves."

By the middle of the 18th century, the differential and integral calculus had been sufficiently developed by the immediate successors of Newton and Leibnitz that the way was open for more profound applications to geometry. Indeed, it is only from this moment that one may properly

* The properties of curves and surfaces "at a point" are those properties that depend only on an arbitrarily small neighborhood of the point. Properties of this sort are defined in terms of the derivatives (at the given point) of the functions occurring in the equations of the curve or surface. It is for this reason that differential geometry imposes conditions guaranteeing that the differential calculus is applicable; it is required that the curve or surface be defined by functions with a sufficient number of derivatives.

speak of a theory of curves and surfaces. For surfaces, and for curves in space, the analogous problems are immeasurably richer in content than for plane curves, so that with the passage of time these problems outgrew the framework of a simple application of analysis to geometry and led to the formation of an independent theory. During the second half of the 18th century, many mathematicians shared in building up the elements of this theory: Clairaut, Euler, Monge, and others, among whom Euler must be considered as the founder of the general theory of surfaces. The first comprehensive work on curves and surfaces was the book of Monge "Application of analysis to geometry," published in 1795.* From the investigations of these mathematicians, and, in particular, from the book of Monge, we can easily understand the upsurge of interest in differential geometry. This upsurge was due to the demands of mechanics, physics, and astronomy, i.e., in the final analysis to the needs of technology and industry, for which the available results of elementary geometry were completely insufficient.

The classical work of Gauss (1777-1855) in the theory of surfaces is also related to practical questions. His "General investigations concerning curved surfaces," published in 1827, is basic for the differential geometry of surfaces as an independent branch of mathematics. His general methods and problems, discussed later in §4, originated to a great degree in the practical needs of map making. The problem of cartography consists of finding as exact a representation as possible of parts of the surface of the earth on a plane. A completely exact representation here is impossible, the mutual relations of various lengths being necessarily distorted because of the curvature of the earth. Thus one has the problem of finding the most nearly exact methods possible. The drawing of maps goes back to remote antiquity, but the creation of a general theory is an achievement of recent times and would not have been possible without the general theory of surfaces and the general methods of mathematical analysis. We note that one of the difficult mathematical problems of cartography was investigated by P. L. Čebyšev (1821-1894), who obtained important results relating to nets of curved lines on surfaces. His investigations also arose from purely practical problems.

The general questions of deforming one surface so that it can be mapped on another still constitute one of the main branches of geometry. Important results in this direction were obtained in 1838 by F. Minding (1806-1885), professor at the University of Dorpat (now Tartu).

* Gaspard Monge (1746-1828) was not only an outstanding scientist but also an active French revolutionary (minister of naval affairs, and then director of the manufacture of cannon and powder). He followed the path, characteristic of the French bourgeois of the time, from Jacobin to adherent of the emperor Napoleon.

By the second half of the last century, the theory of curves and surfaces was already well established in its basic features, provided we are speaking of "classical differential geometry" in contrast with the newer directions discussed later in §5. The basic equations in the theory of curves, namely the so-called Frenet formulas, had already been obtained, and in 1853 K. M. Peterson (1828-1881), a student of Minding's at Tartu University, discovered and investigated in his dissertation the basic equations of the theory of surfaces, rediscovered 15 years later and published by the Italian mathematician Codazzi, with whose name these equations are usually associated. Peterson, after graduating from the university at Tartu, lived and worked in Moscow, as a teacher in a gymnasium. Though he never held any academic position corresponding to his outstanding scientific achievements, he was nevertheless one of the founders of the Moscow Mathematical Society and of the journal "Matematičeskiĭ Sbornik," published in Moscow from 1866 up to the present day. The Moscow school of differential geometry begins with Peterson.

The results to date of the "classical" differential geometry were summarized by the French geometer Darboux in his four-volume "Lectures on the general theory of surfaces," issued from 1887 to 1896. In the present century classical differential geometry continues to be studied, but the center of interest in curves and surfaces has largely shifted to new directions in which the class of figures under study has been even more widely extended.

§2. The Theory of Curves

Various methods of defining curves in differential geometry. From analysis and analytic geometry we are accustomed to the idea of defining curves by means of equations. In a rectangular coordinate system on the plane, a curve may be given either by the equation

$$y = f(x),$$

or by the more general equation

$$F(x, y) = 0.$$

However, this method of definition is suitable only for a plane curve, i.e., a line in the plane. We also require a method of writing equations of space curves not lying in any plane. An example of such a curve may be seen in the helix (figure 2).

For the purposes of differential geometry, and for many other questions as well, it is most convenient to represent a curve as the trace of a continuous motion of a point. Of course, the given curve may have originated in some entirely different way, but we can always think of it as the path of a point moving along it.

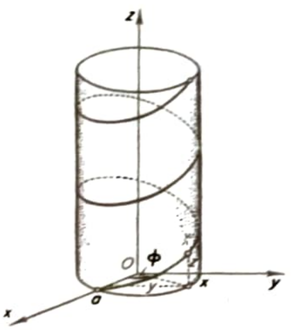

FIG. 2.

Let us assume that we have a fixed Cartesian coordinate system in space. If a moving point X traces out a curve from time $t = a$ to $t = b$, then the coordinates of this moving point are given by the functions of the time $x(t)$, $y(t)$, and $z(t)$; the flight of an airplane or a projectile are examples. Conversely, if we are initially given the functions $x(t)$, $y(t)$, and $z(t)$, we may let them define the coordinates of a moving point X, which traces out some curve. Consequently, curves in space may be given by three equations of the form

$$x = x(t), \quad y = y(t), \quad z = z(t).$$

In the same way a plane curve is defined by two equations

$$x = x(t), \quad y = y(t).$$

This is the most general manner of defining curves.

As an example we consider the helix. It is produced by the spiral motion of a point that revolves uniformly around a straight line, the axis of the helix, and at the same time moves uniformly in a direction parallel to this axis. Let us take the axis of the helix as the axis Oz and suppose that at time $t = 0$ the point lies on the axis Ox. We now wish to find how its coordinates depend on the time. If the motion parallel to the axis Oz has velocity c, then obviously the distance travelled in this direction at time t will be

$$z = ct.$$

Also, if ϕ is the angle of rotation around the axis Oz and a is the distance from the point to this axis, then, as can be seen in figure 2,

$$x = a \cos \phi, \quad y = a \sin \phi.$$

Since the rotation is uniform, the angle ϕ is proportional to time; that is, $\phi = \omega t$, where ω is the angular velocity of the rotation. In this manner we get

$$x = a \cos \omega t, \quad y = a \sin \omega t, \quad z = ct.$$

So these are the equations of the helix, which as t changes will be traced out by the moving point.

Of course the variable t or, as it is usually called, the parameter, need not be thought of as representing the time. Also, the given parameter t may be replaced by another; for example we may introduce a parameter u by the formula $t = u^3$, or, in general, by $t = f(u)$.* In geometry the most natural choice of parameter is the length s of the arc of the curve measured from some fixed point A on it. Every possible value of the length s represents a corresponding arc AX. Thus the position of X is fully determined by the value of s and the coordinates of the point X are given by the functions of arc length s

$$x = x(s), \quad y = y(s), \quad z = z(s).$$

All these ways of defining curves, as well as other possible ones,† open up the possibility of numerical computation. Only when curves have been defined by equations can their properties be investigated by mathematical analysis.

In the differential geometry of plane curves, there are three basic concepts: length, tangent, and curvature. For space curves, there are in addition the osculating plane and the torsion. We now proceed to explain the meaning and significance of these concepts.

Length. Everyone has in mind a natural idea of what is meant by length, but this idea must be converted into an exact definition of the length of a mathematical curve, a definition with a specific numerical character, which will enable us to compute the length of a curve with any desired degree of accuracy and consequently to argue about lengths in a rigorous way. The same remarks apply to all mathematical concepts. The transition from informal ideas to exact measurements and definitions represents the transition from a prescientific understanding of objects to

* Here, strictly speaking, it is necessary that the function f be monotone.

† A curve in space may also be given as the intersection of two surfaces, defined by the equations: $F(x, y, z) = 0$, $G(x, y, z) = 0$, i.e., the curve is given by this pair of equations. In theoretical discussions a curve is most frequently given by a variable vector, i.e., the position of the point X of the curve is defined by the vector $r = \overrightarrow{OX}$, extending from the origin to this point. As the vector r changes, its end point X moves along the given curve (figure 3).

a scientific theory. The need for a precise definition of length arose in the final analysis from the requirements of technology and the natural sciences, whose development demanded investigation of the properties of lengths, areas, and other geometric entities.

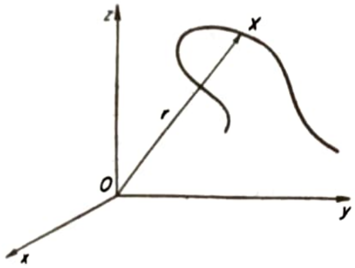

FIG. 3.

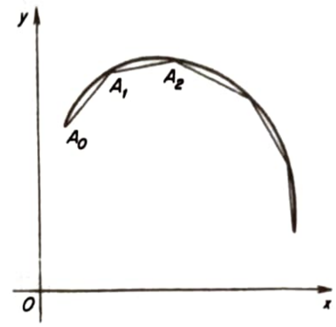

FIG. 4.

A simple and most useful definition of length is the following: The *length* of a curve is the limit of the length of broken lines inscribed in the curve under the condition that their vertices cluster closer and closer together on the curve.

This definition arises naturally from our everyday methods of measuring. On the curve we take a sequence of points $A_0, A_1, A_2, \cdots$ (figure 4) and measure the distances between them. The sum of these distances (which is the length of the broken line) expresses approximately the length of the curve. In order to define the length more exactly, it is natural to take the points A closer together, so that the broken line follows the twists of the curve more closely. Finally, the exact value of the length is defined as the limit of these approximations as the points A are chosen arbitrarily close together.* Thus the earlier definition of length is a generalization, based on taking finer and finer steps, of a completely practical manner of measuring length.

From this definition of length, it is easy to derive a formula for computing lengths when the curve is given analytically. We note, however, that mathematical formulas are useful for more than just computation.

* The existence of the indicated limit, i.e., the length of the curve, is not initially clear, even for curves lying in a bounded domain. If the curve is very twisted, its length may be very great, and it is possible mathematically to construct a plane curve which is so "twisted" that none of its arcs has a finite length since the lengths of broken lines inscribed in it increase beyond all bounds.

They are a brief statement of theorems that establish connections between different mathematical entities. The theoretical significance of such connections may far exceed the computational value of the formula. For example, the importance of the Pythagorean theorem, expressed by the formula

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

is not confined to the computation of the square of the hypotenuse c but lies chiefly in the fact that it expresses a relation among the sides of a right triangle.

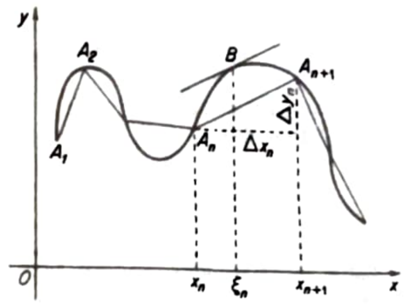

FIG. 5.

Let us now introduce a formula for the length of a plane curve, given in Cartesian coordinates by the equation $y = f(x)$, assuming that the function $f(x)$ has a first derivative.

We inscribe a broken line in the curve (figure 5). Let A_n, A_{n+1} be two of its adjacent vertices with coordinates x_n, y_n and x_{n+1}, y_{n+1}. The line segment A_nA_{n+1} is the hypotenuse of a right triangle the legs of which are equal to

$$\Delta x_n = |x_{n+1} - x_n|, \quad \Delta y_n = |y_{n+1} - y_n|.$$

Thus, by the Pythagorean theorem,

$$\overline{A_nA_{n+1}} = \sqrt{(\Delta x_n)^2 + (\Delta y_n)^2} = \sqrt{1 + \left(\frac{\Delta y_n}{\Delta x_n}\right)^2}\, \Delta x_n .$$

It is easy to see that if the straight line drawn through the points A_n and A_{n+1} is translated parallel to itself, then at the instant when the line leaves the curve it will assume the position of a tangent to this curve at some point B, i.e., on the arc of the curve A_nA_{n+1}, there is at least one point at which the tangent has the same direction as the chord A_nA_{n+1}. (This obvious conclusion can easily be given a rigorous proof.)

Thus we may replace the ratio $\Delta y_n/\Delta x_n$ by the slope of the tangent at B, i.e., by the derivative $y'(\xi_n)$, where ξ_n is the abscissa of the point B. Now the length of one link of the broken line is expressed by

$$\overline{A_nA_{n+1}} = \sqrt{1 + y'^2(\xi_n)}\, \Delta x_n .$$

The entire length of the broken line is the sum of the lengths of its pieces. Denoting the addition by the symbol Σ, we have

$$S_n = \sum \sqrt{1 + y'^2(\xi_n)}\, \Delta x_n .$$

To obtain the length of the curve, we must pass to the limit under the condition that the greatest of the values Δx_n tends to zero,

$$s = \lim_{\Delta x \to 0} \sum \sqrt{1 + y'^2(\xi_n)}\, \Delta x_n .$$

But this limit is exactly the integral defined in Chapter II, namely the integral of the function $\sqrt{1 + y'^2}$. Thus the length of a plane curve is expressed by the formula

$$s = \int_a^b \sqrt{1 + y'^2}\, dx, \tag{1}$$

where the limits of integration a and b are the values of x at the ends of the arc of the curve.

The corresponding, but somewhat different, formula for the length of a space curve is derived in basically the same way.

The actual computation of a length by means of these formulas is, of course, not always simple. Thus the calculation of the circumference of a circle from formula (1) is rather complicated. However, as we have said, the interest of formulas is not confined to computation; in particular, formula (1) is also important for investigating the general properties of length, its relations with other concepts, and so forth. We will have an opportunity to make use of formula (1) in Chapter VIII.

Tangent. The tangent to a plane curve was already considered in Chapter II. Its meaning for a space curve is completely analogous. In order to define the tangent at a point A, we choose a point X on the curve, distinct from A, and consider the secant AX. Then we allow X to approach A along the curve. If the secant AX converges to some limiting position, then the straight line in this limiting position is called the tangent at the point A.*

If we distinguish between the initial point and the end point of the curve and thereby establish an order in which the points of the curve

* The limiting position of the secant may not exist, as can be seen from the example in figure 13, Chapter II. The curve represented by $y = x \sin 1/x$ oscillates near zero in such a way that the secant OA, as A approaches O, constantly oscillates between the straight lines OM and OL.

are traversed, then we may say which of the points A and X comes first and which comes second. (For example, if a train travels from Moscow to Vladivostok, then Omsk obviously precedes Irkutsk.) So we may define a direction along the secant from the first point to the second. The limit of such "directed secants" gives us a "directed tangent." In figure 6, the arrow shows the direction in which the point A is passed through. For the motion of a point along the curve, the velocity at each instant is directed along the tangent to the curve.

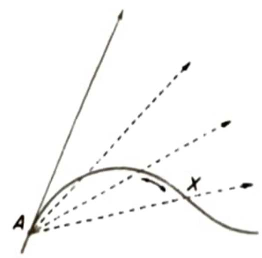

FIG. 6.

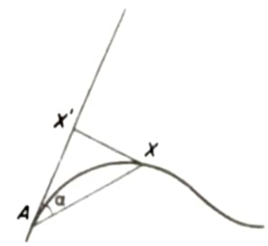

FIG. 7.

The tangent has an important geometric property: Near the point of tangency the curve departs less, in a well-defined sense, from this straight line than from any other. In other words, the distance from the points of the curve to the tangent is very small in comparison with their distance to the point of tangency. More precisely, the ratio XX'/AX (figure 7) tends to zero as X approaches A.* So a small segment of the curve may be replaced by a corresponding segment of the tangent with an error that is small in comparison with length of the segment. This procedure often allows us to simplify proofs, since in a passage to the limit it gives completely exact results.

It is interesting to observe that for a curve which is not a straight line, i.e., does not have a direction in the elementary sense, we have been able, by associating it with a straight line, to define its direction at each point. Thus the concept of direction has been extended; it has been given a meaning which it did not previously have. This new concept of direction reflects the actual nature of motion along a curve; at each instant the point is moving in some definite direction, which changes continuously.

* This result follows immediately from the definition of the tangent itself. Evidently, as is shown in figure 7, $XX'/AX = \sin \alpha$, where α is the angle between the tangent and the secant AX. Thus, as $\alpha \to 0$, XX'/AX also tends to zero.

Curvature. To be able to judge by eye whether a path, a thin rod, or a line in a drawing is more or less curved it is not necessary to be a mathematician. But for even the simplest problems of mechanics, a casual glance is not sufficient; we need an exact quantitative description of the curvature. This is obtained by giving precise expression to our intuitive impression of the curvature as the rapidity of change of direction of the curve.

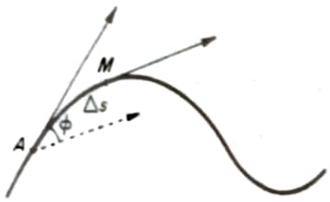

FIG. 8.

Let A be a point on the curve and M a point near A (figure 8). The angle between the tangents at these points expresses how much the curve has changed direction in the segment from A to M. Let us denote this angle by ϕ. The average rate of change of direction (more precisely, the average change per unit length of path along the segment AM of length Δs) will obviously be $\phi/\Delta s$. Then the curvature, namely the rate of change of direction of the curve at the point A itself, is naturally defined as the limit of the ratio $\phi/\Delta s$ as $M \to A$; in other words, as $\Delta s \to 0$. Thus the curvature is defined by the formula

$$k = \lim_{\Delta s \to 0} \frac{\phi}{\Delta s}.$$

As a particular example, let us consider the curvature of the circumference of a circle (figure 9). Obviously, the angle ϕ between the radii OA and OM is equal to the angle ϕ between the tangents at the points A and M, since the tangents are perpendicular to the radii. The arc AM, subtending the angle ϕ, has length $\Delta s = \phi r$, so that

$$\frac{\phi}{\Delta s} = \frac{1}{r}.$$

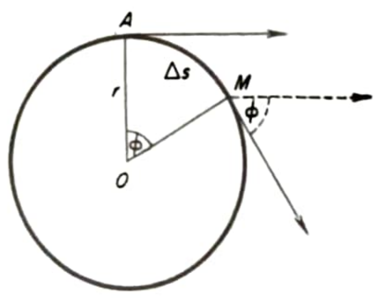

FIG. 9.

This means that the ratio $\phi/\Delta s$ is constant, so that the curvature of the circumference of a circle, as the

limiting value of this ratio, is equal at all points to the reciprocal of the radius.*

Let us derive the formula for the curvature of a plane curve given by the equation $y = f(x)$. As the initial point for arc length we take a fixed point N (figure 10). The angle ϕ between the tangents at the points A and M is obviously equal to the difference in the angle of inclination of the tangents at A to M.

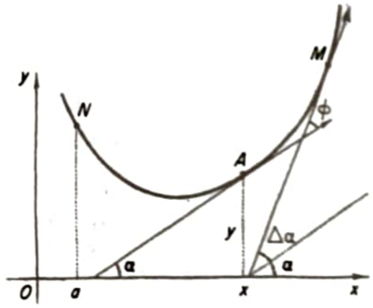

FIG. 10.

$$\phi = |\Delta\alpha|.$$

Since the angle α may decrease, we take the absolute value $|\Delta\alpha|$. We are interested in the value

$$k = \lim_{\Delta s \to 0} \frac{\phi}{\Delta s} = \lim_{\Delta s \to 0} \frac{|\Delta\alpha|}{\Delta s} = \lim_{\Delta x \to 0} \frac{\dfrac{|\Delta\alpha|}{\Delta x}}{\dfrac{\Delta s}{\Delta x}} = \frac{|\alpha'|}{s'}.$$

The length of the arc of the curve NA is expressed by the integral

$$s = \int_a^x \sqrt{1 + y'^2}\, dx,$$

so that

$$s' = \sqrt{1 + y'^2}.$$

* We note that in general the concept of the curvature of a curve at a point may be defined by comparing the curve with the circumference of a certain circle, which plays the role of a model or standard for the curvature. For in fact, the curvature of the given curve proves to be equal to the reciprocal of the radius of the (unique) circle which fits the curve most closely in the neighborhood of the point.

the osculating circle

It remains to find α'. We know that $\tan \alpha = y'$; thus $\alpha = \arctan y'$. Differentiating this last equation with respect to x, we get

$$\alpha' = \frac{1}{1 + y'^2} y''.$$

Thus, finally

$$k = \frac{|\alpha'|}{s'} = \frac{|y''|}{(1 + y'^2)^{3/2}}.$$

The corresponding formulas for other methods of representing plane and space curves are given in the usual courses in analysis or differential geometry.

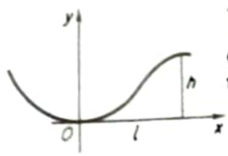

FIG. 11.

This formula allows us to give another geometric interpretation of curvature, which is useful in many questions. Namely, the curvature of a curve at a point is given by the formula

$$k = \lim_{l \to 0} \frac{2h}{l^2},$$

where h is the distance of a second point on the curve to the tangent at the given point and l is the length of the segment of the tangent between the point of tangency and the projection on the tangent of the other point on the curve (figure 11).

To prove this we choose a rectangular coordinate system such that the origin falls at the given point of the curve and the axis Ox is tangent to the curve at this point (figure 11). (For simplicity we assume that the curve is plane.) Then $y' = 0$ and $k = |y''|$. Expanding the function $y = f(x)$ by Taylor's formula, we get $y = \frac{1}{2} y'' x^2 + \epsilon x^2$ (where we have taken into account that $y' = 0$). Here $\epsilon \to 0$ as $x \to 0$. Hence it follows that $k = |y''| = \lim_{l \to 0} 2|y|/x^2$, and thus, since $|y| = h$, $x^2 = l^2$, we have

$$k = \lim_{l \to 0} \frac{2h}{l^2}.$$

This formula shows that the curvature describes the rate at which the curve leaves the tangent.

Let us now turn to some very important applications of curvature to problems of mechanics.

First we consider the following problem. Let a flexible string be stretched over a support (figure 12) in such a way that the string remains in one plane. We wish to find the pressure of the string on the support at every point, or to be more exact, to define the limit

$$p = \lim_{\Delta s \to 0} \frac{P}{\Delta s}, \tag{2}$$

where P is the magnitude of the force $\boldsymbol{P}$ acting on the support along a piece of length Δs containing the given point. We assume for simplicity that the magnitude T of the tension $\boldsymbol{T}$ is the same at all points of the string.

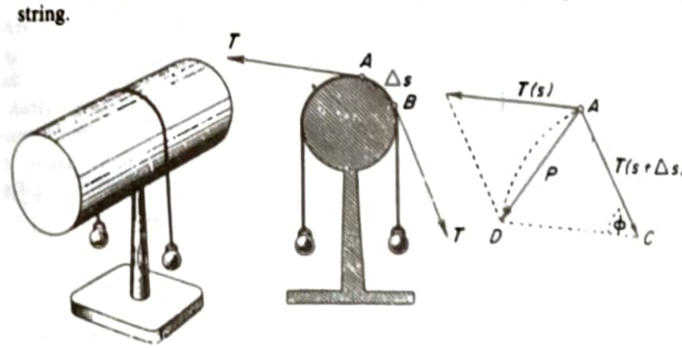

FIG. 12.

Now consider the point A and a segment of the string AB.* On this segment AB of length Δs, in addition to the reaction of the support, only two external forces are acting, namely the tensions at the ends, which are equal in magnitude and are directed along the tangents at the ends of the segment. Thus the force $\boldsymbol{P}$ exerted by the string on the support is equal to the geometric sum of the tensions at the ends. As can be seen from figure 12, the vector $\boldsymbol{P}$ is the base AD of the isosceles triangle CAD. The two equal sides of this triangle have length T and the angle at the vertex C is equal to the change of direction of the tangent in passing from A to B.

With decreasing Δs the angle ϕ decreases and the angle between $\boldsymbol{P}$ and the tangent at the point A approaches a right angle. Thus the pressure is perpendicular to the tangent.

To find the magnitude of the pressure, we make use of the fact that a small arc of the circumference has approximately the same length as

* It would be more natural to choose a segment with the point A in its interior; this would not change the result but would make the computation somewhat more complicated.

the chord subtending it. Thus we replace the length of the chord AD, i.e., the magnitude P, by the length $T\phi$ of the arc AD. Then by formula (2) we get

$$p = \lim_{\Delta s \to 0} \frac{P}{\Delta s} = \lim_{\Delta s \to 0} \frac{T\phi}{\Delta s} = T \lim_{\Delta s \to 0} \frac{\phi}{\Delta s} = Tk.$$

Hence the pressure at each point is equal to the product of the curvature and the tension on the string and is exerted perpendicularly to the tangent at this point.

Consider a second problem. Let a mathematical point (i.e., a very small body) move along a plane curve with a velocity of constant magnitude v. What is its acceleration at a given point A? By definition, the acceleration is equal to the limit of the ratio of the change in velocity (during the time Δt) to the increment Δt of the time. The velocity involves not only magnitude but also direction, i.e., we consider the change in the velocity vector. Therefore the mathematical problem of finding the magnitude of the acceleration consists of finding the limit

$$w = \lim_{\Delta t \to 0} \frac{|\, \boldsymbol{v}(t + \Delta t) - \boldsymbol{v}(t) \,|}{\Delta t},$$

where $\boldsymbol{v}(t)$ is the velocity at the point A itself, and $|\, \boldsymbol{v}(t + \Delta t) - \boldsymbol{v}(t)|$ is the length of the vector difference of the velocities. The limit which concerns us may also be represented as

$$\lim_{\Delta s \to 0} \frac{|\, -\boldsymbol{v}(t) + \boldsymbol{v}(t + \Delta t) \,|}{\Delta s} \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t},$$

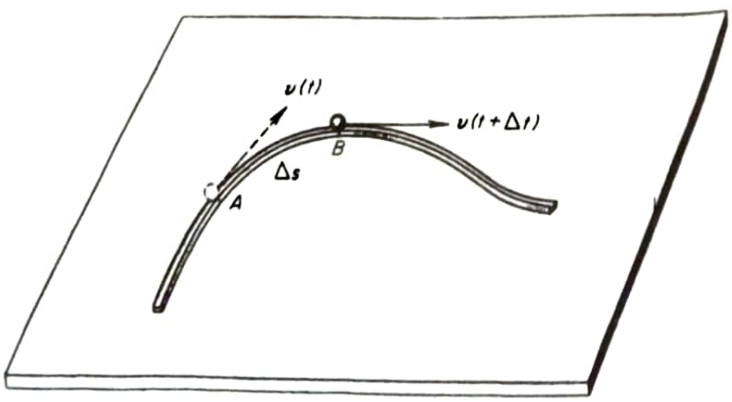

FIG. 13.

where Δs is the length of the arc AB traversed during time Δt. Turning to figure 13 and noting that the velocity at each point is directed along the tangent while remaining constant in magnitude, we see geometrically that finding the sum $-\boldsymbol{v}(t) + \boldsymbol{v}(t + \Delta t)$ is identical with finding the vector $\boldsymbol{P}$ in the preceding problem. So we may avail ourselves of the result there and, replacing tension by velocity, write

$$\lim_{\Delta s \to 0} \frac{|\, -\boldsymbol{v}(t) + \boldsymbol{v}(t + \Delta t) \,|}{\Delta s} = vk.$$

Moreover, $\lim_{\Delta t \to 0} \Delta s/\Delta t = v$. So we have the final result that the acceleration of a body in uniform motion along the curve is equal to the product of the curvature and the square of the velocity

$$w = kv^2 \qquad (3)$$

and is directed along the normal to the curve, i.e., along a straight line perpendicular to the tangent.

Our recourse here to a geometric analogy, enabling us to use the solution of the problem of the pressure exerted by a string in order to solve a problem of the acceleration of a particle, shows once again how useful it is to make an abstraction from the particular concrete properties of a phenomenon to corresponding mathematical concepts and results; for we can then make use of these results in the most varied situations.

We also note that the curvature, which from a mechanical point of view reflects the change in the direction of motion, is seen to be closely connected with the forces causing this change. The equation which expresses this connection is easily derived if we multiply equation (3) by the mass m of the moving point. We have

$$F_n = mw = v^2mk.$$

Here F_n is the magnitude of the normal component of the force acting on the point.

Osculating plane. Although a space curve does not lie in one plane, still with each point A of the curve it is possible, as a rule, to associate a plane P which in the neighborhood of this point lies closer to the curve than any other plane. This plane is called the *osculating plane* of the curve at the point.

Naturally the osculating plane, as the plane closest to the given curve, passes through the point A and contains the tangent T to the curve. But there are many planes containing the point A and the straight line T.

In order to choose from among them the one plane that least deviates from the curve, we investigate the deviation of the curve from the tangent. For this purpose let us see how the curve runs along the tangent T; in other words, let us project our curve onto the *normal plane* Q, which is perpendicular to T at the point A (figure 14). The projection on the plane Q of a segment of our curve containing A forms a new curve, indicated in figure 14 by a dotted line. Usually it has a cusp at the point A. If the curve so obtained has a tangent N at the point A, then the plane P determined by T and N will naturally be closest to the original curve in the neighborhood of the point A, i.e., it will be the osculating plane at the point A. It may be shown that when the functions defining the original curve have second derivatives and the curvature of the curve at the point A is not zero, then the osculating plane necessarily exists, and its equation may be expressed very simply in terms of the first and second derivatives of the functions defining the curve.

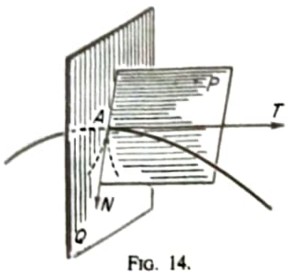

FIG. 14.

We saw earlier that the properties of the tangent allow us to consider a small segment of a plane curve as though it were straight, thereby making an error which is small in comparison with the length of the segment; similarly the properties of the osculating plane allow us to consider a small segment of a space curve as though it were a plane curve, namely its projection on the osculating plane, and here the error will be small in comparison with the square of the length of the segment of the curve.

There are many straight lines in space that are perpendicular to the tangent; they form the normal plane at the given point of the curve. Among these straight lines there is one, the line N, which lies in the osculating plane. This line is called the *principal normal* to the curve. Usually we also fix a direction for it, namely the direction of the concavity of the projection of the curve on the osculating plane. The principal normal plays the same role for a space curve as the ordinary (unique) normal for a plane curve. In particular, if a thin string under tension T is stretched in the form of a space curve over a support, then the pressure of the string on the support has at each point the magnitude Tk and is directed along the principal normal. If a material point is moving along a space curve with a velocity of constant magnitude v, then its acceleration is equal to kv^2 and is directed along the principal normal.

Torsion. From point to point along a curve the position of the osculating plane will probably change. Just as the rate of change of direction of the tangent characterized the curvature, so the rate of change of direction of the osculating plane characterizes a new quantity, the *torsion* of the curve. Here, as in the case of curvature, the rate is taken with respect to arc length; that is, if ψ is the angle between the osculating planes at a fixed point A and at a nearby point X, and if Δs is the length of the arc AX, then the torsion τ at the point A is defined as the limit*

$$\tau = \lim_{\Delta s \to 0} \frac{\psi}{\Delta s}.$$

The sign of the torsion depends on the side of the curve toward which the osculating plane turns as it moves along the curve.

We may imagine the osculating curve as the blade of a fan with the two lines, the tangent and the principal normal, drawn on it. At each moment the tangent is turning in the direction of the normal at a rate determined by the curvature, while the osculating plane rotates around the tangent with a speed and direction determined by the torsion.

The simplest results of the theory of differential equations may be used to prove a fundamental theorem that states, roughly speaking, that two curves with the same curvature and the same torsion are identical with each other. Let us make this idea clearer. If we move along the curve to various distances A from our initial point, we will arrive at points where the curvature k and the torsion τ will have various values, depending on s. Thus $k(s)$ and $\tau(s)$ will be certain well-defined functions of the arc length s.

The theorem in question states that if two curves have identical curvature and torsion as functions of arc length, then the curves are identical (i.e., one of them may be rigidly moved so as to coincide with the other). In this manner curvature and torsion as functions of arc length define a curve completely except for its position in space; they describe all the properties of the curve by stating the relationship between its length, its curvature, and its torsion. In this way the three concepts constitute a sort of ultimate basis for questions concerning curves. With their help we can also express the simplest concepts in the theory of surfaces, to which we now turn.

* It may be shown that a helix has the same torsion at all its points and consequently that we may define the torsion of a curve by comparing the curve with the (unique) helix which best approximates the curve in the neighborhood of the given point. The torsion also characterizes the way in which a given space curve differs from a plane curve. With a certain analogy to curvature, it characterizes the rate at which the curve leaves its osculating plane.

Of course, the theory of curves has not been exhausted by our present remarks. There are many other concepts relating to curves: special types of curves, families of curves, the position of curves on surfaces, questions of the form of a curve as a whole, etc. These questions and the methods of answering them are connected with almost every branch of mathematics. The range of problems that may be solved by the theory of curves is extremely rich and varied.

§3. Basic Concepts in the Theory of Surfaces

The basic methods of defining a surface. If we wish to study surfaces by means of analysis we must, of course, define them analytically. The simplest way is by an equation

$$z = f(x, y),$$

in which x, y, and z are Cartesian coordinates of a point lying on the surface. Here the function $f(x, y)$ need not necessarily be defined for all x, y; its domain may have various shapes. Thus, the surface illustrated in figure 15 is given by the function $f(x, y)$ defined inside an annulus. Examples of surfaces given by equations of the form $z = f(x, y)$ are also familiar from analytic geometry. We know, for example, that the equation $z = Ax + By + C$ represents a plane, and $z = x^2 + y^2$ a paraboloid of revolution (figure 16). For the application of differential calculus it is necessary that the function $f(x, y)$ have first, second, and sometimes even higher derivatives. A surface given by such an equation is called *regular*. Geometrically this means (though not quite precisely) that

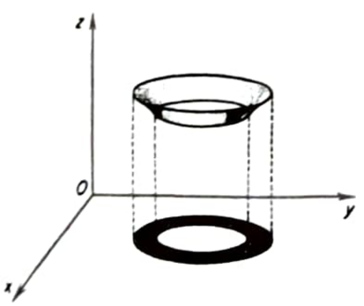

FIG. 15.

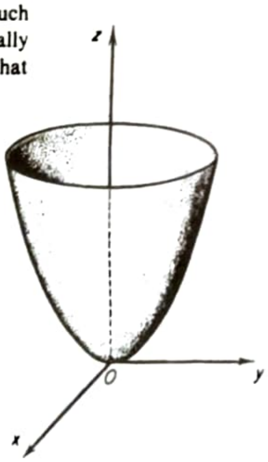

FIG. 16.

the surface curves continuously without breaks or other singularities. Surfaces that do not have this property, for example, those with cusps, breaks, or other singularities, require a new kind of investigation (cf. §5).

However, not every surface, even without singularities, can be entirely represented by an equation of the form $z = f(x, y)$. If every pair of values of x, y in the domain of $f(x, y)$ gives a completely determined z, then every straight line parallel to the axis Oz must intersect the surface at no more than one point (figure 17). Even such simple surfaces as

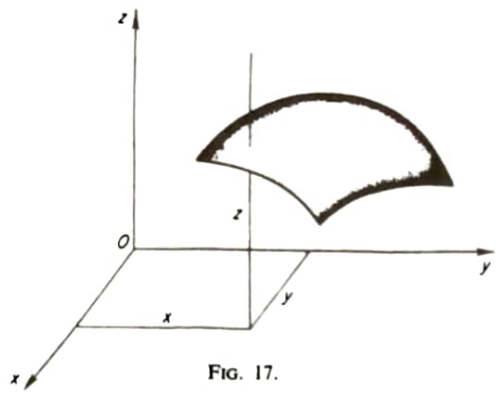

FIG. 17.

spheres or cylinders cannot be represented in the large by an equation of the form $z = f(x, y)$. In these cases the surface is defined in some other manner, for example by an equation of the form $F(x, y, z) = 0$. Thus a sphere of radius R with center at the origin has the equation

$$x^2 + y^2 + z^2 = R^2.$$

The equation $x^2 + y^2 = r^2$ gives a cylinder of radius r.

So when the investigation is concerned only with small segments of the surface, as is usually the case in classical differential geometry, the definition of a surface by an equation $z = f(x, y)$ is perfectly general, since every sufficiently small segment of a smooth surface can be represented in this form. We take this way as basic, and leave other methods of defining surfaces to be considered later in §§4 and 5.

Tangent plane. Just as at each point a smooth *curve* has a tangent line which is close to the curve in a neighborhood of the point, so also surfaces may have, at each of their points, a *tangent plane*.

The exact definition is as follows. A plane P, passing through a point M on a surface F, is said to be tangent to the surface F at this point if the angle α between the plane P and the secant MX, drawn from M to a point X of the surface, converges to zero as the point X approaches the point M (figure 18). All tangents to curves passing through the point M and lying on the surface obviously lie in the tangent plane.

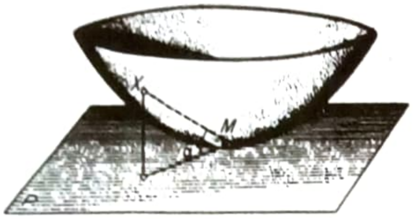

FIG. 18.

A surface F is called *smooth* if it has a tangent plane at each point and if, as we pass from point to point, the position of this plane varies continuously.

Near the point of tangency, the surface departs very little from its tangent plane: If the point X approaches the point M along the surface, then the distance of the point X from the tangent plane becomes smaller and smaller, even in comparison with its distance from the point M (the reader can easily verify this by considering how X approaches M in figure 18). In this way, the surface near the point M may be said to merge into the tangent plane. In the first approximation a small segment or, as it is called, an "element" of the surface may be replaced by a segment of the tangent plane. The perpendicular to the tangent plane which passes through the point of tangency acts as a perpendicular to the surface at this point and is called a *normal*.

This possibility of replacing an element of the surface by a segment of the tangent plane is useful in many situations. For example, the reflection of light on a curved surface takes place in the same way as the reflection on a plane, i.e., the direction of the reflected ray is defined by the usual law of reflection: The incident ray and the reflected ray lie in one plane together with the normal to the surface and they make equal angles with this normal (figure 19), just as if the reflection were occurring in the tangent plane. Similarly for the refraction of light in a curved surface, each ray is refracted by an element of the surface with the usual law of refraction, just as if the element were plane. These facts are the basis for all calculations of reflection and refraction of light in optical apparatus. Further, for example, solid bodies in contact with each other have a common tangent plane at their point of contact. The bodies are in contact over an element of their surface, and the pressure

of one body on the other, in the absence of friction, is directed along the normal at the point of contact. This is also true when the bodies are tangent at more than one point, in which case the pressure is directed along the respective normals at each point of contact.

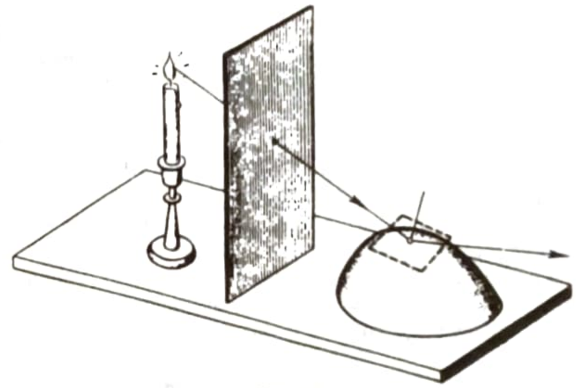
FIG. 19.

The replacement of elements of a surface by segments of the tangent planes can also serve as the basis of a definition of the area of various surfaces. The surface is decomposed into small pieces $F_1, F_2, \cdots, F_n$ and each piece is projected onto a plane tangent to the surface at some point of this piece (figure 20). We thus obtain a number of plane regions $P_1, P_2, \cdots, P_n$, the sum of whose areas gives an approximation to the

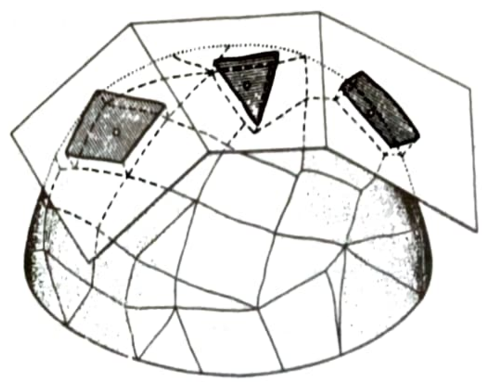
FIG. 20.

area of the surfaces. The area of the surface itself is defined as the limit of the sums of the areas of the segments $P_1, P_2, \cdots, P_n$ under the condition that the partitions of the surface become finer.* From this we can derive an exact expression for the area in the form of a double integral.

These remarks clearly demonstrate the significance of the concept of the tangent plane. However, in many questions the approximate representation of an element of a surface by means of a plane is inadequate and it is necessary to consider the curvature of the surface.

Curvature of curves on a surface. The curvature of a surface at a given point is characterized by the rate at which the surface leaves its tangent plane. But in different directions, the surface may leave its tangent plane at different rates. Thus the surface illustrated in figure 21 leaves the plane P in the direction OA at a faster rate than in the direction OB. So it is natural to define the curvature of a surface at a given point by means of the set of curvatures of all curves lying in the surface and passing through the given point in different directions.

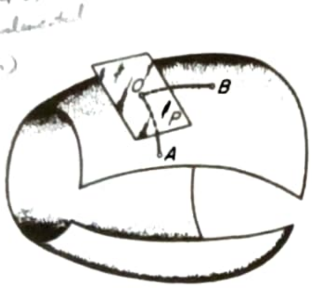

FIG. 21.

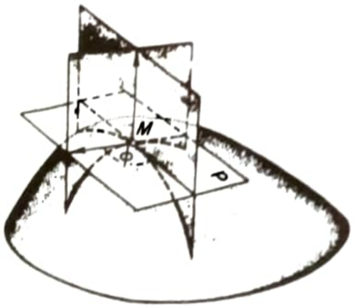

FIG. 22.

This is done as follows. We construct the tangent plane P through the point M and choose a specific direction for the normal (figure 22). Then we consider curves which are sections of the surface cut by planes passing through the normal at the point M; these curves are called *normal sections*. The curvature of a normal section is given a sign, which is plus if the section is concave in the direction of the normal and minus if it is concave in the opposite direction. Thus, in a surface which is saddle-shaped, as illustrated in figure 23 with the arrow indicating the

* This is exactly the expression for the area which was used in §1, Chapter VIII.

direction of the normal to the surface, the curvature of the section MA is positive and that of the section MB is negative.

A normal section is defined by the angle ϕ by which its plane is rotated from some initial ray in the tangent plane (figure 22). If we know the curvature of the normal section $k(\phi)$ in terms of the angle ϕ, we will have a rather complete picture of the behavior of the surface in the vicinity of the point M.

A surface may be curved in many different ways and thus it would appear that the dependence of the curvature k on the angle ϕ may be arbitrary. In fact this is not so. For the surfaces studied in differential geometry, there exists a simple law, due to Euler, that establishes the connection between the curvatures of the normal sections passing through a given point in various directions.

It is shown that at each point of a surface there exist two particular directions such that

1. They are mutually perpendicular;

2. The curvatures k_1 and k_2 of the normal sections in these directions are the smallest and largest values of the curvatures of all normal sections;*

3. The curvature $k(\phi)$ of the normal section rotated from the section with curvature k_1 by the angle ϕ is expressed by the formula

$$k(\phi) = k_1 \cos^2 \phi + k_2 \sin^2 \phi. \tag{4}$$

Such directions are called the *principal directions* and the curvatures k_1 and k_2 are called the *principal curvatures of the surface* at the given point.

This theorem of Euler shows that in spite of the diversity of surfaces, their form in the neighborhood of each point must be one of a very few completely defined types, with an accuracy to within magnitudes of the second order of smallness in comparison with the distance from the given point. In fact, if k_1 and k_2 have the same sign, then the sign of $k(\phi)$ is constant and the surface near the point has the form illustrated in figure 22. If k_1 and k_2 have opposite signs, for example $k_1 > 0$ and $k_2 < 0$, then the curvature of the normal section obviously changes sign. This is seen from the fact that for $\phi = 0$ the curvature $k = k_1 > 0$ and for $\phi = \pi/2$ we have $k = k_2 < 0$.

From formula (4) for $k(\phi)$, it is not difficult to prove that as ϕ changes

* In the particular case $k_1 = k_2$ the curvature of all sections is the same; as, for example, on a sphere.

from 0 to π the sign of $k(\phi)$ changes twice,* so that near the point the surface has a saddle-shaped form (figure 23).

When one of the numbers k_1 and k_2 is equal to zero, the curvature always has the same sign, except for the one value of ϕ, for which it vanishes. This occurs, for example, for every point on a cylinder (figure 24).

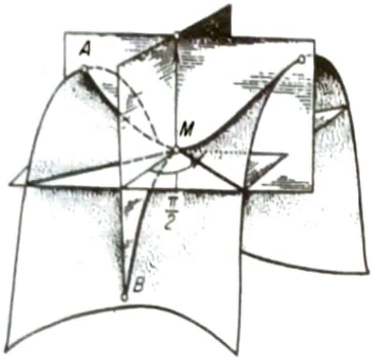

FIG. 23.

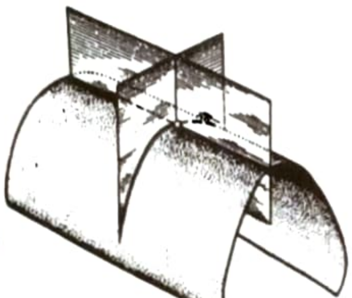

FIG. 24.

In the general case the surface near such point has a form close to that of a cylinder.

Finally, for $k_1 = k_2 = 0$ all normal sections have zero curvature. Near such a point the surface is especially "close" to its tangent plane. Such points are called *flat points*. One example of such a point is given in figure 25 (the point M). The properties of a surface near a flat point may be very complicated.

FIG. 25.

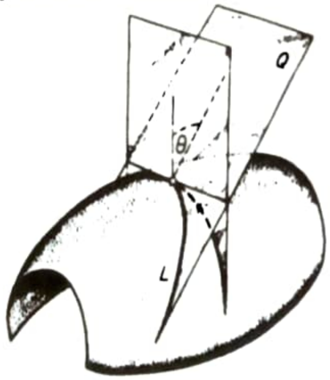

FIG. 26.

* It is a simple matter to show that $k(\phi) = k_1 \cos^2 \phi + k_2 \sin^2 \phi$ vanishes for $\phi = \text{arc tan} \sqrt{-k_1/k_2}$ and $\phi = \pi - \text{arc tan} \sqrt{-k_1/k_2}$, changing sign the first time from plus to minus and the second from minus to plus.

Let us now consider a section of the surface cut by an arbitrary plane Q (figure 26) not passing through the normal. The curvature k_L of such a curve L, as Meusnier showed,* is connected by a simple relation with the curvature k_N of the normal section in the same direction, i.e., the one that intersects the tangent plane in the same straight line. This connection is expressed by the formula

$$k_L = \frac{|k_N|}{\cos \theta},$$

where θ is the angle between the normal and the plane Q. The correctness of this formula may be visualized very conveniently on a sphere.

Finally, the curvature of *any* curve lying in the surface and having the plane Q as its osculating plane may be shown to be identical with the curvature of the intersection of Q with the surface.

Thus, if we know k_1 and k_2, the curvature of any curve in the surface is defined by the direction of its tangent and the angle between its osculating plane and the normal to the surface. Consequently, the character of the curvature of a surface at a given point is defined by the two numbers k_1 and k_2. Their absolute values are equal to the curvatures of two mutually perpendicular normal sections, and their signs show the direction of the concavity of the respective normal sections with respect to a chosen direction on the normal.

Let us now prove the theorems of Euler and Meusnier mentioned earlier.

1. For the proof of Euler's theorem we need the following lemma. If the function $f(x, y)$ has continuous second derivatives at a given point, then the coordinate axes may be rotated through an angle α such that in the new coordinate system the mixed derivative $f_{x'y'}$ will be equal to zero at this point.† We recall that after rotation of axes the new variables x', y' are connected with x and y by the formulas

$$x = x' \cos \alpha - y' \sin \alpha; \quad y = x' \sin \alpha + y' \cos \alpha$$

(cf. Chapter III, §7). For the proof of the lemma we note that

$$\frac{\partial x}{\partial x'} = \cos \alpha, \quad \frac{\partial y}{\partial x'} = \sin \alpha, \quad \frac{\partial x}{\partial y'} = -\sin \alpha, \quad \frac{\partial y}{\partial y'} = \cos \alpha.$$

* Meusnier (1754–1793) was a French mathematician, a student of Monge; he was a general in the revolutionary army and died of wounds received in battle.

† We will denote partial derivatives by subscripts; for example, in place of $\partial f/\partial x$ we write f_x, in place of $\partial^2 f/\partial y^2$ we write f_{yy}, etc.

Computing the derivative $f_{x'y'}$ by the chain rule, we arrive after some calculation at the result

$$f_{x'y'} = f_{xy}\cos 2\alpha + \tfrac{1}{2}(f_{yy} - f_{xx})\sin 2\alpha,$$

from which it readily follows that for

$$\cot 2\alpha = \frac{1}{2}\frac{f_{xx} - f_{yy}}{f_{xy}}$$

we will have

$$f_{x'y'} = 0.$$

We now consider the surface F, given by the equation $z = f(x, y)$, in which the origin is at the point M under consideration and the axes Ox and Oy are so chosen in the tangent plane that $f_{xy}(0, 0) = 0$. In the surface P we take an arbitrary straight line making an angle ϕ with the

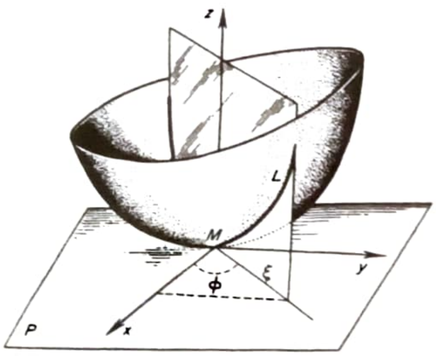

FIG. 27.

axis Ox and consider the normal section L in the direction of this straight line (figure 27). From the formula derived in §2, the curvature of L at the point M, taking its sign into account, is equal to

$$k_L = \lim_{\xi\to 0}\frac{2f(x, y)}{\xi^2}.$$

Here $f(x, y)$ is the distance (again taking its sign into account) of a point on L to the chosen straight line. Expanding $f(x, y)$ by Taylor's formula

(Chapter II, §9) and noting that $f_x(0, 0) = f_y(0, 0) = 0$ (since the axes Ox and Oy lie in the tangent plane) we get

$$f(x, y) = \tfrac{1}{2}(f_{xx}x^2 + f_{yy}y^2) + \epsilon(x^2 + y^2),$$

where $\epsilon \to 0$ as $x \to 0$, $y \to 0$. For a point on L, we have $x = \xi\cos\phi$, $y = \xi\sin\phi$, $\xi^2 = x^2 + y^2$ (figure 27), and thus

$$k_L = \lim_{\xi\to 0}\frac{f_{xx}\xi^2\cos^2\phi + f_{yy}\xi^2\sin^2\phi + 2\epsilon\xi^2}{\xi^2} = f_{xx}\cos^2\phi + f_{yy}\sin^2\phi.$$

Putting $\phi = 0$, $\phi = \pi/2$, we find that f_{xx} and f_{yy} are the curvatures k_1 and k_2 of the normal sections in the direction of the axes Ox and Oy. Thus the formula derived is actually Euler's formula: $k = k_1\cos^2\phi + k_2\sin^2\phi$. The fact that k_1 and k_2 are the maximal and minimal curvatures also follows from this formula.

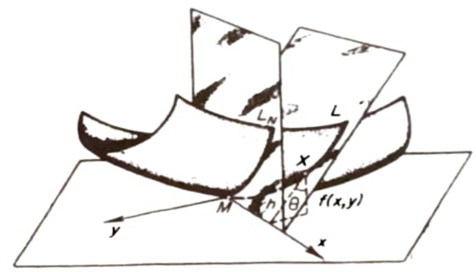

FIG. 28.

2. For the proof of Meusnier's theorem we consider a normal section L_N and a section L whose plane forms an angle θ with the plane of the section L_N, as in figure 28. The axes Ox and Oy lie in the tangent plane, and we also take the axis Ox to be tangent to the curves L_N and L at the origin. The distance $h(x, y)$ to the Ox axis of a point X on L with coordinates x, y, $f(x, y)$ is obviously equal to $h(x, y) = |f(x, y)|/\cos\theta$ (figure 28). Using Taylor's formula, we express the curvature k_L of the curve L in the following manner:

$$k_L = \lim_{x\to 0}\frac{2h(x, y)}{x^2} = \lim_{x\to 0} 2\frac{|f(x, y)|}{x^2\cos\theta}$$
$$= \lim_{x\to 0}\frac{|f_{xx}x^2 + 2f_{xy}xy + f_{yy}y^2 + 2\epsilon(x^2 + y^2)|}{x^2\cos\theta}, \quad (5)$$

where $\epsilon \to 0$ as $x, y \to 0$. Since the axis Ox is tangent to the curve L, obviously $\lim_{x\to 0} y/x = 0$. Thus, taking the limit in formula (5), we get

$$k_L = \frac{|f_{xx}|}{\cos\theta}.$$

But for the chosen coordinate system the curve L_N has the equation $z = f(x, 0)$, for which $|k_N| = |f_{xx}|$. Thus $k_L = |k_N|/\cos\theta$ and Meusnier's theorem is proved.

Mean curvature. In many questions of the theory of surfaces, the most important role is played not by the principal curvatures themselves but by certain quantities dependent on them, namely the *mean curvature* and the *Gaussian* or *total curvature* of the surface at a given point. Let us examine them in detail.

The mean curvature of a surface at a given point is the average of the principal curvatures

$$K_{av} = \tfrac{1}{2}(k_1 + k_2).$$

As an example of the usefulness of this concept, we consider the following mechanical problem. We assume that over the surface of some body F there is stretched a taut elastic rubber film. We ask about the pressure exerted by this film on each point of the surface of F.

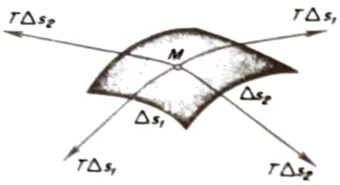

FIG. 29.

The pressure at a point M is measured by the force exerted by the film on a segment of the surface of unit area containing the point M; to be more exact, the pressure "at the point" M is defined as the limit of the ratio of this force to the area of the segment as the latter shrinks to the point M.

We surround the point M on the surface with a small curvilinear rectangle whose sides have lengths Δs_1 and Δs_2 and are perpendicular to the first and second principal directions at M (figure 29).* On each side of the rectangle there is exerted a force that is proportional (from the assumed uniformity of the tension) to the length of the side and the tension T acting on the film. Thus, on the sides perpendicular to the first principal direction, there are exerted forces that are approximately equal to $T\Delta s_1$ and have the direction of the tangent to the surface. Similarly, forces equal to $T\Delta s_2$ act on the other pair of sides of the rectangle. In order to find the pressure at the point M, we must divide the resultant of these four forces by the area of the rectangle (approximately equal to $\Delta s_1 \Delta s_2$) and pass to the limit for $\Delta s_1, \Delta s_2 \to 0$. Let us begin by dividing the resultant of the first two forces by $\Delta s_1 \Delta s_2$.

If we examine the rectangle from its side (figure 30), we see that these forces are directed along tangents to the curve of the first normal section and that the distance between their points of application is exactly Δs_2. So we have the same problem here as in §2 for the pressure of a string on a support. Using the earlier result,

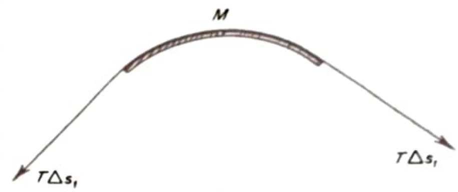

FIG. 30.

we find that the desired limit is equal to $k_1 T$, where k_1 is the curvature of the first normal section. With a similar expression for the other two forces, we obtain the formula:

$$P_M = T(k_1 + k_2) = 2TK_{av}.$$

This result has many important consequences. Let us consider an example.

It is known that the surface film of a liquid is under a tension that is the same in all directions on the surface. For a mass of liquid bounded by a curved surface, this tension, by the previous result, exerts a pressure on the surface which is proportional to its mean curvature at the given point.

So in drops of very small diameter the pressures are very large, a fact that hinders the formation of such drops. In a cooling vapor the drops begin to form, as a rule, around specks of dust and around charged particles. In a completely pure, slightly cooled vapor, the formation of drops is delayed. But if, for example, a particle passes through the vapor at high speed, causing ionization of the molecules, then around the ions formed in its path there will momentarily appear small drops of vapor, constituting a visible track of the particle. This is the basis for construction of the Wilson chamber, widely used in nuclear physics for observing the motions of various charged particles.

* Our reasoning here is not rigorous. However, by making estimates of the errors introduced, it is possible to give a rigorous proof of the result.

Since the pressure exerted by a liquid is the same in all directions, a drop of liquid in the absence of other sources of pressure must assume a form for which at all points of the surface the mean curvature is the same. In the experiment of Plateau, we take two liquids of the same specific weight, so that a clot of one of them will float in equilibrium in the other. It may be assumed that the floating liquid is acted on only by surface tension,* and it turns out that the "floating" liquid always takes the form of a sphere. This result suggests that every closed surface with constant mean curvature is a sphere, a theorem that is in fact true, although the strict mathematical proof of it is very difficult.

It is possible to approach the question from still another side. In view of the fact that the surface tension tends to decrease the area of the surface, while the volume of the liquid cannot change, it is natural to expect that the floating mass of liquid will have the smallest surface for a given volume. It can be proved that a body with this property is a sphere.

The relation between the lateral pressure of the film and its mean curvature can also be used to determine the form of a soap film suspended in a contour. Since the lateral pressure over the surface of the film, being directed along the normal to the surface, is not opposed by any reaction of the support (the support in this case is simply not there), it must be equal to zero, so that for the desired surface we have the condition

$$K_{av} = 0. \tag{6}$$

From the analytic expression for mean curvature, we obtain a differential equation, and the problem consists of solving this equation under the condition that the desired surface passes through the given contour.† There have been many investigations of this difficult problem.

The same equation (6) arises from the problem of finding the surface of least area bounded by a given contour. From a physical point of view, the identity of these two problems is a natural one, since the film tends to decrease its area and reaches a position of stable equilibrium only when it attains the minimal area possible under the given conditions. Surfaces of zero mean curvature, by reason of their connection with this problem, are called *minimal*.

The mathematical investigation of minimal surfaces is of great interest, partly because of their wide variety of essentially different shapes, as

* The increase of pressure with depth may be ignored, since it is the same for both liquids because of their having the same specific weight. So on their common boundary the additional internal and external pressures caused by the depth are neutralized by each other.

† For a surface given by the equation $z = z(x, y)$, equation (6) assumes the form

$$(1 + z_y'^2)z''_{xx} - 2z'_x z'_y z''_{xy} + (1 + z_x'^2)z''_{yy} = 0.$$

discovered by experiments with soap film. Figure 31 illustrates two soap films suspended from different contours.

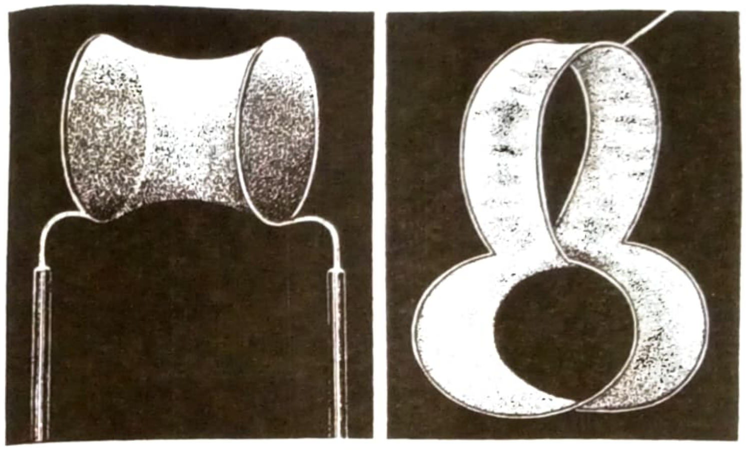

FIG. 31.

Gaussian curvature. The *Gaussian curvature* of a surface at a given point is the product of the principal curvatures

$$K = k_1 k_2 .$$

The sign of the Gaussian curvature defines the character of the surface near the point under consideration. For $K > 0$ the surface has the form of a bowl (k_1 and k_2 have the same sign) and for $K < 0$, when k_1 and k_2 have different signs, the surface is like a saddle. The remaining cases, discussed earlier, correspond to zero Gaussian curvature. The absolute value of the Gaussian curvature gives the degree of curvature of the surface in general, as a sort of abstraction from the various curvatures in different directions. This becomes particularly clear if we consider a different definition of Gaussian curvature, which does not depend on investigating curves on the surface.

Let us consider a small segment G of the surface F, containing the point M in its interior, and at each point of this segment let us erect a normal to the surface.

If we translate the initial points of all these normals to one point, then they fill out a solid angle (figure 32). The size of this solid angle will depend on the area of the segment G and on the extent to which the surface is curved on this segment. Thus the degree of curvature of

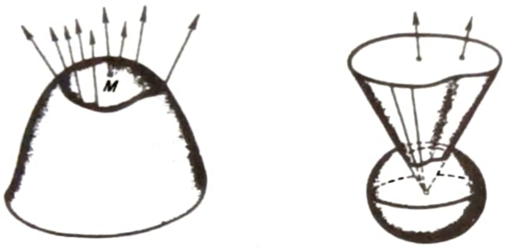

FIG. 32.

the segment G may be characterized by the ratio of the size of the solid angle to the area of G; so it is natural to define the curvature of the surface at a given point as the limit of this ratio when the segment G shrinks to the point M.* It turns out that this limit is equal to the absolute value of the Gaussian curvature at the point M.

The most remarkable property of the Gaussian curvature, which explains its great significance in the theory of surfaces, is the following. Let us suppose that the surface has been stamped out from a flexible but inextensible material, say a very thin sheet of tin, so that we can bend it into various shapes without stretching or tearing it. During this process the principal curvatures will change but, as Gauss showed, their product k_1k_2 will remain unchanged at every point. This fundamental result shows that two surfaces with different Gaussian curvatures are inherently distinct from each other, the distinction consisting of the fact that if we deform them in every possible way, without stretching or tearing, we can never superpose them on each other. For example, a segment of the surface of a sphere can never be distorted so as to lie on a plane or on the surface of a sphere of different radius.

We have now considered certain basic concepts in the theory of surfaces. As for the methods used in this theory, they consist, as was stated previously, primarily in the application of analysis and above all of

* To measure the solid angle itself, we construct a sphere of unit radius with center at its vertex. The area of the region in which the sphere intersects the solid angle is then taken as the size of the solid angle (figure 32).

differential equations. Simple examples of the use of analysis are to be found in the proofs for the theorems of Euler and Meusnier. For more complicated questions, we require a special method of relating problems in the theory of surfaces to problems in analysis. This method is based on the introduction of so-called curvilinear coordinates and was first widely used in the work of Gauss on problems of the type discussed in the following section.

§4. Intrinsic Geometry and Deformation of Surfaces

Intrinsic geometry. As indicated previously, a deformation of a surface is defined as a change of shape that preserves the lengths of all curves lying in the surface. For example, rolling up a sheet of paper into a cylindrical tube represents, from the geometric point of view, a deformation of part of the plane, since in fact the paper undergoes practically no stretching, and the length of any curve drawn on it is not changed by its being rolled up. Certain other geometric quantities connected with the surface are also preserved; for example, the area of figures on it. All properties of a surface that are not changed by deformations make up what is called the *intrinsic geometry* of the surface.

But just which are these properties? It is clear that in a deformation only those properties can be preserved which in the final analysis depend entirely on lengths of curves, i.e., which may be determined by measurements carried out on the surface itself. A deformation is a change of shape preserving the length of curves, and any property which cannot change under *any* deformation must be definable in one way or another in terms of length. Thus intrinsic geometry is simply called *geometry on a surface*. The very meaning of the words "intrinsic geometry" is that it studies intrinsic properties of the surface itself, independent of the manner in which the surface is embedded in the surrounding space.* Thus, for example, if we join two points on a sheet of paper by a straight line and then bend the paper (figure 33), the segment becomes a curve but its property of being the shortest of all lines joining the given points on the surface is preserved; so this property belongs to intrinsic geometry. On the other hand, the curvature of this line will depend on how the paper was bent and thus is not a part of intrinsic geometry.

In general, since the proofs of plane geometry make no reference to the properties of the surrounding space, all its theorems belong to the

* We note that the ideas of intrinsic geometry have led to a wide generalization of the mathematical concept of space and have thereby played a very important role in contemporary physics; for details see Chapter XVII.

intrinsic geometry of any surface obtainable by deformation of a plane. One may say that plane geometry is the intrinsic geometry of the plane.

Another example of intrinsic geometry is familiar to everyone, namely geometry on the surface of a sphere, with which we usually have to deal

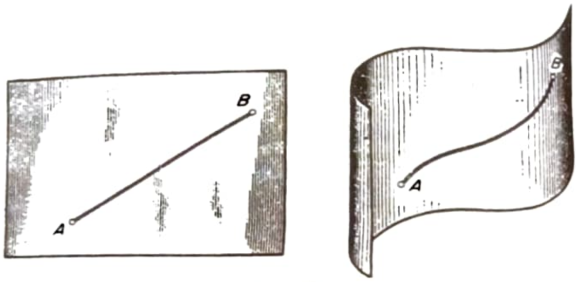

FIG. 33.

in making measurements on the surface of the earth. This example is a particularly good one to illustrate the essential nature of intrinsic geometry; because of the large radius of the earth, any immediately visible area of its surface appears to us as part of a plane, so that the deviations from plane geometry observable in the measurements of large distances impresses us as resulting not from the curvature of the earth's surface in space but from the inherent laws of "terrestrial geometry," expressing the geometric properties of the surface of the earth itself.

It remains to note that the idea of studying intrinsic geometry occurred to Gauss in connection with the problems of geodesy and cartography. Both these applied sciences are concerned in an essential way with the intrinsic geometry of the earth's surface. Cartography deals, in particular, with distortions in the ratios of distances when part of the surface of the earth is mapped on a plane and thus with distinguishing between plane geometry and the intrinsic geometry of the surface of the earth.

The intrinsic geometry of any surface may be pictured in the same way. Let us imagine that on a given surface there exist creatures so small that within the limits of their range of vision the surface appears to be plane (we know that a sufficiently small segment of any smooth surface differs very little from a tangent plane); then these creatures will not notice that the surface is curved in space, but in measuring large distances they will nevertheless convince themselves that in their geometry certain

nonplanar laws prevail, corresponding to the intrinsic geometry of the surface on which they live. That these laws are actually different for different surfaces may easily be seen from the following simple discussion. Let us choose a point O on the surface and consider a curve L such that the distance of each of its points from the point O, measured on the surface (i.e., along the shortest curve connecting this point to the point O) is equal to a fixed number r (figure 34). The curve L, from the

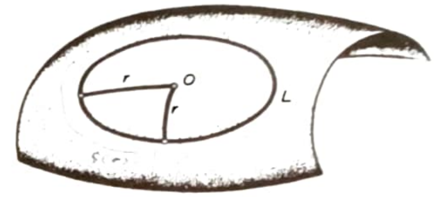

FIG. 34.

point of view of the intrinsic geometry of the surface, is simply the circumference of a circle of radius r. A formula expressing the length $s(r)$ in terms of r is part of the intrinsic geometry of the given surface. But such a formula may vary widely in character, depending on the nature of the surface: Thus on a plane, $s(r) = 2\pi r$; on a sphere of radius R, as can easily be shown, $s(r) = 2\pi R \sin r/R$; on the surface illustrated in figure 35, beginning with a certain value of r, the length of the circumference with center O and radius r is at first independent of r but then begins to decrease. Consequently, all these surfaces have different intrinsic geometries.

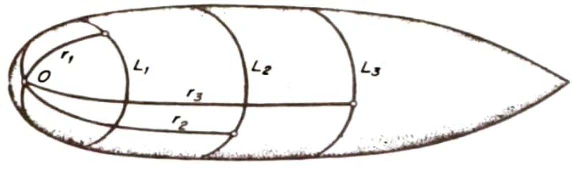

FIG. 35.

The basic concepts of intrinsic geometry. To illustrate the wide range of concepts and theorems in intrinsic geometry, we may turn to plane geometry which, as we have seen, is the intrinsic geometry of the plane. Its subject matter consists of plane figures and their properties, which are usually expressed in the form of relations among basic geometric quantities such as length, angle, and area. For a rigorous proof that angle and area belong to the intrinsic geometry of the plane, it is necessary to show that they can be expressed in terms of length. But this is certainly so; in fact, an angle may be computed if we know the length of the sides of a triangle containing it, and the area of a triangle can also be computed in terms of its sides, while to compute the area of a polygon we need only divide it into triangles.

In considering plane geometry as the intrinsic geometry of the plane, there is no need to restrict ourselves to ideas learned in school. On the contrary, we may develop it as far as we like and study many new problems, provided only they can be stated, in the final analysis, in terms of length. Thus, in plane geometry we may successively introduce the length of a curve, the area of a surface bounded by curves, and so forth; they are all a part of the intrinsic geometry of the plane.

The same concepts are introduced in the intrinsic geometry of an arbitrary surface. The length of a curve is the initial concept; the definition of angles and areas is somewhat more complicated. If the intrinsic geometry of a given surface differs from plane geometry, we cannot use the customary formulas to define an angle or an area in terms of length. However, as we have seen, a surface near a given point differs little from its tangent plane. Speaking more precisely, the following is true: If a small segment of a surface containing a given point M is projected on the tangent plane at this point, then the distance between points, measured on the surface, differs from the distance between their projections by an infinitesimal of higher than the second order in comparison with distances from the point M. Thus in defining geometric quantities

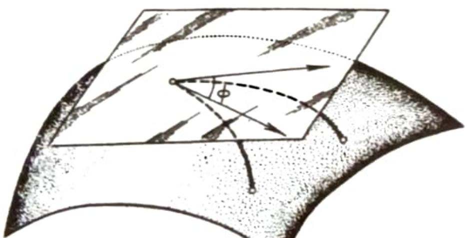

FIG. 36.

at a given point of a surface by taking a limit in which infinitesimals occur of order no higher than the second, we may replace a segment of the surface by its projection on the tangent plane. Thus the quantities determined by measurement in the tangent plane turn out to belong to the intrinsic geometry of the surface. This possibility of considering a small segment of the surface as a plane is the basis of the definitions of all the concepts of intrinsic geometry.

As an example let us consider the definitions of angle and area. Following the general principle, we define the angle between curves on a surface as the angle between their projections on the tangent plane (figure 36). Obviously the angle defined in this manner is identical with the angle between the tangents to the curves. The definition of area given in §3 is based on the same principle. Finally, in order that the tendency of a curve to twist in space may be defined "within" the surface itself, we introduce the concept of "geodesic curvature" the name being reminiscent of measurements on the surface of the earth. The *geodesic curvature* of a curve at a given point is defined as the curvature of its projection on the tangent plane (figure 37).

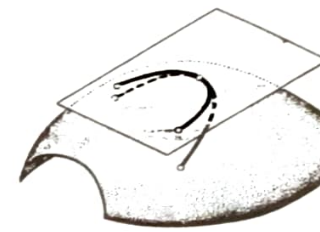

FIG. 37.

In this manner we see that the basic concepts of plane geometry may be introduced into the intrinsic geometry of an arbitrary surface.

In any arbitrary surface it is also easy to define figures analogous to the basic figures on the plane. For example, we have been dealing previously with circumferences of circles, which are defined precisely as in the case of the plane. Similarly, we may define the analogue of a line segment, namely a geodesic segment, as the shortest curve on the surface joining two given points. Further, it is natural to define a triangle as a figure bounded by three geodesic segments and similarly for a polygon, and so forth. Since the properties of all these figures and magnitudes depend on the surface, there exist in this sense infinitely many different intrinsic geometries. But intrinsic geometry, as a special branch of the theory of surfaces, pays particular attention to certain general laws holding for the intrinsic geometry of any surface and makes clear how these laws are expressed in terms of the quantities which characterize a given surface.

Thus, as we have noted earlier, one of the most important characteristics of a surface, its Gaussian curvature, is not changed by deformation, i.e., depends only on the intrinsic geometry of the surface. But it turns out that in general the Gaussian curvature already characterizes, to a remarkable degree, the extent to which the intrinsic geometry of the surface near a given point differs from plane geometry. As an example let us consider on a surface a circle L of very small radius r, with center at a given point O. On a plane the length $s(r)$ of its circumference is expressed by the formula $s(r) = 2\pi r$. On a surface differing from a plane, the dependence of the circumference on the radius is different; here the deviation of $s(r)$ from $2\pi r$, depends essentially, for small r, on the Gaussian curvature K at the center of the circle, namely;

$$s(r) = 2\pi r - \frac{\pi}{3} Kr^3 + \epsilon r^3,$$

where $\epsilon \to 0$ as $r \to 0$. In other words, for small r the circumference may be computed by the usual formula if we disregard terms of the third degree of smallness, and in this case the error (with accuracy to terms of higher than the third order) is proportional to the Gaussian curvature. In particular, if $K > 0$, then the circumference of a circle of small radius is smaller than the circumference of a circle with the same radius in a plane, and if $K < 0$, it is larger. These latter facts are easy to visualize: Near a point with positive curvature the surface has the shape of a bowl so that circumferences are reduced, whereas near a point with negative curvature the circumference, being situated on a "saddle," has a wavelike shape and is thus considerably lengthened (figure 38).

FIG. 38.

From the theorem just mentioned, it follows that a surface with varying Gaussian curvature is extremely inhomogeneous from a geometric point of view; the properties of its intrinsic geometry change from point to point. The general character of the problems of intrinsic geometry causes

it to resemble plane geometry, but this inhomogeneity, on the other hand, makes it profoundly different from plane geometry. On the plane, for example, the sum of the angles of a triangle is equal to two right angles; but on an arbitrary surface the sum of the angles of a triangle, (with geodesics for sides) is undetermined even if we are told that it lies on a known surface and has sides of given length. However, if we know the Gaussian curvature K at every point of the triangle, then the sum of its angles, α, β, γ, can be computed by the formula

$$\alpha + \beta + \gamma = \pi + \iint K \, d\sigma,$$

where the integral is taken over the surface of the triangle. This formula contains as a special case the well-known theorems on the sum of the angles of a triangle in the plane and on the unit sphere. In the first case $K = 0$ and $\alpha + \beta + \gamma = \pi$, while in the second $K = 1$ and $\alpha + \beta + \gamma = \pi + S$, where S is the area of the spherical triangle.

It may be proved that every sufficiently small segment of a surface with zero Gaussian curvature may be deformed, or, as it is customary to say, developed into a plane, since it has the same intrinsic geometry as the plane. Such surfaces are called *developable*. And if the Gaussian curvature is near zero, then although the surface cannot be developed into a plane, still its intrinsic geometry differs little from plane geometry, which indicates once again that the Gaussian curvature acts as a measure of the extent to which the intrinsic geometry of a surface deviates from plane geometry.

Geodesic lines. In the intrinsic geometry of a surface the role of straight lines is played by geodesic lines, or, as they are usually called, "geodesics."

A straight line in a plane may be defined as a line made up of intervals overlapping one another. A geodesic is defined in exactly the same way, with geodesic segments taking the place of intervals. In other words, a *geodesic* is a curve on a surface such that every sufficiently small piece of it is a shortest path. Not every geodesic is a shortest path in the large, as may be noted on the surface of a sphere, where every arc of a great circle is a geodesic, although this arc will be the shortest path between its end points only if it is not greater than a semicircle. A geodesic, as we see, may even be a closed curve.

To illustrate certain important properties of geodesics, let us consider the following mechanical model.* On the surface F let there be stretched

* As noted previously, our reasoning here is not a strict proof of the properties of geodesic curves. It is given only to illustrate the most important of these properties.

a rubber string with fixed ends (figure 39).* The string will be in equilibrium when it has the shortest possible length, since any change in its position will then involve an increase of length, which could be produced only by external forces. In other words, the string will be in equilibrium if it is lying along a geodesic. But for equilibrium, it is necessary that the elastic forces on each segment of the string be counterbalanced by the resistance of the surface, directed along the normal to it. (We assume that the surface is smooth and that there is no friction between it and the string.)

FIG. 39.

But it was proved in §2 that the pressure on the support caused by the tension of the string is directed along the principal normal to the curve along which the string lies. Thus we are led to the following result: The principal normal to a geodesic at each point coincides in direction with the normal to the surface. The converse of this theorem is also true: Every curve on a regular surface which has this property is a geodesic.

This property of a geodesic allows us to deduce the following important fact: If a material point is moving on a surface in such a way that there are no forces acting on it except for the reaction of the surface, then it follows a geodesic. For, as we know from §2, the normal acceleration of a point is directed along the principal normal to the trajectory and since the reaction of the surface is the only force acting on the point, the principal normal to the trajectory is identical with the normal to the surface, so that from the preceding theorem the trajectory is a geodesic. This last property of geodesics increases their resemblance to straight lines. Just as the motion of a free point, because of inertia, is along a straight line, so the motion of a point forced to stay on a surface, but not affected by external forces, will be along a geodesic.†

From the same property of geodesics comes the following theorem. If two surfaces are tangent along a curve that is a geodesic on one of them, then this curve will also be a geodesic on the other. For at each point of the curve, the surfaces have a common tangent plane and consequently a common normal, and since the curve is a geodesic on one of the surfaces, this normal coincides with the principal normal to the curve, so that on the second surface also the curve will be a geodesic.

* A stretched string will not remain on a surface unless the surface is convex; so in order not to make exceptions, it is better to imagine that the surface is in two layers, with the string running between them.

† Here by "external" forces we mean all forces except the reaction of the surface.

From these results follow two further intuitive properties of geodesic curves. In the first place, if an elastic rectangular plate (for example a steel ruler) lies with its median line completely on a surface, then it is tangent to this surface along a geodesic. (Evidently the line of contact is a geodesic on the ruler, so that it must be a geodesic also on the surface.) Second, if a surface rolls along a plane in such a way that the point of contact traces a straight line on the plane, then the trace of this straight line on the surface is a geodesic.* Both these properties are readily demonstrated on a cylinder, where it is easy to convince oneself by experiment that the median line of a straight plane strip lying on the cylinder (figure 40) coincides with either a generator or the circumference of a

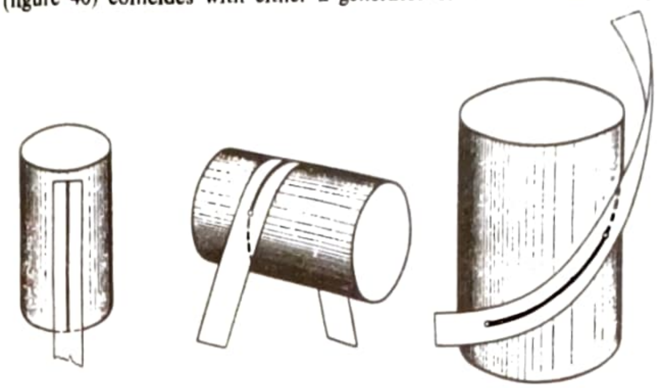

FIG. 40.

circle or a helix, and it is not difficult to prove that a geodesic curve on a cylinder can be only one of these three. The same curves will be traced out on a cylinder if we roll it on a plane on which we have drawn a straight line in chalk.

The analogy between geodesics and straight lines in a plane may be supplemented by still another important property, taken directly from the definition of a geodesic. Namely, straight lines in the plane may be defined as curves of zero curvature and geodesics on a surface as curves of zero geodesic curvature. (We recall that the geodesic curvature is the curvature of the projection of the curve on the tangent plane, cf. figure 37.) It is quite natural that our present definition of a geodesic should coincide with the earlier one; for if at every point of the curve the curvature of

* This proposition does not differ essentially from the preceding one, since the rolling of a surface on a plane is equivalent in a well-defined sense to the unwinding of a plane strip along the surface.

its projection on the tangent plane is equal to zero, then the curve departs from its tangent essentially in the direction of the normal to the surface, so that the principal normal to the curve is directed along the normal to the surface and the curve is a geodesic in the original sense. Conversely, if a curve is a geodesic, then its principal normal, and so also its deviation from the tangent line, are directed along the normal to the surface, so that in projecting on the tangent plane we get a curve in which the deviation from the tangent is essentially smaller than for the original curve, and the curvature of the projection so formed turns out to be equal to zero.

The course of a geodesic may vary widely for different surfaces. As an example, in figure 41 we trace some geodesics on a hyperboloid of revolution.

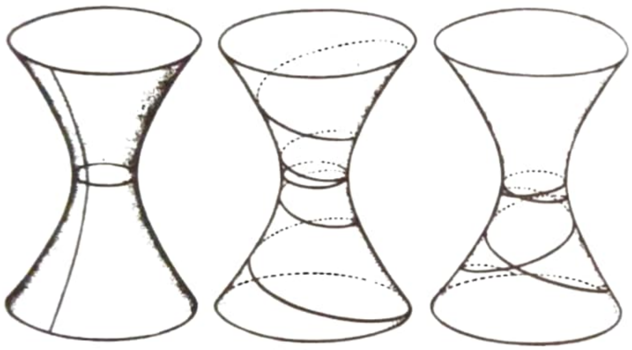

FIG. 41.

Deformation of surfaces. Since intrinsic geometry studies the properties of surfaces that are invariant under deformation, it naturally investigates these deformations themselves. The theory of deformation of surfaces is one of the most interesting and difficult branches of geometry and includes many problems which, although simple to state, have not yet been finally solved.

Certain questions about the deformation of surfaces were already considered by Euler and Minding, but general results for arbitrary surfaces were not derived until later.

In the general theory of deformation, we first of all raise the question whether deformation is possible for all surfaces and, if so, to what extent.

For analytic surfaces, i.e., surfaces defined by functions of the coordinates that can be expanded in a Taylor series, this question was solved at the end of the last century by the French mathematician Darboux. In particular, he showed the following: If on such a surface we consider any geodesic and assign in space an arbitrary (analytic) curve with the same length, and with curvature nowhere equal to zero, then a sufficiently narrow strip of the surface, containing the given geodesic, can be deformed so that the geodesic coincides with the given curve.* This theorem shows that a strip of the surface may be deformed rather arbitrarily. However, it has been proved that if a geodesic is to be transformed into a preassigned curve, then the surface may be deformed in no more than two ways. For example, if the curve is plane, then the two positions of the surface will be mirror images of each other in the plane. If the geodesic is a straight line, then this last proposition is not true, as can be shown by deforming a cylindrical surface.

We have defined a deformation as a transformation of the surface that preserves the lengths of all curves on the surface. Here we have considered only the final result of the transformation; the question of what happens to the curve during the process did not enter. However, in considering a surface as made from a flexible but unstretchable material, it is natural to consider a continuous transformation, at each instant of which the lengths remain unchanged (physically this corresponds to the unstretchability of the material). Such transformations are called *continuous* deformations.

At first glance it may seem that every deformation can be realized in a continuous manner, but this is not so. For example, it has been shown that a surface in the form of a circular trough (figure 42), does not admit

FIG. 42.

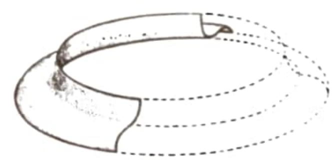

FIG. 43.

continuous deformations (this explains, among other things, the familiar fact that a pail with a curved rim is considerably stronger than one with a plain rim) although deformations of such a surface are possible: for

* The case of transforming a geodesic into a curve with zero curvature is excluded, since it is easy to show that for surfaces of positive Gaussian curvature this is impossible.

example, one may cut the trough along the circle on which it rests on a horizontal plane and replace one half of it by its mirror image in this plane (compare figure 43 with figure 42; to aid visualization we have drawn only the left half of the surface). It is intuitively clear that the impossibility of a continuous deformation is due to the circular shape of the trough; for a straight trough such a deformation can be performed continuously.

If we restrict ourselves to a sufficiently small segment of the surface, then there are no obvious hindrances to its continuous deformation, and we might expect that every deformation of a small segment of the surface can be realized by a continuous transformation, followed perhaps by a mirror reflection. This is in fact true, but only under the condition that on the given small segment of the surface the Gaussian curvature never vanishes (excepting the case that it vanishes everywhere). But if the Gaussian curvature vanishes at isolated points, then, as N. V. Efimov showed in 1940, even arbitrarily small segments of a regular surface may not admit any continuous deformation without loss of regularity. For example, the surface defined by the equation $z = x^9 + \lambda x^6 y^3 + y^9$, where λ is a transcendental number, has the property that no segment containing the origin, no matter how small it may be, admits sufficiently regular continuous deformations. Efimov's theorem is a new and somewhat unexpected result in classical differential geometry.

In addition to these general questions about deformation, a great deal of attention is being paid to special types of deformation of surfaces.

The connection of the intrinsic geometry of a surface with the form of the surface in space. We already know that certain properties of a surface, and of the figures on it, are defined by the intrinsic geometry of the surface even though these properties are very closely related to other properties that depend on how the surface is embedded in the surrounding space, properties that are, as they say, "extrinsic" to the surface. For example, the principal curvatures are extrinsic properties of a surface, but their product (the Gaussian curvature) is intrinsic. Another example, in order that the principal normal of a curve lying on a surface should coincide with the normal to the surface, it is necessary and sufficient that this curve have a property defined by its intrinsic geometry, namely that it be a geodesic.

Consequently, the intrinsic geometry of a surface will determine its space form only to a certain extent.

The dependence of the space form of a surface on its intrinsic geometry may be expressed analytically in the form of equations containing certain quantities that characterize the intrinsic geometry and certain other

quantities that characterize the way in which the curved surface is embedded in space. One of these equations is the formula expressing the Gaussian curvature in intrinsic terms and is due to Gauss. Two other such equations are those of Peterson and Codazzi, mentioned in §1.

The equations of Gauss, Peterson, and Codazzi completely express the connection between the intrinsic geometry of a surface and the character of its curvature in space, since all possible interrelations between intrinsic and extrinsic properties of an arbitrary surface are included, at least in implicit form, in these equations.

Since the form of a surface in space is not completely defined by its intrinsic geometry, we naturally ask, What extrinsic properties must still be assigned in order to determine the surface completely? It turns out that if two surfaces have the same intrinsic geometry and if, at corresponding points and in corresponding directions, the curvatures of the normal sections of these surfaces have the same sign, then the surfaces are congruent; that is, they can be translated so as to coincide with each other. We note that Peterson discovered this theorem 15 years earlier than Bonnet, with whose name it is usually associated.

Analytic apparatus in the theory of surfaces. The systematic application of analysis to the theory of surfaces led to the building up of an analytic apparatus especially suitable for this purpose. The decisive step in this direction was taken by Gauss, who introduced the method of representing surfaces by so-called curvilinear coordinates. This method is a natural generalization of the idea of Cartesian coordinates on the plane and is closely connected with the intrinsic geometry of the surface, for which the presentation of the surface by an equation of the form $z = f(x, y)$ is not convenient. The inconvenience consists of the fact that the x, y coordinates of a point on the surface change when the surface is deformed. To eliminate this difficulty, the coordinates are chosen on the surface itself; they define each point by two numbers u and v, which are associated with the given point and remain associated with it even after deformation of the surface. The space coordinates x, y, z of the point will in each case be functions of u and v. The numbers u and v defining the point on a surface are called its *curvilinear coordinates*. The choice of name is to be understood as follows: If we fix the value of one of these coordinates, say v, and vary the other, then we get a coordinate curve on the surface. The coordinate curves form a curvilinear net on the surface, similar to the coordinate net on a plane. We note that the familiar method of describing the position of a point on the surface of the earth by means of longitude and latitude consists simply of introducing curvilinear coordinates on the surface of a sphere; the coordinate net in this case consists of circles, namely the meridians and parallels* (figure 44). To describe the spatial position of a surface by means of curvilinear coordinates, we need to define the position of each point in terms of u and v, for example by giving, as a function of u and v, the vector $\boldsymbol{r} = \boldsymbol{r}(u, v)$, issuing from some fixed origin to the points on the surface and called the radius vector of the surface. (This is equivalent to giving the x, y, and z components of the vector $\boldsymbol{r}$ as functions of u and v.)† To define a curve lying on a given surface, we need to give the coordinates u, v as functions of one parameter t; then the radius vector to a point moving along this curve is expressed as a composite function

$$\boldsymbol{r}[u(t), v(t)].$$

For vector functions the concepts of derivative and differential may be generalized word for word; from the definition of the derivative as the limit of $\Delta \boldsymbol{r}/\Delta t$ when $\Delta t \to 0$ ($\boldsymbol{r}$ is a function of the parameter t) it follows

FIG. 44.

FIG. 45.

* It is characteristic that geographic coordinates and their practical applications were known long in advance of Descartes' introduction of the usual coordinates in the plane.

† Of course Gauss did not use vector notation, but defined the three coordinates x, y, z of the points of the surface separately as functions of u and v. Vectors, which were introduced as a result of the work of Hamilton and Grassmann, were at first used widely in physics and only later (in fact, in the 20th century) became the traditional apparatus for analytic and differential geometry.

at once that the derivative of the radius vector of a curve is a vector directed along the tangent to the curve (figure 45). For vector functions the basic properties of ordinary derivatives are still valid; for example, the chain rule

$$\frac{d\boldsymbol{r}[u(t), v(t)]}{dt} = \frac{\partial \boldsymbol{r}}{\partial u}\frac{du}{dt} + \frac{\partial \boldsymbol{r}}{\partial v}\frac{dv}{dt} = \boldsymbol{r}_u u'_t + \boldsymbol{r}_v v'_t, \qquad (7)$$

where $\boldsymbol{r}_u$ and $\boldsymbol{r}_v$ are the partial derivatives of the vector function $\boldsymbol{r}(u, v)$.

The length of a curve, as can be shown, is expressed by the integral

$$s = \int \sqrt{x'^2(t) + y'^2(t) + z'^2(t)}\, dt.$$

Thus, the differential of the length of a curve is equal to

$$ds = \sqrt{x'^2(t) + y'^2(t) + z'^2(t)}\, dt.$$

But since $x'(t)$, $y'(t)$, and $z'(t)$ are components of the vector $d\boldsymbol{r}/dt = \boldsymbol{r}'_t$, we may write $ds = |\boldsymbol{r}'_t|\, dt$, where $|\boldsymbol{r}'_t|$ denotes the length of the vector $\boldsymbol{r}'_t$. For curves lying on a surface, we get from (7)

$$ds = |\boldsymbol{r}_u u'_t + \boldsymbol{r}_v v'_t|\, dt.$$

Computing the square of the length of the vector on the right we obtain, by the rules of vector algebra,*

$$ds^2 = [\boldsymbol{r}_u^2 u_t'^2 + 2\boldsymbol{r}_u \boldsymbol{r}_v u'_t v'_t + \boldsymbol{r}_v^2 v_t'^2]\, dt^2.$$

Passing to differentials and introducing the notation

$$\boldsymbol{r}_u^2 = E(u, v), \ \boldsymbol{r}_u \boldsymbol{r}_v = F(u, v), \ \boldsymbol{r}_v^2 = G(u, v),$$

we have

$$ds^2 = E\, du^2 + 2F\, du\, dv + G\, dv^2.$$

We see that the square of the differential of arc length on a surface is a quadratic form in the differentials du and dv with coefficients depending on the point of the surface. This form is called the *first fundamental quadratic form* of the surface. Given the coefficients E, F, and G of this

* The square of the length of a vector is the scalar product of the vector with itself, and for scalar multiplication (cf. Chapter III, §9) the usual rules hold for the removal of brackets.

form at each point on a surface we may compute the length of any curve on the surface by the formula

$$s = \int_{t_1}^{t_2} \sqrt{Eu_t'^2 + 2Fu_t'v_t' + Gv_t'^2}\, dt,$$

so that its intrinsic geometry is thereby completely determined.

We show, as an example, how to express angle and area in terms of E, F, and G. Let two curves issue from a given point, one of them given by the equations $u = u_1(t)$, $v = v_1(t)$ and the other by the equations $u = u_2(t)$, $v = v_2(t)$. Then the tangents to these curves are given by the vectors

$$\boldsymbol{r}_1 = \boldsymbol{r}_u \frac{du_1}{dt} + \boldsymbol{r}_v \frac{dv_1}{dt},$$

$$\boldsymbol{r}_2 = \boldsymbol{r}_u \frac{du_2}{dt} + \boldsymbol{r}_v \frac{dv_2}{dt}.$$

The cosine of the angle between these vectors is equal to the scalar product $\boldsymbol{r}_1\boldsymbol{r}_2$ divided by the product of the lengths $r_1 r_2$

$$\cos\alpha = \frac{\boldsymbol{r}_1\boldsymbol{r}_2}{r_1 r_2}$$

$$= \frac{\boldsymbol{r}_u^2 \dfrac{du_1}{dt}\dfrac{du_2}{dt} + \boldsymbol{r}_u\boldsymbol{r}_v\left(\dfrac{du_1}{dt}\dfrac{dv_2}{dt} + \dfrac{du_2}{dt}\dfrac{dv_1}{dt}\right) + \boldsymbol{r}_v^2 \dfrac{dv_1}{dt}\cdot\dfrac{dv_2}{dt}}{r_1 r_2}.$$

Recalling that $\boldsymbol{r}_u^2 = E, \boldsymbol{r}_u\boldsymbol{r}_v = F, \boldsymbol{r}_v^2 = G$, we get

$$\cos\alpha = \frac{E\dfrac{du_1}{dt}\dfrac{du_2}{dt} + F\left(\dfrac{du_1}{dt}\dfrac{dv_2}{dt} + \dfrac{du_2}{dt}\dfrac{dv_1}{dt}\right) + G\dfrac{dv_1}{dt}\dfrac{dv_2}{dt}}{\sqrt{E\left(\dfrac{du_1}{dt}\right)^2 + 2F\dfrac{du_1}{dt}\dfrac{dv_1}{dt} + G\left(\dfrac{dv_1}{dt}\right)^2}\sqrt{E\left(\dfrac{du_2}{dt}\right)^2 + 2F\dfrac{du_2}{dt}\dfrac{dv_2}{dt} + G\left(\dfrac{dv_2}{dt}\right)^2}}.$$

To obtain a formula for area, we consider a curvilinear rectangle bounded by the coordinate curves $u = u_0$, $v = v_0$, $u = u_0 + \Delta u$, $v = v_0 + \Delta v$, and we take as an approximation to it the parallelogram lying in the tangent plane and bounded by the vectors $\boldsymbol{r}_u \Delta u$, $\boldsymbol{r}_v \Delta v$, tangent to the coordinate curves (figure 46). The area of this parallelogram is

$\Delta s = |\boldsymbol{r}_u||\boldsymbol{r}_v| \Delta u\, \Delta v \sin\phi$, where ϕ is the angle between $\boldsymbol{r}_u$ and $\boldsymbol{r}_v$. Since $\sin\phi = \sqrt{1 - \cos^2\phi}$, it follows that $\Delta s = |\boldsymbol{r}_u||\boldsymbol{r}_v| \Delta u\, \Delta v \sqrt{1 - \cos^2\phi}$ $= \sqrt{\boldsymbol{r}_u^2\boldsymbol{r}_v^2 - |\boldsymbol{r}_u|^2|\boldsymbol{r}_v|^2\cos^2\phi}\, \Delta u\, \Delta v$. Recalling that $\boldsymbol{r}_u^2 = E$, $\boldsymbol{r}_v^2 = G$, $|\boldsymbol{r}_u|\cdot|\boldsymbol{r}_v|\cos\phi = \boldsymbol{r}_u\boldsymbol{r}_v = F$, we get $\Delta s = \sqrt{EG - F^2}\, \Delta u\, \Delta v$. Summing up the areas of the parallelograms and taking the limit as $\Delta u, \Delta v \to 0$ we obtain the formula for area $S = \iint_D \sqrt{EG - F^2}\, du\, dv$, where the integration is taken over the domain D of the variables u and v which describe the given segment of the surface.

In this way, curvilinear coordinates are very convenient for studying the intrinsic geometry of a surface.

It also turns out that the manner in which a curved surface is embedded in the surrounding space can be characterized by a certain quadratic form in the differentials du, dv. Thus if $\boldsymbol{n}$ is a unit vector normal to the surface at the point M, and $\Delta\boldsymbol{r}$ is the increment in the radius vector to

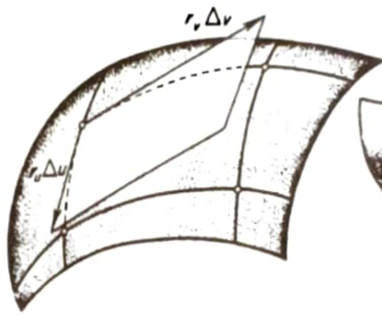

FIG. 46.

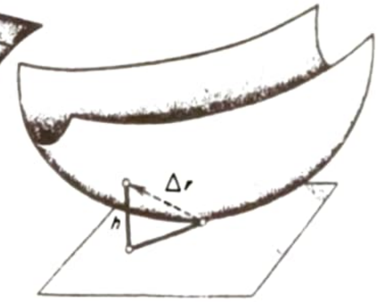

FIG. 47.

the surface as we move from this point, then the deviation h of the surface from the tangent plane (figure 47) is equal to $\boldsymbol{n}\,\Delta\boldsymbol{r}$. Expanding the increment $\Delta\boldsymbol{r}$ by Taylor's formula, we get

$$h = \boldsymbol{n}\, d\boldsymbol{r} + \tfrac{1}{2}\boldsymbol{n}\, d^2\boldsymbol{r} + \epsilon(du^2 + dv^2),$$

where $\epsilon \to 0$ as $\sqrt{du^2 + dv^2} \to 0$. Since the vector $d\boldsymbol{r}$ lies in the tangent plane, we have $\boldsymbol{n}\, d\boldsymbol{r} = 0$. The last term, $\epsilon(du^2 + dv^2)$ is small in comparison with the squares of the differentials du and dv. There remains the principal term $\frac{1}{2}\boldsymbol{n}\, d^2\boldsymbol{r}$. Thus twice the principal part of h, namely $\boldsymbol{n}\, d^2\boldsymbol{r}$, is a quadratic form with respect to du and dv

$$\boldsymbol{n}\, d^2\boldsymbol{r} = \boldsymbol{n}\boldsymbol{r}_{uu}\, du^2 + 2\boldsymbol{n}\boldsymbol{r}_{uv}\, du\, dv + \boldsymbol{n}\boldsymbol{r}_{vv}\, dv^2.$$

This form describes the character of the deviation of the surface from the tangent plane. It is called the *second fundamental quadratic form* of the surface. Its coefficients, which depend on u and v, are usually written:

$$\boldsymbol{n}\boldsymbol{r}_{uu} = L, \quad \boldsymbol{n}\boldsymbol{r}_{uv} = M, \quad \boldsymbol{n}\boldsymbol{r}_{vv} = N.$$

Knowing the second fundamental quadratic form, we can compute the curvature of any curve on a surface. Thus, applying the formula $k = \lim_{l \to 0} 2h/l^2$, we obtain the result that the curvature of the normal section in the direction corresponding to the ratio du/dv is equal to

$$k_n = \frac{\boldsymbol{n}\, d^2\boldsymbol{r}}{ds^2} = \frac{L\, du^2 + 2M\, du\, dv + N\, dv^2}{E\, du^2 + 2F\, du\, dv + G\, dv^2}.$$

If the curve is not a normal section, then by Meusnier's theorem it is sufficient to divide the curvature of the normal section in the same direction by the cosine of the angle between the principal normal to the curve and the normal to the surface.

The introduction of the second fundamental quadratic form provides an analytic approach to the study of how the surface is curved in space. In particular, one may derive the theorems of Euler and Meusnier, the expressions for the Gaussian and mean curvature, and so forth, in a purely analytic way.

Peterson's theorem, mentioned earlier, shows that the two quadratic forms, taken together, define a surface up to its position in space, so that the analytic study of any properties of a surface consists of the study of these forms. In conclusion, we note that the coefficients of the two quadratic forms are not independent; the connection mentioned earlier between the intrinsic geometry of a curved surface and the way in which it is embedded in space is expressed analytically by three relations (the equations of Gauss-Codazzi) between the coefficients of the first and the second fundamental quadratic forms.

§5. New Developments in the Theory of Curves and Surfaces

Families of curves and surfaces. Even though the basic theory of curves and surfaces was to a large degree complete by the middle of the last century, it has continued to develop in several new directions, which greatly extend the range of figures and properties investigated in contemporary differential geometry. There is one of these developments whose origins go back to the beginning of differential geometry, namely the theory of "families" or of continuous collections of curves and surfaces, but this theory may be considered new in the sense that its more profound aspects were not investigated until after the basic theory of curves and surfaces was already completely developed.

In general a continuous collection of figures is called an *n-parameter family* if each figure of the collection is determined by the values of n parameters and all the quantities characterizing the figure (in respect to its position, form, and so forth) depend on these parameters in a manner which is at least continuous. From the point of view of this general definition, a curve may be considered as a one-parameter family of points and a surface as a two-parameter family of points. The collection of all circles in the plane is an example of a three-parameter family of curves, since a circle in the plane is determined by three parameters: the two coordinates of its center and its radius.

The simplest question in the theory of families of curves or surfaces consists of finding the so-called envelope of the family. A surface is called the *envelope* of a given family of surfaces if at each of its points it is tangent to one of the surfaces of the family and is in this way tangent to every one of them. For example, the envelope of a family of spheres of equal radius with centers on a given straight line will be a cylinder (figure 48), and the envelope of such spheres with centers on all points of a given plane will consist of two parallel planes. The envelope of a family of curves is defined similarly. Figure 49 diagrams jets of water issuing from a fountain at various angles; in any one plane they form a family of curves, which may be considered approximately as parabolas; their envelope stands out clearly as the general contour of the cascade of water. Of course, not every family of curves or surfaces has an envelope; for example, a family of parallel straight lines does not have one. There exists a simple general method of finding the envelope of any family; for a

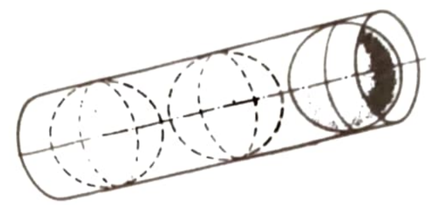

FIG. 48.

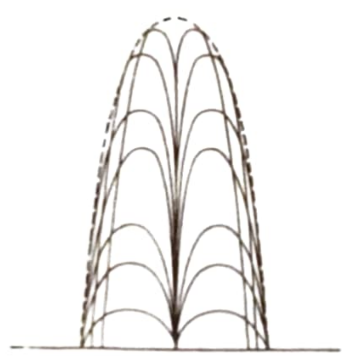

FIG. 49.

family of curves in the plane this method was given by Leibniz.

Every curve is obviously the envelope of its tangents, and in exactly the same way every surface is the envelope of its tangent planes. Incidentally, this fact provides a new method of defining a curve or a surface by giving the family of its tangent lines or planes. For some problems this method turns out to be the most convenient.

Generally speaking, the tangent planes of a surface are different at different points, so that the family of tangents to the surface is obviously a two-parameter one. But in some cases, for example, a cylinder, it is one parameter. It can be shown that the following remarkable theorem holds. A one-parameter family of tangent planes occurs only for those surfaces that are developable into a plane, i.e., those in which any sufficiently small segment may be deformed into a plane segment; these are the developable surfaces noted in §4. Every analytic surface of this kind consists of segments of straight lines and is either cylindrical (parallel straight lines) or conical (straight lines passing through one point), or consists of the tangents to some space curve.

The theory of envelopes is particularly useful in engineering problems, for example in the theory of transmissions. We consider two gears A and B. To study their motion relative to each other, we may assume that gear A is stationary and gear B moves around it (figure 50). Then the contour of a cog on gear B, as it assumes various positions, traces out a family of curves in the plane of gear A, and the contour of gear A must at all times be tangent to them, i.e., must be the envelope of the family. Of course, this is not a complete statement of the situation, since in an actual transmission this engagement must be transferred from one pair of cogs to the next, but this condition is nevertheless the basic one which must be satisfied by every type of gear.

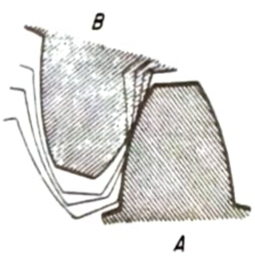

FIG. 50.

As we have said, the question of envelopes is a relatively simple one, solved long ago, in the theory of families of curves and surfaces. This theory is just as rich in interesting problems as, let us say, the theory of surfaces itself. Especially well developed is the theory of "congruences," i.e., two-parameter families of various curves (and in particular of straight lines: the so-called "straight-line" congruences). In this theory one applies essentially the same methods as in the theory of surfaces.

The theory of straight-line congruences originated in the paper of Monge, "On excavations and fills," the title of which already shows that

Monge undertook the investigation for practical purposes; the main idea was to find the most convenient way of transporting earth from an excavation to a fill.

The systematic development of the theory of congruences, beginning in the middle of the last century, is due in large measure to its connection with geometric optics; the set of rays of light in a homogeneous medium at any time constitutes a straight-line congruence.

Nonregular surfaces and geometry "in the large." The theory of curves and surfaces (and of families of them), as it had been constructed by the end of the last century, is usually called classical differential geometry; it has the following characteristic features.

First, it considers only "sufficiently smooth" (i.e., regular) curves and surfaces, namely those which are defined by functions with a sufficient number of derivatives. Thus, for example, surfaces with cusps or edges, such as polyhedral surfaces or the surface of a cone, are either excluded from the argument or are considered only on the parts where they remain smooth.

Second, classical differential geometry pays especial attention to properties of sufficiently small segments of curves and surfaces (geometry "in the small") and nowhere considers properties of an entire closed surface (geometry "in the large").

Typical examples, illustrating the distinction between geometry "in the small" and "in the large" are provided by the deformation of surfaces. For example, already in 1838 Minding showed that a sufficiently small segment of the surface of a sphere can be deformed, and this is a theorem "in the small." At the same time, he expressed the conjecture that the entire sphere cannot be deformed. This theorem was proved by other mathematicians as late as 1899. Incidentally, it is easy to confirm by experiment that a sphere of flexible but inextensible material cannot be deformed. For example, a ping-pong ball holds its shape perfectly well although the material it is made from is quite flexible. Another example, mentioned in §4, is the tin pail; it is rigid in the large, thanks to the presence of a curved flange, but separate pieces of it can easily be bent out of shape. As we see, there is an essential difference between properties of surfaces "in the small" and "in the large."

Other characteristic examples are provided by the theory of geodesics, discussed in §4. A geodesic "in the small," i.e., on a small segment of the surface, is a shortest path, but "in the large" it may not be so at all; for example, it may even be a closed curve, as was pointed out earlier for great circles of a sphere.

The reader will readily note that the theorems on geodesics formulated in §4 are basically theorems "in the small." Questions on the behavior of geodesic curves throughout their whole course will belong to geometry "in the large." It is known, for example, that on a regular surface two sufficiently adjacent points can be joined by a unique geodesic, remaining entirely in a certain small neighborhood of two points. But if we consider geodesics that during their course may depart as far as we like from the two points, then by a theorem of Morse any pair of points on a closed surface may be joined by an infinite number of geodesics. Thus, two points A and B on the lateral surface of a curved cylinder may be joined by very different geodesics: it is sufficient to consider helices which run from A to B but wind around the cylinder a different number of times. The theorem of Poincaré on closed geodesics, stated in §5 of Chapter XVIII, and proved by Ljusternik and Šnirelman, also belongs to geometry "in the large."

The proofs for these theorems, as for many theorems of geometry "in the large," were inaccessible with the usual tools of classical differential geometry and required the invention of new methods.

When these problems of geometry "in the large" were inevitably attracting the attention of mathematicians, the restriction to regular surfaces could no longer be maintained, if only because we are continually encountering surfaces that are not regular but have discontinuous curvature; for example, convex lenses with a sharp edge, and so forth. Moreover, there are many analytic surfaces that cannot be extended in any natural way without acquiring "singularities" in the form of edges or cusps and thus becoming nonregular.

Thus, a segment of the surface of a cone cannot be extended in a natural way without leading to the vertex, a cusp where the smoothness of the surface is destroyed.

This last result is only a particular case of the following remarkable theorem. Every developable surface other than a cylinder will lead, if naturally extended, to an edge (or a cusp in the case of a cone) beyond which it cannot be continued without losing its regularity.

Thus there is a profound connection between the behavior of a surface "in the large" and its singularities. This is the reason why the solution of problems "in the large" and the study of surfaces with "singularities" (edges, cusps, discontinuous curvature and the like) must be worked out together.

Similar new directions were taken in analysis. For example, the qualitative theory of differential equations mentioned in §7 of Chapter V, studies the properties of solutions of a differential equation in its entire domain of definition, i.e., "in the large," paying particular attention to "singularities," i.e., to violations of regularity, and to singular points of

the equation. Moreover, contemporary analysis includes the study of nonregular functions which did not occur in classical analysis (cf. Chapter XV) and thereby provides geometry with a new means of studying more general surfaces. Finally, in the calculus of variations, where we are usually looking for curves or surfaces with some extremal property, it sometimes happens that the limit curve, for which the extreme is attained, is not regular. For such problems it is necessary that the class of curves or surfaces under consideration should be closed (that is, should include all its limit curves or surfaces), a fact which necessarily led to the study of at least the simplest nonregular curves and surfaces. In a word, the new directions taken by geometry did not originate in isolation but in close connection with the whole development of mathematics.

The turning of attention to problems "in the large" and nonregular surfaces began about 50 years ago and was shared by many mathematicians. The first essential step was taken by Hermann Minkowski (1864-1909), who laid the foundation for an extensive branch of geometry, the theory of convex bodies. Incidentally, one of the questions which started Minkowski on his investigations was the problem of regular lattices, which is closely connected with the theory of numbers and geometric crystallography.

A body is called *convex* if through each point of its surface we may pass a plane that does not intersect the body, i.e., at any point of its surface the body may rest on a plane (figure 51). A convex body is defined

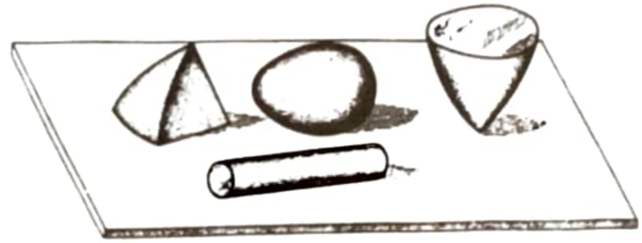

FIG. 51.

by its surface alone, so that for the most part it makes no difference whether we speak of the theory of convex bodies or of closed convex surfaces. The general theorems on convex bodies are proved, as a rule, without any additional assumptions about the smoothness or "regularity" of their surfaces. Thus these theorems are usually concerned with the whole convex body or surface, so that the restrictions of classical differential geometry are automatically removed. However, the two theories (of convex bodies and of nonregular surfaces) were at first very little connected with each other, the combination of the two taking place considerably later.

Beginning in 1940, A. D. Aleksandrov developed the theory of general curves and surfaces, including both the regular surfaces of classical differential geometry and also such nonsmooth surfaces as polyhedra, arbitrary convex sets, and others. In spite of the great generality of this theory, it is chiefly based on intuitive geometric concepts and methods, although it also makes essential use of contemporary analysis. One of the basic methods of the theory consists of approximating general surfaces by means of polyhedra (polyhedral surfaces). This device in its simplest form is known to every schoolboy, for example, in computing the area of the lateral surface of a cylinder as the limit of the areas of prisms. In a number of cases the method produces strong results that either cannot be derived in another way or else, if they are to be proved by an analytic method, require the introduction of complicated ideas. Its essential feature consists of the fact that the result is first obtained for polyhedra and is then extended to general surfaces by a limit process.

One of the beginnings of the theory of general convex surfaces was the theorem on the conditions under which a given evolute (cf. figure 52) may be pasted together to form a convex polyhedron. This theorem, completely elementary in its formulation, has a nonelementary proof and leads to far-reaching corollaries for general convex surfaces. The reader is, of course, familiar with the pasting together of a polyhedral surface from segments; for example, the assembling of a cube from the cross-shaped pattern in figure 52, or of a cylinder from a rectangle and two circles.

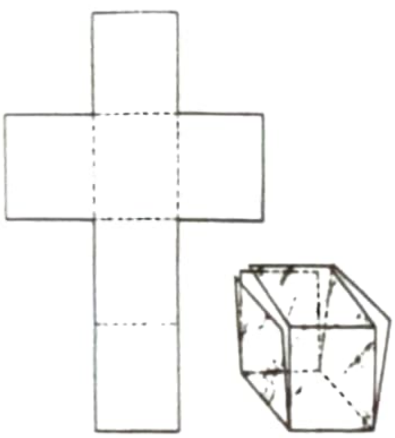

FIG. 52.

This simple example of assembling surfaces from segments of them is converted into a general method of "cutting apart and pasting together," which has produced profound results in various questions of the theory of surfaces and has found practical applications.

Deep-lying results in this theory were obtained by A. V. Pogorelov. In particular, he showed that every closed convex surface cannot be deformed as a whole with preservation of its convexity. This result, achieved in 1949, completes the efforts of many well-known mathematicians, who for the preceding 50 years had tried to prove it but had been successful only under various additional hypotheses. The results of Pogorelov, in conjunction with the "method of pasting together," not only provided a complete solution for the problem, but almost completely cleared up the whole question of the deformability or nondeformability of closed and nonclosed convex surfaces. They also established a close connection between the new theory and "classical" differential geometry.

In this way a theory of surfaces was constructed that included the classical theory as well as the theory of polyhedra, of arbitrary convex surfaces, and of very general nonconvex surfaces. Lack of space does not allow us to discuss in detail the results or the still unsolved problems of the theory, although this could readily be done, since they are for the most part quite easily visualized and, in spite of the difficulty of exact proofs, do not require any special knowledge.

In §4, in speaking of the deformation of surfaces, we had in mind deformations of a regular (continuously curved) surface that preserved its regularity. But in the theorem of Pogorelov, on the contrary, there is no requirement of regularity for either the initial or the deformed surface, although the requirement of convexity is imposed on both surfaces.

It is obvious that deformation of a sphere, for example, becomes possible if we allow breaks in the surface and violation of the convexity. It is sufficient to cut out a segment of the surface and then replace it after the deformation; that is, so to speak, to push a segment of the surface into the interior. Considerably more unexpected is the result obtained recently by the American mathematician Nash and the Dutch mathematician Kuiper. They showed that if we preserve only the smoothness of a surface and allow the appearance of any number of sharp jumps in the curvature of the surface (i.e., if we eliminate any requirement of continuity, boundedness, or even existence of the second derivatives of the functions defining the surface) then it turns out to be possible to deform the surface as a whole with a very great degree of arbitrariness. In particular a sphere may be deformed into an arbitrarily small ball, which has a smooth surface consisting of very shallow wavelike creases. Some idea of a deformation of this sort may be gained by the easily imagined possibility of rumpling up into almost any shape a spherical cover made of very soft cloth. On the other hand, a small celluloid ball behaves very differently. The elastic material of its surface resists not only extension but also sharp bending, so that such a ball is very rigid.

Differential geometry of various groups of transformations. At the beginning of this century, there arose from classical differential geometry a series of new developments based on one general idea, namely the study of properties of curves, surfaces, and families of curves and surfaces which remain invariant under various types of transformations. Classical differential geometry investigated properties invariant under translation, but of course there is nothing to prevent us from considering other geometric transformations. For example, a *projective transformation* is one in which straight lines remain straight, and projective geometry, which has been in existence for a long time, studies those properties of figures that remain invariant under projective transformations. Ordinary projective geometry remains similar, in the problems it investigates, to the usual elementary and analytic geometry, whereas "projective differential geometry" (the theory of curves, surfaces, and families developed at the beginning of the present century) is similar to classical differential geometry, except that it studies properties that are invariant under projective transformations. Fundamental in this last direction were the contributions of the American Wilczynski, the Italian Fubini, and the Czech mathematician, Čech.

In the same way arose "affine differential geometry," which studies the properties of curves, surfaces, and families invariant under affine transformations, i.e., under transformations that not only take straight lines into straight lines but also preserve parallelism. The work of the German mathematician Blaschke and his students developed this branch of geometry into a general theory. Let us also mention "conformal geometry," in which one studies the properties of figures invariant under transformations that do not change the angles between curves.

In general, the possible "geometries" are very diverse in character, since essentially any group of transformations may serve as the basis of a "geometry," which then studies just those properties of figures that are left unchanged by the transformations of the group. This principle for the definition of geometries will be discussed further in Chapter XVII.

Other new directions in differential geometry are being successfully developed by Soviet geometers, S. P. Finikov, G. F. Laptev, and others. But in our present outline it is not possible to give an account of all the various investigations that are taking place nowadays in the different branches of differential geometry.

Suggested Reading

H. Busemann, *Convex surfaces*, Interscience, New York, 1958.

H. S. M. Coxeter, *Introduction to geometry*, Wiley, New York, 1961.

H. G. Eggleston, *Convexity*, Cambridge University Press, New York, 1958.

D. Hilbert and S. Cohn-Vossen, *Geometry and the imagination*, Chelsea, New York, 1952.

I. M. Yaglom and V. G. Boltyanskiĭ, *Convex figures*, Holt, Rinehart and Winston, New York, 1961.