This is a direct continuation of the author's textbooks *Lectures in Geometry. I. Analytical Geometry* and *Lectures in Geometry. II. Linear Algebra.* The text includes sections on general topology. The concept of submanifolds is explained, Sard's and Whitney's theorems are proved, the theory of differential forms and their integrals, and elementary differential geometry, i. e. the theory of curves (Frene's formulas) and the theory of surfaces, are presented.

M. POSTNIKOV

LECTURES IN GEOMETRY

SEMESTER III

SMOOTH MANIFOLDS

MIR PUBLISHERS
MOSCOW

Translated from the Russian
by Vladimir Shokurov

First published 1989
Revised from the 1987 Russian edition

На английском языке

Printed in the Union of Soviet Socialist Republics

ISBN 5-3-000910-8

CONTENTS

PREFACE TO THE RUSSIAN EDITION

Geometry has been and remains the Cinderella of the curriculum at the Moscow University's Mathematics and Mechanics faculty. Never once during the last fifty years has the curriculum contained a course on the foundations of geometry or algebraic curves or transformation groups or even projective geometry (if we do not count the scraps in the first-semester courses in analytic geometry which may only be given under special circumstances, and nobody cares when the lecturer curtails them up or even drops them altogether). A student might well graduate from the faculty—with honours!—with no idea about Lobachevskian geometry, Caley-Klein ideas in the foundations of geometry, or the properties of algebraic curves or Lie groups.

Some twelve years ago the overflow into the calculus course of geometric material due to the ever increasing implementation of geometric methods led to the creation in the second year of a new course with the *ad hoc* name "Smooth manifolds and differential geometry". This course was delivered at a rate of a lecture a week and it was hoped that the course would free lecturers from presenting the extraneous geometric material. The course, however, was not well thought-out, and the parallel courses in calculus and differential equations were not coordinated with it. As a result the lecturers on calculus did not derive any advantage, and ridiculous as it may seem, integration of the differential forms on manifolds and Stokes' formula were discussed twice in as much detail but from slightly different points of view, in the two concurrent courses.

Delivering the geometry course in the third semester did not allow the generalizing and unifying role played by geometric concepts in modern mathematics to be brought out since to do so it is necessary for the main analytic courses to have been covered.

These and various more particular considerations led to the transfer of geometry to the third year (the fifth and sixth semesters). It has immediately become clear that this also had disadvantages.

A necessary constituent part of any course in geometry is the theory of curves and surfaces in three-dimensional Euclidean space, which is important both in its own right and as a source of examples and analogues for Riemannian geometry and the geometry of affine connections. By the third year, this material is too elementary (by this time students have already acquired a knack and a taste for more complicated constructions and concepts) and for it to play its propaedeutic role one cannot pass too quickly onto Riemannian geometry.

It is clear that this theory must be presented no later than the third semester (or perhaps earlier, as suggested by me in the first Russian edition of Semester II of these Lectures, even in the second semester). Moreover, a third-year geometry course does not help lecturers presenting second-year analysis (which I am sure will soon lead to the abolition of the course in geometry in the third semester and may be, alas, to its ousting from the schedule of the curriculum).

The radical solution is, of course, to overhaul the traditional system of mathematical courses. However, since there is an acute struggle between the departments for hours and courses such a review, which will have to be carried out sooner or later, is at present not possible, and a temporary solution would be the return of the geometry course in the third or fourth semester with the presentation of integration topics clearly distributed between the courses in calculus and geometry, each passing on the baton to the other as it were.

The following distribution of topics is suggested. After the integral of functions over domains in $\mathbb{R}^n$ has been discussed in calculus, the geometry lecturers cover the integration of densities and forms on manifolds. Simultaneously the calculus lecturer illustrates the general theory by particular cases of line and surface integrals of the first (density) and the second (form) kind. During this time, the generalized Stokes theorem is discussed which in the calculus course is immediately rendered concrete in the form of the Green, Gauss-Ostrogradskii and Stokes formulas. This duet, in which the general melody sometimes drifts, sometimes merges, ends in the apotheosis of vector analysis with elements of potential theory where the course in calculus changes freely into the theory of multidimensional improper integrals and the geometry course into cohomology theory. All this, of course, requires close coordination between the lecturers, which is not easy to achieve.

This book has, like its predecessors*, grown out of lectures given at the mathematics and mechanics faculty of Moscow University in different years. It is not, however, a recording of any particular course, but is instead a realization of the proposed geometry syllabus for the third semester. It can, however, certainly be used as the text for the fifth semester course.

The textbook is intended as a normal course presented at two lectures a week. The number of lectures (29) arises because although the winter semester formally contains 18 weeks, in practice it is impossible to deliver more than 11 to 15 weeks of lectures. The course can be used, however, even if the curriculum assigns only one or one and a half lectures a week (11 to 15 and respectively 16 to 22 lectures).

To be able to estimate the time required for a syllabus, I have tried to make each lecture in the book correspond to a two-hour lecture. Repeating material from other courses and considering examples, in written form, requires much more time. This accounts for difference in the volume of the lectures and the unexpectedly large size of Lectures 3, 11, and 20.

* M. Postnikov. *Lectures in Geometry*: *Semester* I. *Analytic Geometry*. (Mir Publishers, Moscow, 1981), *Semester* II. *Linear Algebra and Differential Geometry*. Mir Publishers, Moscow, 1982). (Referred to as I and II in what follows.)

The book focusses on smooth manifolds, and general topological facts and ideas are not separately presented being interspersed in the text.

In recent years a rather strange view of smooth manifolds has become widespread, a view surprisingly shared by some respected and competent mathematicians. Since a smooth manifold can be regarded as the result of the natural attempt to generalize axiomatically the simple idea of a manifold as a subset of a Euclidean space defined by a system of functionally independent equations, it is argued that the generalization does not actually lead to new objects because of the Whitney embedding theorem and so manifolds should be defined as subsets of that kind and that the general concept of a manifold is just an example of an axiomatic construction which inevitably arises in following a concept to its conclusion but one which it is then better to forget. I cannot share this opinion because in practice—for example, in mechanics—manifolds tend to appear in an abstract form, unembedded in a Euclidean space, and their forced embedding (with great arbitrariness!) introduces an additional structure that is sometimes useful but often having no relevance to the crux of the matter. The adherents of the former opinion appeal to Poincaré, who has shared it. In fact Poincaré clearly understood the necessity of having a general concept of manifold and dwelt on pasting together the charts of an atlas. Referring to the extremes of axiomatization is also wrong, since in reality manifolds were not introduced as the result of "natural attempt to generalize the simple notion of a manifold given by equations" but as an answer to the need to clearly explicate the notion arising in mathematical investigation. A consistent execution of the same principles would throw mathematics a hundred years back, since from this point of view, for example, all linear algebra in its present form has no right to exist, being based as it is, on the concept of a vector space which could be said to "have arisen as a result of a natural attempt to generalize the simple idea of the space $\mathbb{R}^n$", (which is as false as it is for manifolds), whereas the isomorphism theorem shows that "the generalization does not actually lead to new objects" (which, though true, does not deprive the concept of a vector space of its value). In this book, therefore, manifolds are defined in the usual way, on the basis of an atlas, while subsets of Euclidean spaces only appear as examples.

The problems in this book are mainly quite trivial and intended exclusively for a reader to test himself. Some more difficult problems are given in small print. Auxiliary material on algebra or calculus is also given in small print.

The first five lectures are only indirectly related to the theory of smooth manifolds, mainly being devoted to elementary differential geometry. The theory of curves (Frenet formulas) is followed by the first and second quadratic forms of a surface, the Weingarten formulas are derived, and Gauss's theorem on the invariance of total curvature is proved. Everything not directly related to the Gauss theorem has been omitted (the Meusnier theorem and the Euler theorem, geodesics, asymptotic curves, lines of curvature, and the like). When delivering lectures in the second year this material had sometimes to be postponed until the middle of the semester (so as to satisfy the needs of the course on differential equations, by introducing the general theory of smooth manifolds as soon as possible). Although this did remove some repetitions (for example, it was then unnecessary to define the differential of a smooth mapping twice, first for surfaces and then in the general case), it was barely justified methodologically (as it links elementary differential geometry, which is local in nature, to the theory of manifolds).

The theory of manifolds begins in Lecture 6. The first ten lectures (from the sixth to the fifteenth) are devoted to basic geometric notions and theorems of the theory of manifolds. In the shorter 11-lecture course one may omit seven of these lectures, reducing the first five lectures to four and sacrificing Lectures 8, 9, and 10 (which treat in the main the topological theory of dimension and the Tychonoff theorems), and Lecture 10 in a 16-lecture course. The remaining lectures in this group (particularly those on the Sard and the Whitney theorems) must, in

my opinion, be kept in the course under all circumstances.

An 11-lecture course actually stops here. It turns out to be, however, possible to save, by slightly reducing and condensing the presentation, about an hour and a half of lecture time in presenting some of the material in Lectures 16 and 17. As to the theory of differential forms (Lectures 18, 19, and 20), it must be put off in the shortened course until the next semester (or left to the calculus lectures).

In a 16-lecture course it is possible to finish the course with Lecture 20, which demonstrates various ways of computing de Rham cohomology groups using the example of a sphere. This means that the "integration" Lectures 24 to 29 are excluded from the course and their material is thus left to the calculus course.

In Lectures 21 to 23 an attempt is made to expound the theory of homologies and cohomologies (up to spectral sequences!) in a form suitable for the compulsory course. This is made possible by changing the generally accepted point of view and giving up the treatment of simplicial homology theory, which is alleged to be geometrically obvious. I am pleased to note that a similar approach, at a more advanced level, is accepted in *Differential Forms in Algebraic Topology* by R. Bott and L. W. Tu, which must be read by anyone who wants to become acquainted with the basic ideas and constructions of the classical homology theory in a bright and up-to-date presentation. When time is lacking it is possible to omit the second half of Lecture 22 and all of Lecture 23.

Finally, the concluding Lectures 24 to 29, which, if desired, can be partially interchanged with Lectures 21 to 23, deal with integration. Here the presentation is deliberately incomplete (for example, nothing is said about additive functions of a set), since these lectures reflect only part of the general picture, and omit what relates to calculus. Lecture 28 can be left entirely to the calculus lecturer. It is also possible to confine oneself to just one lecture, Lecture 29, which is virtually independent of the previous four lectures.

PREFACE TO THE ENGLISH EDITION

This book is actually Semester 3 of my *Lectures in Geometry*. It starts a new subject, however, and is therefore independent of the previous two semesters.

The book has two major features that distinguish it from other textbooks on elementary smooth manifold theory. Firstly, a lot of space is allotted to topological dimensional theory, the most geometry-oriented branch of general topology, and an acquaintance with it will bring joy to lovers of elegant mathematical constructions which provide deep insights. Secondly, I've ventured, in this elementary course, to present the basic notions of the theory of spectral sequences, a tool whose power and significance is becoming increasingly clear in current studies. This is done without first expounding general (co)homology theory.

It is hoped that the two topics will appeal to the interested English reader.

The symbol □ signifies the end of a proof of a theorem.

December 31, 1988 *M. Postnikov*

Lecture I

Simple lines in the plane · Giving lines by an equation · Whitney's theorem · Jordan curves · Smooth and regular curves · Nonparametrized curves · Natural parameter

There are several different approaches to explicate the idea of a line, which yield different results. In simple cases, however, all the approaches give virtually the same result.

Let us first consider lines in the plane.

A set Γ in the plane is said to be a *graph* if there is a system of (Euclidean or affine) x-, y-coordinates and a differentiable (alternatively, continuous) function f: $I \to \mathbb{R}$ defined on the (closed, half-open or open) interval I of the real axis $\mathbb{R}$ such that a point with coordinates x and y is in Γ if and only if $x \in I$ and $y = f(x)$. Intuitively, all graphs, are, of course, lines.

A point p_0 of a set C in the plane is said to be *simple* if there is an open disk U with centre at p_0 such that the intersection $U \cap C$ is a graph.

The set C is said to be *connected* if it cannot be divided into two sets having the property that each limiting point of one set does not belong to the other. (This graphically means that the set consists of one piece.)

The set C in the plane is said to be a *simple line* if it is connected and consists of only simple points.

Problem 1. Prove that any graph is connected (and hence is a simple line).

The various ways of explicating the notion of a line differ mainly by nonsimple points that are allowed. We shall avoid discussing these questions once and for ever by agreeing to consider only simple lines.

A simple line may (or may not) have *end points*. There may be at most two of them. A simple line with two end points (having the shape of a bent closed interval of the number line) is called *closed*, and that with one end point (having the shape of a bent half-open interval of the number line) is called *half-open*. A simple line without end points may have the shape of a bent open interval of the number line or that of a bent circle. In the former case it is called *open*, and in the latter case *closed*. (Thus the term "closed" as applied to simple lines has two meanings! This historically established ambiguity must be constantly kept in mind.)

The usual way of giving lines in the plane is to define them by equations of the form

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

where x, y are coordinates (affine or rectangular) in the plane, and F is a function of x, y. [The statement that a set $\mathscr{L}$ is given by equation (1) implies by definition that a point p of the plane is in $\mathscr{L}$ if and only if its coordinates x, y satisfy equation (1). When F is treated as a function in the plane this means that $p \in \mathscr{L}$ if and only if $F(p) = 0$.]

To obtain lines (in the sense of an explicit definition) it is certainly necessary to impose on F certain conditions. First of all it is natural to require that F be *continuous*. [If discontinuous functions are allowed, then equations of the form (1) may define an arbitrary set A of points of the plane; it suffices to take as F the function $1 - \chi$, where χ is the so-called *characteristic function* of A equal to unity for the points of A and to zero outside A.]

Recall from the course in calculus that a point p of the plane (or generally of an arbitrary metric—Euclidean in particular—space) is said to be an *interior point* of A if there is $\varepsilon > 0$ such that the ε-neighbourhood of that point (an open ball—a disk in the plane—of radius ε with centre at p) is entirely in A. A set consisting of only interior points is called *open*. A set C is said to be *closed* if its complement is open or, equivalently, if for any

convergent sequence of points $p_n \in C$ its limit $\lim p_n$ is also in C.

A subset of an **affine** (real and finite-dimensional) space $\mathscr{A}$ is said to be *closed* (*open*) if it is closed (open) with respect to some Euclidean metric on $\mathscr{A}$.

Problem 2. Show that if a subset of $\mathscr{A}$ is closed (open) with respect to one Euclidean metric on $\mathscr{A}$, then it is closed (open) with respect to any other as well.

Remark 1. A closed simple line in the plane is a closed (and bounded) set (for each of the two meanings of the term "closed line"). On the contrary, none of the simple lines—the open one included—is an open set. Moreover, there are open lines (for example, the graph of a tangent) which are closed sets (necessarily unbounded).

Remark 2. It can be easily seen that *none of the simple lines has a single interior point*. Therefore Cantor suggested to consider arbitrary closed sets without interior points as lines in the plane. This definition offers some advantages but it is too general for most mathematical theories (while failing to cover, say, open simple lines).

It is obvious that for every continuous function F on a metric space $\mathscr{X}$ the set of all points in which F is zero is closed. This means that *an equation of the form* (1) *with a continuous function F can specify only closed sets of the plane*.

Problem 3. Show that, conversely, for any closed set C of $\mathscr{X}$ there is a continuous function F on $\mathscr{X}$ such that $p \in C$ if and only if $F(p) = 0$. [**Hint.** Consider on $\mathscr{X}$ the function

$$F(p) = \inf_{q \in C} \rho(p, q), \quad p \in \mathscr{X},$$

a distance from p to C; here ρ is a metric in $\mathscr{X}$.]

A point of a set (1) is said to be *nonsingular* if at this point both partial derivatives

$$\frac{\partial F}{\partial x} \text{ and } \frac{\partial F}{\partial y}$$

of F exist, are continuous, and at least one of them is nonzero. The other points of set (1) are called *singular points*.

The **implicit-function theorem** known from calculus states that in the neighbourhood of any nonsingular point each set (1) is a graph, i.e. *every nonsingular point is a simple point*. (The converse is false. For example, when $F(x, y) = x(x^2 + y^2)$ the set (1) consists of the points of the axis of ordinates $x = 0$ and hence all its points are simple. At the same time the point $(0, 0)$ is its singular point.)

It follows that the *set of all nonsingular points of every set* (1) *is a union of simple lines* (joining, in general, at singular points). If therefore the number of singular points is not very large, for example, it is finite, then the set of form (1) corresponds quite well to an intuitive idea of lines. (And it is quite appropriate to call them so.) If, however, there are many singular points, then the situation is quite different. Namely, as was shown by the American mathematician Whitney, an equation of the form (1) with an infinitely differentiable function F can define any closed subset of the plane.

Whitney's theorem is related to an arbitrary finite-dimensional point affine space $\mathcal{A}$. Every function F on that space can be regarded, on choosing a reference point O, as a function on the associated vector space $\mathcal{V}$ and hence, after choosing in $\mathcal{V}$ a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, as a function on a Euclidean space $\mathbb{R}^n$. A function F is said to be *smooth function of class* C^∞ or C^∞-function if, as a function on $\mathbb{R}^n$, it has continuous partial derivatives of all orders. (It is clear that if this condition is satisfied for one choice of the frame $O\mathbf{e}_1, \ldots \mathbf{e}_n$, then it is so for any other choice of it.)

Theorem 1 (Whitney's theorem). *For any closed subset C of an affine space $\mathcal{A}$ there is a C^∞-function F on $\mathcal{A}$ such that $p \in C$ if and only if $F(p) = 0$.*

The proof of Theorem 1 rests on the following lemma, which is of interest:

Lemma 1. *There is a C^∞-monotonic function $\alpha\colon \mathbb{R} \to \mathbb{R}$ such that:*

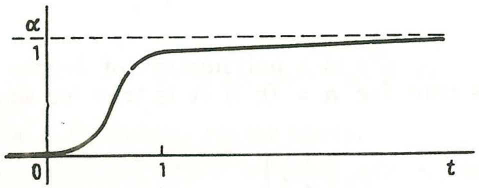
1° $0 \leqslant \alpha(t) < 1$ *for any* $t \in \mathbb{R}$,

2° $\alpha(t) = 0$ *if and only if* $t \leqslant 0$.

Proof. Put

$$\alpha(t) = \begin{cases} e^{-1/t} & \text{if } t > 0, \\ 0 & \text{if } t \leqslant 0. \end{cases} \tag{2}$$

Function (2) is clearly monotonic and has Properties 1° and 2°. Moreover, when $t \neq 0$ this function is obviously infinitely differentiable. We therefore must only prove that it is infinitely differentiable for $t = 0$ as well.



Graph of a function α

To this end recall that a function f defined in the neighbourhood of $t = 0$ is differentiable at that point if there are limits

$$\lim_{t \uparrow 0} \frac{f(t) - f(0)}{t}, \quad \lim_{t \downarrow 0} \frac{f(t) - f(0)}{t} \tag{3}$$

(the left and right derived numbers of f at $t = 0$) and if these limits are equal.

On the other hand, if f is differentiable in the neighbourhood of $t = 0$, except possibly for the point itself, and if there are limits

$$\lim_{t \uparrow 0} f'(t), \quad \lim_{t \downarrow 0} f'(t), \tag{4}$$

then, as it follows directly from the Lagrange theorem on finite increments, limits (3) exist and are equal to limits (4).

Since for all functions $f = \alpha^{(n)}$ the left-hand limits (4) obviously exist and are zero (for when $t < 0$ these functions are identically zero) it follows that to prove

Lemma 1 it suffices to establish that for any $n \geqslant 0$ the limit

$$\lim_{t \downarrow 0} \alpha^n(t) = \lim_{t \downarrow 0} (e^{-1/t})^{(n)}$$

exists and is equal to zero.

But it is easy to see that for any $n \geqslant 0$ there is a formula

$$(e^{-1/t})^{(n)} = e^{-1/t} p_n\left(\frac{1}{t}\right),$$

where $p_n = p_n(T)$ is a polynomial of degree $2n$. [This formula is true for $n = 0$; if it is true for some $n \geqslant 0$, then

$$(e^{-1/t})^{(n+1)} = \left[e^{-1/t} p_n\left(\frac{1}{t}\right)\right]'$$

$$= e^{-1/t}\left[\frac{1}{t^2} p_n\left(\frac{1}{t}\right) - \frac{1}{t^2} p_n'\left(\frac{1}{t}\right)\right]$$

$$= e^{-1/t} p_{n+1}\left(\frac{1}{t}\right),$$

where $p_{n+1}(T) = T^2 p_n(T) - T^2 p_n'(T)$ is a polynomial of degree $2n+2$.]

Therefore

$$\lim_{t \downarrow 0} (e^{-1/t})^{(n)} = \lim_{t \to +\infty} \frac{p_n(t)}{e^t} = 0,$$

which was to be proved. □

Corollary 1. *For any closed interval* $[a, b]$ *of the axis* $\mathbb{R}$ *there is a* C^∞*-function* $\beta: \mathbb{R} \to \mathbb{R}$ *such that* $0 \leqslant \beta(t) \leqslant 1$ *for all* $t \in \mathbb{R}$ *and*

$$\beta(t) = \begin{cases} 1 & \text{if } t \leqslant a, \\ 0 & \text{if } t \geqslant b. \end{cases}$$

Proof. It suffices to put

$$\beta(t) = \frac{\alpha(b-t)}{\alpha(b-t) + \alpha(t-a)},$$

where α is the function in Lemma 1. □

Remark 3. We can construct C^∞-functions with a more complicated behaviour in a similar way. For example, for any numbers $a < c < d < b$ the formula

$$\lambda(t) = \frac{\alpha(B - |t - C|)}{\alpha(B - |t - C|) + \alpha(|t - D| - A)},$$

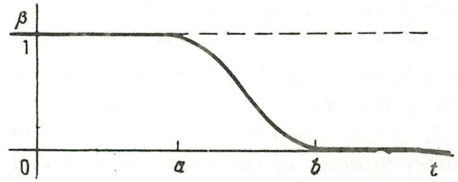
where

$$A = \frac{d-c}{2}, \quad B = \frac{b-a}{2}, \quad C = \frac{b+a}{2}, \quad D = \frac{d+c}{2},$$

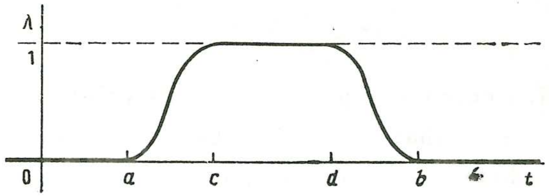
defines a C^∞-function equal to zero outside $[a, b]$ and to unity on $[c, d]$. We shall need such a function in Lecture 15.

Let $\mathscr{V}$ be a Euclidean vector space.

Notation. For any $r > 0$ we shall denote by $\mathbb{B}_r^{\mathscr{V}}$ (or simply $\mathbb{B}_r$) a ball of radius r of $\mathscr{V}$ with centre at $\mathbf{0}$, i.e. the



Graph of a function β



Graph of a function λ

set of all vectors $\mathbf{x} \in \mathscr{V}$ for which $|\mathbf{x}| \leqslant r$. The corresponding *open ball* (a set of vectors $\mathbf{x} \in \mathscr{V}$ for $|\mathbf{x}| < r$) will be denoted by $\mathring{\mathbb{B}}_r^{\mathscr{V}}$.

For $\mathscr{V} = \mathbb{R}^n$ we shall write, as a rule, $\mathbb{B}_r^n$ rather than $\mathbb{B}_r^{\mathbb{R}^n}$.

In a Euclidean point space $\mathscr{A}$ a ball of radius r with centre at p will be denoted by $\mathbb{B}_r^{\mathscr{A}}(p)$ (and open ball by $\mathring{\mathbb{B}}_r^{\mathscr{A}}(p)$).

These symbols will be constantly used throughout this course.

Corollary 2. *For any point p_0 of a Euclidean point space $\mathscr{A}$ and any $r > 0$ there is a function $f: \mathscr{A} \to \mathbb{R}$ such that $0 \leqslant f \leqslant 1$ on $\mathscr{A}$ and*

$$f(p) = \begin{cases} 1 \text{ if and only if } p \in \mathbb{B}_r(p_0), \\ 0 \text{ if and only if } p \notin \mathring{\mathbb{B}}_{2r}(p_0). \end{cases}$$

Proof. It suffices to put

$$f(p) = \beta(|\mathbf{x}|),$$

where $\mathbf{x} = \mathbf{p_0 p}$ is the radius vector of the point p counted off from p_0 and β is the function in Corollary 1, constructed for $[r, 2r]$. □

On choosing in $\mathscr{A}$ a system of rectangular coordinates we call $p \in \mathscr{A}$ a *rational point* if all its coordinates are rational numbers. A ball $\mathring{\mathbb{B}}_r(p)$ will be called *rational* if its centre p and radius r are rational.

Lemma 2. *Every open set $U \subset \mathscr{A}$ is a union of a countable (or finite) number of rational balls, i.e. there are rational points $q_1, \ldots, q_m, \ldots$ and rational numbers $r_1, \ldots, r_m, \ldots$ such that*

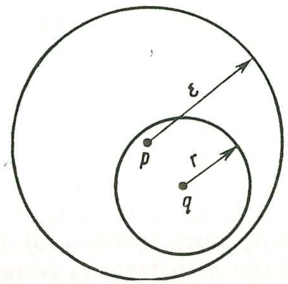
$$U = \bigcup_{m=1}^{\infty} \mathring{\mathbb{B}}_{r_m}(q_m). \tag{5}$$

Proof. Under the hypothesis, for any point $p \in U$ there is $\varepsilon > 0$ such that $\mathring{\mathbb{B}}_\varepsilon(p) \subset U$. Consider a rational ball $\mathring{\mathbb{B}}_r(q)$, where q is a rational point such that $\rho(p, q) < \varepsilon/2$ and r is a rational number such that $\rho(p, q) < r < \varepsilon/2$ (the existence of q and r is ensured by the fact that any real number may be approximated as much as desired by a rational one). Since $\rho(p, q) < r$, we have $p \in \mathring{\mathbb{B}}_r(q)$, and since

$$\rho(x, p) \leqslant \rho(x, q) + \rho(p, q) < 2r < \varepsilon$$

for any point $x \in \mathring{\mathbb{B}}_r(q)$, we have $\mathring{\mathbb{B}}_r(q) \subset \mathring{\mathbb{B}}_\varepsilon(p) \subset U$.

We thus see that any point $p \in U$ is contained in some rational ball $\mathring{\mathbb{B}}_r(q) \subset U$. This means that the set U is a union of rational balls of the form $\mathring{\mathbb{B}}_r(q)$ which are constructed for all possible points $p \in U$. But the set of



all rational balls of a space $\mathscr{A}$ is clearly countable. The number of distinct balls of the form $\mathring{\mathbb{B}}_r(q)$ is therefore at most countable. Denoting them by $\mathring{\mathbb{B}}_{r_m}(q_m)$ we obtain expansion (5). □

We are now ready to prove Theorem 1.

Proof of Theorem 1. We may obviously assume without loss of generality that the affine space $\mathscr{A}$ under consideration is Euclidean and hence the complement $U = \mathscr{A} \setminus C$ of the set C allows a representation of the form (5). Let f_m be the function in Corollary 2 to Lemma 1 corresponding to q_m and $r_m/2$. This function (and hence each of its partial derivatives) is identically zero outside the compact (closed and bounded) set $\mathbb{B}_{r_m}(q_m)$. For any $k \geqslant 0$ therefore there is a number $c_m^k > 0$ such that the absolute value of each partial derivative of order k of f_m is at most c_m^k throughout $\mathscr{A}$. Let

$$c_m = \max(1, c_m^0, c_m^1, \ldots, c_m^m).$$

Consider a functional series

$$\sum_{m=1}^{\infty} \frac{f_m}{2^m c_m}. \tag{6}$$

Since by construction $c_m \geqslant 1$ and $f_m \leqslant 1$, this series is majorized by the numerical convergent series

$$\sum_{m=1}^{\infty} \frac{1}{2^m}. \tag{7}$$

Consequently, series (6) converges to some function $F: \mathcal{A} \to \mathbb{R}$. If $p \in C$, then $p \notin \mathbb{B}_{r_m}(q_m)$ for every $m \geqslant 1$ and hence $f_m(p) = 0$. But if $p \notin C$ (i.e. $p \in U$), then there is $m \geqslant 1$ such that $p \in \mathbb{B}_{r_m}(q_m)$. Then $f_m(p) \neq 0$ and hence $F(p) \neq 0$. Thus $F(p) = 0$ if and only if $p \in C$.

On the other hand, since for any $m \geqslant k$ each partial derivative of the m-th term of series (6) of order k obviously does not exceed the m-th term of series (7), on differentiating (6) k times we obtain a series all terms of which, except possibly for the first k terms, are also majorized by terms of series (7) and which is, consequently, uniformly convergent. According to the well-known series-differentiation theorem the sum of series (6) is infinitely differentiable (and each of its partial derivatives is the sum of a series made up of the corresponding partial derivatives of series (6)).

This completes the proof of Theorem 1. □

Another approach to the concept of a line, usually associated with the name of the French mathematician Jordan, is based on the concept of a line as a trajectory of a moving point. Lines in the sense of Jordan will be called curves.

According to Jordan, by a *curve* in an n-dimensional affine space $\mathcal{A}$ an arbitrary continuous mapping

$$\gamma: I \to \mathcal{A} \tag{8}$$

is meant, where I is some interval of the number line $\mathbb{R}$ (open, half-open or closed), i.e. after choosing the reference point in $\mathcal{A}$, a continuous vector function

$$\mathbf{r} = \mathbf{r}(t), \quad t \in I \tag{9}$$

which assumes values in the associated vector space $\mathcal{V}$.

In affine coordinates $x^1, \ldots x^n$ the Jordan curve (8) is given by continuous numerical functions

$$x^1 = x^1(t), \ldots, x^n = x^n(t), \quad t \in I. \tag{10}$$

Equations (9) and (10) are called *parametric equations* of curve (8) (a *vector* and a *coordinate* one, respectively).

We stress that curves—in contrast to lines!—are not sets but mappings.

In practice, however, it is convenient to handle curves, at least from a terminological standpoint, as if they were sets. For example, for any $t_0 \in I$ a point $p_0 = \gamma(t_0)$ of $\mathcal{A}$ is called a *point of the curve* (8) *corresponding to the value of the parameter* t_0, and it is also said that *for* $t = t_0$ *curve* (8) *passes through* p_0. When the interval I is closed ($[a, b]$) points $\gamma(a)$ and $\gamma(b)$ are called the *end points* of the curve (8). Curve (8) is also said to join the point $\gamma(a)$ to the point $\gamma(b)$ and so on and so forth.

When $\gamma(a) = \gamma(b)$ curve (8) may be thought of as a continuous mapping of a circle. Such curves are called *closed.*

When it is required to stress the difference between a curve and the set of its points, the latter is called the *support* of the curve. Thus the support of curve (8) is nothing but the image $\gamma(I)$ of the interval I under mapping (8).

In general, the support of a curve may have the shape very far from the intuitive idea of a line. For example, it may have interior points or even fill a square, as shown by the example of a Peano curve.

Curve (8) is said to be *simple* if it is, first, an injective mapping $I \to \mathcal{A}$, i.e. $\gamma(t_1) = \gamma(t_2)$, $t_1, t_2 \in I$ if and only if $t_1 = t_2$, and, second, a reciprocal continuous mapping (also known as a *moneomorphic* mapping), i.e. such that if for a sequence $\{t_m\}$ of points of an interval I there is a point $\tau \in I$ such that $\lim \gamma(t_m) = \gamma(\tau)$, then $\{t_m\}$ converges (and $\lim t_m = \tau$). A closed curve

$$\gamma: [a, b] \to \mathcal{A}, \quad \gamma(a) = \gamma(b) \tag{11}$$

is said to be *simple* if $\gamma(t_1) = \gamma(t_2)$ for $t_1 < t_2$ if and only if $t_1 = a$ and $t_2 = b$.

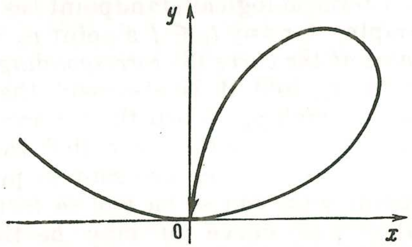
A typical example of an injective rather than moneomorphic mapping of an open interval into $\mathscr{A}$ is the curve

$$x=\frac{3t}{1+t^3}, \quad y=\frac{3t^2}{1+t^3}, \quad -1<t<+\infty$$

(the "cut-off folium of Descartes").

Problem 4. Prove that for $I = [a, b]$ any injective mapping $I \to \mathscr{A}$ is moneomorphic.

The supports of simple curves are called *simple arcs.*



The cut-off folium of Descartes

In general, simple arcs already correspond to an intuitive idea of a line; in any event it follows from the theorem on the topological invariance of dimension (see Lecture 8 below) that they have no interior points (for $n>1$). At the same time their structure is quite complicated.

Example 1. Let $x = x(t)$, $y = y(t)$, $0 \leqslant t \leqslant 1$ be parametric equations of the Peano curve in the plane. Then the equations

$$x = x(t), \quad y = y(t), \quad z = t, \quad 0 \leqslant t \leqslant 1$$

will give in space a simple arc whose projection onto the x, y plane is a square. Figuratively speaking, this means that a square area could be completely covered by a roof which is nevertheless a line rather than a surface!

Recall (from calculus) that a real function given on (a, b) is said to be a *smooth C^r-function*, where r is either a natural number or ∞, if it has continuous derivatives of all orders $\leqslant r$ (for $r = \infty$ this means by definition that there are continuous partial derivatives of all orders; see above).

In accordance with this we shall say that (8) given on $I = (a, b)$ is a *smooth C^r-curve* if all functions (10) are C^r-functions. Since the derivatives

$$x^i(t)' = \frac{dx^i(t)}{dt}, \quad i = 1, \ldots, n,$$

of functions (10) are the coordinates of a vector

(12) $$\mathbf{r}'(t) = \lim_{h\to 0} \frac{\mathbf{r}(t+h)-\mathbf{r}(t)}{h},$$

this condition means that continuous derivatives of all orders $\leqslant r$ for $r \neq \infty$ (continuous derivatives of all orders for $r = \infty$) exist for vector function (9).

In what follows we shall always consider the number r to be sufficiently large for all differentiations we need to have meaning, and no reference to the class C^r will be made as a rule.

When the interval I has end points (i.e. when either $I = [a, b]$, or $I = [a, b)$ or $I = (a, b]$, curve (8) is said to be smooth if it bounds some smooth curve (of the given class C^r) defined on some large interval $I' \supset I$.

Problem 5. Prove that this is equivalent to the fact that (for $r \neq \infty$) functions (10) have continuous derivatives of all orders $\leqslant r$ on (a, b) and the corresponding one-sided derivatives at points $t = a$ and/or $t = b$.

A closed curve (11) is said to be smooth if in addition the one-sided derivatives at $t = a$ and $t = b$ coincide.

Vector (12) is called a *tangent vector* to the smooth curve (11) at t. Somewhat loosely, it is also called a tangent vector at a *point* $\gamma(t)$. (For simple curves this terminology is quite valid, though.)

In Lecture 15 we prove Sard's theorem from which it follows in particular that the *support of a smooth curve has no interior points* (and is even the so-called set of measure zero). Since the projection of a smooth curve is obviously a smooth curve, it follows that the phenomenon described in Example 1 is impossible in the class of smooth curves.

It is interesting that a smooth curve may have breaks.

Example 2. A curve in the plane with equations

(13) $$x = \alpha(t), \ y = \alpha(-t), \quad -\infty < t < +\infty,$$

where α is the function in Lemma 1, has a support which consists of two coordinate half-lines $x = 0$, $y \geqslant 0$ and $x \geqslant 0$, $y = 0$ meeting at right angles!

Curve (8) (or (1)) is called *regular at* t_0 if $\mathbf{r}'(t_0) \neq 0$. A curve regular at all points is called *regular*.

Notice that curve (13) is not regular at the break point $t = 0$. This is not accidental, since it is known from calculus that the *support of a simple curve* (8) *regular at* t_0 *has at* $\gamma(t_0)$ *a single tangent* ($\mathbf{r}'$, (t_0) serving as the direction vector).

The two curves

(14) $$\gamma: I \to \mathcal{A}, \quad \gamma^*: I^* \to \mathcal{A},$$

where I and I^* the intervals of the same type (both are closed, both are open or both are half-open), are called *equivalent* if there is a smooth (C^r-) function

(15) $$\varphi: I^* \to I,$$

with a derivative nonzero everywhere, which maps I^* onto I and such that $\gamma^* = \gamma \circ \varphi$, i.e. such that

(16) $$\gamma^*(t^*) = \gamma(\varphi(t^*)) \text{ for any } t^* \in I^*.$$

It is also said that the function φ brings about a *change of parameter* on the curve γ.

The equivalence classes of curves are called *nonparametrized curves.* To stress the difference between curves and nonparametrized curves the former are sometimes called *parametrized curves.*

A nonparametrized curve is said to be *smooth*, *simple* or *regular* if it is the equivalence class of a smooth, simple or regular parametrized curve. Since a curve equivalent to a smooth, simple or regular curve is obviously also smooth or respectively simple and regular, this definition is correct.

If curves (14) are connected by relation (16), where φ is in general an arbitrary function, then the supports of these curves coincide. *Equivalent curves* therefore *have the same support* (which is called the support of the corresponding nonparametrized curve), but the converse is in general false.

In the class of simple and regular curves, however, the situation is more satisfactory.

Proposition 1. *If both curves* (14) *are simple and regular, then they have the same support if and only if they are equivalent.*

Proof. If curves (14) are simple and have the same support, then a continuous (why?) mapping $\varphi = \gamma^{-1} \circ \gamma^*$ of I^* onto I is correctly defined. It is only necessary therefore to prove that the mapping φ is smooth and that its derivative is everywhere nonzero.

Let t_0^* be an arbitrary point on I^* and let $t_0 = \varphi(t_0^*)$. In that case, if $p_0 = \gamma^*(t_0^*)$, then $p_0 = \gamma(t_0)$ and the support of curves (14) has a single tangent at p_0. If $\mathbf{r} = \mathbf{r}(t)$ and $\mathbf{r} = \mathbf{r}^*(t^*)$ are vector parametric equations of curves (14), then $\mathbf{r}'(t_0)$ and $\mathbf{r}^{*\prime}(t_0^*)$ will be the direction vectors of the tangent. Therefore *these vectors are collinear.*

Since the curve γ is regular at t_0, we have $\mathbf{r}'(t_0) \neq 0$. Therefore if

$$x^1 = f^1(t), \ldots, x^n = f^n(t), \quad t \in I,$$

and

$$x^1 = g^1(t^*), \ldots, x^n = g^n(t^*), \quad t^* \in I^*,$$

are the coordinate parametric equations of curves (14), then it may be assumed without loss of generality that $\frac{df^1}{dt}(t_0) \neq 0$ and therefore, by virtue of $\mathbf{r}'(t_0)$ and $\mathbf{r}^{*\prime}(t_0^*)$ being collinear, that $\frac{dg^1}{dt}(t_0^*) \neq 0$.

But if $\frac{df^1}{dt}(t_0) \neq 0$, then by the **inverse function theorem** known from calculus f^1 is locally invertible, i.e. there exists an interval (a, b) on the x-axis which contains $x_0 = f^1(t_0)$, and a function $t = h(x)$ mapping that interval onto some other interval (α, β) of the axis t, containing t_0 (and contained in I), such that

$$h(f^1(t)) = t \text{ for any point } t \in (\alpha, \beta).$$

The smooth function h belongs to the same class C^r as the smooth function f^1 and its derivative is nonzero at x_0.

By construction $g^1(t_0^*) = f^1(t_0) = x_0 \in (a, b)$. Therefore, there is an interval (α^*, β^*) on t^*, containing t_0^* and contained in I^*, such that $g^1(t^*) \in (a, b)$ for any point $t^* \in (\alpha^*, \beta^*)$. Consequently, defined on (α^*, β^*) is a function

$$h \circ g^1 : t^* \mapsto h(g^1(t^*)) \tag{17}$$

which assumes values on (α, β). This function belongs to smooth functions of class C^r and has the property that its derivative is nonzero at t_0^*.

On the other hand, under the hypothesis

$$f^i(\varphi(t^*)) = g^i(t^*)$$

for any point $t^* \in I^*$ and any $i = 1, \ldots, n$; in particular, for $t^* \in (\alpha^*, \beta^*)$ and $i = 1$. Therefore $\varphi(t^*) = (h \circ g^1)(t^*)$ for $t^* \in (\alpha^*, \beta^*)$, i.e. function (17) is a restriction of the function φ to the interval (α^*, β^*). Hence φ belongs to the smoothness function of class C^r on (α^*, β^*) and its derivative is nonzero at t_0^*. Since t_0^* is an arbitrary point on I^*, and intervals of the form (α^*, β^*) cover the whole of this interval, this proves that φ is of class C^r on the entire interval I^* and that its derivative is nonzero everywhere on I^*.

This completes the proof of Proposition 1. □

Proposition 1 implies that simple regular curves are uniquely defined up to equivalence by their supports (and may therefore be identified with them). These supports are called *regular simple arcs*. A regular simple curve whose support is a regular simple arc $\mathcal{L}$ is called a *parametrization* of $\mathcal{L}$. As a rule we shall identify regular simple arcs with their parametrizations (considered up to equivalence).

Remark 4. Although generalizing Proposition 1 to arbitrary curves by more general changes of the parameter (for example, with vanishing derivatives) seems natural at first sight, alas it is one of the many far-fetched problems that are meaningless.

For $n = 2$ (in the plane) any graph is obviously a regular simple curve. Moreover, it can be shown (try to do it!) that *in the plane regular simple arcs are precisely simple lines*. Thus, in relation to simple lines all explicit definitions of the intuitive idea of a line lead to the same result.

Problem 6. Show that *in the plane every regular curve* (8) *is locally equivalent to a graph*, i. e. for any point $t_0 \in I$ there is its neighbourhood $(a, b) \subset I$ in $\mathbb{R}$ such that the curve $\gamma|_{(a,b)}$ is equivalent to a curve with equations of the form $x = t$, $y = f(t)$ (and is hence a regular simple curve).

Problem 7. Give an example which shows that a regular curve, which is an injective mapping, cannot nevertheless be a simple curve (and may even have a whole interval of nonsimple points)e.

If $\mathscr{A}$ is a Euclidean space, then for any smooth curve (8) on I a function

$$t \mapsto |\mathbf{r}'(t)|, \quad t \in I,$$

is defined, which is the length of a tangent vector $\mathbf{r}'(t)$. This function is trivially continuous, and hence when $I = [a, b]$ there is an integral

$$s = \int_a^b |\mathbf{r}'(t)|\, dt$$

of this function over $[a, b]$. As shown in calculus, this integral is equal to the limit of the length of the broken lines which are a refinement of curve (8), i.e. to the *length of curve* (8).

Now let I be an arbitrary interval, and let $t_0 \in I$. Then the formula

$$s(t) = \int_{t_0}^{t} |\mathbf{r}'(t)|\, dt, \quad t \in I, \tag{18}$$

defines on I a smooth function mapping I onto some other interval J of the s-axis, which contains the point 0. This function is called the *arc length*. (Notice that it may assume negative values as well.)

If $s(t) = t - t_0$, then the parameter t is called natural. Allowing the inaccuracy generally accepted in calculus we may thus say that the *parameter t is natural if it is an arc length.*

The property of a parameter to be natural is equivalent to the identity $s'(t) = 1$. Since by definition $s'(t) =$

$|\mathbf{r}'(t)|$, we see therefore that the *parameter t on curve* (8) *is natural if and only if*

$$|\mathbf{r}'(t)| = 1 \text{ for all } t \in I.$$

In particular, we see that a *curve referred to a natural parameter is trivially regular.*

Conversely, let curve (8) be regular. Then $|\mathbf{r}'(t)| > 0$ for all $t \in I$, and therefore function (18) is monotonic and an inverse function

$$\varphi: J \to I$$

is defined for it. The curve

$$\gamma_1 = \gamma \circ \varphi: J \to \mathcal{A}$$

is equivalent to the curve γ and we have for it

$$\mathbf{r}_1'(s) = \mathbf{r}'(t) \frac{d\varphi}{ds}(s) = \mathbf{r}'(t) \frac{1}{s'(t)},$$

where $t = \varphi(s)$. Since $s'(t) = |\mathbf{r}'(t)|$ it follows that

$$|\mathbf{r}_1'(s)| = 1 \text{ for all } s \in J,$$

i.e. that the parameter s on the curve γ_1 is natural.

We thus see that *every regular curve is equivalent to a curve referred to a natural parameter.*

Therefore, since we restrict ourselves to regular (and in addition, simple) curves, all curves under consideration may be, without loss of generality, referred to a natural parameter. It is important to bear in mind that *for a regular simple curve the natural parameter is defined up to a transformation of the form*

$$t \mapsto \pm t + t_0$$

(i.e. up to a starting point and the direction of the measurement chosen).

In what follows the natural parameter will, as a rule, be denoted by s.

Differentiation with respect to s will be denoted by a point:

$$\dot{\mathbf{r}}(s) = \frac{d\mathbf{r}(s)}{ds}, \quad \ddot{\mathbf{r}}(s) = \frac{d^2\mathbf{r}(s)}{ds^2}, \quad \ldots, \text{ etc.}$$

As we have already seen, the parameter s naturalness is equivalent to the identity

$$|\dot{\mathbf{r}}(s)| = 1 \qquad \text{for all } s.$$

In this connection it is good to have in mind the following lemma (in which s is, of course, not the natural parameter):

Lemma 3. *Let $\mathbf{u} = \mathbf{u}(s)$ be a vector-valued smooth function such that $|\mathbf{u}(s)| = 1$ for all s. Then*

$$\mathbf{u}(s)\,\dot{\mathbf{u}}(s) = 0 \qquad \textit{for every } s. \tag{19}$$

Proof. The equation $|\mathbf{u}(s)| = 1$ is equivalent to $\mathbf{u}(s)^2 = 1$. But it is easy to see that for a scalar (as well as for a vector) multiplication of vectors the usual product differentiation formula remains. In particular,

$$(\mathbf{u}^2)^{\cdot} = \dot{\mathbf{u}}\mathbf{u} + \mathbf{u}\dot{\mathbf{u}} = 2\mathbf{u}\dot{\mathbf{u}}.$$

Therefore, if $\mathbf{u}^2 = 1$, then $\mathbf{u}\dot{\mathbf{u}} = 0$. □

Corollary. *For any curve $\mathbf{r} = \mathbf{r}(s)$ referred to the natural parameter there is a formula*

$$\dot{\mathbf{r}}(s)\,\ddot{\mathbf{r}}(s) = 0 \qquad \textit{for every } s. \tag{20}$$

Lecture 2

Curves in the plane · Frenet formulas for a space curve · Projections of a curve onto the coordinate planes of the canonical frame · Frenet formulas for a curve in an n-dimensional space · The existence and uniqueness of a curve with given curvature

Let

(1) $$\gamma: I \to \mathcal{A}$$

be an arbitrary regular and simple curve in an n-dimensional Euclidean space $\mathcal{A}$. As we know from the preceding lecture, curve (1) without loss of generality may be referred to the natural parameters s.

Let $\mathbf{r} = \mathbf{r}(s)$ be a vector parametric equation of curve (1) and let

(2) $$\mathbf{t}(s) = \dot{\mathbf{r}}(s)$$

be its tangent vector. Since s is the natural parameter, vector (2) is a unit vector and the vector

$$\dot{\mathbf{t}}(s) = \ddot{\mathbf{r}}(s)$$

is orthogonal to it:

$$\mathbf{t}(s)\,\dot{\mathbf{t}}(s) = 0 \text{ for all } s.$$

Definition 1. The length $|\dot{\mathbf{t}}(s)|$ of a vector $\dot{\mathbf{t}}(s)$ is denoted by $k(s)$ (or simply k) and called the *curvature* of the curve (1) at s (or $\mathbf{r}(s)$).

For example, for a plane curve

$$k(s) = \sqrt{\ddot{x}^2(s) + \ddot{y}^2(s)},$$

where $x = x(s)$ and $y = y(s)$ are coordinate parametric equations of curve (1) in a Euclidean coordinate system x, y.

By the curvature of a curve referred to an arbtrary parameter is meant the curvature of an equivalent curve referred to the natural parameter. The formula for this curvature (which can be obtained by simple but rather cumbersome calculations using nothing but formulas for differentiation of functions) has, even for plane curves, quite a complicated form:

$$k = \left| \frac{x''y' - y''x'}{[(x')^2 + (y')^2]^{3/2}} \right|.$$

Obviously, the number k is the instantaneous velocity of rotation of a unit vector $\mathbf{t}$. The greater the curvature of the curve, the greater the velocity. Hence the term "curvature".

In an oriented plane the so-called *relative curvature* k_{rel} equal to the curvature k can be considered when (for $k \neq 0$) vectors $\mathbf{t}$ and $\dot{\mathbf{t}}$ form a positively oriented basis for the plane, and to $-k$ otherwise. We shall need this curvature in Lecture 4.

Example 1. If

$$x = x_0 + sl, \quad y = y_0 + sm, \quad \text{where } l^2 + m^2 = 1,$$

i.e. if the curve in question is a straight line, then $\ddot{x} = 0$ and $\ddot{y} = 0$. Therefore $k = 0$ for all s, i.e. as was to be expected, the *curvature of a straight line is identically zero.*

Since linear functions are the only functions whose second derivative is identically zero, the converse is also true, i.e. a *curve whose curvature is identically zero is a straight line* (or its segment).

The point $\mathbf{r}_0 = \mathbf{r}(s_0)$ on a curve $\mathbf{r} = \mathbf{r}(s)$ is said to be the *point of rectification* if $k(s_0) = 0$.

Example 2. Parametric equations of a circle of radius R in the natural parameter s are, obviously, of the form

$$x = R\cos\frac{s}{R}, \quad y = R\sin\frac{s}{R}.$$

Since

$$\ddot{x} = -\frac{1}{R}\cos\frac{s}{R}, \quad \ddot{y} = -\frac{1}{R}\sin\frac{s}{R},$$

for a circle

$$k(s) = \frac{1}{R} \qquad \text{for all } s.$$

Thus the *curvature of a circle is constant and equal to the inverse of its radius.*

From the general Theorem 1 to be proved below the converse follows, namely a *curve with constant curvature is a circle* (or its arc).

If for some curve $k(s_0) \neq 0$, then a number $R(s_0) = \frac{1}{k(s_0)}$ is defined which is called the *radius of curvature* of the curve at the point in question.

A curve $\mathbf{r} = \mathbf{r}(s)$ is called a *generic curve* (*a curve of general type*) if there are no points of rectification on it, i.e. if $k(s) \neq 0$ for all s. At each point of such a curve the unit vector

$$\mathbf{n}(s) = \frac{\dot{\mathbf{t}}(s)}{k(s)}$$

is defined which is directed along the *normal* to the curve (i.e. along the straight line passing through the point of tangency at right angles to the tangent).

For any s the vectors $\mathbf{t}(s)$ and $\mathbf{n}(s)$ form an orthonormal basis which is called *Frenet's moving basis* for a given generic curve.

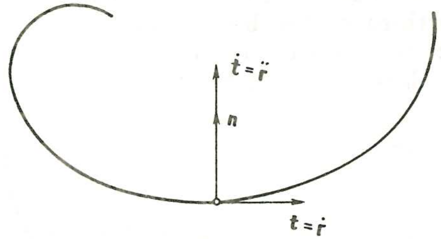
By definition

$$\dot{\mathbf{t}}(s) = k(s)\,\mathbf{n}(s).$$

Let us find a similar formula for $\dot{\mathbf{n}}(s)$. Let

$$\dot{\mathbf{n}}(s) = \alpha(s)\,\mathbf{t}(s) + \beta(s)\,\mathbf{n}(s)$$

be a decomposition of $\dot{\mathbf{n}}(s)$ in terms of the basis $\mathbf{t} = \mathbf{t}(s)$, $\mathbf{n} = \mathbf{n}(s)$. Since $\mathbf{tn} = 0$, we have $\mathbf{t}\dot{\mathbf{n}} + \dot{\mathbf{t}}\mathbf{n} = 0$ (we use again the fact that for a scalar product of vectors the formula for product differentiation is valid), and therefore $\alpha = \mathbf{t}\dot{\mathbf{n}} = -\dot{\mathbf{t}}\mathbf{n} = -k$. On the other hand, according



The Frenet basis for a plane curve

to Lemma 2, Lecture 1, $\beta = \mathbf{n}\dot{\mathbf{n}} = 0$. This proves that for *any generic curve* formulas

$$\text{(3)} \qquad \begin{aligned} \dot{\mathbf{t}} &= k\mathbf{n} \\ \dot{\mathbf{n}} &= -k\mathbf{t} \end{aligned}$$

hold (we omit the hint about the argument s) which describe the instantaneous rotation of the moving basis.

Formulas (3) are called *Frenet's formulas for a plane curve.*

Remark 1. In an oriented plane the Frenet basis can also be defined for curves with points of rectification, taking as $\mathbf{n}(s)$ a vector forming together with $\mathbf{t}(s)$ a positively oriented basis for the plane. Then in formulas (3) a relative curvature k_{rel} appears instead of the curvature k.

For curves in a three-dimensional space (referred to rectangular coordinates x, y, z) the formula for curvature is as follows:

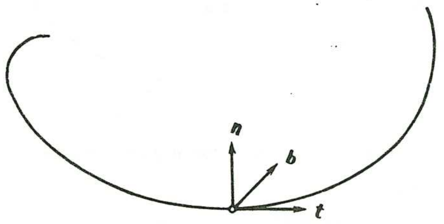
$$k = \sqrt{\ddot{x}^2 + \ddot{y}^2 + \ddot{z}^2}.$$

Curve (1) for $n = 3$, as for $n = 2$, is called a *generic curve* if $k(s) \neq 0$ for all s. For such a curve a unit vector

$$\mathbf{n}(s) = \frac{\dot{\mathbf{t}}(s)}{k(s)}$$

is defined which is called the *vector of principal normal* to the curve.

But now we can (assuming the space to be oriented) introduce a third vector, $\mathbf{b}(s)$, which forms together with $\mathbf{t}(s)$ and $\mathbf{n}(s)$ a positively oriented orthonormal basis $\mathbf{t}(s)$, $\mathbf{n}(s)$, $\mathbf{b}(s)$ (i.e. such that $\mathbf{b}(s) = \mathbf{t}(s) \times \mathbf{n}(s)$).



The Frenet basis for a space curve

This vector is called a *binormal vector*, and the basis $\mathbf{t}(s)$, $\mathbf{n}(s)$, $\mathbf{b}(s)$ *is Frenet's moving basis* for a given generic curve.

By construction (to simplify the formulas we omit the argument s)

$$\dot{\mathbf{t}} = k\mathbf{n}.$$

Moreover, since $\mathbf{b} = \mathbf{t} \times \mathbf{n}$, we have

$$\dot{\mathbf{b}} = \dot{\mathbf{t}} \times \mathbf{n} + \mathbf{t} \times \dot{\mathbf{n}} = \mathbf{t} \times \dot{\mathbf{n}}$$

from which it follows that $\dot{\mathbf{b}}\mathbf{t} = 0$. Since by Lemma 2, in Lecture 1, $\dot{\mathbf{b}}\mathbf{b} = 0$, this proves that $\dot{\mathbf{b}}$ is collinear with $\mathbf{n}$, i.e. there is a number $\varkappa = \varkappa(s)$ such that

$$\dot{\mathbf{b}} = -\varkappa\mathbf{n}.$$

The number $\varkappa(s)$ is called the *torsion* of a given curve at $\mathbf{r}(s)$. It is the rate of turn of a binormal vector.

On differentiating $\mathbf{nt} = 0$ and $\mathbf{nb} = 0$ we immediately get $\dot{\mathbf{n}}\mathbf{t} = -\mathbf{n}\dot{\mathbf{t}} = -k$ and $\dot{\mathbf{n}}\mathbf{b} = -\mathbf{n}\dot{\mathbf{b}} = \varkappa$. Since, in addition, $\dot{\mathbf{n}}\mathbf{n} = 0$ (Lemma 2 in Lecture 1), this proves that

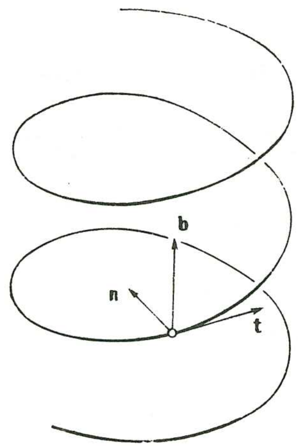
$$\dot{\mathbf{n}} = -k\mathbf{t} + \varkappa\mathbf{b}.$$

Thus, *for any generic curve we have*

$$\text{(4)} \qquad \begin{aligned} \dot{\mathbf{t}} &= k\mathbf{n}, \\ \dot{\mathbf{n}} &= -k\mathbf{t} + \varkappa\mathbf{b}, \\ \dot{\mathbf{b}} &= -\varkappa\mathbf{n}. \end{aligned}$$

These formulas are called *Frenet's formulas for a space curve.*

Example 3. If a curve $\mathbf{r} = \mathbf{r}(s)$ is in a plane Π, then the vectors $\dot{\mathbf{r}}(s)$ and $\ddot{\mathbf{r}}(s)$ are parallel to that plane (for this is



A circular helix

the case for increments $\mathbf{r}(s + \Delta s) - \mathbf{r}(s)$ and $\dot{\mathbf{r}}(s + \Delta s) - \dot{\mathbf{r}}(s)$ of the vectors $\mathbf{r}(s)$ and $\dot{\mathbf{r}}(s)$). Therefore $\mathbf{t}(s)$, $\mathbf{n}(s) \parallel \Pi$ and hence $\mathbf{b}(s) \perp \Pi$. This proves that $\mathbf{b}(s) =$

const and therefore $\varkappa(s) = 0$ for all s. Conversely, let $\varkappa(s) = 0$ for all s and hence $\mathbf{b}(s) = \mathbf{b}_0 = \text{const}$. Then $(\mathbf{r}(s)\,\mathbf{b}_0)^{\cdot} = \dot{\mathbf{r}}(s)\,\mathbf{b}_0 = \mathbf{t}(s)\,\mathbf{b}_0 = 0$ for all s and therefore $\mathbf{r}(s)\,\mathbf{b}_0 = \text{const}$. This means that the curve $\mathbf{r} = \mathbf{r}(s)$ is in the plane $\mathbf{r}\mathbf{b}_0 = \text{const}$. Thus *a curve in space is a plane curve if and only if its torsion is identically zero.*

Example 4. A *circular helix* is the path described by a point moving with constant velocity around a generator of a right circular cylinder rotating uniformly about its axis. The parametric equations of the circular helix are of the form

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

Since $x' = -a\sin t$, $y' = a\cos t$, $z' = b$ we have

$$s' = \sqrt{(x')^2 + (y')^2 + (z')^2} = \sqrt{a^2 + b^2}$$

and hence $s = ct$, where $c = \sqrt{a^2 + b^2}$. Therefore

$$x = a\cos\frac{s}{c}, \quad y = a\sin\frac{s}{c}, \quad z = \frac{b}{c}s.$$

But then

$$\dot{x} = -\frac{a}{c}\sin\frac{s}{c}, \quad \dot{y} = \frac{a}{c}\cos\frac{s}{c}, \quad \dot{z} = \frac{b}{c},$$

$$\ddot{x} = -\frac{a}{c^2}\cos\frac{s}{c}, \quad \ddot{y} = -\frac{a}{c^2}\sin\frac{s}{c}, \quad \ddot{z} = 0$$

and hence

$$k = \sqrt{\ddot{x}^2 + \ddot{y}^2 + \ddot{z}^2} = \frac{a}{c^2} = \text{const}.$$

Besides,

$$\mathbf{t} = \left(-\frac{a}{c}\sin\frac{s}{c}, \frac{a}{c}\cos\frac{s}{c}, \frac{b}{c}\right),$$

$$\mathbf{n} = \left(-\cos\frac{s}{c}, -\sin\frac{s}{c}, 0\right)$$

and

$$\mathbf{b} = \mathbf{t}\times\mathbf{n} = \left(\frac{b}{c}\sin\frac{s}{c}, -\frac{b}{c}\cos\frac{s}{c}, \frac{a}{c}\right).$$

Therefore

$$\dot{\mathbf{b}} = \left(\frac{b}{c^2}\cos\frac{s}{c}, \frac{b}{c^2}\sin\frac{s}{c}, 0\right) = -\frac{b}{c^2}\mathbf{n}$$

and consequently

$$\varkappa = \frac{b}{c^2} = \text{const}.$$

Thus the *curvature and torsion of a circular helix are constant.*

According to the general Theorem 1 to be proved below, and conversely, *every curve whose curvature and torsion are constant is a circular helix* (or its arc).

Remark 2. Note the difference in the treatment of the concept of a generic curve in the plane and in space. To achieve unity, for curves in the plane we must consider relative rather than absolute curvature. Cf. Remark 1.

To investigate the behaviour of an arbitrary space curve $\mathbf{r} = \mathbf{r}(s)$ near one of its points we choose the origin O at that point, take the moving basis $\mathbf{t}_0$, $\mathbf{n}_0$, $\mathbf{b}_0$ at O as the coordinate basis $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ and count off the natural parameter s from 0. Then

$$\mathbf{r}(0) = 0, \quad \dot{\mathbf{r}}(0) = \mathbf{t}_0 = \mathbf{i}, \quad \ddot{\mathbf{r}}(0) = k_0\mathbf{n}_0 = k_0\mathbf{j},$$

$$\dddot{\mathbf{r}}(0) = \dot{k}_0\mathbf{n}_0 + k_0\dot{\mathbf{n}}_0 = -k^2\mathbf{i} + \dot{k}_0\mathbf{j} + k_0\varkappa_0\mathbf{k},$$

where k_0, $\dot{k}_0$ and $\varkappa_0$ are the values of the functions k, $\dot{k}$ and $\varkappa$ for $s = 0$. Hence, using the Taylor formula,

$$\mathbf{r}(s) = \mathbf{r}(0) + s\dot{\mathbf{r}}(0) + \frac{s^2}{2}\ddot{\mathbf{r}}(0) + \frac{s^3}{6}\dddot{\mathbf{r}}(0) + \ldots$$

$$= (s + \ldots)\,\mathbf{i} + \left(\frac{k_0}{2}s + \ldots\right)\mathbf{j} + \left(\frac{k_0\varkappa_0}{6}s^3 + \ldots\right)\mathbf{k}.$$

This means that near O our curve is given by the parametric equations

$$x = s + \ldots$$

$$y = \frac{k_0}{2}s^2 + \ldots$$

$$z = \frac{k_0\varkappa_0}{6}s^3 + \ldots$$

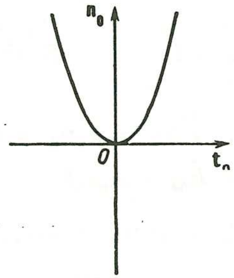
If $k_0 \neq 0$, $\varkappa_0 \neq 0$, then the projection of the curve onto the plane $O\mathbf{i}\mathbf{j} = O\mathbf{t}_0\mathbf{n}_0$ (the plane is called the *osculating* *plane* of a space curve at O) approximately coincides with the parabola

$$x = s, \quad y = \frac{k_0}{2} s^2,$$

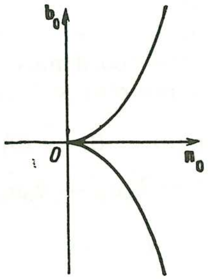
its projection onto the plane $O\mathbf{jk} = O\mathbf{n}_0\mathbf{b}_0$ (called the *normal plane* to a space curve at O) coincides with the semicubical parabola

$$y = \frac{k_0}{2} s^2, \quad z = \frac{k_0\varkappa_0}{6} s^3,$$

and, finally, its projection onto the plane $O\mathbf{ik} = O\mathbf{t}_0\mathbf{b}_0$



Projection onto an osculating plane



Projection onto a normal plane

(called a *rectifying plane* of a space curve at O) coincides with the cubical parabola

$$x = s, \quad z = \frac{k_0\varkappa_0}{2} s^3.$$

This gives a fairly clear idea of how a space curve is constructed near any of its points (at which curvature and torsion are nonzero).

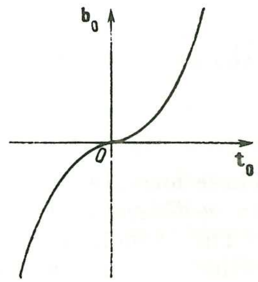
Let us now consider the general case of a Euclidean space of an arbitrary dimension $n \geqslant 2$.

A curve $\mathbf{r} = \mathbf{r}(s)$ (referred to the natural parameter) in an n-dimensional oriented Euclidean space is called *a generic curve* if for any s the vectors

(5) $\quad \dot{\mathbf{r}}(s), \ \ldots, \ \overset{(n-1)}{\mathbf{r}}(s)$

are linearly independent.

By applying to vectors (5) the Gram-Schmidt orthogonalization process we obtain an orthonormal family of vectors $\mathbf{t}_1(s), \ldots, \mathbf{t}_{n-1}(s)$. Let $\mathbf{t}_n(s)$ be a vector (uniformly defined) extending that family to a positively oriented orthonormal basis



Projection onto a rectifying plane

(6) $\quad \mathbf{t}_1(s), \ \ldots, \ \mathbf{t}_{n-1}(s), \quad \mathbf{t}_n(s).$

Definition 2. Basis (6) is called *Frenet's moving basis* of a generic curve at a point $\mathbf{r}(s)$.

Let

$$\dot{\mathbf{t}}_i = \sum_{j=1}^{n} \alpha_{ij}\mathbf{t}_j, \quad i = 1, \ldots, n$$

(we omit the argument s to simplify the formulas). Since by construction the vector $\mathbf{t}_i$, $i = 1, \ldots, n-1$, is expressed linearly in terms of the vectors $\dot{\mathbf{r}}, \ldots, \overset{(i)}{\mathbf{r}}$, the vector $\dot{\mathbf{t}}_i$ is expressed linearly in terms of the vectors $\dot{\mathbf{r}}, \ldots, \overset{(i+1)}{\mathbf{r}}$. Since the latter vectors can be expressed linearly in terms of $\mathbf{t}_1, \ldots, \mathbf{t}_{i+1}$, this proves that $\alpha_{ij} = 0$ provided that $j > i + 1$.

On the other hand, since $\mathbf{t}_i\mathbf{t}_j = \delta_{ij}$, we have $\dot{\mathbf{t}}_i\mathbf{t}_j + \mathbf{t}_i\dot{\mathbf{t}}_j = 0$, i.e.

$$\alpha_{ij} + \alpha_{ji} = 0.$$

Therefore $\alpha_{ii} = 0$ and $\alpha_{ij} = 0$ for $j < i - 1$.

Thus only the coefficients $\alpha_{i,i+1} = -\alpha_{i+1,i}$ can be nonzero. Setting

$$k_1 = \alpha_{12}, \quad k_2 = \alpha_{23}, \ \ldots, \ k_{n-1} = \alpha_{n-1,n}$$

we therefore see that the following formulas hold

$$(7)\qquad \begin{aligned} \dot{t}_1 &= k_1 t_2, \\ \dot{t}_2 &= -k_1 t_1 + k_2 t_3, \\ &\ldots\ldots\ldots\ldots \\ \dot{t}_{n-1} &= -k_{n-2} t_{n-2} + k_{n-1} t_n, \\ \dot{t}_n &= -k_{n-1} t_{n-1}. \end{aligned}$$

These formulas are called *Frenet's formulas for a curve in an n-dimensional space.*

The functions $k_1 = k_1(s), \ldots, k_{n-1} = k_{n-1}(s)$ are called the *curvatures* of a curve. They are defined only for a generic curve.

In the formulas

$$(8)\qquad t_i = \beta_{i1}\dot{r} + \ldots + \beta_{ii}\overset{(i)}{r}_i, \quad i = 1, \ldots, n-1,$$

resulting from applying the Gram-Schmidt orthogonalization process, the coefficients β_{ii} are positive. Therefore in the reverse formulas

$$(9)\qquad \overset{(i)}{r} = \gamma_{i1} t_1 + \ldots + \gamma_{ii} t_i$$

the coefficients $\gamma_{ii} = \beta_{ii}^{-1}$ are also positive. Differentiating formulas (8) we get

$$\dot{t}_i = \dot{\beta}_{i1}\dot{r} + (\dot{\beta}_{i2} + \beta_{i1})\ddot{r} + \ldots + (\dot{\beta}_{ii} + \beta_{i,\,i-1})\overset{(i)}{r} + \beta_{ii}\overset{(i+1)}{r},$$
$$i = 1, \ldots, n-1.$$

On replacing here (for $i < n - 1$) the vectors $\dot{r}, \ldots, \overset{(i+1)}{r}$ by expressions (9) we must get formulas (7). This shows that

$$k_i = \beta_{ii}\gamma_{i+1,i+1}, \quad i = 1, \ldots, n-2.$$

It follows, in particular, that *for any curve of the general type the curvatures*

$$k_1, \ldots, k_{n-2}$$

are positive. The curvature k_{n-1} (the analogue of torsion), on the other hand, may have any sign.

Now we show that any $n - 1$ functions

$$(10)\qquad k_1(s) > 0, \quad \ldots, \quad k_{n-2}(s) > 0, \quad k_{n-1}(s)$$

(given on some interval (a, b) of the axis $\mathbb{R}$) may serve as curvatures for a curve (regular but generally not simple) and that these curvatures uniquely (up to congruence) define the curve.

Let for definiteness $a < 0 < b$.

Theorem 1. *Let $n - 1$ smooth functions* (10), *all positive except possibly the last, be given on an arbitrary interval (a, b). Then for any initial point $O \in \mathcal{A}$ and for any positively oriented orthonormal basis $i_1, \ldots, i_n$ there exists one and only one generic curve $r = r(s)$, $a < s < b$ of the general type having the following two properties:*

1° *the given functions* (10) *are the curvatures of the curve,*

2° *for $s = 0$ we have*

$$r(0) = 0, \quad t_1 = (0) = i_1, \quad \ldots, \quad t_n(0) = i_n.$$

Proof. We carry out the proof in four stages.

Stage 1. At this stage we use the **unique existence theorem** for solutions (UES) of linear ordinary differential equations.

Theorem (UES). *Let m^2 smooth functions $A_{ij}(s)$, $i, j = 1, \ldots, m$ be given on an interval (a, b) and let $x_1^{(0)}, \ldots, x_m^{(0)}$ be arbitrary numbers. Then there is one and only one family of smooth functions $x_1(s), \ldots, x_m(s)$, $a < s < b$, having the following two properties:*

1° *identically in terms of s, $a < s < b$, the relations*

$$(11)\qquad \begin{aligned} \dot{x}_1 &= A_{11}x_1 + \ldots + A_{1m}x_m, \\ &\ldots\ldots\ldots\ldots \\ \dot{x}_m &= A_{m1}x_1 + \ldots + A_{mm}x_m \end{aligned}$$

hold,

2° *for $s = 0$ we have*

$$x_1(0) = x_0^{(0)}, \quad \ldots, \quad x_m(0) = x_m^{(0)}. \quad \square$$

We shall apply this theorem to relations (7) which for the given functions $k_1, \ldots, k_{n-1}$ are equations of the form (11) for $m = n^2$ coordinate vectors $t_1, \ldots, t_n$. Thus, according to the UES theorem, there is one and only one family of vector-valued functions $\mathbf{t}_1(s), \ldots, \mathbf{t}_{n-1}(s)$, $a < s < b$, on (a, b) such that

1° for any s relations (7) are satisfied,

2° for $s = 0$ equations

$$\mathbf{t}_1(0) = \mathbf{i}_1, \ldots, \mathbf{t}_n(0) = \mathbf{i}_n \tag{12}$$

hold.

Stage 2. We consider scalar products $\mathbf{t}_i\mathbf{t}_j$, $i, j = 1, \ldots, n$. According to relations (7), for these products we have

$$\begin{aligned}(\mathbf{t}_i\mathbf{t}_j)^\cdot &= \dot{\mathbf{t}}_i\mathbf{t}_j + \mathbf{t}_i\dot{\mathbf{t}}_j \\ &= (-k_{i-1}\mathbf{t}_{i-1} + k_i\mathbf{t}_{i+1})\,\mathbf{t}_j \\ &\quad + \mathbf{t}_i\,(-k_{j-1}\mathbf{t}_{j-1} + k_j\mathbf{t}_{j+1})\end{aligned}$$

(we assume that $\mathbf{t}_0 = 0$ and $\mathbf{t}_{n+1} = 0$), i.e. the equations

$$\begin{aligned}(\mathbf{t}_i\mathbf{t}_j)^\cdot = {}&-k_{i-1}(\mathbf{t}_{i-1}\mathbf{t}_j) + k_i(\mathbf{t}_{i+1}\mathbf{t}_j) \\ &-k_{j-1}(\mathbf{t}_i\mathbf{t}_{j-1}) + k_j(\mathbf{t}_i\mathbf{t}_{j+1})\end{aligned} \tag{13}$$

which may be regarded as equations of the form (11) for $m = \frac{n(n+1)}{2}$ functions $\mathbf{t}_i\mathbf{t}_j$. By the UES theorem therefore there is one and only one set of these functions having the property that for $s = 0$ they are equal to $\delta_{ij} = \mathbf{i}_i\mathbf{i}_j$ (i.e. to zero if $i \neq j$ and to unity if $i = j$).

On the other hand, a direct verification shows that the functions $\mathbf{t}_i\mathbf{t}_j$ identically equal to δ_{ij} satisfy equations (13). (Indeed, when $i \neq j - 1, j + 1$ all the terms of the sum $-k_{i-1}\delta_{i-1,j} + k_i\delta_{i+1,j} - k_{j-1}\delta_{i,j-1} + k_j\delta_{i,j+1}$ are zero and when $i = j - 1, j + 1$ the sum has only two terms which are nonzero but cancel in pairs.) Hence, for all s there are, by virtue of the UES theorem, equations $\mathbf{t}_i\mathbf{t}_j = \delta_{ij}$, $i, j = 1, \ldots, n$, implying that for any s, $a < s < b$, the vectors $\mathbf{t}_1, \ldots, \mathbf{t}_n$ form an orthonormal basis.

Since for $s = 0$ that basis coincides with a positively oriented basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$, the basis $\mathbf{t}_1, \ldots, \mathbf{t}_n$ is positively oriented for any s, $a < s < b$, too.

Stage 3. We compose consecutive derivatives for the vector $\mathbf{t}_1$

$$\mathbf{t}_1, \dot{\mathbf{t}}_1, \ddot{\mathbf{t}}_1, \ldots, \overset{(n-1)}{\mathbf{t}_1} \tag{14}$$

and apply the Gram-Schmidt orthogonalization process to them. Since $\mathbf{t}_1$ is a unit vector, we need not do anything in the first step of the process. Since by Lemma 2 in Lecture 1 $\dot{\mathbf{t}}_1$ is orthogonal to $\mathbf{t}_1$, in the second step we must only normalize it. However, since we have proved $\mathbf{t}_2$ to be a unit vector and under the hypothesis $k_1 > 0$, according to the first of the relations (7) $|\dot{\mathbf{t}}_1| = k_1$. In the second step therefore we obtain the vector

$$\mathbf{t}_2 = \frac{\dot{\mathbf{t}}_1}{k_1}.$$

In the third step we should consider the vector

$$\ddot{\mathbf{t}}_1 = (k_1\mathbf{t}_2)^\cdot = \dot{k}_1\mathbf{t}_2 + k_1\dot{\mathbf{t}}_2 = -k_1^2\mathbf{t}_1 + \dot{k}_1\mathbf{t}_2 + k_1k_2\mathbf{t}_3,$$

subtract from it the linear combination of vectors $\mathbf{t}_1$ and $\mathbf{t}_2$ so as to obtain a vector orthogonal to those vectors and then normalize the vector. But since according to what has been proved the vectors $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ form an orthonormal family and under the hypothesis $k_1k_2 > 0$, the procedure will obviously yield the vector $\mathbf{t}_3$.

It is clear that this reasoning is of a general character so that at each step of the orthogonalization process we obtain the corresponding vector $\mathbf{t}_i$, $i = 1, \ldots, n - 1$. This proves that the family of vectors $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{n-1}$ is uniquely characterized as an orthonormal family of vectors obtained from family (14) by applying Gram-Schmidt orthogonalization.

Stage 4. Let

$$\mathbf{r}(s) = \int_0^s \mathbf{t}_1(s)\,ds, \quad a < s < b. \tag{15}$$

Then $\mathbf{r}(0) = 0$ and $\dot{\mathbf{r}}(s) = \mathbf{t}_1(s)$, i.e. for $s = 0$ the curve $\mathbf{r} = \mathbf{r}(s)$, $a < s < b$ passes through the point O and for any s the vector $\mathbf{t}_1(s)$ is tangent to it. But for every curve the first $n - 1$ vectors of the moving basis are vectors obtained from the first $n - 1$ derivatives of the tangent vector by the Gram-Schmidt orthogonalization process. According to the foregoing therefore those vectors coincide with $\mathbf{t}_1, \ldots, \mathbf{t}_{n-1}$.

As to the last vector of the moving basis, it is uniquely characterized as a unit vector which together with the first $n - 1$ vectors forms a positively oriented basis. Since the basis $\mathbf{t}_1, \ldots, \mathbf{t}_n$, as we have seen, is positively oriented, the vector $\mathbf{t}_n$ should be the required vector.

Thus, we have proved that for any s the vectors $\mathbf{t}_1(s), \ldots, \mathbf{t}_n(s)$ form the moving basis of the curve $\mathbf{r}=\mathbf{r}(s)$. Since for these vectors the Frenet formulas (7) are valid, the functions $k_i(s)$, $i=1, \ldots, n-1$ appearing in the formulas must be the curvatures of $\mathbf{r} = \mathbf{r}(s)$.

This completes the proof of the existence of a curve $\mathbf{r} = \mathbf{r}(s)$ having Properties 1° and 2°.

The uniqueness of the curve follows from the fact that according to the UES theorem, the moving basis $\mathbf{t}_1(s), \ldots, \mathbf{t}_n(s)$ is uniquely defined by equations (7) and the initial conditions (12) and the radius vector $\mathbf{r}(s)$ is uniquely defined (by formula (15)) by the relation $\dot{\mathbf{r}}(s) = \mathbf{t}_1(s)$ and the initial condition $\mathbf{r}(0) = 0$. □

Lecture 3

Elementary surfaces and their parametrizations · Examples of surfaces · Tangent plane and tangent subspace · Smooth mappings of surfaces and their differentials · Diffeomorphisms of surfaces · The first quadratic form of a surface · Isometries · Beltrami's first differential parameter · Examples of computation of first quadratic forms · Developable surfaces

An intuitive idea of a surface is explicated by analogy with that of a line, but the explication is more difficult that that for a line. We shall restrict our discussion to the analogue of an idea of an open simple regular arc (though considering more general surfaces when discussing particular examples).

To introduce this analogue we begin with an arbitrary continuous mapping of the form

(1) $$\gamma\colon\ U \to \mathscr{A},$$

where $\mathscr{A}$ is some Euclidean (or only affine) space of dimension $n \geqslant 3$ and U is a *convex* (i.e. containing every straight segment whose ends are in the set) open subset of a Euclidean plane $\mathbb{R}^2$ (a two-dimensional analogue of the interval $I = (a, b)$). Once the reference point O has been chosen in $\mathscr{A}$, mapping (1) is given by a continuous vector function

(2) $$\mathbf{r} = \mathbf{r}(u, v), \quad (u, v) \in U$$

which assumes values in the associated vector space $\mathscr{V}$ and when, in addition, a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is chosen in $\mathscr{V}$, (1) is given by n continuous numerical functions

(3) $$x^1 = x^1(u, v), \ \ldots, \ x^n = x^n(u, v),$$

the coordinates in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of the vector $\mathbf{r}(u, v)$.

Mapping (or map) (1) is said to be a C^r-*mapping*, where r is some natural number or a symbol ∞, if every function (3) or equivalently vector function (2) has continuous partial derivatives of all orders $\leqslant r$ (recall that when $r = \infty$ this implies the existence of partial derivatives of all orders). In what follows we shall assume the number r to be once and for ever fixed and sufficiently large. [In this lecture any number $r \geqslant 1$ will suit, but in, say, Lecture 5 we shall have to require that $r \geqslant 3$.]

In particular, for the smooth mapping (1) partial derivatives

(4) $$\mathbf{r}_u = \frac{\partial \mathbf{r}(u, v)}{\partial u}, \quad \mathbf{r}_v = \frac{\partial \mathbf{r}(u, v)}{\partial v}$$

of the vector function (2) are defined. Smooth mapping (1) is said to be *regular* if at every point $(u, v) \in U$ partial derivatives (4) are linearly independent.

Definition 1. Mapping (1) is said to be a *parametrization* if it is:

1° smooth,

2° regular, and

3° moneomorphic (injective and has the property that if a sequence of points $\gamma(u_n, v_n)$, $(u_n, v_n) \in U$, of $\mathcal{A}$ converges to a point of the form $\gamma(a, b)$, where $(a, b) \in U$, then the sequence of points $(u_n, v_n) \in U$ also converges (to the point (a, b), by virtue of continuity).

Problem 1. Prove that if mapping (1) is smooth and regular, then for any point $(u_0, v_0) \in U$ there is its neighbourhood $V \subset U$ on which this mapping is moneomorphic (is a parametrization).

Definition 2. A subset $\mathcal{X}$ of a space $\mathcal{A}$ is said to be an *elementary surface* if there is a parametrization $\gamma: U \to \mathcal{A}$ (called in this case a *parametrization of the surface* $\mathcal{X}$) such that $\gamma(U) = \mathcal{X}$. It is also said that $\mathcal{X}$ is the *support* of a parametrization γ.

Elementary surfaces are the two-dimensional analogues of simple regular arcs (and parametrizations are the analogues of simple regular curves).

Remark 1. In another terminological scheme, which we shall also use now and again, sometimes without stating this explicitly, parametrizations (1) themselves are called the elementary surfaces. For lines we have avoided such homonyms by distinguishing between curves and lines. Unfortunately, there is no such pair of generally accepted terms for the two-dimensional case.

Since, as a rule, we consider only elementary surfaces in what follows we shall call elementary surfaces simply surfaces.

Since the parametrization $\gamma: U \to \mathcal{A}$ of an arbitrary elementary surface $\mathcal{X}$ is an injective mapping, for any point $p \in \mathcal{X}$ there are unique numbers u and v having the property that $(u, v) \in U$ and $\gamma(u, v) = p$. These numbers are called the *coordinates* of p in the given parametrization. Traditionally additional terms are used for these coordinates, they are called *curvilinear* or *local*, although no other coordinates on the surface are usually considered, and therefore these terms are not necessary.

Speaking loosely, numbers u and v are often called the coordinates *on a surface* (1); this is a manifestation of a general tendency to confuse in usage surfaces and their parametrizations.

Every curve in U with parametric equations

(5) $$u = u(t), \quad v = v(t), \quad t \in I,$$

is mapped by parametrization (1) of a surface $\mathcal{X}$ to a curve

(6) $$\mathbf{r} = \mathbf{r}(u(t), v(t)), \quad t \in I,$$

of a space $\mathcal{A}$. Curve (6) is said to *lie* on $\mathcal{X}$ and equations (5) to be its *parametric equations in coordinates* u *and* v.

In particular, curves u = const and v = const (which are the images of coordinate lines in U) are called the *coordinate lines on the surface* $\mathcal{X}$ and their totality is the *coordinate network*.

For any open subsets $U, U^* \subset \mathbb{R}^2$ every mapping

(7) $$\varphi: U^* \to U$$

is given by a pair of functions

(8) $$u = u(u^*, v^*), \quad v = v(u^*, v^*), \quad (u^*, v^*) \in U^*$$

having the property that for any point $(u^*, v^*) \in U^*$ the point $(u, v) = (u(u^*, v^*), v(u^*, v^*))$ is in U. Mapping (7) is said to be a C^r-*mapping*, where r is a natural number or a symbol ∞, if for $r \neq \infty$ functions (8) have continuous partial derivatives of all orders $\leqslant r$. Smooth mapping (7) is said to be a diffeomorphism if it is bijective and the inverse mapping

(8) $$\varphi^{-1}: U \to U^*$$

is also smooth.

Two parametrizations

(9) $$\gamma: U \to \mathcal{A} \quad \text{and} \quad \gamma^*: U^* \to \mathcal{A}$$

are said to be *equivalent* if there is a diffeomorphism (7) such that

(10) $$\gamma^* = \gamma \circ \varphi.$$

Since giving a parametrization of a surface $\mathscr{X}$ is equivalent to giving curvilinear coordinates on $\mathscr{X}$, diffeomorphism (7) is also said to perform a *change of coordinates* on $\mathscr{X}$ (or to specify a *transition* from coordinates u, v to coordinates u^*, v^*).

Problem 2. Prove that relation (10) is an equivalence in the general algebraic sense (is reflexive, symmetric and transitive) and hence we should speak about classes of equivalent parametrizations.

It is clear that *equivalent parametrizations have the same support*. Conversely, we can easily show that *parametrizations* (9) *having the same support are equivalent*. This means that elementary surfaces are bijectively associated with equivalence classes of their parametrizations and therefore can be identified with them.

Problem 3. Prove the last statement. (Notice that for this to be true all the three conditions 1° to 3° of Definition 1 are essential. Cf. the proof of Proposition 1 of Lecture 1.)

In Lecture 15 we prove the general proposition of which this statement (as well as Proposition 1 of Lecture 1) is a special case.

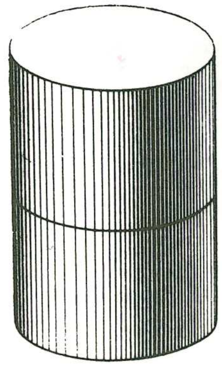
Examples of surfaces.

For clarity we restrict ourselves to surfaces in a three-dimensional Euclidean space. The coordinates x, y, z will be assumed to be rectangular.

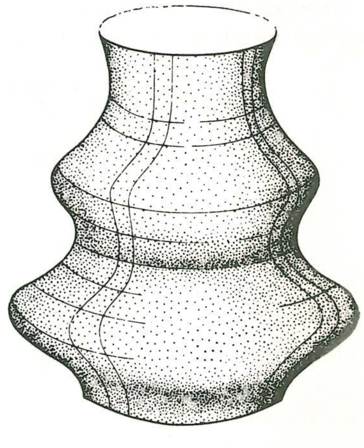
Example 1. The equations

(11) $$x = R\cos u, \quad y = R\sin u, \quad z = v$$

give a right circular cylinder in a three-dimensional Euclidean space. This is not an elementary surface in the sense of Definition 2, since for $-\infty < u < +\infty$ each point of the cylinder is covered an infinite (countable) number of times by the points of the plane $\mathbb{R}^2$. To obtain an elementary surface the cylinder should be cut through along its element, i.e. in (11) the parameter u should satisfy the inequalities $0 < u < 2\pi$. The entire cylinder, however, is covered by two such cut cylinders.



A circular cylinder



A surface of revolution

The coordinate network on cylinder (11) consists of vertical straight lines $u = \text{const}$ and horizontal circles $v = \text{const}$.

Example 2. Let $x = x(v)$, $z = z(v)$ be a simple regular curve in the plane Oxz not intersecting the Oz-axis. The surface with parametrization

(12) $$x = x(v)\cos u, \quad y = x(v)\sin u, \quad z = z(v)$$

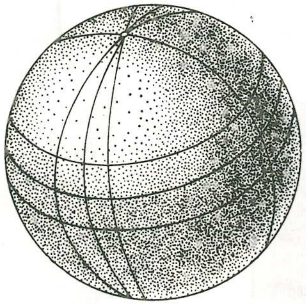
is called a *surface of revolution* and the curve $x = x(v)$, $z = z(v)$ is its *profile*. Graphically, surface (12) is obtained by rotating its profile about the Oz-axis.

The regularity of parametrization (12), i.e. linear independence of vectors

$$\mathbf{r}_u = (-x(v) \sin u, \ x(v) \cos u, \ 0),$$
$$\mathbf{r}_v = (x'(v) \cos u, \ x'(v) \sin u, \ z'(v))$$

is ensured by the regularity of the profile (i.e. by the condition $x'(v)^2 + z'(v)^2 = 1$) and by the fact that the profile does not intersect the Oz-rotation axis (i.e. by $x(v) \neq 0$).

The coordinate network on surface (12) consists of plane curves which are rotations of the profile about the Oz-axis



A sphere

(they are called *meridians*) and circles perpendicular to them (*parallels*). To make surface (12) elementary, it should be cut through along a meridian.

A cylinder is a surface of revolution whose profile is a straight line $x = R$, $z = v$.

A sphere

$$x = R \cos v \cos u, \quad y = R \cos v \sin u, \quad z = R \sin v$$

of radius R with centre at O is a surface of revolution of a circle with profile $x = R \cos v$, $z = R \sin v$. The coordinates u and v on the sphere are the well-known "geographical coordinates", longitude and latitude, and the coordinate curves are geographical meridians and parallels.

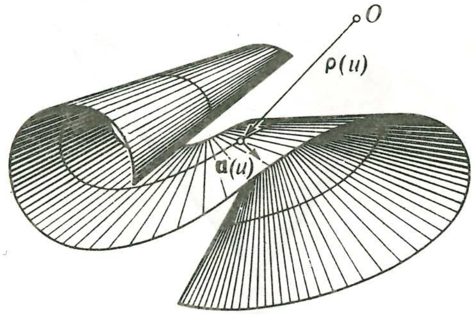
Strictly speaking, the profile of a sphere is only its semicircle $-\pi < v < +\pi$ (which excludes the poles).

To obtain an elementary surface it is necessary to exclude one meridian (the "line of change of dates").

Example 3. A surface $\mathbf{r} = \mathbf{r}(u, v)$ is said to be a *ruled surface* if

$$\mathbf{r}(u, v) = \boldsymbol{\rho}(u) + v\mathbf{a}(u), \tag{13}$$

where $\boldsymbol{\rho}(u)$ and $\mathbf{a}(u)$ are vector-valued functions having the property (ensuring regularity) that the vectors



A ruled surface

$\boldsymbol{\rho}'(u) + v\mathbf{a}'(u)$ and $\mathbf{a}(u)$ are linearly independent for all u and v under consideration (so that, in particular, $\mathbf{a}(u) \neq 0$ for all u). A straight line with a direction vector $\mathbf{a}(u_0)$, which passes through the point with radius vector $\boldsymbol{\rho}(u_0)$ is a coordinate curve $u = u_0 = \text{const}$. Thus, graphically, a ruled surface is generated by various positions of a moving straight line. Cf. Definition 1, Lecture I.23.

It is clear that we may assume without loss of generality the vector $\mathbf{a}(u)$ to be a unit vector:

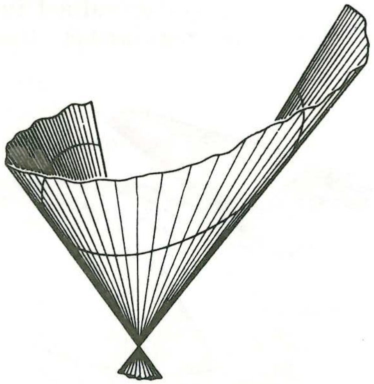
$$|\mathbf{a}(u)| = 1 \text{ for all } u.$$

If $\boldsymbol{\rho}'(u) = 0$ for all u, i.e. $\boldsymbol{\rho}(u) = \text{const}$, then, after translation of the origin, we obtain instead of (13) an equation of the form

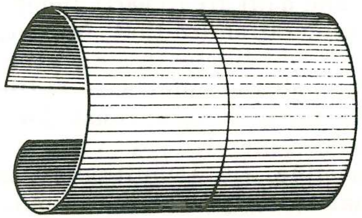
$$\mathbf{r} = v\mathbf{a}(u). \tag{14}$$

It is a cone whose directrix is a regular space curve $\mathbf{r} = \mathbf{a}(u)$. [In (14) it must be assumed that $v > 0$ or $v < 0$ (for $v = 0$ is a singular point of the cone which divides it into two nappes). But if the generators of the cone intersect its directrix at several points, additional constraints have to be introduced on v.]

If $\mathbf{a}'(u) = 0$ for all u, i.e. $\mathbf{a}(u) = \text{const}$, then surface (14) is a cylinder with directrix $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ (generally a space one).



A cone

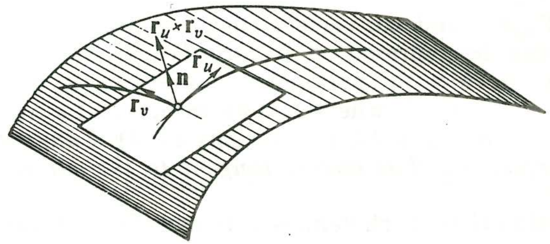


A cylinder

If the vector $\boldsymbol{\rho}'$ is not identically zero, then, going over if necessary to a smaller domain in $\mathbb{R}^2$, we may assume that $\boldsymbol{\rho}'(u) \neq 0$ for all u. Then $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ is a regular curve in space and we may assume that u is the natural parameter (arc length) on that curve. Cone (14) may also be specified by an equation of form (13) with $\boldsymbol{\rho}'(u) \neq 0$. To do this it is sufficient to put $\boldsymbol{\rho}(u) = \mathbf{a}(u)$ in (13) (if $\mathbf{a}'(u) \neq 0$).

If $\mathbf{a}(u)$ is a tangent vector $\boldsymbol{\tau}(u)$ of a curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$, then surface (13) is called a *surface of tangents*. A *surface of principal normals* and a *surface of binormals* are defined similarly.

Notice that for a surface of tangents all points of the curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ are singular points of the surface at which the regularity condition fails. (They form the so-called *edge of regression* of the surface of tangents.)



A tangent plane

Again let $\mathscr{X}$ be an arbitrary (elementary) surface in an n-dimensional affine space $\mathscr{A}$ with parametrization $\mathbf{r} = \mathbf{r}(u, v)$ and let p_0 be an arbitrary point of $\mathscr{X}$ and $\mathbf{r}_0 = \mathbf{r}(u_0, v_0)$ be its radius vector. By the regularity condition the values

$$\mathbf{r}_{u_0} = \frac{\partial \mathbf{r}}{\partial u}(u_0, v_0), \quad \mathbf{r}_{v_0} = \frac{\partial \mathbf{r}}{\partial v}(u_0, v_0)$$

of partial derivatives of the vector function $\mathbf{r} = \mathbf{r}(u, v)$ at the point (u_0, v_0) are linearly independent, i.e. the bivector $\mathbf{r}_{u_0} \wedge \mathbf{r}_{v_0}$ is nonzero. A two-dimensional plane, which has a direction bivector $\mathbf{r}_{u_0} \wedge \mathbf{r}_{v_0}$ and passes through the point p_0 is therefore defined in $\mathscr{A}$. The vector parametric equation of that plane has the form

$$\mathbf{r} = \mathbf{r}_0 + a\mathbf{r}_{u_0} + b\mathbf{r}_{v_0}, \tag{15}$$

where a and b are parameters.

Definition 3. Plane (15) is called the *tangent plane* of a surface $\mathscr{X}$ (or *to a surface* $\mathscr{X}$) at p_0. The corresponding subspace of the associated vector space $\mathscr{V}$ (consisting of vectors of the form $a\mathbf{r}_{u_0} + b\mathbf{r}_{v_0}$) is called a *tangent* *subspace* and is denoted by $\mathbf{T}_{p_0}\mathscr{X}$ and its vectors are *tangent vectors* of $\mathscr{X}$ at p_0. To stress its two-dimensional character, the tangent subspace is often called a *tangent plane*.

This terminology is justified by the fact that for any curve (5) on $\mathscr{X}$ passing (say, for $t = t_0$) through p_0 its tangent vector

(16) $$\mathbf{r}'(t_0) = u'(t_0)\mathbf{r}_{u_0} + v'(t_0)\mathbf{r}_{v_0}$$

is in $\mathbf{T}_{p_0}\mathscr{X}$, and conversely any vector $a\mathbf{r}_u + b\mathbf{r}_v$ in $\mathbf{T}_{p_0}\mathscr{X}$ can be represented in form (16) (it suffices to consider a curve with parametric equations $u = u_0 + at$, $v = v_0 + bt$, $t \in I$, where I is an interval of a t-axis such that $(u_0 + at, v_0 + bt) \in U$ for any $t \in I$). *Tangent vectors of a surface are thus vectors tangent to the curves on that surface.*

Relation (10) which defines a change of coordinates on $\mathscr{X}$ can be written in radius vectors as a formula

$$\mathbf{r}^*(u^*, v^*) = \mathbf{r}(u(u^*, v^*), v(u^*, v^*))$$

differentiation of which gives the relations

(17) $$\mathbf{r}^*_{u^*} = \frac{\partial u}{\partial u^*}\mathbf{r}_u + \frac{\partial v}{\partial u^*}\mathbf{r}_v, \qquad \mathbf{r}^*_{v^*} = \frac{\partial u}{\partial v^*}\mathbf{r}_u + \frac{\partial v}{\partial v^*}\mathbf{r}_v.$$

It follows from relations (17) that the *vectors* $\mathbf{r}^*_{u^*}$ and $\mathbf{r}^*_{v^*}$ *are linearly equivalent to* $\mathbf{r}_u$ and $\mathbf{r}_v$ and hence generate the same subspace. This proves that *for any point* $p \in \mathscr{X}$ *the vector space* $\mathbf{T}_p\mathscr{X}$ *is defined correctly* (is independent of the choice of parametrization $\mathbf{r} = \mathbf{r}(u, v)$). When changing the parametrization in $\mathbf{T}_p\mathscr{X}$ only the basis $\mathbf{r}_u$, $\mathbf{r}_v$ is changed, by formulas (17).

The vectors of space $\mathbf{T}_p\mathscr{X}$ are usually denoted by $d\mathbf{r}$ and their coordinates (in the basis $\mathbf{r}_u$, $\mathbf{r}_v$) by du and dv (these coordinates being written to the right of the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$). Thus, in the notation

(18) $$d\mathbf{r} = \mathbf{r}_u du + \mathbf{r}_v dv$$

for any vector $d\mathbf{r}$ of $\mathbf{T}_p\mathscr{X}$. It follows directly from (17) that the coordinates du, dv are connected with du^*, dv^* in the basis $\mathbf{r}^*_{u^*}$, $\mathbf{r}^*_{v^*}$ by the relations

(19) $$du = \frac{\partial u}{\partial u^*}du^* + \frac{\partial u}{\partial v^*}dv^*, \qquad dv = \frac{\partial v}{\partial u^*}du^* + \frac{\partial v}{\partial v^*}dv^*$$

which formally coincide with the formulas for differentials known from calculus (which is the basic argument for notation (18)).

Now let us consider together with the surface $\mathscr{X}$ its parametrization $\gamma\colon U \to \mathscr{A}$ given by the vector function $\mathbf{r} = \mathbf{r}(u, v)$, $(u, v) \in U$, another elementary surface $\hat{\mathscr{X}}$ with parametrization $\hat{\gamma}\colon \hat{U} \to \mathscr{A}$ given by the vector function $\mathbf{r} = \hat{\mathbf{r}}(\hat{u}, \hat{v})$, $(\hat{u}, \hat{v}) \in \hat{U}$.

Since γ and $\hat{\gamma}$ are injective mappings, every mapping $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ uniquely defines the mapping $\bar{f}\colon U \to \hat{U}$ which sastisfies the relation

(20) $$f \circ \gamma = \hat{\gamma} \circ \bar{f}$$

and uniquely defines the mapping f. Graphically relation (20) means that in the diagram

(21) $$\begin{array}{ccc} \mathscr{X} & \xrightarrow{f} & \hat{\mathscr{X}} \\ \gamma\uparrow & & \uparrow\hat{\gamma} \\ U & \xrightarrow{\bar{f}} & \hat{U} \end{array}$$

going from the bottom left-hand corner to the top right-hand corner along with the two possible paths leads to the same result. [Such diagrams are called *commutative*.]

The mapping $\bar{f}$ is said to *represent* a mapping f in parametrizations γ and $\hat{\gamma}$ and the functions

(22) $$\hat{u} = \hat{u}(u, v), \quad \hat{v} = \hat{v}(u, v), \quad (u, v) \in U$$

which specify $\bar{f}$ are said to *give* f *in coordinates* u, v *and* u^*, v^*.

A mapping f is said to be *smooth* if so is $\bar{f}$, i.e. if so are functions (22). This definition is correct (if $\bar{f}$ is smooth for one choice of parametrizations γ and $\hat{\gamma}$, then it will be so for any other).

We associate every smooth mapping $f: \mathscr{X} \to \hat{\mathscr{X}}$ and any point $p \in \mathscr{X}$ with a linear mapping

$$\mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\hat{p}}\hat{\mathscr{X}}, \quad \hat{p} = f(p) \tag{23}$$

of tangent spaces carrying vector (18) of the space $\mathbf{T}_p\mathscr{X}$ into the vector

$$d\hat{\mathbf{r}} = \hat{\mathbf{r}}_{\hat{u}}\, d\hat{u} + \hat{\mathbf{r}}_{\hat{v}}\, d\hat{v}$$

of the space $\mathbf{T}_{\hat{p}}\hat{\mathscr{X}}$, where in full accord with the differential calculus formulas

$$\begin{aligned} d\hat{u} &= \frac{\partial \hat{u}}{\partial u} du + \frac{\partial \hat{u}}{\partial v} dv, \\ d\hat{v} &= \frac{\partial \hat{v}}{\partial u} du + \frac{\partial \hat{v}}{\partial v} dv. \end{aligned} \tag{24}$$

Mapping (23) is defined for the given parametrizations γ, $\hat{\gamma}$ of surfaces $\mathscr{X}$, $\hat{\mathscr{X}}$ and therefore the question arises as to whether it is correct, i.e. independent of the choice of these parametrizations.

Suppose, for example, we have replaced the parametrization γ by another parametrization γ^*: $U^* \to \mathscr{A}$ of $\mathscr{X}$. By definition $\gamma^* = \gamma \circ \varphi$, where $\varphi: U^* \to U$ is some diffeomorphism given by the functions

$$u = u(u^*, v^*), \quad v = v(u^*, v^*).$$

For the corresponding bases $\mathbf{r}_u$, $\mathbf{r}_v$ and $\mathbf{r}_{u^*}$, $\mathbf{r}_{v^*}$ of $\mathbf{T}_p\mathscr{X}$ formulas (17) hold and for the corresponding coordinates du, dv and du^*, dv^* of the tangent vectors formulas (19) do. In coordinates u^*, v^*, and $\hat{u}$, $\hat{v}$ the mapping f is obviously specified by the functions

$$\begin{aligned} \hat{u}^*(u^*, v^*) &= \hat{u}(u(u^*, v^*), \ v(u^*, v^*)), \\ \hat{v}^*(u^*, v^*) &= \hat{v}(u(u^*, v^*), \ v(u^*, v^*)) \end{aligned}$$

and hence the vector $d\mathbf{r}$ of $\mathbf{T}_p\mathscr{X}$, which has coordinates du^*, dv^* in the basis $\mathbf{r}_{u^*}$, $\mathbf{r}_{v^*}$, will carry mapping (23) constructed using the parametrizations γ^*, $\hat{\gamma}$ into the vector $d\hat{\mathbf{r}}^*$ of $\mathbf{T}_{\hat{p}}\hat{\mathscr{X}}$ which has coordinates

$$d\hat{u}^* = \frac{\partial \hat{u}^*}{\partial u^*} du^* + \frac{\partial \hat{u}^*}{\partial v^*} dv^*,$$

$$d\hat{v}^* = \frac{\partial \hat{v}^*}{\partial u^*} du^* + \frac{\partial \hat{v}^*}{\partial v^*} dv^*$$

in the basis $\hat{\mathbf{r}}_{\hat{u}}$, $\hat{\mathbf{r}}_{\hat{v}}$. On the other hand, on substituting in formulas (24) for the coordinates $d\hat{u}$, $d\hat{v}$ of vector $d\hat{\mathbf{r}}$ expressions (19) for the coordinates du, dv in terms of du^*, dv^* and taking into account the fact that according to the rules for differentiating composite functions known from calculus the equations

$$\frac{\partial \hat{u}}{\partial u}\frac{\partial u}{\partial u^*} + \frac{\partial \hat{u}}{\partial v}\frac{\partial v}{\partial u^*} = \frac{\partial \hat{u}^*}{\partial u^*}, \qquad \frac{\partial \hat{u}}{\partial u}\frac{\partial u}{\partial v^*} + \frac{\partial \hat{u}}{\partial v}\frac{\partial v}{\partial v^*} = \frac{\partial \hat{u}^*}{\partial v^*},$$

$$\frac{\partial \hat{v}}{\partial u}\frac{\partial u}{\partial u^*} + \frac{\partial \hat{v}}{\partial v}\frac{\partial v}{\partial u^*} = \frac{\partial \hat{v}^*}{\partial u^*}, \qquad \frac{\partial \hat{v}}{\partial u}\frac{\partial u}{\partial v^*} + \frac{\partial \hat{v}}{\partial v}\frac{\partial v}{\partial v^*} = \frac{\partial \hat{v}^*}{\partial v^*}$$

hold, we immediately see that $d\hat{u} = d\hat{u}^*$ and $d\hat{v} = d\hat{v}^*$, i.e. that $d\hat{\mathbf{r}} = d\hat{\mathbf{r}}^*$. This shows that $d\hat{\mathbf{r}}$ is independent of the choice of parametrization γ. Similarly it can be shown (do it!) that the vector is also independent of the choice of parametrization $\hat{\gamma}$. Consequently, *mapping* (23) *is defined correctly*.

Definition 4. Mapping (23) is called the *differential* of a mapping f at a point p (or the *principal linear part* of f) and is denoted by $(df)_p$ (or $\mathbf{T}_p f$).

Problem 4. An arbitrary curve (5) on $\mathscr{X}$ is carried by f into a curve

$$\hat{u} = \hat{u}(u(t), v(t)), \quad \hat{v} = \hat{v}(u(t), v(t)), \quad t \in I \tag{25}$$

on $\hat{\mathscr{X}}$. Let for $t = t_0$ the curve (5) pass through p_0 of $\mathscr{X}$. Show that the differential $(df)_{p_0}$ of f carries the tangent vector to the curve (5) at p_0 into the tangent vector to the curve (25) at $f(p_0)$.

A mapping $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ is called a *diffeomorphism* if so is the mapping $\overline{f}$. (This definition is correct.) Functions (22) which specify a diffeomorphism have, as is known from calculus, the property that their Jacobian

$$\begin{vmatrix} \dfrac{\partial \hat{u}}{\partial u} & \dfrac{\partial \hat{u}}{\partial v} \\ \dfrac{\partial \hat{v}}{\partial u} & \dfrac{\partial \hat{v}}{\partial v} \end{vmatrix}$$

is nonzero. This means (see formulas (24)) that the *differential* $(df)_p$ *of the diffeomorphism* f *at every point* $p \in \mathscr{X}$ *is an isomorphism of a vector space* $\mathbf{T}_p\mathscr{X}$ *onto* $\mathbf{T}_{\hat{p}}\hat{\mathscr{X}}$.

For $\hat{U} = U$, the formulas $\hat{u} = u$, $\hat{v} = v$ obviously define a diffeomorphism. This diffeomorphism is a mapping which *preserves the coordinates of vectors in another system of coordinates.*

It is interesting to note that *for any diffeomorphism* $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ *and any parametrization* $\gamma\colon U \to \mathscr{A}$ *of* $\mathscr{X}$ *there is a parametrization* $\gamma^*\colon U \to \mathscr{A}$ *of* $\hat{\mathscr{X}}$ *such that in* γ *and* γ^* *the diffeomorphism* f *is a mapping which preserves the coordinates of vectors in another system of coordinates.*

Indeed, let $\hat{\gamma}\colon \hat{U} \to \mathscr{A}$ be an arbitrary parametrization of $\hat{\mathscr{X}}$ and let the mapping f be represented by the diffeomorphism $\overline{f}\colon U \to \hat{U}$ in γ and $\hat{\gamma}$. Consider a composite mapping $\gamma^* = \hat{\gamma} \circ \overline{f}$. Since $\overline{f}$ is a diffeomorphism, this mapping is a parametrization of $\hat{\mathscr{X}}$ equivalent to $\hat{\gamma}$. Since $f \circ \gamma = \hat{\gamma} \circ \overline{f} = \gamma^* \circ \mathrm{id}$, where id is the identity mapping, in γ and γ^* the deffeomorphism f is represented by id and hence preserves the coordinates of vectors in another system of coordinates. □

As we shall see later, this property of diffeomorphisms often significantly facilitates their study.

Now we shall assume a space $\mathscr{A}$ containing a given surface $\mathscr{X}$ to be *Euclidean*. Then a Euclidean structure arises in every tangent subspace $\mathbf{T}_p\mathscr{X}$, the square of the length of an arbitrary vector (18) of that subspace being expressed as follows:

$$dr^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2, \tag{26}$$

where

$$E = \mathbf{r}_u^2, \quad F = \mathbf{r}_u\mathbf{r}_v, \quad G = \mathbf{r}_v^2$$

are metric coefficients of the basis $\mathbf{r}_u$, $\mathbf{r}_v$.

Definition 5. A quadratic form (26) of variables du and dv is called the *first quadratic form* of a surface $\mathscr{X}$. It is usually denoted by I or I (dr). Thus by definition

$$\mathrm{I} = d\mathbf{r}^2.$$

Note that the coefficients E, F, and G of the first quadratic form depend on $p \in \mathscr{X}$ and in the coordinates u and v are the smooth functions

$$E = E(u, v), \quad F = F(u, v), \quad G = G(u, v)$$

of u and v.

Remark 2. The expression $E\,du^2 + 2F\,du\,dv + G\,dv^2$ is usually understood as a quadratic form. From the modern point of view it should be treated as a quadratic functional on a space $\mathbf{T}_p\mathscr{X}$ whose value on vector (18) is $E\,du^2 + 2F\,du\,dv + G\,dv^2$.

The length $|\mathbf{r}'(t)|$ of the tangent vector $\mathbf{r}'(t)$ of curve (6) on $\mathscr{X}$ can be expressed in view of (16) as follows:

$$|\mathbf{r}'(t)| = \sqrt{\mathbf{r}'(t)^2} = \sqrt{(u'(t)\,\mathbf{r}_u + v'(t)\,\mathbf{r}_v)^2}$$
$$= \sqrt{Eu'(t)^2 + 2F\,u'(t)\,v'(t) + G\,v'(t)^2},$$

where, of course, E, F, and G are regarded as functions of t:

$$E = E(u(t), v(t)), \quad F = F(u(t), v(t)),$$
$$G = G(u(t), v(t)).$$

For the length s of curve (6) this yields the formula

(27) $$s=\int_a^b \sqrt{E u'(t)^2+2F u'(t) v'(t)+G v'(t)^2}\, dt$$

which can be rewritten in the following handy mnemonic form:

$$s=\int_L \sqrt{E\, du^2+2F\, du\, dv+G\, dv^2},$$

where L designates curve (6).

The formula

(28) $$ds^2 = E\, du^2 + 2F\, du\, dv + G\, dv^2$$

is of a still more conventional form (in shorter notation $ds^2 = \mathrm{I}\,(d\mathbf{r})$ or $ds^2 = d\mathbf{r}^2$) which means that the *first quadratic form* I *specifies the square* ds^2 *of the element of length.* [It should be remembered that both the formulation and formula (28) are conventional in character, they serve only as shortened expressions of formula (27).]

By the general rules of linear algebra, for the angle θ between two tangent vectors

$$d\mathbf{r} = \mathbf{r}_u du + \mathbf{r}_v\, dv \quad \text{and} \quad \delta\mathbf{r} = \mathbf{r}_u \delta u + \mathbf{r}_v \delta v$$

we have

$$\cos\theta=\frac{d\mathbf{r}\, \delta\mathbf{r}}{|\, d\mathbf{r}\, |\, |\, \delta\mathbf{r}\, |},$$

i.e.

(29) $$\cos\theta=\frac{E\, du\, \delta u+F\, (du\, \delta v+\delta u\, dv)+G\, dv\, \delta v}{\sqrt{E\, du^2+2F\, du\, dv+G\, dv^2}\,\sqrt{E\, \delta u^2+2F\, \delta u\, \delta v+G\, \delta v^2}}$$

which can be arbitrarily written in the following mnemonic form:

$$\cos\theta=\frac{\mathrm{I}\,(d,\, \delta)}{\sqrt{\mathrm{I}\,(d)}\,\sqrt{\mathrm{I}\,(\delta)}}.$$

The angle between two tangent vectors

$$\mathbf{r}'\,(t_0) = u'\,(t_0)\, \mathbf{r}_{u_0} + v'\,(t_0)\mathbf{r}_{v_0} \quad \text{and}$$
$$\mathbf{r}_1'\,(t_0) = u_1'\,(t_0)\mathbf{r}_{1u_0} + v_1'\,(t_0)\, \mathbf{r}_{1v_0}$$

at p_0 is called the *angle between two curves*

$$\mathbf{r} = \mathbf{r}\,(u\,(t),\ \ v\,(t)) \quad \text{and} \quad \mathbf{r}_1 = \mathbf{r}_1\,(u_1\,(t),\ \ v_1\,(t))$$

on $\mathscr{X}$, which for $t = t_0$ pass through the same point p_0. We have

$$\cos\theta=\frac{\mathbf{r}'\,(t_0)\, \mathbf{r}_1'\,(t_0)}{|\, \mathbf{r}'\,(t_0)| \cdot |\, \mathbf{r}_1'\,(t_0)\, |},$$

where

$$\mathbf{r}'\,(t_0)\, \mathbf{r}_1'\,(t_0)$$
$$=E_0u'\,(t_0)\, u_1'\,(t_0)+F_0\,(u'\,(t_0)\, v_1'\,(t_0)+u_1'\,(t_0)\, v'\,(t_0))$$
$$+G_0v'\,(t_0)\, v_1'\,(t_0),$$
$$|\mathbf{r}'\,(t_0)|=\sqrt{E_0u'\,(t_0)^2+2F_0u'\,(t_0)\, v'\,(t_0)+G_0v'\,(t_0)^2},$$
$$|\mathbf{r}_1'\,(t_0)|=\sqrt{E_0u_1'\,(t_0)^2+2F_0u_1'\,(t_0)\, v_1'\,(t_0)+G_0v_1'\,(t_0)^2}$$

and

$$E_0 = E\,(u_0,\ v_0) = E\,(u\,(t_0),\ v\,(t_0)),$$
$$F_0 = F\,(u_0,\ v_0) = F\,(u\,(t_0),\ v\,(t_0)),$$
$$G_0 = G\,(u_0,\ v_0) = G\,(u\,(t_0),\ v\,(t_0))$$

are the values of the coefficients of the first quadratic form at p_0.

In particular, for the cosine of the angle between two coordinate lines u = const and v = const, we get

$$\cos\theta=\frac{F}{\sqrt{E}\,\sqrt{G}},$$

from which it follows that the *coordinate lines* u = const *and* v = const *are orthogonal if and only if* $F = 0$.

Thus *we can compute the lengths of curves on a surface and the angles between them knowing only the first quadratic form of that surface*, i.e. the Euclidean structures on all tangent planes.

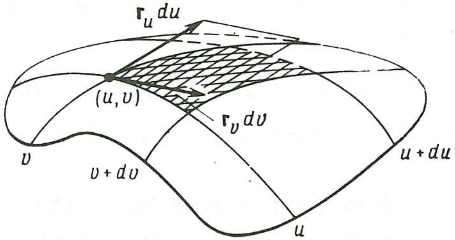
Known in calculus is the concept of *area* of an arbitrary part D of a surface $\mathscr{X}$ as the limit of areas that approximate that part of polyhedral surfaces. This limit (if any) is expressed by the integral

$$\iint_D \sqrt{\Gamma\,(\mathbf{r}_u,\ \mathbf{r}_v)}\, du\, dv,$$

where

$$\Gamma(\mathbf{r}_u, \mathbf{r}_v) = \begin{vmatrix} \mathbf{r}_u^2 & \mathbf{r}_u\mathbf{r}_v \\ \mathbf{r}_u\mathbf{r}_v & \mathbf{r}_v^2 \end{vmatrix} = \begin{vmatrix} E & F \\ F & G \end{vmatrix} = EG - F^2$$

is the *Gram determinant* or *Gramian* of vectors $\mathbf{r}_u$ and $\mathbf{r}_v$. An intuitive idea of this formula is that the area of an infinitesimal curvilinear parallelogram of a coordinate network with vertex at a point (u, v) and sides du and dv is approximately equal to that of a parallelogram in



a tangent plane with sides $\mathbf{r}_u du$ and $\mathbf{r}_v dv$ (see the figure). On the other hand, it is easy to prove (see Lemma 1 of Lecture II.26) that this area is equal to $\sqrt{\Gamma(\mathbf{r}_u, \mathbf{r}_v)}\, du\, dv$. In calculus it is said that an *element of a surface area is equal to* $\sqrt{EG - F^2}\, du\, dv$.

Thus the first quadratic form allows us to compute areas on $\mathscr{X}$.

For brevity a linear isomorphism of Euclidean vector spaces is called an isometry. By definition (see Definition 7 of Lecture II.14) the linear isomorphism $\varphi: \mathscr{V} \to \mathscr{V}_1$ is an isometry if

(30) $$\mathbf{xy} = \varphi\mathbf{x}\cdot\varphi\mathbf{y}$$

for any vectors $\mathbf{x}, \mathbf{y} \in \mathscr{V}$. For condition (30) to hold it suffices that $\mathbf{x} = \mathbf{y}$, i.e. a *linear mapping* $\varphi: \mathscr{V} \to \mathscr{V}_1$ *having the property that*

(31) $$\mathbf{x}^2 = (\varphi\mathbf{x})^2 \text{ for any vector } \mathbf{x} \in \mathscr{V}$$

is an isometry. (To prove this it suffices to apply relation (31) to the vector $\mathbf{x} + \mathbf{y}$.)

In coordinates the mapping φ is written by the following linear formulas

$$y^i = a^i_j x^j$$

and the scalar squares $\mathbf{x}^2$, $\mathbf{x} \in \mathscr{V}$, and $\mathbf{y}^2$, $\mathbf{y} \in \mathscr{V}_1$, by the quadratic forms

$$\mathbf{x}^2 = g_{ij}x^i x^j, \quad \mathbf{y}^2 = h_{ij}y^i y^j.$$

For this notation equation (31) implies that substituting expressions $y^i = a^i_j x^j$ in the form $h_{ij}y^i y^j$ yields $g_{ij}x^i x^j$, i.e. that there are

$$g_{ij} = h_{kl}a^k_i a^l_j.$$

In particular, for Euclidean two-dimensional spaces $\mathbf{T}_p\mathscr{X}$, $\mathbf{T}_{\hat{p}}\hat{\mathscr{X}}$ and the differential $(df)_p: \mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\hat{p}}\hat{\mathscr{X}}$, $\hat{p} = f(p)$, of the diffeomorphism $f: \mathscr{X} \to \hat{\mathscr{X}}$ the fact that conditions (31) hold at all points $p \in \mathscr{X}$ implies that substituting in the first quadratic form $\hat{E}\, d\hat{u}^2 + 2\hat{F}\, d\hat{u}\, d\hat{v} + \hat{G}\, d\hat{v}^2$ of $\hat{\mathscr{X}}$ expressions (24) for $d\hat{u}$ and $d\hat{v}$ (and in its coefficients $\hat{E}$, $\hat{F}$, $\hat{G}$, expressions (22) for $\hat{u}$ and $\hat{v}$) yields the first quadratic form $E\, du^2 + 2F\, du\, dv + G\, dv^2$ of $\mathscr{X}$, which is identical in u and v to

$$E(u, v) = \hat{E}(\hat{u}, \hat{v})\left(\frac{\partial\hat{u}}{\partial u}\right)^2 + 2\hat{F}(\hat{u}, \hat{v})\frac{\partial\hat{u}}{\partial u}\frac{\partial\hat{v}}{\partial u} + \hat{G}(\hat{u}, \hat{v})\left(\frac{\partial\hat{v}}{\partial u}\right)^2,$$

$$F(u, v) = \hat{E}(\hat{u}, \hat{v})\frac{\partial\hat{u}}{\partial u}\frac{\partial\hat{u}}{\partial v} + \hat{F}(\hat{u}, \hat{v})\left(\frac{\partial\hat{u}}{\partial u}\frac{\partial\hat{v}}{\partial v} + \frac{\partial\hat{u}}{\partial v}\frac{\partial\hat{v}}{\partial u}\right) + \hat{G}(\hat{u}, \hat{v})\frac{\partial\hat{v}}{\partial u}\frac{\partial\hat{v}}{\partial v},$$

(32)

$$G(u, v) = \hat{E}(\hat{u}, \hat{v})\left(\frac{\partial\hat{u}}{\partial v}\right)^2 + 2\hat{F}(\hat{u}, \hat{v})\frac{\partial\hat{u}}{\partial v}\frac{\partial\hat{v}}{\partial v} + \hat{G}(\hat{u}, \hat{v})\left(\frac{\partial\hat{v}}{\partial v}\right)^2.$$

Thus *for a diffeomorphism* $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ *linear mappings*

$$(df)_p\colon \mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\hat{p}}\hat{\mathscr{X}}, \quad \hat{p} = f(p)$$

are isometries for all points $p \in \mathscr{X}$ *if and only if equations* (32) *hold identically in* u *and* v.

For the diffeomorphism $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ preserving the coordinates of vectors in another system of coordinates this condition implies that in the coordinates under consideration the *first quadratic forms of* $\mathscr{X}$ *and* $\hat{\mathscr{X}}$ *coincide* (or to be exact, they differ only in the notation of variables).

Definition 6. A diffeomorphism $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ is said to be an *isometric mapping* of a surface $\mathscr{X}$ onto $\hat{\mathscr{X}}$ (or simply an *isometry*) if for any curve

$$u = u(t), \quad v = v(t), \quad a \leqslant t \leqslant b, \tag{33}$$

on $\mathscr{X}$ its image

$$\hat{u} = \hat{u}(t), \quad \hat{v} = \hat{v}(t), \quad a \leqslant t \leqslant b$$

for $\hat{\mathscr{X}}$, where $\hat{u}(t) = \hat{u}(u(t), v(t))$ and $\hat{v}(t) = \hat{v}(u(t), v(t))$, has the same length, i.e. if

$$\int_a^b \sqrt{E(t)\,u'(t)^2 + 2F(t)\,u'(t)\,v'(t) + G(t)\,v'(t)^2}\,dt$$

$$= \int_a^b \sqrt{\hat{E}(t)\,\hat{u}'(t)^2 + 2\hat{F}(t)\,\hat{u}'(t)\,\hat{v}'(t) + \hat{G}(t)\,\hat{v}'(t)^2}\,dt,$$

where

$$E(t) = E(u(t), v(t)), \quad F(t) = F(u(t), v(t)),$$
$$G(t) = G(u(t), v(t))$$

and similarly

$$\hat{E}(t) = \hat{E}(\hat{u}(t), \hat{v}(t)), \quad \hat{F}(t) = \hat{F}(\hat{u}(t), \hat{v}(t)),$$
$$\hat{G}(t) = \hat{G}(\hat{u}(t), \hat{v}(t)).$$

Surfaces for which there is at least one isometric mapping $\mathscr{X} \to \hat{\mathscr{X}}$ are called *isometric*.

Proposition 1. *A diffeomorphism* $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ *is an isometry if and only if for any point* $p \in \mathscr{X}$ *so is its differential*

$$(df)_p\colon \mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\hat{p}}\hat{\mathscr{X}}, \quad \hat{p} = f(p).$$

Proof. According to the above remarks we may assume without loss of generality that the diffeomorphism f acts as a function defined by equating coordinates. Then for any curve (33) on $\mathscr{X}$ its image on $\hat{\mathscr{X}}$ under the diffeomorphism f has *the same* parametric equations (33) and the statement that the differential $(df)_p$ of f is an isometry at every point $p \in \mathscr{X}$ will mean that the first quadratic forms of $\mathscr{X}$ and $\hat{\mathscr{X}}$ differ only in the notation of variables. Formulas (27) will therefore be identical for both curves and hence the lengths of these curves will be identical. Consequently, the diffeomorphism f will be an isometry.

Conversely, let a diffeomorphism $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ preserving the coordinates of vectors in another system of coordinates be an isometry. This means that for any smooth functions $u = u(t)$, $v = v(t)$, $a \leqslant t \leqslant b$ with the property that $(u(t), v(t))$, $\in U$, $a \leqslant t \leqslant b$, we have

$$\int_a^b \sqrt{E(t)\,u'(t)^2 + 2F(t)\,u'(t)\,v'(t) + G(t)\,v'(t)^2}\,dt$$

$$= \int_a^b \sqrt{\hat{E}(t)\,u'(t)^2 + 2\hat{F}(t)\,u'(t)\,v'(t) + \hat{G}(t)\,v'(t^2)}\,dt.$$

Differentiating this identity with respect to b (and replacing b by t) we obtain, after squaring, the identity

$$E(t)\,u'(t)^2 + 2F(t)\,u'(t)\,v'(t) + G(t)\,v'(t)^2$$
$$= \hat{E}(t)\,u'(t)^2 + 2\hat{E}(t)\,u'(t)\,v'(t) + \hat{G}(t)\,v'(t)^2.$$

In particular, this identity must hold for linear functions of the form

$$u(t) = u_0 + \alpha t, \quad v(t) = v_0 + \beta t, \quad |t| < \varepsilon,$$

where (u_0, v_0) is an arbitrary point of a domain U, α and β are arbitrary numbers, and $\varepsilon > 0$ is a sufficiently small positive number. But in this case, after substituting $t = 0$ it becomes

$$E_0\alpha^2 + 2F_0\alpha\beta + G_0\beta^2 = \hat{E}_0\alpha^2 + 2\hat{F}_0\alpha\beta + \hat{G}_0\beta^2,$$

where $E_0, \ldots, \hat{G}_0$ are the values of the functions $E, \ldots, \hat{G}$ at (u_0, v_0), and is therefore possible, in view of α and β being arbitrary numbers, if and only if $E_0 = \hat{E}_0$, $F_0 = \hat{F}_0$ and $G_0 = \hat{G}_0$, i.e. if $E = \hat{E}$, $F = \hat{F}$ and $G = \hat{G}$ everywhere in U. Hence linear mappings $(df)_p$ are isometries. □

Corollary 1. *On isometric surfaces the corresponding curves intersect at the same angles, and the corresponding domains have the same areas.* □

Corollary 2. *Two surfaces are isometric if and only if local coordinates can be chosen on them in which the first quadratic forms of those surfaces coincide.* □

Of course, this criterion for isometry is extremely ineffective (how can we guess if there are local coordinates it provides for?). Our final goal (to be achieved in Lecture 5) is to make this criterion effective, but to do this we shall have to go quite a long way.

Assuming a surface to be made from a flexible but unstretchable material and bending it arbitrarily we shall not change the lengths of curves on the surface and hence get an isometric surface. Based on this graphical representation the founders of the theory of surfaces in the 19th century called isometries the *bendings*. This terminology has partly survived till the present day, but nowadays bendings are usually understood in a narrower sense, as isometries that can be connected with the identity transformation by a continuous family of isometries. All mathematicians were sure for a long time that in a local situation, i.e. in a sufficiently small neighbourhood of an arbitrary point, any isometry is a bending in this sense. Relatively recently, however, N. V. Efimov has shown that this is wrong by constructing an appropriate counterexample.

Suppose a surface is inhabited by intelligent creatures that can measure lengths, areas, and angles, but cannot reach out of the space. Then with any bending of the surface all their geometry will remain the same and they will simply fail to notice this bending. For this reason isometric surfaces are said to have the same *intrinsic geometry*.

Here is one important example of an intrinsic and geometric construction.

According to the general results of linear algebra (see Lecture II.14) for any Euclidean space $\mathscr{V}$ with metric tensor g_{ij} the conjugate space $\mathscr{V}^*$ is a Euclidean space with metric tensor g^{ij} whose components form a matrix $\| g^{ij} \|$ inverse to the matrix $\| g_{ij} \|$ of the components of g_{ij}. For any covector $\xi = (\xi_1, \ldots, \xi_n)$ in $\mathscr{V}^*$ its length $| \xi |$ is thus defined whose square is expressed as follows:

$$| \xi | = g^{ij}\xi_i\xi_j.$$

In our case, for a two-dimensional Euclidean space $\mathbf{T}_p\mathscr{X}$ the matrix $\| g_{ij} \|$ is of the form

$$\begin{Vmatrix} E & F \\ F & G \end{Vmatrix}$$

and hence $\| g^{ij} \|$ is of the form

$$\frac{1}{EG-F^2}\begin{Vmatrix} G & -F \\ -F & E \end{Vmatrix}.$$

For the length $| \xi |$ of an arbitrary covector ξ on $\mathbf{T}_p\mathscr{X}$ we have

$$|\xi|^2 = \frac{G\xi^2 - 2F\xi\eta + E\eta^2}{EG-F^2},$$

where ξ, η are the coordinates of that covector.

An example of the covector ξ is the gradient $\left(\frac{\partial\varphi}{\partial u}, \frac{\partial\varphi}{\partial v}\right)$ of an arbitrary smooth function $\varphi = \varphi(u, v)$ on $\mathscr{X}$. The square of the length of that covector is called *Beltrami's differential operator* of the first kind of the function φ and is denoted by $\Delta_1\varphi$. Thus by definition

$$\Delta_1\varphi = \frac{G\varphi_u^2 - 2F\varphi_u\varphi_v + E\varphi_v^2}{EG-F^2}.$$

This construction is invariant, i.e. independent of the choice of parametrization, and for any isometry $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ we have

(34) $$\Delta_1(\varphi \circ f) = \Delta_1\varphi \circ f$$

(check it!). In this sense it belongs to the intrinsic geometry of the surface $\mathscr{X}$.

Consider in conclusion some examples of computing the first quadratic form of surfaces in a three-dimensional Euclidean space. In these examples the surfaces are not elementary as a rule. But they are easily reduced to elementary ones by cuttings and restrictions of the domains of definition of parametrizations.

Example 4. The plane Oxy has the parametric equation $\mathbf{r} = u\mathbf{i} + v\mathbf{j}$ in coordinates $u = x$ and $v = y$. Therefore $\mathbf{r}_u = \mathbf{i}$, $\mathbf{r}_v = \mathbf{j}$ and hence $E = 1$, $F = 0$, $G = 1$, i.e. for a plane,

(35) $$\mathrm{I} = du^2 + dv^2.$$

(The result is easy to foresee without any computations.)

An open subset of a plane (considered as a surface in space) also has the same quadratic form.

Example 5. For a circular cylinder

$$\mathbf{r} = R\cos u\cdot\mathbf{i} + R\sin u\cdot\mathbf{j} + v\cdot\mathbf{k}$$

we have $\mathbf{r}_u = -R\sin u\cdot\mathbf{i} + R\cos u\cdot\mathbf{j}$ and $\mathbf{r}_v = \mathbf{k}$. Therefore

$$E = \mathbf{r}_u^2 = R^2, \quad F = \mathbf{r}_u\mathbf{r}_v = 0, \quad G = \mathbf{r}_v^2 = 1,$$

i.e. for a cylinder

$$\mathrm{I} = R^2\,du^2 + dv^2.$$

Introducing a new coordinate $u_1 = Ru$ (and denoting u_1 by u) we transform this to form (35).

Thus, there are coordinates in which the first quadratic form of a plane and that of a cylinder coincide! This does not mean yet that the plane and the cylinder (of course, cut in; see Example 1 above) are isometric, since for a cylinder the coordinates range over only some region in $\mathbb{R}^2$ and so the cut-in cylinder is only isometric to a part of the plane. We express this by saying that the cylinder and the plane are *locally isometric*.

An isometric mapping of a cut-in cylinder onto a flat region is produced graphically by gradually unbending it.

Example 6. For a surface of revolution

$$\mathbf{r} = x(v)\cos u\cdot\mathbf{i} + x(v)\sin u\cdot\mathbf{j} + z(v)\cdot\mathbf{k}$$

we have

$$\mathbf{r}_u = -x(v)\sin u\cdot\mathbf{i} + x(v)\cos u\cdot\mathbf{j},$$
$$\mathbf{r}_v = x'(v)\cos u\cdot\mathbf{i} + x'(v)\sin u\cdot\mathbf{j} + z'(v)\,\mathbf{k}.$$

Hence

$$E = x(v)^2\sin^2 u + x(v)^2\cos^2 u = x(v)^2,$$
$$F = -x(v)\sin u\cdot x'(v)\cos u + x(v)\cos u\cdot x'(v)\sin u = 0,$$
$$G = x'(v)^2\cos^2 u + x'(v)^2\sin^2 u + z'(v)^2 = x'(v)^2 + z'(v)^2,$$

so that for the surface of revolution

$$\mathrm{I} = x(v)^2du^2 + (x'(v)^2 + z'(v)^2)\,dv^2.$$

It is graphically obvious that the meridians and parallels of any surface of revolution are orthogonal. The equation $F = 0$ could therefore be foreseen without any computations as well.

When the profile $x = x(v)$, $z = z(v)$ of a surface of revolution is referred to the natural parameter $v = s$ (and therefore $x'(v)^2 + z'(v)^2 = 1$) the form I is particularly simple:

$$\mathrm{I} = x(v)^2du^2 + dv^2.$$

In particular, we see that the *first quadratic form of a sphere of radius 1 is of the form*

(36) $$\mathrm{I} = \cos^2 v\,du^2 + dv^2.$$

Cartographic experience shows that no portion of a sphere however small can be bent into a plane. This means that no transformation of coordinates can convert form (36) into form (35). But how is this to be proved? The answer will be given in Lecture 5.

Example 7. The curve formed by a chain of uniform density hanging freely from two fixed points not in the same vertical line is called a *catenary* and a surface of revolution whose profile is a catenary is called a *catenoid.*

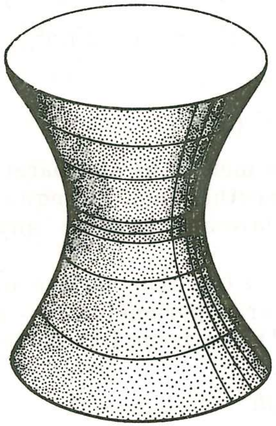
In mechanics (statics) a catenary is the graph of a hyperbolic cosine. Thus for a catenoid $x(v) = \cosh v$, $z(v) = v$ and hence

$$x(v)^2 = \cosh^2 v \text{ and } x'(v)^2 + z'(v)^2 = \sinh^2 v + 1 = \cosh^2 v.$$

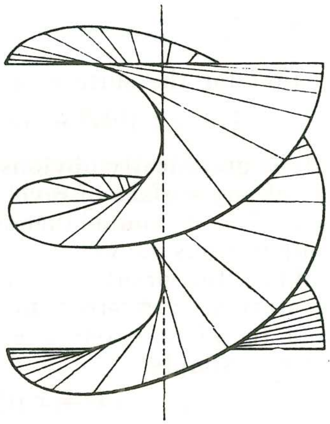
Thus for the catenoid

$$\text{(37)} \qquad \mathbf{I} = \cosh^2 v\,(du^2 + dv^2).$$

Example 8. Let a straight line perpendicular to the axis Oz rotate uniformly near it remaining perpendicular



A catenoid



A helicoid

to it in rotation and simultaneously ascending in helical motion (to a height proportional to the angle of rotation). The ruled surface generated by a moving straight line is called a *helicoid.* It has the form of a helical ramp for cars to drive up.

If v is the parameter on the straight line and u is the angle of rotation, then the helicoid will have the equation

$$\mathbf{r} = v \cos u \cdot \mathbf{i} + v \sin u \cdot \mathbf{j} + u \cdot \mathbf{k}.$$

Therefore

$$\mathbf{r}_u = -v \sin u \cdot \mathbf{i} + v \cos u \cdot \mathbf{j} + \mathbf{k},$$
$$\mathbf{r}_v = \cos u \cdot \mathbf{i} + \sin u \cdot \mathbf{j}$$

and hence

$$E = 1 + v^2, \quad F = 0, \quad G = 1.$$

Thus for a helicoid

$$\mathbf{I} = (1 + v^2)\,du^2 + dv^2.$$

Let us transform this form by introducing new coordinates u_1, v_1 related to the coordinates u, v as follows:

$$u = u_1, \quad v = \sinh v_1.$$

Then

$$1 + v^2 = 1 + \sinh^2 v_1 = \cosh^2 v_1,$$
$$du = du_1, \quad dv = \cosh v_1 dv_1$$

and therefore (we drop the indices in the new coordinates)

$$\mathbf{I} = \cosh^2 v\,(du^2 + dv^2),$$

which coincides with form (37).

This proves that the *catenoid and the helicoid are locally isometric* (more exactly, the catenoid cut-in along a meridian is isometric to a part $0 < u < 2\pi$ of the helicoid), there being an isometry which sends meridians of the catenoid into rectilinear generators of the helicoid.

Example 9. For an arbitrary ruled surface

$$\text{(38)} \qquad \mathbf{r} = \boldsymbol{\rho}(u) + v\mathbf{a}(u),$$

where (see Example 3) $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ is a regular curve referred to the natural parameter and $\mathbf{a}(u)$ is a vector function such that $|\mathbf{a}(u)| = 1$ for all u, denoting differentiation with respect to u by a dot, we have

$$\mathbf{r}_u = \dot{\boldsymbol{\rho}} + v\dot{\mathbf{a}}, \quad \mathbf{r}_v = \mathbf{a}.$$

Since $\dot{\boldsymbol{\rho}}^2 = 1$, and $\mathbf{a}^2 = 1$ and $\mathbf{a}\dot{\mathbf{a}} = 0$, we have

$$E = 1 + 2v\dot{\boldsymbol{\rho}}\dot{\mathbf{a}} + v^2\dot{\mathbf{a}}^2, \quad F = \dot{\boldsymbol{\rho}}\mathbf{a}, \quad G = 1.$$

If, in particular, $\mathbf{a} = \dot{\boldsymbol{\rho}}$ (a surface of tangents), then $\dot{\boldsymbol{\rho}}\mathbf{a} = \mathbf{a}^2 = 1$ (i.e. $F = 1$) and $\dot{\boldsymbol{\rho}}\dot{\mathbf{a}} = 0$ and $\dot{\mathbf{a}}^2 = k^2$, where k is the curvature of the curve $\rho = \rho(u)$ (i.e. $E = 1 + k^2v^2$). Thus for a surface of tangents

$$\mathrm{I} = (1 + k^2v^2)\, du^2 + 2du\,dv + dv^2. \tag{39}$$

But if $\mathbf{a}(u)$ is the binormal vector of the curve $\rho = \rho(u)$, then $\dot{\rho}\mathbf{a} = 0$, $\ddot{\rho}\mathbf{a} = 0$ and $\dot{\mathbf{a}}^2 = \varkappa^2$, where $\varkappa$ is the torsion of the curve $\rho = \rho(u)$. Hence for a surface of binormals

$$\mathrm{I} = (1 + \varkappa^2v^2)\, du^2 + dv^2.$$

We thus see that the first quadratic form of a surface of tangents depends only on the curvature of a given curve and that the first quadratic form of a surface of binormals depends only on the torsion of the surface.

For surfaces of tangents this implies that *every surface of tangents is locally isometric to a plane*. Indeed, consider a plane curve with the same curvature $k = k(u)$ (such a curve exists by virtue of Theorem 1 of Lecture 2). The first quadratic form of the surface of tangents for that curve is of form (39). But, on the other hand, it is clear that a surface of tangents of a plane curve is, outside its singular points, a domain in the plane. There exists therefore a change of coordinates transforming the first quadratic form $dx^2 + dy^2$ of the plane into form (39). (This change of coordinates has the form

$$x = x(u) + x'(u)\, v, \quad y = y(u) + y'(u)\, v,$$

where $x(u)$ and $y(u)$ are functions such that $x'(u)^2 + y'(u)^2 = 1$ and $x''(u)^2 + y''(u)^2 = k(u)^2$.) □

This isometry can be carried out by continuous bending, gradually deforming the curve $\rho = \rho(u)$ into a plane curve.

For this reason surfaces of tangents are called *developable surfaces* (or developables) (development into a plane is meant).

If $\mathbf{a}(u) = \rho(u)$, then surface (38) is a cone with vertex at the origin (and the curve $\rho = \rho(u)$ is the intersection of the cone and the unit sphere $|\rho| = 1$). In this case we have

$$\ddot{\rho}\mathbf{a} = \rho^2 = 1, \quad \dot{\mathbf{a}}^2 = 1, \quad \dot{\rho}\mathbf{a} = 0,$$

from which for the first quadratic form we get

$$\mathrm{I} = (1 + v)^2 du^2 + dv^2.$$

Here the change of coordinates $(u, v) \mapsto (u, 1 + v)$ suggests itself, which converts the last form into a slightly simpler form

$$\mathrm{I} = v^2du^2 + dv^2. \tag{40}$$

Now let us introduce new coordinates

$$x = v\cos u, \quad y = v\sin u.$$

Then

$$dx = -v\sin u\, du + \cos u\, dv,$$
$$dy = v\cos u\, du + \sin u\, dv$$

and hence

$$dx^2 + dy^2 = v^2du^2 + dv^2.$$

This proves that the *cone is also locally isometric to a plane* (more exactly, each nappe of a cone cut-through along a generator is isometric to some plane sector; the fact is intuitively obvious). For this reason cones are also reckoned among developables.

Notice that form (10) is nothing but the first quadratic form of a plane referred to polar coordinates $r = v$ and $\varphi = u$.

Finally, if the vector $\mathbf{a}(u)$ is constant (and therefore $\dot{\mathbf{a}} = 0$), then surface (38) is a cylinder. We may assume without loss of generality that the directrix $\rho = \rho(u)$ of the cylinder is a plane curve whose plane is orthogonal to the vector $\mathbf{a}$ (and hence $\rho\mathbf{a} = 0$ and $\dot{\rho}\mathbf{a} = 0$). Therefore, as for a circular cylinder (Example 3),

$$\mathrm{I} = du^2 + dv^2.$$

For this reason all cylinders are also reckoned among developables.

In Lecture 4 we shall show that among ruled surfaces only developables (i.e. cylinders, cones, and surfaces of tangents) are locally isometric to a plane, and in Lecture 5 it will be shown that developables exhaust all the surfaces of a three-dimensional space which are locally isometric to a plane.

Lecture 4

Normal vector · Surface as the graph of a function · Normal sections · The second quadratic form of a surface · The Dupin indicatrix · Principal, total and mean curvatures · The second quadratic form of a graph · Ruled surfaces of zero curvature · Surfaces of revolution

In this lecture we shall take a closer look at surfaces in a three-dimensional oriented Euclidean space. As a rule we shall study surfaces only locally, i.e. in sufficiently small neighbourhoods of their points. We may therefore assume, without loss of generality, all surfaces to be elementary.

So let $\mathscr{X}$ be an arbitrary elementary surface in a three-dimensional oriented Euclidean space $\mathscr{A}$ and let $\mathbf{r} = \mathbf{r}(u, v)$ be its arbitrary parametrization. Then at every point $p \in \mathscr{X}$ there is a single vector $\mathbf{n}$ of unit length which is perpendicular to the tangent plane and forms together with vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ a positively oriented basis for the space. This vector is given as follows:

$$\mathbf{n} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}$$

and is called the *normal vector* to $\mathscr{X}$ at p (and the basis $\mathbf{r}_u$, $\mathbf{r}_v$, $\mathbf{n}$ is called the *moving basis* of the surface). See the figure on page 68.

By definition the square of the length $|\mathbf{r}_u \times \mathbf{r}_v|^2$ of a vector product is equal to the area of the parallelogram constructed on $\mathbf{r}_u$ and $\mathbf{r}_v$, i.e. (see Lecture 3) to the

Gramian $\Gamma(\mathbf{r}_u, \mathbf{r}_v) = EG - F^2$ of $\mathbf{r}_u$ and $\mathbf{r}_v$. Hence

$$\mathbf{n} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{\sqrt{EG - F^2}} .$$

It is this formula that is usually used to compute $\mathbf{n}$.

Choosing in $\mathscr{A}$ rectangular coordinates x, y, and z with the origin O at a point p and the Oz-axis directed along the vector $\mathbf{n}$, consider a Jacobian matrix

$$\begin{Vmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \end{Vmatrix}$$

whose rows are the coordinates of $\mathbf{r}_u$ and $\mathbf{r}_v$. The condition imposed on Oz implies, in particular, that

$$\begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix} = 1.$$

Therefore according to the **inverse mapping theorem** known from calculus, in some neighbourhood of the point (0, 0) the coordinates x and y can be expressed in terms of the coordinates u and v. Substituting these expressions in the vector function $\mathbf{r} = \mathbf{r}(u, v)$ we obtain a parametrization of the surface $\mathscr{X}$ (or more precisely some part of the surface, which contains the point p) of the form

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

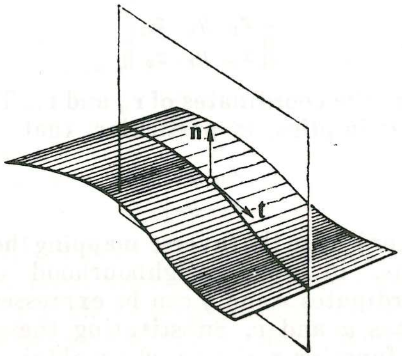
By definition this means that *locally* (in a neighbourhood of p) *the surface $\mathscr{X}$ is the graph of the function $z = z(x, y)$.*

Every plane Π passing through the Oz-axis of the coordinate system involved is called the *normal plane* of $\mathscr{X}$ at p and its intersection $\Pi \cap \mathscr{X}$ with the surface $\mathscr{X}$ is the *normal section* of $\mathscr{X}$ at p. The direction bivector of every normal plane Π is of the form $\mathbf{t} \wedge \mathbf{n}$, where $\mathbf{t}$ is some nonzero tangent vector of $\mathscr{X}$, defined up to proportionality, at p (a vector of the space $\mathbf{T}_p\mathscr{X}$). The vector $\mathbf{t}$ uniquely defines the normal plane Π and we shall denote the plane by Π_t.

6—0737

If x, y, z are chosen so that $\mathbf{t}$ is directed along Ox, then Π_t is the coordinate plane Oxz and the normal section $\mathscr{L}_t = \Pi \cap \mathscr{X}$ has the equation $z = z(x, 0)$ (in coordinates x, z in the plane Oxz). This proves that (in some neighbourhood of p) every normal section is locally a graph and hence a simple singular arc.

A tangent to the graph $z = z(x, 0)$ is, of course, in the tangent plane to the surface $z = z(x, y)$ and is at the



Normal section

same time in the normal plane. It is therefore directed along $\mathbf{t}$, i.e. the tangent vector to the normal section at p is proportional to $\mathbf{t}$.

Since $\mathbf{t}$ is only defined by the normal section only up to proportionality, without loss of generality we may assume it to be a unit vector. Referring the normal section to the natural parameter we may therefore consider the unit vector $\mathbf{t}$ as the tangent vector to the normal section at p.

Thinking of the normal section as a curve on Π_t we may speak of the relative curvature k_{rel} of the curve at p with respect to the orientation of Π_t given by the bivector $\mathbf{t} \wedge \mathbf{n}$. Denoting this curvature by $k(\mathbf{t})$ we thus define some function $\mathbf{t} \mapsto k(\mathbf{t})$ on unit vectors of the tangent space $\mathbf{T}_p\mathscr{X}$. We find the expression for that function in terms of the coordinates of $\mathbf{t}$.

Let

(1) $$u = u(s), \quad v = v(s), \quad |s| < s_0,$$

be the parametric equation of the normal section $\mathscr{L}_t$ on $\mathscr{X}$ where s is the natural parameter counted off from p. As a curve in space the section $\mathscr{L}_t$ has the vector parametric equation $\mathbf{r} = \mathbf{r}(u(s), v(s))$ and hence the tangent vector of the curve, $\dot{\mathbf{r}}$, is expressed as follows:

(2) $$\dot{\mathbf{r}} = \mathbf{r}_u \dot{u} + \mathbf{r}_v \dot{v},$$

where as ever $\dot{u}$ and $\dot{v}$ are the derivatives of functions (1) with respect to s.

The derivative $\ddot{\mathbf{r}}$ of vector (2) is orthogonal to it, and parallel to the plane Π_t for $s = 0$. For $s = 0$ therefore it is proportional to $\mathbf{n}$. The corresponding proportionality factor (equal to the scalar product $\ddot{\mathbf{r}}\mathbf{n}$) is (see the definition of relative curvature in Lecture 2) nothing but relative curvature $k(\mathbf{t})$. Thus

(3) $$k(\mathbf{t}) = \ddot{\mathbf{r}}\mathbf{n}.$$

Since

$$\begin{aligned}\ddot{\mathbf{r}} &= \dot{\mathbf{r}}_u \dot{u} + \mathbf{r}_u \ddot{u} + \dot{\mathbf{r}}_v \dot{v} + \mathbf{r}_v \ddot{v} \\ &= (\mathbf{r}_{uu}\dot{u} + \mathbf{r}_{uv}\dot{v})\dot{u} + (\mathbf{r}_{uv}\dot{u} + \mathbf{r}_{vv}\dot{v})\dot{v} + \mathbf{r}_u \ddot{u} + \mathbf{r}_v \ddot{v} \\ &= \mathbf{r}_{uu}\dot{u}^2 + 2\mathbf{r}_{uv}\dot{u}\dot{v} + \mathbf{r}_{vv}\dot{v}^2 + \mathbf{r}_u \ddot{u} + \mathbf{r}_v \ddot{v}\end{aligned}$$

and $\mathbf{r}_u\mathbf{n} = 0$, $\mathbf{r}_v\mathbf{n} = 0$, this proves that

$$k(\mathbf{t}) = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2,$$

where

(4) $$L = \mathbf{r}_{uu}\mathbf{n}, \quad M = \mathbf{r}_{uv}\mathbf{n}, \quad N = \mathbf{r}_{vv}\mathbf{n}.$$

Thus function $\mathbf{t} \mapsto k(\mathbf{t})$ can be easily extended to all possible tangent vectors $d\mathbf{r} \neq 0$ assuming by definition that

$$k(d\mathbf{r}) = k\left(\frac{d\mathbf{r}}{|d\mathbf{r}|}\right).$$

Since $|d\mathbf{r}|^2 = d\mathbf{r}^2 = ds^2$, where

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

(see formula (28), Lecture 3) and the coordinates $\dot{u}$ and $\dot{v}$ of the vector $\frac{d\mathbf{r}}{|d\mathbf{r}|} = \frac{d\mathbf{r}}{ds}$ are equal to $\frac{du}{ds}$ and $\frac{dv}{ds}$, we have

$$k(d\mathbf{r}) = \frac{L\,du^2 + 2M\,du\,dv + N\,dv^2}{E\,du^2 + 2F\,du\,dv + G\,dv^2}. \tag{5}$$

[Notice that the equations $\dot{u} = \frac{du}{ds}$ and $\dot{v} = \frac{dv}{ds}$ formally coincide with the equations, known from calculus, for derivatives as the ratios of the differentials. However, they have another intentional meaning in our case since du and dv are the coordinates of the tangent vector $d\mathbf{r}$ rather than differentials and ds is the length of the vector.]

Definition 1. The quadratic form

$$L\,du^2 + 2M\,du\,dv + N\,dv^2$$

is called the *second quadratic form* of a surface $\mathscr{X}$ and is denoted by **II**.

In this notation formula (5) can be written as follows:

$$k = \frac{\mathrm{II}}{\mathrm{I}}. \tag{6}$$

For the coefficients of form II, besides formulas (4) there are also formulas

$$L = -\mathbf{r}_u\mathbf{n}_u,\quad M = -\mathbf{r}_u\mathbf{n}_v = -\mathbf{r}_v\mathbf{n}_u,\quad N = -\mathbf{r}_v\mathbf{n}_v. \tag{7}$$

Indeed, since $\mathbf{r}_u\mathbf{n} = 0$, we have $\mathbf{r}_{uu}\mathbf{n} + \mathbf{r}_u\mathbf{n}_u = 0$ and $\mathbf{r}_{uv}\mathbf{n} + \mathbf{r}_u\mathbf{n}_v = 0$, i.e. $L = -\mathbf{r}_u\mathbf{n}_u$ and $M = -\mathbf{r}_u\mathbf{n}_v$. Similarly, since $\mathbf{r}_v\mathbf{n} = 0$ we have $\mathbf{r}_{vu}\mathbf{n} + \mathbf{r}_v\mathbf{n}_u = 0$ and $\mathbf{r}_{vv}\mathbf{n} + \mathbf{r}_v\mathbf{n}_v = 0$, i.e. $M = -\mathbf{r}_v\mathbf{n}_u$ and $N = -\mathbf{r}_v\mathbf{n}_v$. □

Introducing the vector

$$d\mathbf{n} = \mathbf{n}_u du + \mathbf{n}_v dv \tag{8}$$

we can write form II (by virtue of (7)) as

$$\mathrm{II} = -d\mathbf{r}\,d\mathbf{n},$$

in a way similar to $\mathrm{I} = d\mathbf{r}^2$ for form I.

Remark 1. By analogy the *third quadratic form* $\mathrm{III} = d\mathbf{n}^2$ could also be introduced. But, as is to be shown below, it can be linearly expressed in terms of forms I and II and therefore gives nothing new. The coefficients L, M, N of form II are also denoted by D, D', D''.

To represent the function $\mathbf{t} \mapsto k(\mathbf{t})$ graphically the French mathematician Dupin suggested considering on a tangent plane the curve (now called the *Dupin indicatrix*) that results when for any unit tangent vector $\mathbf{t}$ a segment of length $|k(\mathbf{t})|^{-1/2}$ is laid off from the point of tangency p (taken as the origin O on the tangent plane) in the direction of that vector. Denote by x and y the coordinates (in the coordinate system $O\mathbf{r}_u\mathbf{r}_v$) of the end point of the segment; then its length is expressed as follows:

$$|x\mathbf{r}_u + y\mathbf{r}_v| = \sqrt{\mathrm{I}(x, y)}.$$

Since the curvature $k(\mathbf{t})$ is expressed by formula (6), which in the present notation has the form

$$k(\mathbf{t}) = \frac{\mathrm{II}(x, y)}{\mathrm{I}(x, y)},$$

we obtain for the Dupin indicatrix the equation

$$\sqrt{\mathrm{I}(x, y)} = \sqrt{\frac{\mathrm{I}(x, y)}{\mathrm{II}(x, y)}},$$

i.e. the equation

$$|\mathrm{II}(x, y)| = 1.$$

This proves that the *Dupin indicatrix is a curve with equation*

$$|Lx^2 + 2Mxy + Ny^2| = 1.$$

When $LN - M^2 > 0$ the curve (more precisely, the set of its real points) is an ellipse with equation

$$Lx^2 + 2Mxy + Ny^2 = \varepsilon, \tag{9}$$

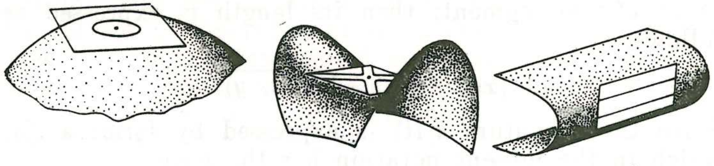
where $\varepsilon = +1$ if $L > 0$ and $\varepsilon = -1$ if $L < 0$. Accordingly a point of the surface at which $LN - M^2 > 0$ is called *elliptic*.

At an elliptic point all curvatures $k(\mathbf{t})$ have the same sign (coinciding with that of L). Among them, there is one maximum k_1 and one minimum k_2 curvatures (unless they all coincide, i.e. unless the Dupin indicatrix is a circle) corresponding to the directions of the minor and major axes of ellipse (9).

When $LN - M^2 < 0$ the Dupin indicatrix consists of two hyperbolas

$$Lx^2 + 2Mxy + Ny^2 = \pm 1 \tag{10}$$

with common asymptotes and therefore a point of the surface at which $LN - M^2 < 0$ is called *hyperbolic*. In the direction of the real axis of one of the hyperbolas (10)



(1) (2) (3)

The Dupin indicatrix
(1) At an elliptic point
(2) At a hyperbolic point
(3) At a parabolic point

the curvature k (t) attains its maximum value $k_1 > 0$. As the vector **t** is rotated the curvature first decreases to zero when the vector **t** assumes the asymptotic direction, and then while continuing to decrease attains its minimum value $k_2 < 0$ when the direction of **t** coincides with that of the real axis of the other hyperbola (i.e. with the direction of the imaginary axis of the first hyperbola).

When $LN - M^2 = 0$ a point of the surface is called *parabolic*. At such a point the Dupin indicatrix has the equation

$$(\sqrt{|L|}\, x + \sqrt{|N|}\, y)^2 = 1 \tag{11}$$

and therefore is a pair of parallel lines (provided $L \neq 0$ or $N \neq 0$). In the direction of these lines the curvature k (t) is equal to zero, in the perpendicular direction it reaches its maximum value (in magnitude) while maintaining throughout the same sign. But if $L = 0$, $N = 0$ (and therefore $M = 0$), the curvature k (t) is identically as a function of **t** equal to zero (and the Dupin indicatrix is not defined).

Notice that at elliptic and parabolic points the Dupin indicatrix is a second degree curve, and at hyperbolic points it is a quartic curve.

In each of the three cases the function k (t) attains *twice* its maximum k_1 and minimum k_2 (unless it is identically zero).

Problem 1. Prove that

$$k(\mathbf{t}) = k_1 \cos^2 \varphi + k_2 \sin^2 \varphi,$$

where φ is the angle made by the vector **t** and the direction in which the curvature is equal to k_1. This formula is known as the *Euler formula*.

Definition 2. Numbers k_1 and k_2 are called the *principal curvatures* of the surface at the point under consideration. Their product

$$K = k_1 k_2$$

is called the *total* (or *Gaussian*) *curvature* and their half-sum

$$H = \frac{k_1 + k_2}{2}$$

is called *mean curvature.*

According to the foregoing, $K > 0$ at an elliptical point, $K < 0$ at a hyperbolic point, and $K = 0$ at a parabolic point.

We stress that H and K depend on the point $p \in \mathscr{X}$, i.e. are functions on $\mathscr{X}$. As functions of local coordinates these functions are smooth.

To find principal curvatures one could seek for the principal directions of the second degree curves (9) and (10) (there is no problem with curve (11)) and then find their canonical equations. This method, however, involves lengthy computations because the coordinates x and y are not rectangular. Therefore we shall proceed in a different way, directly applying the basic formula (6).

According to this formula curvature k_2 is the smallest value of the function

$$\frac{\mathrm{II}(x, y)}{\mathrm{I}(x, y)} = \frac{Lx^2 + 2Mxy + Ny^2}{Ex^2 + 2Fxy + Gy^2}$$

of two variables x and y, with $(x, y) \neq (0, 0)$. Hence

$$\frac{\mathrm{II}(x, y)}{\mathrm{I}(x, y)} \geqslant k_2$$

for all $(x, y) \neq (0, 0)$, the equality holding at least at one point (x, y). Since $\mathrm{I}(x, y) > 0$ when $(x, y) \neq (0, 0)$, this inequality is equivalent to the inequality

$$\mathrm{II}(x, y) - k_2 \mathrm{I}(x, y) \geqslant 0$$

which means that the quadratic form $\mathrm{II} - k_2\mathrm{I}$ with matrix

$$\begin{Vmatrix} L-k_2E & M-k_2F \\ M-k_2F & N-k_2G \end{Vmatrix}$$

is nonnegative at all points $(x, y) \neq (0, 0)$ and zero at least at one of them.

Similarly, the number k_1 is characterized by the fact that the quadratic form $\mathrm{II} - k_1\mathrm{I}$ is everywhere positive and zero at least at one point $(x, y) \neq (0, 0)$.

But it is easy to see (directly or on the basis of the general theory of quadratic forms over the field $\mathbb{R}$; see Lecture II.12) that a *quadratic form in two variables is everywhere nonpositive or nonnegative and zero at least at one point* $(x, y) \neq (0, 0)$ *if and only if its rank is less than two, i.e. if the determinant of its matrix is zero.*

This proves that the *principal curvatures* k_1, k_2 *are the roots of the equation*

$$\begin{vmatrix} L-kE & M-kF \\ M-kF & N-kG \end{vmatrix} = 0,$$

i.e. of the equation

$$(EG - F^2)k^2 - (EN + GL - 2FM)\,k + (LN - M^2) = 0.$$

In particular, it follows (by virtue of Viète's formulas) that

$$K = \frac{LN-M^2}{EG-F^2}, \quad H = \frac{1}{2}\,\frac{EN+GL-2FM}{EG-F^2}.$$

The first of these formulas will find an important application in our next lecture.

Suppose that the coordinates x, y, and z in a space $\mathscr{A}$ have been chosen so that the surface under consideration is the graph of a function $z = z(x, y)$, with $z(0, 0) = 0$ and the normal vector at the point $(0, 0)$ being the unit vector $\mathbf{k}$ of the Oz-axis (see above).

Then x and y are, in a neighbourhood of the point $(0, 0)$, the coordinates u and v on the surface, with

$$\mathbf{r}_u = (1, 0, z_x), \quad \mathbf{r}_v = (0, 1, z_y).$$

Since at $(0, 0)$ the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ are parallel by the condition for the coordinate plane Oxy, it follows that

$$(z_x)_0 = 0 \quad \text{and} \quad (z_y)_0 = 0$$

(with the subscript 0 we write the values of partial derivatives at $(0, 0)$) and hence the expansion of the function $z(x, y)$ into a Taylor series begins with quadratic terms

$$z = \frac{1}{2}(rx^2 + 2sxy + ty^2) + \ldots,$$

where

$$r = (z_{xx})_0, \quad s = (z_{xy})_0, \quad t = (z_{yy})_0$$

(surface of Monge).

On the other hand, since

$$\mathbf{r}_{uu} = (0, 0, z_{xx}), \; \mathbf{r}_{uv} = (0, 0, z_{xy}), \; \mathbf{r}_{vv} = (0, 0, z_{yy}),$$

at $(0, 0)$ we have

$$L = r, \quad M = s, \quad N = t.$$

Thus, in this case, the *second quadratic form of a surface differs only by a constant factor of* $\frac{1}{2}$ *from the sum* $z_2(x, y)$ *of the quadratic terms in the Taylor series of the function* $z(x, y)$.

Since near the point $(0, 0)$ a surface $z = z(x, y)$ differs but little from a surface $z = z_2(x, y)$ and since for $rt - s^2 > 0$ the latter surface is an elliptic paraboloid and for $rt - s^2 < 0$ it is a hyperbolic paraboloid, this proves that an *arbitrary surface differs but little from an elliptic paraboloid near the elliptic point and from a hyperbolic paraboloid near the hyperbolic point.*

This gives quite a satisfactory idea of the behaviour of a surface near nonparabolic points.

When $rt - s^2 = 0$ but either $r \neq 0$ or $s \neq 0$, the surface $z = z_2(x, y)$ is a parabolic cylinder. *About a parabolic point for which either $L \neq 0$ or $N \neq 0$ the surface therefore differs but little from the parabolic cylinder.*

As to the behaviour of a surface near a parabolic point at which $L = 0$ and $N = 0$ (and hence $M = 0$) nothing definite can be said; it may generally be quite complex.

The computations also show that at (0, 0) we have $E = 1$, $F = 0$ and $G = 0$ for the surface under consideration, from which it follows that at this point

$$K = rt - s^2 \quad \text{and} \quad H = \frac{r+t}{2}.$$

Moreover, immediate calculations (which can be done mentally if we notice that for functions f and g having the property that $f(0) = 0$ and $g(0) = 1$ the derivative $(fg)'$ of their product fg assumes at zero the value $f'(0)$) shows that at (0, 0)

$$\mathbf{n}_u = (-r, -s, 0), \quad \mathbf{n}_v = (-s, -t, 0)$$

and hence

$$\mathbf{n}_u^2 = r^2 + s^2, \quad \mathbf{n}_u\mathbf{n}_v = s(r+t), \quad \mathbf{n}_v^2 = s^2 + t^2.$$

It follows that

$$\mathbf{n}_u^2 = 2HL - KE, \quad \mathbf{n}_u\mathbf{n}_v = 2HM - KF, \quad \mathbf{n}_v^2 = 2HN - KG,$$

i.e. that

$$\mathrm{III} = 2H\mathrm{II} - K\mathrm{I},$$

where III is the third quadratic form of a surface introduced in Remark 1. Thus form III is indeed linearly expressed in terms of forms I and II.

For the ruled surface

$$\mathbf{r} = \boldsymbol{\rho}(u) + v\mathbf{a}(u), \tag{12}$$

as we already know,

$$E = 1 + 2v\dot{\boldsymbol{\rho}}\dot{\mathbf{a}} + v^2\dot{\mathbf{a}}^2, \quad F = \dot{\boldsymbol{\rho}}\mathbf{a}, \quad G = 1$$

(we assume as usual that the parameter u on the curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ is natural and the vector $\mathbf{a}(u)$ is a unit vector). Further

$$\mathbf{r}_u = \dot{\boldsymbol{\rho}} + v\dot{\mathbf{a}}, \quad \mathbf{r}_v = \mathbf{a},$$

$$\mathbf{r}_u \times \mathbf{r}_v = \dot{\boldsymbol{\rho}} \times \mathbf{a} + v(\dot{\mathbf{a}} \times \mathbf{a}),$$

$$\mathbf{n} = \frac{\dot{\boldsymbol{\rho}} \times \mathbf{a} + v(\dot{\mathbf{a}} \times \mathbf{a})}{\sqrt{EG - F^2}},$$

$$\mathbf{r}_{uu} = \ddot{\boldsymbol{\rho}} + v\ddot{\mathbf{a}}, \quad \mathbf{r}_{uv} = \dot{\mathbf{a}}, \quad \mathbf{r}_{vv} = 0,$$

$$L = \frac{(\ddot{\boldsymbol{\rho}} + v\ddot{\mathbf{a}})(\dot{\boldsymbol{\rho}} \times \mathbf{a} + v(\dot{\mathbf{a}} \times \mathbf{a}))}{\sqrt{EG - F^2}}, \quad M = \frac{\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}}}{\sqrt{EG - F^2}}, \quad N = 0,$$

$$LN - M^2 = -\frac{(\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}})^2}{EG - F^2},$$

and therefore

$$K = -\frac{(\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}})^2}{(EG - F)^2} \leqslant 0.$$

Thus the total curvature of an arbitrary ruled surface is nonpositive at any of its points, i.e. a *ruled surface has no elliptic points.*

When the surface is a cylinder ($\dot{\mathbf{a}} = 0$), a cone ($\mathbf{a} = \boldsymbol{\rho}$ and therefore $\dot{\mathbf{a}} = \dot{\boldsymbol{\rho}}$) or a surface of tangents ($\mathbf{a} = \dot{\boldsymbol{\rho}}$), the formula obtained yields $K = 0$. Thus the *total curvature of every developable is zero* (at any point).

Conversely, if $K = 0$, then $\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}} = 0$, i.e. the vectors $\dot{\boldsymbol{\rho}}$, $\mathbf{a}$, $\dot{\mathbf{a}}$ are coplanar. If the vector $\dot{\mathbf{a}}(u)$ is not identically zero, i.e. if surface (12) is not a cylinder, then, going over if necessary, to a smaller neighbourhood, we may assume that $\dot{\mathbf{a}}(u) \neq 0$ for all u. The vectors $\mathbf{a}$ and $\dot{\mathbf{a}}$ are therefore linearly independent (they are nonzero and orthogonal) and hence the vector $\dot{\boldsymbol{\rho}}$ is expressed linearly in terms of $\mathbf{a}$ and $\dot{\mathbf{a}}$:

$$\dot{\boldsymbol{\rho}} = \lambda\mathbf{a} + \mu\dot{\mathbf{a}},$$

where $\lambda = \lambda(u)$, $\mu = \mu(u)$ are some functions of u.

Let

$$u_1 = u, \quad v_1 = v + \mu(u).$$

Since the Jacobian of this transformation is equal to 1, the numbers u_1 and v_1 are also, after changing to a smaller neighbourhood, the coordinates on surface (12). In these coordinates the equation of the surface is of the form

$$\mathbf{r} = \boldsymbol{\rho}_1(u) + v\mathbf{a}(u)$$

(we omit the subscripts in u_1 and v_1), where

$$\boldsymbol{\rho}_1(u) = \boldsymbol{\rho}(u) - \mu(u)\,\mathbf{a}(u),$$

and so,

$$\dot{\boldsymbol{\rho}}_1 = \dot{\boldsymbol{\rho}} - \dot{\mu}\mathbf{a} - \mu\dot{\mathbf{a}} = (\lambda - \dot{\mu})\,\mathbf{a}.$$

If $\dot{\boldsymbol{\rho}}_1$ is identically zero (i.e. $\lambda = \dot{\mu}$), then the equation of the surface is of the form

$$\mathbf{r} = \text{const} + v\mathbf{a}(u)$$

and therefore the surface is a cone. Otherwise we may assume, diminishing if necessary the neighbourhood, that $\dot{\boldsymbol{\rho}}_1(u) \neq 0$ for all u. Changing to the natural parameter (and changing if necessary the sign of v) we see that $\dot{\boldsymbol{\rho}}_1 = \mathbf{a}$, i.e. that the surface under consideration is a surface of tangents.

Thus we have proved the following proposition.

Proposition 1. *A ruled surface has zero total curvature at every point,*

$$K = 0,$$

if and only if it is a developable. □

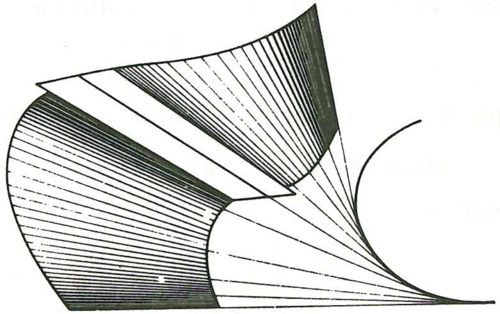
We have also established that developables are characterized by the condition

$$\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}} = 0$$

which is, as is easily seen, equivalent to the collinearity of the vectors $\dot{\boldsymbol{\rho}} \times \mathbf{a}$ and $\dot{\mathbf{a}} \times \mathbf{a}$. But the collinearity of these vectors implies that the vector

$$\mathbf{r}_u \times \mathbf{r}_v = \dot{\boldsymbol{\rho}} \times \dot{\mathbf{a}} + v(\dot{\mathbf{a}} \times \mathbf{a})$$

is, up to proportionality, independent of v, i.e. the corresponding unit vector $\mathbf{n}$ is independent of v. This proves that *developables can be distinguished among all the ruled*



A developable of tangents

surfaces by the property that at all the points of each rectilinear generator such a surface has the same tangent plane.

For an arbitrary surface of revolution

$$\mathbf{r} = x(v)\cos u\cdot\mathbf{i} + x(v)\sin u\cdot\mathbf{j} + z(v)\cdot\mathbf{k}$$

we have

$$\mathbf{r}_u = -x(v)\sin u\cdot\mathbf{i} + x(v)\cos u\cdot\mathbf{j},$$
$$\mathbf{r}_v = x'(v)\cos u\cdot\mathbf{i} + x'(v)\sin u\cdot\mathbf{j} + z'(v)\,\mathbf{k}$$

and hence $E = x(v)^2$, $F = 0$, $G = 1$ (we assume that $x'(v)^2 + z'(v)^2 = 1$; see Lecture 3). Therefore

$$\mathbf{r}_u \times \mathbf{r}_v = x(v)\,z'(v)\cos u\cdot\mathbf{i} + x(v)\,z'(v)\sin u\cdot\mathbf{j} - x(v)\,x'(v)\,\mathbf{k},$$
$$\mathbf{n} = z'(v)\cos u\cdot\mathbf{i} + z'(v)\sin u\cdot\mathbf{j} - x'(v)\,\mathbf{k},$$
$$\mathbf{r}_{uu} = -x(v)\cos u\cdot\mathbf{i} - x(v)\sin u\cdot\mathbf{j},$$
$$\mathbf{r}_{uv} = -x'(v)\sin u\cdot\mathbf{i} + x'(v)\cos u\cdot\mathbf{j},$$
$$\mathbf{r}_{vv} = x''(v)\cos u\cdot\mathbf{i} + x''(v)\sin u\cdot\mathbf{j} + z''(v)\,\mathbf{k},$$
$$L = \mathbf{r}_{uu}\mathbf{n} = -x(v)\,z'(v), \quad M = \mathbf{r}_{uv}\mathbf{n} = 0,$$

$$N = \mathbf{r}_{vv}\mathbf{n} = x''(v)\,z'(v) - z''(v)\,x'(v) = -\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix},$$

$$\frac{LN-M^2}{EG-F^2} = \frac{z'(v)}{x(v)}\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix}.$$

This proves that for a surface of revolution

$$K = \frac{z'(v)}{x(v)}\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix}.$$

Example 1. For a sphere of radius R we have

$$x(v) = R\cos\frac{v}{R}, \quad z(v) = R\sin\frac{v}{R}$$

and therefore

$$x'(v) = -\sin\frac{v}{R}, \quad z'(v) = \cos\frac{v}{R},$$

$$x'' = -\frac{1}{R}\cos\frac{v}{R}, \quad z''(v) = -\frac{1}{R}\sin\frac{v}{R},$$

$$K = \frac{z'(v)}{x(v)}\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = \frac{1}{R^2}.$$

Thus the *total curvature of a sphere of radius R is constant and equal to $1/R^2$.*

The result is intuitively obvious.

The following example is more interesting.

Example 2. A surface of revolution with profile

$$x(v) = R\sin(v), \quad z(v) = R\left(\ln\tan\frac{v}{2} + \cos v\right),$$

$$0 < v < \frac{\pi}{2}$$

(it is the so-called *tractrix*) is termed a *pseudosphere* (R is a *pseudoradius*). For this surface

$$x'(v) = R\cos v, \quad z'(v) = \frac{R}{\sin v} - R\sin v = R\frac{\cos^2 v}{\sin v}$$

and hence

$$x'(v)^2 + z'(v)^2 = R^2\cot^2 v.$$

Since $x'(v)^2 + z'(v)^2 \neq 1$, the above general formula cannot be immediately applied. First it is necessary to go over to the natural parameter of the profile.

We have

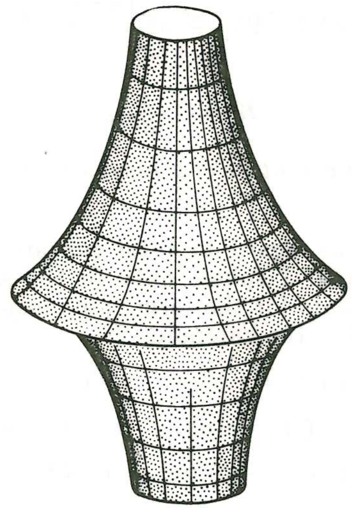
$$s = -R\int_{\frac{\pi}{2}}^{v} \cot v\,dv = -R\ln\sin v$$

and hence

$$\sin v = e^{-\frac{s}{R}}, \quad \cos v = \sqrt{1 - e^{-2\frac{s}{R}}},$$

$$\tan\frac{v}{2} = e^{\frac{s}{R}} - \sqrt{e^{\frac{s}{R}} - 1}.$$

Thus in terms of the natural parameter (which is again de-



A pseudosphere

noted by v) the tractrix will be given by the functions

$$x(v) = Re^{-\frac{v}{R}},$$

$$z(v) = R\ln\left(e^{\frac{v}{R}} - \sqrt{e^{2\frac{v}{R}} - 1}\right) + R\sqrt{1 - e^{-2\frac{v}{R}}}.$$

We calculate:

$$x'(v) = -e^{-\frac{v}{R}}, \quad z'(v) = -\sqrt{1-e^{-2\frac{v}{R}}},$$

$$x''(v) = \frac{1}{R} e^{-\frac{v}{R}}, \quad z''(v) = -\frac{e^{-2\frac{v}{R}}}{R\sqrt{1-e^{-2\frac{v}{R}}}},$$

$$\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = -\frac{e^{-\frac{v}{R}}}{R\sqrt{1-e^{-2\frac{v}{R}}}},$$

$$\frac{z'(v)}{x(v)} \begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = -\frac{1}{R^2}.$$

Thus

$$K = -\frac{1}{R^2},$$

so that the *total curvature of a pseudosphere is constant and equal to* $-\frac{1}{R^2}$.

We see that in regard to total curvature the pseudosphere differs from a sphere only in the sense of curvature. This accounts for the term "pseudosphere".

Example 3. For the catenoid

$$x(v) = \cosh v, \quad z(v) = v,$$

$$x'(v) = \sinh v, \quad z'(v) = 1,$$

$$x'(v)^2 + z'(v)^2 = \cosh^2 v,$$

and therefore we must again pass to the natural parameter

$$s = \int_0^v \cosh v\, dv = \sinh v.$$

Again denoting this parameter by v we obtain the functions

$$x(v) = \sqrt{1+v^2}, \quad z(v) = \ln(v + \sqrt{1+v^2}).$$

Therefore

$$x'(v) = \frac{v}{\sqrt{1+v^2}}, \quad z'(v) = \frac{1}{\sqrt{1+v^2}},$$

$$x''(v) = \frac{1}{(1+v^2)^{3/2}}, \quad z''(v) = -\frac{v}{(1+v^2)^{3/2}},$$

$$\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = -\frac{1}{1+v^2},$$

and hence

$$K = -\frac{1}{(1+v^2)^2}.$$

It is interesting to compare the curvature of the catenoid with that of the isometric helicoid.

For the helicoid we have equation (12) with

$$\boldsymbol{\rho}(u) = u \cdot \mathbf{k} \text{ and } \mathbf{a}(u) = \cos u \cdot \mathbf{i} + \sin u \cdot \mathbf{j}.$$

Therefore

$$\dot{\boldsymbol{\rho}} = \mathbf{k}, \quad \dot{\mathbf{a}} = -\sin u \cdot \mathbf{i} + \cos u \cdot \mathbf{j},$$

$$E = 1 + 2v\dot{\boldsymbol{\rho}}\dot{\mathbf{a}} + v^2\dot{\mathbf{a}}^2 = 1 + v^2,$$

$$F = \dot{\boldsymbol{\rho}}\mathbf{a} = 0, \quad G = 1,$$

$$EG - F^2 = 1 + v^2,$$

$$\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}} = \begin{vmatrix} 0 & 0 & 1 \\ \cos u & \sin u & 1 \\ -\sin u & \cos u & 0 \end{vmatrix} = 1,$$

and hence

$$K = -\frac{1}{(1+v^2)^2}.$$

We have obtained the same result as that for the catenoid! This means that *when the catenoid is bent into the helicoid the total curvatures at the corresponding points coincide.*

What happens to the mean curvature?

For the catenoid $E = 1 + v^2$, $F = 0$, $G = 1$. In addition

$$L = -x(v)z'(v) = -1,\ M = 0,$$

$$N = -\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = \frac{1}{1+v^2},$$

and therefore

$$EN + GL - 2FM = 0,$$

i.e.

$$H = 0.$$

Thus *the mean curvature of the catenoid is equal to zero.*

For the helicoid

$$\dot{\rho} \times \mathbf{a} = \sin u \cdot \mathbf{i} - \cos u \cdot \mathbf{j},\quad \mathbf{a} \times \mathbf{a} = -\mathbf{k},$$

$$\ddot{\rho} = 0,\quad \ddot{\mathbf{a}} = -\cos u \cdot \mathbf{i} - \sin u \cdot \mathbf{j},$$

$$(\ddot{\rho} + v\ddot{\mathbf{a}})(\dot{\rho} \times \mathbf{a} + v(\dot{\mathbf{a}} \times \mathbf{a})) = 0$$

and, in addition, as we have already seen,

$$E = 1 + v^2,\quad F = 0,\quad G = 1,$$

$$EG - F^2 = 1 + v^2,\quad \dot{\rho}\mathbf{a}\dot{\mathbf{a}} = 0.$$

Therefore

$$L = 0,\quad M = \frac{1}{\sqrt{1+v^2}},\quad N = 0,$$

and hence

$$EN + LG - 2FM = 0,$$

i.e.

$$H = 0.$$

Thus the *mean curvature of the helicoid is also zero.*

The example of catenoid and helicoid suggests that total and mean curvatures are preserved under bending (isometry). It turns out that the hypothesis is true for the total curvature (we shall show this in our next lecture) whereas for the mean curvature it is false. Indeed, for a plane the mean curvature is zero while for a circular cylinder of radius R developable into a plane it is obviously equal to $1/2R$.

As to the reasons why the catenoid and helicoid have equal mean curvatures, we are deprived of the possibility of discussing them in this book.

Lecture 5

Weingarten formulas · Coefficients of connection · The Gauss theorem · Explicit formula for Gaussian curvature · The necessary and sufficient conditions of isometry · Surfaces of constant curvature

For the moving basis $\mathbf{r}_u$, $\mathbf{r}_v$, $\mathbf{n}$ of an arbitrary surface

(1) $$\mathbf{r} = \mathbf{r}(u, v)$$

formulas can be written, similar to Frenet's formulas for curves, that yield an expansion of the derivatives

$$\mathbf{r}_{uu},\ \mathbf{r}_{uv},\ \mathbf{r}_{vv},\ \mathbf{n}_u,\ \mathbf{n}_v$$

of the vectors of the moving basis with respect to that same basis.

Since $\mathbf{n}^2 = 1$ and hence $\mathbf{n}\mathbf{n}_u = 0$ and $\mathbf{n}\mathbf{n}_v = 0$, the vectors $\mathbf{n}_u$ and $\mathbf{n}_v$ are only expanded with respect to the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$, so that

$$\mathbf{n}_u = \alpha\mathbf{r}_u + \beta\mathbf{r}_v,$$

$$\mathbf{n}_v = \alpha_1\mathbf{r}_u + \beta_1\mathbf{r}_v.$$

Multiplying the first of these formulas by $\mathbf{r}_u$ and $\mathbf{r}_v$ we obtain two relations:

$$-L = \mathbf{r}_u\mathbf{n}_u = \alpha\mathbf{r}_u^2 + \beta\mathbf{r}_u\mathbf{r}_v = \alpha E + \beta F,$$

$$-M = \mathbf{r}_v\mathbf{n}_u = \alpha\mathbf{r}_u\mathbf{r}_v + \beta\mathbf{r}_v^2 = \alpha F + \beta G,$$

from which it follows that

$$\alpha = \frac{FM - GL}{EG - F^2},\quad \beta = \frac{FL - EM}{EG - F^2}.$$

The coefficients of the second formula are calculated in the same way:

$$\alpha_1 = \frac{FN-GM}{EG-F^2}, \quad \beta_1 = \frac{FM-EN}{EG-F^2}.$$

Further, since by definition

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

and since under the hypothesis $\mathbf{r}_u\mathbf{n} = 0$, $\mathbf{r}_v\mathbf{n} = 0$, the coefficients of $\mathbf{n}$ in the decomposition of vectors $\mathbf{r}_{uu}$, $\mathbf{r}_{uv}$, $\mathbf{r}_{vv}$ in terms of the basis $\mathbf{r}_u$, $\mathbf{r}_v$, $\mathbf{n}$ are equal to L, M, N respectively.

We thus have

$$\begin{aligned} \mathbf{r}_{uu} &= \Gamma^1_{11}\mathbf{r}_u + \Gamma^2_{11}\mathbf{r}_v + L\mathbf{n}, \\ \mathbf{r}_{uv} &= \Gamma^1_{12}\mathbf{r}_u + \Gamma^2_{12}\mathbf{r}_v + M\mathbf{n}, \\ \mathbf{r}_{vv} &= \Gamma^1_{22}\mathbf{r}_u + \Gamma^2_{22}\mathbf{r}_v + N\mathbf{n}, \\ \mathbf{n}_u &= \frac{FM-GL}{EG-F^2}\mathbf{r}_u + \frac{FL-EM}{EG-F^2}\mathbf{r}_v, \\ \mathbf{n}_v &= \frac{FN-GM}{EG-F^2}\mathbf{r}_u + \frac{FM+EN}{EG-F^2}\mathbf{r}_v, \end{aligned} \tag{2}$$

where Γ^k_{ij}, $i, j, k = 1, 2$, are some functions of u and v. Formerly these functions were denoted by

$$\left\{\begin{matrix} i\ j \\ k \end{matrix}\right\}$$

and called the *Christoffel symbols*. But now they are usually called *connection coefficients*.

Formulas (2) are called *Weingarten formulas*.

To compute connection coefficients Γ^k_{ij} we first find the six products of vectors $\mathbf{r}_{uu}$, $\mathbf{r}_{uv}$, $\mathbf{r}_{vv}$ by vectors $\mathbf{r}_u$ and $\mathbf{r}_v$.

Since $\mathbf{r}_u^2 = E$, we have $2\mathbf{r}_{uu}\mathbf{r}_u = E_u$ and $2\mathbf{r}_{uv}\mathbf{r}_u = E_v$, i.e.

$$\mathbf{r}_{uu}\mathbf{r}_u = \frac{1}{2}E_u \text{ and } \mathbf{r}_{uu}\mathbf{r}_u = \frac{1}{2}E_v.$$

Similarly, since $\mathbf{r}_v^2 = G$ we have

$$\mathbf{r}_{uv}\mathbf{r}_v = \frac{1}{2}G_u \text{ and } \mathbf{r}_{vv}\mathbf{r}_v = \frac{1}{2}G_v.$$

Besides, since $\mathbf{r}_u\mathbf{r}_v = F$, we have $\mathbf{r}_{uu}\mathbf{r}_v + \mathbf{r}_u\mathbf{r}_{uv} = F_u$ and $\mathbf{r}_{uv}\mathbf{r}_v + \mathbf{r}_u\mathbf{r}_{vv} = F_v$, from which it follows that

$$\mathbf{r}_{uu}\mathbf{r}_v = F_u - \frac{1}{2}E_v \text{ and } \mathbf{r}_{vv}\mathbf{r}_u = F_v - \frac{1}{2}G_u.$$

Now multiplying the first three of the formulas (2) by $\mathbf{r}_u$ and $\mathbf{r}_v$ we obtain six relations:

$$\begin{cases} E\Gamma^1_{11} + F\Gamma^2_{11} = \frac{1}{2}E_u, \\ E\Gamma^1_{11} + G\Gamma^2_{11} = F_u - \frac{1}{2}E_v, \end{cases}$$
$$\begin{cases} E\Gamma^1_{12} + F\Gamma^2_{12} = \frac{1}{2}E_v, \\ E\Gamma^1_{12} + G\Gamma^2_{12} = \frac{1}{2}G_u, \end{cases} \tag{3}$$
$$\begin{cases} E\Gamma^1_{22} + F\Gamma^2_{22} = F_v - \frac{1}{2}G_u, \\ F\Gamma^1_{22} + G\Gamma^2_{22} = \frac{1}{2}G_v, \end{cases}$$

from which it is easy to find the coefficients Γ^k_{ij}. (The equations are uniquely solvable since the determinant $EG - F^2$ of each pair of equations is nonzero.)

We see that the *connection coefficients* Γ^k_{ij} *can be expressed in terms of the coefficients of the first quadratic form and of their derivatives.* Hence *they remain unaltered under bendings (isometries) of a surface.*

We shall not need explicit expressions for the coefficients Γ^k_{ij} in terms of the coefficients of the first quadratic form, and so we shall not give them here.

The coefficients of derived equations are connected by three relations resulting from calculating partial derivatives $\mathbf{r}_{uuv}$, $\mathbf{r}_{uvv}$, and $\mathbf{n}_{uv}$ in two different ways by using these formulas. One of these relations was found by Gauss and the other two by Peterson, Mainardi, and Codazzi. We shall consider only Gauss' relation which we shall obtain by calculating the coefficient of $\mathbf{r}_v$ in the expansion of the partial derivative $\mathbf{r}_{uuv}$ in terms of the vectors $\mathbf{r}_u$, $\mathbf{r}_v$, and $\mathbf{n}$.

In this calculation we shall only follow the coefficient of $\mathbf{r}_v$ and only those of its terms which depend on the coefficients of the second quadratic form. All the other terms will be replaced by dots.

We have

$$\begin{aligned}\mathbf{r}_{uuv} &= (\mathbf{r}_{uu})_v = (\Gamma^1_{11}\mathbf{r}_u + \Gamma^2_{11}\mathbf{r}_v + L\mathbf{n})_v \\ &= \ldots + \Gamma^1_{11}\mathbf{r}_{uv} + \ldots + \Gamma^2_{11}\mathbf{r}_{vv} + \ldots + L\mathbf{n}_v \\ &= \ldots + \Gamma^1_{11}(\ldots) + \ldots + \Gamma^2_{11}(\ldots) \\ &\quad + \ldots + L\left(\ldots + \frac{FM-EN}{EG-F^2}\mathbf{r}_v\right) \\ &= \left(L\frac{FM-EN}{FG-F^2} + \ldots\right)\mathbf{r}_v + \ldots\end{aligned}$$

Similarly

$$\begin{aligned}\mathbf{r}_{uuv} &= (\mathbf{r}_{uv})_u = (\Gamma^1_{12}\mathbf{r}_u + \Gamma^2_{22}\mathbf{r}_v + M\mathbf{n})_u \\ &= \left(M\frac{FL-EM}{EG-F^2} + \ldots\right)\mathbf{r}_v + \ldots\end{aligned}$$

Hence

$$L\frac{FM-EN}{EG-F^2} = M\frac{FL-EM}{EG-F^2} + \ldots,$$

where dots denote terms dependent only on the coefficients of the first quadratic form. But

$$M\frac{FL-EM}{EG-F^2} - L\frac{FM-EN}{EG-F^2} = E\frac{LN-M^2}{EG-F^2} = EK.$$

Since $E \neq 0$ (form I is positive definite), this proves that the *total curvature K of a surface is expressed in terms of the coefficients of the first quadratic form (and of their derivatives)*. It follows that the curvature K remains unchanged under bendings. More precisely, if f is an isometry of a surface $\mathscr{X}$ onto a surface $\mathscr{Y}$, then

$$K_{\mathscr{Y}} \circ f = K_{\mathscr{X}}, \tag{4}$$

where $K_{\mathscr{X}}$ and $K_{\mathscr{Y}}$ are the total curvatures on $\mathscr{X}$ and $\mathscr{Y}$ respectively. (Indeed, if f preserves the coordinates of vectors in another system of coordinates, then both sides in (4) differ only in the notation of the coordinates.) This result deserves to be distinguished as a theorem.

Theorem 1 (**Gauss' theorem**). *The total (Gaussian) curvature of a surface remains unchanged under bendings (isometries), i.e. isometric surfaces have the same curvature at corresponding points.* □

Gauss was so delighted with the theorem that he called it the theorema egregium, which means a "brilliant theorem" in Latin.

From Theorem 1 it follows in particular that *no arbitrarily small part of a sphere can be bent into a plane*. Therefore no map gives an accurate representation of the Earth's surface.

An explicit expression of curvature K in terms of the coefficients E, F, and G of the first quadratic form is

$$K = -\frac{1}{4(EG-F^2)^2}\begin{vmatrix} E & E_u & E_v \\ F & F_u & F_v \\ G & G_u & G_v \end{vmatrix} - \frac{1}{2\sqrt{EG-F^2}}\left\{\left(\frac{E_v-F_u}{\sqrt{EG-F^2}}\right)_v - \left(\frac{F_v-G_u}{\sqrt{EG-F^2}}\right)_u\right\}. \tag{5}$$

The other two relations resulting from differentiating the derived equations (and usually called the *Peterson-Codazzi formulas*) are of the form

$$2(EG-F^2)(L_v-M_u) - (EN+GL-2FM)(E_v-F_u) + \begin{vmatrix} E & E_u & L \\ F & F_u & M \\ G & G_u & N \end{vmatrix} = 0,$$

$$2(EG-F^2)(M_v-N_u) - (EN+GL-2FM)(F_v-G_u) + \begin{vmatrix} E & E_v & L \\ F & F_v & M \\ G & G_v & N \end{vmatrix} = 0. \tag{6}$$

To prove these formulas all one needs is patience and care.

We shall prove only formula (5) and only for the case where $E = 1$ and $F = 0$, i.e. where the first quadratic form of a surface is expressed as

(7) $$\mathrm{I} = du^2 + G\,dv^2.$$

In this case equations (3) for Γ_{ij}^k are of the form

$$\Gamma_{11}^1 = 0, \quad \Gamma_{12}^1 = 0, \quad \Gamma_{22}^1 = -\frac{1}{2}G_u,$$

$$G\Gamma_{11}^2 = 0, \quad G\Gamma_{12}^2 = G_u, \quad G\Gamma_{22}^2 = \frac{1}{2}G_v,$$

from which it follows that

$$\Gamma_{11}^1 = 0, \quad \Gamma_{12}^1 = 0, \quad \Gamma_{11}^2 = 0,$$

$$\Gamma_{22}^1 = -\frac{1}{2}G_u, \quad \Gamma_{12}^2 = \frac{1}{2}\frac{G_u}{G}, \quad \Gamma_{22}^2 = \frac{1}{2}\frac{G_v}{G}.$$

Hence

$$\mathbf{r}_{uu} = L\mathbf{n} \quad \text{and} \quad \mathbf{r}_{uv} = \frac{1}{2}\frac{G_u}{G}\mathbf{r}_v + M\mathbf{n}.$$

Since in the case at hand

$$\mathbf{n}_u = -L\mathbf{r}_v - \frac{M}{G}\mathbf{r}_v \quad \text{and} \quad \mathbf{n}_v = -M\mathbf{r}_u - \frac{N}{G}\mathbf{r}_v,$$

it follows that

$$\mathbf{r}_{uuv} = (L\mathbf{n})_v = L_v\mathbf{n} + L\left(-M\mathbf{r}_u - \frac{N}{G}\mathbf{r}_v\right)$$

$$= -LM\mathbf{r}_u - \frac{2N}{G}\mathbf{r}_v + L_v\mathbf{n}$$

and

$$\mathbf{r}_{uvu} = \left(\frac{1}{2}\frac{G_u}{G}\mathbf{r}_v + M\mathbf{n}\right)_u = \frac{1}{2}\left(\frac{G_u}{G}\right)_u\mathbf{r}_v$$

$$+\frac{1}{2}\frac{G_u}{G}\left[\frac{1}{2}\frac{G_u}{G}\mathbf{r}_v + M\mathbf{n}\right]$$

$$+M\mathbf{n} + M\left(-L\mathbf{r}_u - \frac{M}{G}\mathbf{r}_v\right)$$

$$= -LM\mathbf{r}_u + \left[\frac{1}{2}\left(\frac{G_u}{G}\right)_u + \frac{1}{4}\left(\frac{G_u}{G}\right)^2 - \frac{M^2}{G}\right]\mathbf{r}_v$$

$$+\left(\frac{1}{2}\frac{G_u}{G}M + M_u\right)\mathbf{n}$$

and hence

$$-\frac{LN}{G} = \frac{1}{2}\left(\frac{G_u}{G}\right) + \frac{1}{4}\left(\frac{G_u}{G}\right)^2 - \frac{M^2}{G},$$

$$L_v = \frac{1}{2}\frac{G_u}{G}M + M_u.$$

The second equation is now of no interest to us (it is the first of equations (6) for $E=1$ and $F=0$) and the first (since in the case at hand $K = \frac{1}{G}(LN - M^2)$) gives

$$K = -\frac{1}{2}\left(\frac{G_u}{G}\right)_u - \frac{1}{4}\left(\frac{G_u}{G}\right)^2,$$

i.e., as an obvious computation shows,

(8) $$K = -\frac{(\sqrt{G})_{uu}}{\sqrt{G}},$$

which coincides with the result of substituting the values of $E=1$ and $F=0$ in formula (5).

Thus we have proved that the *total curvature of a surface with the first quadratic form* (7) *is expressed by formula* (8).

Let, for example,

(9) $$\mathrm{I} = du^2 + \cos^2 u\; dv^2.$$

Then $\sqrt{G} = \cos u$ and $(\sqrt{G})_{uu} = -\cos u$. Hence $K = 1$, which agrees quite well with the result of Example 1 of Lecture 4 (for (9) is the first quadratic form of a sphere of radius $R=1$; see formula (36) of Lecture 3, in which u and v are interchanged).

Similarly, it can be shown that a surface with the first quadratic form

$$\mathrm{I} = du^2 + \cosh^2 u\; dv^2$$

has the curvature $K = -1$ (cf. Example 2 of Lecture 4).

Remark 1. We stress that all these results hold for surfaces in a *three-dimensional* Euclidean space. For surfaces in a larger, three-dimensional space, however, (5) (or its special form (8) may be taken to be a definition of curvature K.

Remark 2. For six functions

(10) $$E,\ F,\ G,\ L,\ M,\ N$$

given in an open convex set $U \subset \mathbb{R}^2$ to be the coefficients of the first and the second quadratic form of some surface $\mathbf{r} = \mathbf{r}(u, v)$ it is necessary that for these functions, in addition to conditions

$$E > 0, \quad EG - F^2 > 0 \tag{11}$$

of positive definiteness, relations (5) and (6) hold (it is meant that we have substituted $K = \frac{LM - M^2}{EG - F^2}$ in relation (5)). It turns out that *these relations are also sufficient* (for the existence of a regular, but not generally elementary, surface with the given forms I and II). Moreover, functions (10) (satisfying relations (5), (6), and (11)) *define a surface up to movement of space*. These statements are the two-dimensional analogue of the corresponding statements for curves (see Theorem 1 of Lecture 2) and can be proved by a similar method (but instead of the unique existence theorem for solutions of linear ordinary differential equations we use the corresponding theorem for the system of linear partial differential equations).

The Gauss theorem states that for the two surfaces to be isometric it is necessary that the total curvatures be equal. Although this condition is by no means sufficient, it is so strong that making use of it sufficient conditions can be easily obtained. We shall not expound this question and only consider the most important special case of the corresponding theorem.

Let

$$\Delta_1 K = \frac{EK_v^2 - 2FK_uK_v + GK_u^2}{EG - F^2}$$

be Beltrami's first differential parameter of the function K. If the two functions K and $\Delta_1 K$ of u and v are *functionally independent*, i.e. if their Jacobian

$$\begin{vmatrix} \frac{\partial K}{\partial u} & \frac{\partial K}{\partial v} \\ \frac{\partial \Delta_1 K}{\partial u} & \frac{\partial \Delta_1 K}{\partial v} \end{vmatrix}$$

is everywhere nonzero, then they may be taken as new local coordinates on the surface. We call them *Gaussian coordinates*. From the property that the operator Δ_1 is invariant (formula (34) of Lecture 3) and formula (4) it follows directly that for any isometry $f: \mathscr{X} \to \mathscr{Y}$ the equality

$$\Delta_1 K_{\mathscr{Y}} \circ f = \Delta_2 K_{\mathscr{X}} \tag{12}$$

holds. Taken together, formulas (4) and (12) imply that *every isometry is a mapping which preserves the coordinates of vectors in another system of coordinates*. The following theorem is therefore true.

Theorem 2. *Two elementary surfaces on which Gaussian coordinates are defined are isometric if and only if the first quadratic forms in these coordinates coincide.* □

Thus, to determine whether or not two surfaces are isometric it is necessary to introduce (if possible) Gaussian coordinates and calculate in these coordinates the first quadratic forms of the surfaces. If the forms coincide, the surfaces are isometric, but if they differ, the surfaces are not isometric.

Theorem 2 gives no answer when K and $\Delta_1 K$ are functionally dependent, for example, when $\Delta_1 K = 0$ (which occurs, as can be easily figured out, if and ony if $K = \text{const}$). In this case, however, the condition of Theorem 1 proves to be sufficient, i.e. *two elementary surfaces of constant total curvature are isometric if and only if they have the same curvature*. In other words, *any surface of constant total curvature K is locally isometric to a sphere of radius $R = \frac{1}{\sqrt{K}}$ if $K > 0$, to a plane if $K = 0$, and to a pseudosphere of pseudoradius $R = \frac{1}{\sqrt{-K}}$ if $K < 0$.*

To prove this we shall need a lemma which is to be proved in the next semester:

Lemma (Gauss). *On any surface there are local coordinates u, v in which the first quadratic form of that surface is of form (7), the function $G = G(u, v)$ having the property that*

$$G(0, v) = 1 \quad \textit{and} \quad G_u(0, v) = 0 \ \textit{for all } v. \tag{13}$$

By virtue of this lemma we may assume without loss of generality that the first quadratic form of a given surface of constant total curvature K is of form (7) and hence formula (8) holds for K. This formula may be thought of as a differential equation of second degree with constant coefficients relative to the function $\sqrt{G}$. It is known from the theory of differential equations that the general solution of that equation is of the form

$$\sqrt{G}=\begin{cases} A\cos a(u+B) & \text{if } K=a^2>0,\\ Au+B & \text{if } K=0,\\ A\cosh a(u+B) & \text{if } K=-a^2<0,\end{cases}$$

where A and B are arbitrary functions of v. But in view of the first of the conditions (13) we must have

$$\begin{aligned} A\cos aB&=1 \text{ for } K>0,\\ B&=1 \text{ for } K=0,\\ A\cosh aB&=1 \text{ for } K<0,\end{aligned}$$

and in view of the second of the conditions (13), by virtue of the identities

$$(\sqrt{G})_u=\frac{1}{2}\frac{G_u}{\sqrt{G}}=\begin{cases} -Aa\sin a(u+B) & \text{if } K>0,\\ A & \text{if } K=0,\\ Aa\sinh a(u+B) & \text{if } K<0\end{cases}$$

we must have

$$\begin{aligned} Aa\sin aB&=0 \text{ for } K>0,\\ A&=0 \text{ for } K=0,\\ Aa\sinh aB&=0 \text{ for } K<0.\end{aligned}$$

Hence

$$\sqrt{G}=\begin{cases} \cos au & \text{if } K=a^2>0,\\ 1 & \text{if } K=0,\\ \cosh au & \text{if } K=-a^2<0.\end{cases}$$

In the first case we obtain the first quadratic form

$$du^2+\cos^2 au\, dv^2$$

for a sphere of radius $1/a$, in the second case we obtain the first quadratic form

$$du^2+dv^2$$

for a plane, and in the third the first quadratic form

$$du^2+\cosh^2 au\, dv^2$$

for a pseudosphere of pseudoradius $1/a$. □

Here we interrupt the exposition of the theory of surfaces and turn to the main subject of this course, the theory of smooth manifolds.