DIFFERENTIAL GEOMETRY

OF SPECIAL MAPPINGS

Josef Mikeš
et al.

Differential geometry of special mappings

Josef Mikeš
et al.

DIFFERENTIAL GEOMETRY of SPECIAL MAPPINGS

Josef Mikeš
et al.

Olomouc 2019

Reviewed by: prof. RNDr. Miroslav Doupovec, CSc.
doc. RNDr. Miroslav Kureš, CSc.

Authors:

Josef Mikeš
Sándor Bácsó
Vladimir E. Berezovski
Olena Chepurna
Marie Chodorová
Hana Chudá
Stanisław Formella
Michail L. Gavrilchenko
Michael Haddad
Irena Hinterleitner
Marek Jukl
Lenka Juklová
Dzhanybek Moldobaev
Patrik Peška
Lenka Rýparová
Igor G. Shandra
Mohsen Shiha
Dana Smetanová
Vasilij S. Sobchuk
Sergey E. Stepanov
Elena S. Stepanova
Irina I. Tsyganok
Alena Vanžurová

This product is co-financed by grants IGA PrF 2018012 and 2019015 of Palacky University.

First Edition is co-financed by the European Social Fund and the state budget of the Czech Republic, project POST-UP, reg. number CZ.1.07/2.3.00/30.0004.

First Edition

DOI: 10.5507/prf.19.24455365
ISBN 978-80-244-5536-5 (online : PDF)

CONTENTS

Out of nothing I have created
a strange new universe

Dedicated to the memory of

János BOLYAI
1802 – 1860

The Hungarian mathematician and
the founder of non-Euclidean Geometry

He served as a captain in Olomouc
from July 10, 1832 to June 15, 1833

INTRODUCTION

During the last 50 years, many new and interesting results have appeared in the theory of conformal, geodesic, holomorphically projective, F-planar and others mappings and transformations of manifolds with affine connection, Riemannian, Kähler and Riemann-Finsler manifolds. The authors dedicate the present monograph to the exposition of this topic.

Problems connected with this field were considered in many monographs, surveys (pp. 613–619) and dissertation theses (pp. 620–621).

In the theory of geodesic, conformal and holomorphically projective mappings and some generalizations, three main directions have been specified:

- the investigation of general laws and rules;
- the integration of basic equations, and
- the investigations for special spaces.

Recently, new results that were not reflected in the papers mentioned above have been obtained. On the one hand, some results of a general character, on the other hand, results concerning mappings of special manifolds with affine connection and Riemannian spaces, including spaces of constant curvature, Kähler, Einstein spaces, conformally flat spaces, Klingenberg spaces, etc.

Many works have been dedicated to the problem of non-existence of conformal, geodesic and holomorphically projective mappings and transformations, and concircular vector fields in spaces of a special kind. Such problems are often closely related. However, much attention has not been paid to their investigation yet. New results on the integration of basic geodesic mappings equations are considered in the review [10, 12] and in the monograph [11] by A.V. Aminova.

We give the basic concepts of the theory of manifolds with affine connection, Riemannian, Kähler and Riemann-Finsler manifolds, using the notation from [63, 64, 140, 141, 143, 144, 164, 181, 196, 199, 229].

Unless otherwise stated, the investigations are carried out in tensor form, locally, in the class of sufficiently smooth real functions. The dimension n of the spaces under consideration is supposed to be higher than two, as a rule. This fact is not explicitly stipulated in the text. All the spaces are assumed to be connected. Under Riemannian manifolds we mean both positive as well as pseudo-Riemannian manifolds.

The book was edited by J. Mikeš. The book consists of 18 chapters. The first sixt chapters of the book are of introductory character, and include also some historical remarks.

We would like to stress that we use here the classical definition of geodesics, i.e. with a general parameter, which is widely used in applications in theoretical physics. Further note that the definition of the Ricci tensor was splitted, since 1950' its sign is used with an opposite sign, see [196]. We go back to the original notation, L.P. Eisenhart [63].

Some parts of the text are based on several graduate courses on topology, differential geometry, tensor analysis, Riemannian geometry, geodesic mappings, holomorphically mappings and Lie groups given by N.S. Sinyukov, M.L. Gavrilchenko and J. Mikeš at Mechnikov's Odessa State University, and differential geometry and topology by J. Mikeš and A. Vanžurová at Palacky University in Olomouc.[1)]

The authors believe that the text might evoke interest and might be helpful for post-graduate students in mathematics, geometry or physics as well as for research-work specialists in these fields.

We would like to thank to prof. P.I. Kovalev for his corrections.

We wish to express our deep appreciation to our referees, Professors M. Doupovec and M. Kureš.

We are also grateful to M. Závodný, L. Rachůnek and V. Heinz for preparing the figures and the final camera-ready copy of the text.

We appologize to our readers for all pertinent mistakes.

This book is the second edited and extended edition of the monograph *Differential geometry of Special Mappings* which was published in 2015 (by the project POST-UP CZ 1.07/2.3.00/30.0004).

This work was supported by grants IGA PrF 2018012 and IGA PrF 2019015 of Palacky University.

November 2018

Josef Mikeš
Palacky University
Olomouc

[1)] Sinyukov Nikolai Stepanovich, 1925-1992, was the head of the Department of Geometry and Topology of Odessa State University. He was a founder of Odessa school of geometry of geodesic mapping and their generalizations. He was a scientific supervisor of the co-authors S. Bácsó, M.L. Gavrilchenko, J. Mikeš, Dz. Moldobayev, I.G. Shandra, and also L.L. Bezkorovainaya, S.I. Fedishchenko, P.I. Kovalev, I.N. Kurbatova, S.G. Leiko, E.D. Oboznaya, S.A. Pokas, see pp. 620, 621.

Josef Mikeš was a scientific supervisor of V. Berezovskii, O. Chepurna, M. Chodorová, H. Chudá, K. Esenov, M. Haddad, V. Kiosak, J. Křížek, P. Peška, L. Rachůnek, L. Rýparová, A. Sabykanov, M. Shiha, M. Sochor, J. Stránská, M. Trnková, see pp. 620, 621; and also supervisor of approximately 70 master's degree theses, among others L. Pospíšilová, E.N. Sinyukova.

1 CURVES AND SURFACES IN EUCLIDEAN SPACES

First, let us give a brief motivation from Euclidean spaces.[2)]

What shall we understand under a curve? From the "static" point of view, a curve can be considered as a particular point subset in the plane, in the 3-space, or on a surface in the space, which is "one-dimensional" in a certain sense. Conics in elementary geometry, affine or projective algebraic curves in algebraic geometry (considered as null-sets of polynomials) can serve as well-known examples. In practice, we can see various arcs or curve segments on various objects surrounding us: on trees and plants, on our furniture, buildings, bridges, roads etc. On the other hand, in many situations, "curves" are drawn up by "moving object" (imagine a car on a road, an airplane on the sky, a ray of light, a moving particle in physics, various moving parts of machines etc.). Curves arising in such a "dynamic" way can be nicely described by suitable functions.

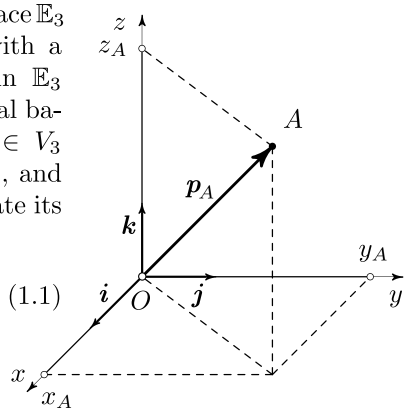
And how surfaces do arise? In real life, "surfaces" are often created by the deformation of flat pieces followed by "glueing together". A football-match ball is sewed from pieces of leather, clothes are sewed from flat pieces of textile material. A car body style is made from flat pieces of metal by deformation. These examples give a pretty good inspiration for creating precise mathematical definitions. In the language of functions, we are able to describe the situation when a curve lies on a surface. Here we will focus on the behavior of the so-called geodesic curves, that is, curves on surfaces or in spaces which, in a way, play an analogous role as straight lines in a plane.

1.1 Vector functions

1.1.1 Coordinates

Consider the three-dimensional Euclidean space $\mathbb{E}_3$ over a real vector space $V_3(\mathbb{R})$ endowed with a dot product "$\cdot$". If we fix a point O in $\mathbb{E}_3$ which is called an *origin*, and an orthonormal basis $\mathcal{E} = \langle \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k} \rangle$ in V_3 then any vector $\boldsymbol{p} \in V_3$ determines a unique point $A = O + \boldsymbol{p} \in \mathbb{E}_3$, and vice versa, for any point $A \in \mathbb{E}_3$, we can create its "radius-vector"

$$\boldsymbol{p}_A = A - O = \overrightarrow{OA} \in V_3. \tag{1.1}$$



[2)] Euclid of Alexandria, 323–285 BC, was a Greek mathematician often referred to as the *founder of geometry* or the *father of geometry.*

Its expression $\boldsymbol{p}_A = x_A\boldsymbol{i} + y_A\boldsymbol{j} + z_A\boldsymbol{k}$ with respect to the basis $\mathcal{E}$ gives rise to the (bijective) map

$$f\colon\ \mathbb{R}^3 \to \mathbb{E}_3, \qquad [x, y, z] \mapsto O + x\,\boldsymbol{i} + y\,\boldsymbol{j} + z\,\boldsymbol{k} \tag{1.2}$$

which is called a *Cartesian*[3] *coordinate system*[4]; $[x_A, y_A, z_A]$ are *Cartesian coordinates of the point* A.

In V_3 we denote *dot*, *cross* and *triple product*[5]: $\boldsymbol{u}\cdot\boldsymbol{v}$, $\boldsymbol{u}\times\boldsymbol{v}$ and $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})$, for which

$$\boldsymbol{u}\cdot\boldsymbol{v} = x_u\,x_v + y_u\,y_v + z_u\,z_v, \quad \boldsymbol{u}\times\boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} \quad \text{and} \quad (\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) = \begin{vmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \\ x_w & y_w & z_w \end{vmatrix}.$$

A *length* of the vector $\boldsymbol{u}$ is defined $|\boldsymbol{u}| = \sqrt{\boldsymbol{u}\cdot\boldsymbol{u}} = \sqrt{x_u^2 + y_u^2 + z_u^2}$.

Obviously, the same construction works in arbitrary dimension n and gives a way how to describe particular point subsets, such as curves and surfaces, by means of vector functions[6].

1.1.2 Vector functions, continuity and differentiability

Let $V_n(\mathbb{R})$ be a real n-dimensional vector space with a dot product, and let $I \subset \mathbb{R}$ be an open interval (often $I = (0, 1)$, but not necessarily). Denote by $|\boldsymbol{v}|$ the *length of vector*, $|\boldsymbol{v}| = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}} = \sqrt{{v^1}^2 + \cdots + {v^n}^2}$.

We say that $\boldsymbol{v}_0 \in V_n$ is a *limit* of a vector function $\boldsymbol{v}\colon\ I \to V_n$ at $t_0 \in I$ if $\lim\limits_{t\to t_0} |\boldsymbol{v}(t) - \boldsymbol{v}_0| = 0$ holds, and we write

$$\lim_{t\to t_0} \boldsymbol{v}(t) = \boldsymbol{v}_0. \tag{1.3}$$

If $\lim\limits_{t\to t_0} \boldsymbol{v}(t) = \boldsymbol{v}_0$ holds (or if it holds for all $t_0 \in I$), $\boldsymbol{v}$ is called *continuous at the point* t_0 (or *continuous on the interval* I, respectively).

If there exists a limit $\lim\limits_{t\to t_0} \dfrac{\boldsymbol{v}(t) - \boldsymbol{v}(t_0)}{t - t_0}$ then this limit is called a *derivation of the vector function* $\boldsymbol{v}(t)$ at t_0, and it is denoted by $\boldsymbol{v}'(t_0)$ or

$$\frac{d\boldsymbol{v}(t_0)}{dt} = \boldsymbol{v}'(t_0) = \lim_{t\to t_0} \frac{\boldsymbol{v}(t) - \boldsymbol{v}(t_0)}{t - t_0}. \tag{1.4}$$

---

[3] René Descartes, 1596–1650, the French mathematician and philosopher, an author of *La Géométrie*. Cartesian means here related to Descartes who was also known as Renatus Cartesius (Latinized form). Descartes was one of the key figures of the Scientific Revolution. His influence in physics and mathematics is also apparent. Originally, Descartes introduced the new idea of specifying the position of a point in a plane using two intersecting (perpendicular) axes as measuring guides; among others, his discovery brings a method of how to connect geometry and algebra. The new method influenced particularly analytic geometry, calculus, and cartography. Later, the Cartesian coordinate system, used in both plane and space geometry, was named after him. The idea of this system was developed in 1637 in two writings by Descartes and independently by Pierre de Fermat (but Fermat did not publish his ideas).

[4] Sometimes, under a coordinate system one means its inverse $f^{-1}\colon\ \mathbb{E}_3 \to \mathbb{R}^3$.

[5] Often called *inner* (*scalar*), *vector* and *mixed product*, respectively.

[6] Under a *vector function* on a set $N \subset \mathbb{R}^m$, $m \in \mathbb{N}$ (with the definition domain N) we mean a mapping $N \to V$ where V is a vector space.

Obviously, $\boldsymbol{v}'(t_0) \in V_n$ is a vector. If $\boldsymbol{v}'(t_0)$ exists for all $t_0 \in I$ we get a new vector function $\boldsymbol{v}' = \boldsymbol{v}'(t)$, $\boldsymbol{v}'\colon I \to V_n$ called a *derivative* of $\boldsymbol{v}$. By iteration, higher order derivates can be introduced.

With respect to an orthonormal basis $\langle \boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n \rangle$ of V_n, we can write

$$\boldsymbol{v}(t) = v^1(t)\,\boldsymbol{e}_1 + \cdots + v^n(t)\,\boldsymbol{e}_n \tag{1.5}$$

for any $t \in I$. The real functions $v^i(t)$, $i = 1, \dots, n$, are called *components* of the vector function; in short, we write $\boldsymbol{v}(t) = (v^1(t), \dots, v^n(t))$.

In terms of components, the following useful description of continuity and differentiability can be given: $\boldsymbol{v}(t)$ is continuous (continuous at $t_0 \in I$, respectively) if and only if all components $v^i(t)$ are continuous (are continuous at $t_0 \in I$, respectively). The derivative of $\boldsymbol{v}$ at t_0 exists if and only if all components have a derivative at t_0, and if this is the case then

$$\boldsymbol{v}'(t_0) = \left(\frac{d\boldsymbol{v}^1(t_0)}{dt}, \dots, \frac{d\boldsymbol{v}^n(t_0)}{dt} \right) \tag{1.6}$$

holds.

Similarly for higher order derivatives. We say that a vector function $\boldsymbol{v}\colon I \to V_n$ is *of the class* C^r on I, in short $\boldsymbol{v} \in C^r(I)$, if all components are of the class C^r (as real functions) on I, that is, are continuously differentiable up to order r. A vector function is *smooth* if derivatives of all orders exist, we write $\boldsymbol{v} \in C^\infty(I)$, and (real) *analytic*, $\boldsymbol{v} \in C^\omega(I)$, if all components are real analytic (i.e. each component possesses derivatives of all orders and agrees with its Taylor series in a neighborhood of every point)[7].

If the following derivatives $\alpha'(t), \boldsymbol{u}'(t), \boldsymbol{v}'(t), \boldsymbol{w}'(t)$ exist, then at a point t we have:

$$\begin{gathered}(\boldsymbol{u} \pm \boldsymbol{v})' = \boldsymbol{u}' \pm \boldsymbol{v}', \quad (\alpha \cdot \boldsymbol{u})' = \alpha' \cdot \boldsymbol{u} + \alpha \cdot \boldsymbol{u}', \\ (\boldsymbol{u} \cdot \boldsymbol{v})' = \boldsymbol{u}' \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{v}', \quad (\boldsymbol{u} \times \boldsymbol{v})' = \boldsymbol{u}' \times \boldsymbol{v} + \boldsymbol{u} \times \boldsymbol{v}', \\ (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})' = (\boldsymbol{u}', \boldsymbol{v}, \boldsymbol{w}) + (\boldsymbol{u}, \boldsymbol{v}', \boldsymbol{w}) + (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}').\end{gathered} \tag{1.7}$$

Let $\boldsymbol{p}(t) \in C^r(\langle t_0, t_0 + h \rangle)$ and $\boldsymbol{p}(t) \in C^{r+1}((t_0, t_0 + h))$, $h > 0$, then *Taylor's formula* of $\boldsymbol{p}(t)$ at t_0 has the following form

$$\boldsymbol{p}(t_0+h) = \boldsymbol{p}(t_0) + \boldsymbol{p}'(t_0)\cdot h + \frac{\boldsymbol{p}''(t_0)}{2!}\cdot h^2 + \cdots + \frac{\boldsymbol{p}^{(r)}(t_0)}{r!}\cdot h^r + \boldsymbol{R}_{(p,t_0)}(h).^{8)} \tag{1.8}$$

---

[7] An analytic function is a function that is locally given by a convergent power series. Another speaking, a function is analytic if it is equal to its Taylor series in some neighborhood of any point.

Function $f(x)=\begin{cases} e^{-\frac{1}{x^2}}, & x \neq 0, \\ 0, & x = 0, \end{cases}$ is an example which belongs to $C^\infty(\mathbb{R})$, and is not $C^\omega(\mathbb{R})$.

This follows from $f(0) = f'(0) = \cdots = f^{(n)}(0) = \cdots = 0$. Evidently, Taylor series at point 0 does not converge to function f.

[8] Remainder $\boldsymbol{R}_{(p,t_0)}(h)$ in the Lagrange and Euler forms have the following forms

$$\boldsymbol{R}_{(p,t_0)}(h) = \frac{h^{r+1}}{(r+1)!} \cdot (x^1(\theta_1), \dots, x^n(\theta_n)), \quad \theta_1, \dots, \theta_n \in (t_0, t_0 + h),$$

$$\boldsymbol{R}_{(p,t_0)}(h) = \boldsymbol{o}(h^r) \text{ with Euler function } \boldsymbol{o}(\tau)\colon \lim_{\tau \to 0} \frac{\boldsymbol{o}(\tau)}{\tau} = 0.$$

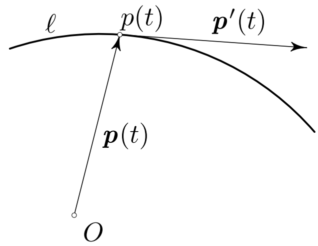
Taylor's series of $\boldsymbol{p}(t)$ at t_0: $\boldsymbol{p}(t_0) + \boldsymbol{p}'(t_0) \cdot h + \dfrac{\boldsymbol{p}''(t_0)}{2!} \cdot h^2 + \cdots + \dfrac{\boldsymbol{p}^{(r)}(t_0)}{r!} \cdot h^r + \cdots$.

1.2 Curves in Euclidean space

1.2.1 Curves and parametrization of curves

Under a *motion* we understand a map $p\colon I \to \mathbb{E}_n$ (usually, satisfying some additional differentiability conditions). Such a map will be later called a *parametrization of a curve*. This concept is a useful tool for describing a motion of a moving particle (geometrically represented by a point) in the Euclidean three-space $\mathbb{E}_3$ as well as in $\mathbb{E}_n$. After a slight modification of a definition, we can also represent motions of points (moving particles) either on surfaces or in more general types of spaces.

Note that in differential geometry, we usually prefer to view curves as *functions* so that we can use methods and results of the Calculus. On the other hand, sometimes it is more geometric to work with curves as point subsets, that is, to consider images of parametrized curves, which in fact represent paths, trajectories of motions. Therefore, we will keep both viewpoints here, and use them alternatively according to the purpose.



If a parametrization (motion) $p\colon I \to \mathbb{E}_n$ is given in the Euclidean space, endowed with a fixed coordinate system with the origin O, the corresponding vector function $\boldsymbol{p}(t)$ is introduced by $p(t) = O + \boldsymbol{p}(t)$, or equivalently, by $\boldsymbol{p}(t) = p(t) - O$. Due to the relationship between p and $\boldsymbol{p}$, we can write in components $p(t) = (p^1(t), \dots, p^n(t))$, $t \in I$ (of course, if $\boldsymbol{p}(t) = (p^1(t), \dots, p^n(t))$, i.e. $p^i(t)$ are real functions on I, which are components of the vector function $\boldsymbol{p}$.

Evidently, if $\boldsymbol{p}'(t)$ is defined then it is independent on the coordinate system (particularly of the choice of the origin O); for t fixed, $\boldsymbol{p}'(t)$ is the velocity of the motion and

$$|\boldsymbol{p}'(t)| = \sqrt{\left(\frac{dp^1}{dt}\right)^2 + \dots + \left(\frac{dp^n}{dt}\right)^2} \tag{1.9}$$

is the corresponding speed.

For a given motion $p\colon I \to \mathbb{E}_n$, we define its derivative $p'(t) = (p(t), \boldsymbol{p}'(t))$[9] for every $t \in I$ (similarly for higher order derivatives). We can say that a parametrization p is *of the class* C^r if $\boldsymbol{p}$ is of the class C^r. If $r \geq 1$ and its particular value is not so important we speak about *differentiability* only.

We say that a differentiable parametrization (a motion) p is *regular* if $\boldsymbol{p}'(t) \neq \boldsymbol{o}$, i.e. the velocity vector of a regular motion is a non-zero vector at any point. If $\boldsymbol{p}'(t_0) = 0$ then the point $p(t_0)$ is called a *singular point* of the parametrization p of curve ℓ.

[9] See 2.1.4., tangent vector of a curve is an element of the tangent bundle $TV \approx \mathbb{R}^n \times \mathbb{R}^n$ of $V \approx TV$.

We say that a motion $p\colon I \to \mathbb{E}_n$ is *simple* if $p(t_1) \neq p(t_2)$ whenever $t_1 \neq t_2$ (in a more geometric language, the trajectory of a simple motion does not intersect itself).

A point subset $\ell \subset \mathbb{E}_n$ (particularly, in a space $\mathbb{E}_3$ if $n = 3$) is sometimes called a *simple curve* (of class C^r) if there is a simple regular motion $p\colon I \to \mathbb{E}_n$ (of class C^r), $p(t) = (p^1(t), \dots, p^n(t))$, such that $p(I) = \ell$. We call p a *parametrization* of ℓ. The following can be checked (the proof is based on the Implicit Function Theorem):

Two maps $p(t)\colon I \to \mathbb{E}_n$ and $\tilde{p}(\tau)\colon J \to \mathbb{E}_n$ are parametrizations of the same simple curve $\ell \in \mathbb{E}_n$ (of the class C^r) if and only if there exists a bijective map (of the class C^r) $\varphi\colon J \to I$, $\varphi(\tau) = t$ such that $\tilde{p} = p \circ \varphi$, i.e. $\tilde{p}(\tau) = p(\varphi(\tau))$, and $d\varphi/d\tau \neq 0$ on J. In short, the function $\varphi(\tau) = t$ is written as $t = t(\tau)$, its inverse φ^{-1} as $\tau = \tau(t)$, and each of them is called a *regular transformation of parameter*. From the formula $\tilde{p}(\tau) = p(\varphi(\tau))$ it follows that the corresponding tangent vectors are parallel.

Particularly, two parametrizations of the same simple curve differ up to a strictly monotonous (differentiable) function, called *regular parameter transformation*. A parametrization $p\colon I \to \mathbb{E}_n$, $s \mapsto p(s) = (p^1(s), \dots, p^n(s))$ of a simple curve $\ell = p(I)$ is called *natural* (also *parametrization by arc length s*) if

$$|\dot{p}(s)| = \sqrt{(\dot{p}^1)^2 + \ldots + (\dot{p}^n)^2} = 1 \qquad (1.10)$$

for all $s \in I$[10]. Here and in what follows, in the case of natural parameter, we denote $\dot{p}^i = \frac{dp^i}{ds}$, $i = 1, \dots, n$, etc.

Recall that for arbitrary parametrization $p(t) \in C^1$, arc length s of $p(t)$ is given by the formula

$$s(t) = \int_{t_0}^{t} \sqrt{\left(dp^1(\tau)/d\tau\right)^2 + \ldots + \left(dp^n(\tau)/d\tau\right)^2}\, d\tau, \qquad t_0 \in I. \qquad (1.11)$$

For any simple curve, a natural parametrization always exists and is determined up to parameter transformations of the form $s \mapsto \pm s + C$, $C \in \mathbb{R}$. From the physical point of view, the motion along a naturally parametrized curve has the (constant) unit speed.[11]

Of course, many well-known “curves” are excluded under such a definition of a simple curve: not only Cartesian Knot or Lemniscate of Bernoulli (because of self-intersection) but also circles and ellipses. That is why we introduce a *curve (of the class C^r)* as a subset $\ell \subset \mathbb{E}_n$ such that for any point $A \in \ell$ there exists its neighborhood $U \subset \mathbb{E}_n$ for which $\ell \cap U$ is a simple curve (of the class C^r). Parametrizations of the intersections $\ell \cap U$ are *local parametrizations* of the curve ℓ.

If $p\colon I \to \mathbb{E}_n$ is a local parametrization of a curve $\ell \subset \mathbb{E}_n$ we say that a straight line determined in $\mathbb{E}_n$ by the point $p(t_0)$ and the vector $p'(t_0)$, $t_0 \in I$, is a *tangent* of ℓ in the point $p(t_0)$.

---

[10] We identify $p \equiv \boldsymbol{p}$.

[11] Moreover, from (1.10) and (1.11) it follows, that for $s_1 < s_2$ the difference $(s_2 - s_1)$ is length of arc between $p(s_1)$ and $p(s_2)$. Evidently, the natural parameter s measures the length of the arc of the curve and hence it is called the *arc*.

For two intersecting curves, we introduce a relation which appears to be a useful tool in what follows. Among others, it helps us to describe a tangent and introduce the notion of an inflex point in a very elegant way.

Let $\ell, \overline{\ell} \subset \mathbb{E}_n$ be two curves with a common point. Suppose they are given by local parametrizations $p(t)$ and $\overline{p}(t)$ respectively, on the same definition domain I for simplicity, and let the common point corresponds to the same parameter t_0, $Q = p(t_0) = \overline{p}(t_0)$.

We say that ℓ and $\overline{\ell}$ have a *contact of k-th order at a point Q* if all derivatives[12)] up to order k coincide for the parameter t_0,

$$\frac{d^i p(t_0)}{dt^i} = \frac{d^i \overline{p}(t_0)}{dt^i}, \qquad i = 1, \ldots, k. \tag{1.12}$$

We get an equivalence relation "to have contact of k-th order in a point" on the class of curves.[13)]

It can be checked that the tangent of ℓ in the point Q is the unique line that has a contact of 1^{st} order with ℓ in Q, and two curves have a contact of the first order in a common point Q if and only if they have a common tangent in Q.

We say that a point $Q = p(s_0)$ is a *point of inflection*, or an *inflex point* of a curve $\ell \subset \mathbb{E}^n$ given by a local natural parametrization $p(s)$ if the tangent in Q has a contact of second order with ℓ in Q. A point $Q = p(s_0)$ is a point of inflection[14)] if and only if $\ddot{p}(s_0) = 0$. As well known, a (simple) curve each point of which is inflex is (a part of) a line.

1.2.2 Frenet frame and Frenet–Serret formulas

Let $p(s)\colon I \to \mathbb{E}_3$ be a natural parametrization of a space curve of the class (at least) C^2. Let us assume a non-inflex point $p(s)$, $\ddot{p}(s) \neq 0$. Denote by $\boldsymbol{t}(s) = \dot{p}(s)$ the unit tangent vector, $\boldsymbol{t}(s) \cdot \boldsymbol{t}(s) = 1$; we easily check that $\dot{\boldsymbol{t}}(s) \cdot \boldsymbol{t}(s) = 0$, i.e. $\dot{\boldsymbol{t}}(s) = \ddot{p}(s)$ is orthogonal to $\boldsymbol{t}(s)$. The straight line determined by the point $p(s)$ and the vector $\dot{\boldsymbol{t}}(s)$ is called the *principal normal* of the curve at the point $p(s)$. There exists a number $k(s) > 0$ such that

$$\dot{\boldsymbol{t}}(s) = k(s) \cdot \boldsymbol{n}(s) \tag{1.13}$$

holds[15)] where $\boldsymbol{n}(s)$ is the unit vector of the principal normal at $p(s)$ for which

$$\boldsymbol{n}(s) = \frac{\dot{\boldsymbol{t}}(s)}{|\dot{\boldsymbol{t}}(s)|}. \tag{1.14}$$

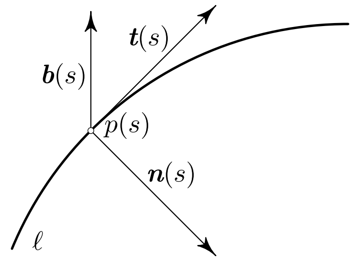
[12)] The definition of the relation is correct: it depends neither on a particular choice of arc length, nor on the fact that we use natural parametrization; it is, in fact sufficient to suppose the existence of some local parametrizations, with the same argument and domain, such that the derivatives in a common point coincide.

[13)] The class of this equivalence relation is called the k-jet; the jet calculus is widely used, see e.g. [109].

[14)] Equivalently, if we assume an arbitrary parametrization $p(t)$, the necessary and sufficient condition for inflex point is: the vectors $p'(t_0)$ and $p''(t_0)$ are collinear (linearly dependent), i.e. they span a 1-dimensional subspace.

[15)] At points of inflection, we define $k = 0$. In this case $\dot{\boldsymbol{t}}(s) = \boldsymbol{o}$.

Further, $\boldsymbol{b}(s) = \boldsymbol{t}(s) \times \boldsymbol{n}(s)$ is the vector field of unit vectors on the *binormals* of the curve. Then $\langle \boldsymbol{t}(s), \boldsymbol{n}(s), \boldsymbol{b}(s) \rangle$ is a (positive) orthonormal basis called the *Frenet frame* or the *moving frame*. Also, it is called *generating Frenet's moving frame* at any point if we take a frame as $\langle p(s); \boldsymbol{t}(s), \boldsymbol{n}(s), \boldsymbol{b}(s) \rangle$.



For a curve $p(s)$ free of points of inflection, the following holds (the so-called *Frenet*, or *Frenet–Serret formulas*[16]):

$$\begin{aligned} \dot{\boldsymbol{t}} &= \phantom{-k\,\boldsymbol{t} +{}} k\,\boldsymbol{n} \\ \dot{\boldsymbol{n}} &= -k\,\boldsymbol{t} \quad + \quad \varkappa\,\boldsymbol{b} \\ \dot{\boldsymbol{b}} &= \phantom{-k\,\boldsymbol{t} +{}} -\varkappa\,\boldsymbol{n} \end{aligned} \tag{1.15}$$

where $k(s)$ and $\varkappa(s)$ are (at least continuous) functions. That is, the first derivatives of (vector) functions forming the Frenet moving frame are expressed by the functions itselves. The Frenet-Serret formulas, in fact, describe the kinematic properties of a particle which moves along a (differentiable) curve in the three-dimensional Euclidean space $\mathbb{E}_3$, more specifically, the formulas describe the derivatives (i.e. the changes during the motion) of the tangent, normal and binormal vectors in terms of each other. At the same time, they introduce two important functions connected with the curve, the so-called the *curvature* $k(s)$ and the *torsion* $\varkappa(s)$ of the curve $p(s)$.

A point $p(s_0)$ is called *planar* if $\varkappa(s_0) = 0$. A curve is contained in some plane if and only if all its points are planar.

For $p(s)$ and $p(t) \in C^2$, the following holds:

$$k(s) = |\ddot{p}| \quad \text{and} \quad k(t) = \frac{|p' \times p''|}{|p'|^3}, \tag{1.16}$$

and for $p(s)$ and $p(t) \in C^3$, the following holds:

$$\varkappa(s) = \frac{(\dot{p}, \ddot{p}, \dddot{p}\,)}{k^2} \quad \text{and} \quad \varkappa(t) = \frac{(p', p'', p''')}{|p' \times p''|^2}. \tag{1.17}$$

Physically, we can think of a space curve in $\mathbb{E}_3$ as being obtained from a straight line by "bending" (curvature) and "twisting" (torsion). The equations

$$k = k(s) > 0, \qquad \varkappa = \varkappa(s) \tag{1.18}$$

are called *natural equations of the curve*.

It can be checked that the local behaviour of the curve can be completely described by k and $\varkappa$. That is, any curve can be given by its natural equations (and is "unique" up to an isometry):

[16] Jean Frédéric Frenet, 1816–1900, described the formulas in his thesis of 1847.
Joseph Alfred Serret, 1819–1885, discovered the same formulas independently in 1851.

Theorem 1.1 (Fundamental Theorem of the local theory of curves, see [4][17]) *Let k, $\varkappa \in C^0$, $k(s) > 0$, $\varkappa(s)$, $s \in I$, be continuous*[18] *functions. Then there exists a regular parametrized curve $p(s)$ defined on I such that s is the arc length, $k(s)$ is the curvature, and $\varkappa(s)$ is the torsion of $p(s)$. Moreover, any other curve $\overline{c}$ satisfying the same conditions differs from c by a rigid motion*[19].

Reconstruction of the curve $p(s)$ of Theorem 1.1 which is defined by the equations (1.17) is the solution of the Frenet-Serret system (1.15) together with equation $\dot{\boldsymbol{p}} = \boldsymbol{t}$ respective unknown vector functions $\boldsymbol{p}(s), \boldsymbol{t}(s), \boldsymbol{n}(s), \boldsymbol{b}(s)$.

It is known, this system of ordinary differential equations has one and only one solution for *initial Cauchy conditions*

$$\boldsymbol{p}(s_0) = (0,0,0),\ \boldsymbol{t}(s_0) = (1,0,0),\ \boldsymbol{n}(s_0) = (0,1,0),\ \boldsymbol{b}(s_0) = (0,0,1). \quad (1.19)$$

1.2.3 Osculating circle, evolute and involute

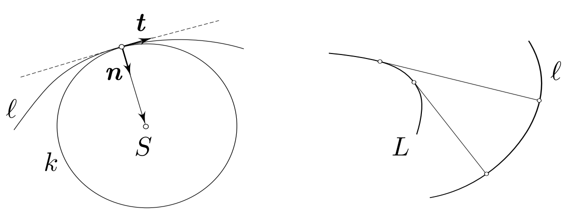
The *osculating circle* k of a curve $\ell \in C^2$ at a given point p on the curve has been traditionally defined as the circle passing through p and a pair of additional points on the curve infinitesimally close to p. Its center S lies on the principal normal line, and its curvature is the same as that of the given curve at that point. This circle, which is the one among all tangent circles at the given point that approaches the curve most tightly, was named circulus osculans (Latin for *kissing circle*) by Leibniz. The osculating circle has the second order contact with a curve.[20]

The center and radius of the osculating circle at a given point are called *center of curvature* and *radius of curvature* of the curve at that point. A geometric construction was described by Isaac Newton in his Principia: *There being given, in any places, the velocity with which a body describes a given figure, by means of forces directed to some common centre: to find that centre.*

A set of curvatures centers of curve ℓ is called *evolute.* If L is evolute of ℓ then ℓ is called *involute* of curve L.[21]

Let $\ell\colon p = p(t)$ and $L\colon p = P(t)$. An equation of evolute L of ℓ has the following form

$$P(t) = p(t) + \frac{1}{k(t)} \cdot \boldsymbol{n}(t). \quad (1.20)$$



[17] Aleksandr Danilovich Aleksandrov, 1912-1999, was a great Russian mathematician.
[18] It can be proven that the assumption is sufficient for the existence of a curve.
[19] That is, there is an orthogonal (linear) transformation R of $\mathbb{E}_3$ with positive determinant and a vector $\boldsymbol{w}$ such that $\overline{\ell} = R \circ \ell + \boldsymbol{w}$.
[20] The point in which exists the third order contact is called a *peak point* of the curve.
[21] There exist alternative definitions of evolute and involute.

1.3 Surfaces in Euclidean space

1.3.1 Surfaces and simple surfaces (patchs)

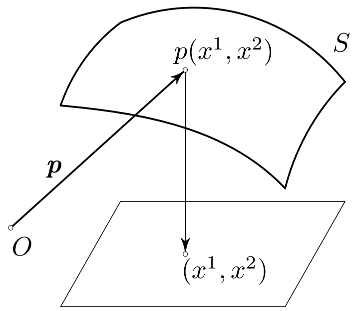
If we wish to parametrize surfaces we meet analogous problems as we have already met for curves since most surfaces in the Euclidean three-space cannot be globally parametrized by a unique function. First, we must choose a suitable collection of "patches" so that each point of the surface belongs to at least one member of them (i.e. we cover a surface by open neighborhoods, with "overlappings"). Then we parametrize each patch "individually" (by an open domain of the plane, that is, by means of two parameters). On a "nice" surface, we can expect "nice transition functions" on overlappings of patches which in fact determine coordinate changes; if the surface is "smooth enough", we get differentiable, or even smooth, transition functions on overlappings. A more precise theory of surfaces can be made if we pass to the concept of a manifold (see below).

Let $D \subset \mathbb{R}^2$ be an open connected subset with coordinates (x^1, x^2), and let

$$p\colon\ D \to \mathbb{E}_3,\ (x^1, x^2) \mapsto p(x^1, x^2) \quad (1.21)$$

be an injective map of the class C^r, $r \geq 1$, from D to $\mathbb{E}_3$. For any point $p(x^1, x^2)$, we can create its radius-vector again,

$$\boldsymbol{p}(x^1, x^2) = p(x^1, x^2) - O \in V_3. \quad (1.22)$$



In this way, we introduce a vector function $\boldsymbol{p}\colon\ D \to V_3$ (for which partial derivatives and C^r-differentiability were already introduced).

Define

$$\boldsymbol{p}_i = \frac{\partial \boldsymbol{p}(x^1, x^2)}{\partial x^i}, \quad i = 1, 2. \quad (1.23)$$

Higher order partial derivatives are introduced by iteration. If the additional requirement (protecting non-existence of singularities) is satisfied: *the vectors* $\boldsymbol{p}_1$ *and* $\boldsymbol{p}_2$ *are linearly independent at any point of* D we say that $S = p(D)$ is a two-dimensional *"surface patch"*, sometimes also called a *simple* or *regular surface* in $\mathbb{E}_3$. D is called the *parameter domain* and the function p is called a *parametrization* of the surface S. Since our further considerations keep local character mostly the notion of a surface patch is quite adequate.

Throughout this subsection and the next section, let S denote a surface in the above sense. Obviously, S can be given by its *vector equation*

$$\boldsymbol{p} = \boldsymbol{p}(x^1,\, x^2). \quad (1.24)$$

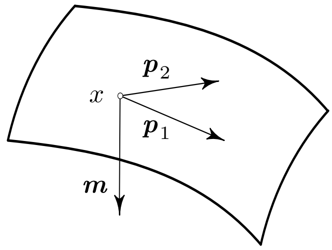
The tangent plane T_xS at the point $x = p(x^1, x^2) \in S$ is an Euclidean plane

$$T_xS = p(x^1, x^2) + \operatorname{span}\{\boldsymbol{p}_1(x^1, x^2), \boldsymbol{p}_2(x^1, x^2)\}. \quad (1.25)$$

There exists a unique perpendicular to $T_x S$ through x, called a *normal* of S at the point x. The vector field

$$\boldsymbol{m} = \frac{\boldsymbol{p}_1 \times \boldsymbol{p}_2}{|\boldsymbol{p}_1 \times \boldsymbol{p}_2|} \tag{1.26}$$

is the so-called *standard unit normal* vector field of the surface patch S.
The frame $\langle \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{m} \rangle$ is assumed to be positive.



1.3.2 First and second fundamental forms of the surfaces

Gauss[22)] introduced the *first* and the *second fundamental form* of the surface S as follows:

$$I = g_{ij}\, dx^i\, dx^j \quad \text{and} \quad II = b_{ij}\, dx^i\, dx^j, \tag{1.27}$$

with

$$g_{ij} = \boldsymbol{p}_i \cdot \boldsymbol{p}_j \quad \text{and} \quad b_{ij} = \boldsymbol{m} \cdot \boldsymbol{p}_{ij}, \tag{1.28}$$

where $\boldsymbol{p}_{ij} \equiv \partial_j \boldsymbol{p}_i \equiv \partial_{ij} \boldsymbol{p} = \dfrac{\partial^2 \boldsymbol{p}}{\partial x^i \partial x^j}$, $i, j = 1, 2$.

In the formula (1.27) and further, we will use the following notation.

Einstein summation convention: In formulas, symbols for sums are omitted, and when the same index, e.g. i, occurs in the same expression twice, once as a superscript and once as a subscript we understand that the expression denotes a sum of members of the same shape but with the index i varying from 1 to n (n either follows from the situation or must be settled). In our case $n = 2$.

1.3.3 Length of curves, angle between curves and area of surfaces

A curve ℓ on a surface S that is given by a parametrization $p\colon\ D \to S \subset \mathbb{E}_3$, can be determined by its *curvolinear* or *inner coordinates* $x^i = x^i(t)$, that is, by some mapping $I \to D$ with components $x^1(t)$, $x^2(t)$.

The *arc length* s of the arc $\widehat{AB}$ of curve $\ell \subset S$ ($A = p(x^i(t_0)); B = p(x^i(t_1))$) is expressed by the integral

$$s = \int_{t_0}^{t_1} \sqrt{g_{ij}(x(t))\, \dot{x}^i(t)\, \dot{x}^j(t)}\ dt. \tag{1.29}$$

Evidently $ds^2 = I$, therefore I and ds^2 are called a *metric of* S, and are often denoted by g. In classical notation, the metric form reads

$$ds^2 = g_{11}\, {dx^1}^2 + 2g_{12}\, dx^1 dx^2 + g_{22}\, {dx^2}^2 = \sum_{i,j=1}^{2} g_{ij}\, dx^i dx^j = g_{ij}\, dx^i dx^j \tag{1.30}$$

which turns S into a Riemannian space $V_2 = (S, g) = (S, ds^2)$, see p. 137.

22) Johann Karl Friedrich Gauss, 1777–1855, was a German scientist and mathematician who contributed to many fields, including analysis, statistics (Gaussian distribution), geodesy and differential geometry (*Gaussian curvature, Theorema Egregium*), number theory (an author of *Disquisitiones Arithmeticae*, finished 1798, published 1801), optics, astronomy, electrostatics.

Let $D \subset \mathbb{R}^2$ be an open domain in S. Under an *area* of the domain $p(D) \subset S$ we understand
$$S = \iint_D \sqrt{g_{11}g_{22} - g_{12}^2}\ dx^1 dx^2. \tag{1.31}$$

Let ℓ_1 and ℓ_2 be curves on surface S which intersect at the point $p_0 = \ell_1 \cap \ell_2$ and which have at p_0 tangent vectors $\boldsymbol{u} = u^\alpha \boldsymbol{p}_\alpha$ and $\boldsymbol{v} = v^\alpha \boldsymbol{p}_\alpha$, respectively. An *angle* α between ℓ_1 and ℓ_2 at p_0 can be calculated from the following formula

$$\cos\alpha = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{|\boldsymbol{u}|\cdot|\boldsymbol{v}|} = \frac{g_{ij}u^i v^j}{\sqrt{g_{ij}u^i u^j}\cdot\sqrt{g_{ij}v^i v^j}}\ . \tag{1.32}$$

The angle between coordinate curves $(x^1 = t, x^2 = x_0^2)$ and $(x^1 = x_0^1, x^2 = t)$ are expressed
$$\cos\alpha = \frac{g_{12}}{\sqrt{g_{11}\cdot g_{22}}}\ . \tag{1.33}$$

From this follows that a criteria of *orthogonal coordinate system* is $g_{12} = 0$.

On arbitrary surface S, it is possible to locally introduce a special orthogonal coordinate system, which is called an *isothermal coordinate system* [70, p. 128], in which the first quadratic form is expressed

$$ds^2 = f(x^1, x^2)\,({dx^1}^2 + {dx^2}^2), \tag{1.34}$$

where $f(x^1, x^2)$ is a function.

Interesting Proof. Let surface S has the first quadratic form $ds^2 = g_{ij}dx^i dx^j$, thus
$$ds^2 = \left(\sqrt{g_{11}}\ dx^1 + \frac{g_{12} + i\cdot G}{\sqrt{g_{11}}}\ dx^2\right)\cdot\left(\sqrt{g_{11}}\ dx^1 + \frac{g_{12} - i\cdot G}{\sqrt{g_{11}}}\ dx^2\right), \tag{1.35}$$
where $G = \sqrt{g_{11}g_{22} - g_{12}^2}$ and i is the complex unit. There exists[23] integrability multiplicator ϱ for $dw = \varrho\left(\sqrt{g_{11}}\ dx^1 + \frac{g_{12} + i\cdot G}{\sqrt{g_{11}}}dx^2\right)$. Obviously, ϱ and w are complex functions of (x^1, x^2), where $w = u + i\cdot v$. From (1.35) it follows
$$ds^2 = \frac{dw}{\varrho}\cdot\frac{d\overline{w}}{\overline{\varrho}} = \frac{1}{\varrho\cdot\overline{\varrho}}\ (du^2 + dv^2) = f(u,v)\,(du^2 + dv^2).^{24)} \qquad \square$$

From the last formula it follows that any surface is locally conformal to Euclidean plane, see p. 272. Gaussian curvature, see p. 38, in isothermal coordinate system has the following form $K = -\frac{\Delta \ln f}{f}$, where Δ is Laplacian.[25]

---

[23] We find the solution of the formula $dw = \varrho(Adx + Bdy)$ where $A(x,y)$, $B(x,y)$ are given functions and $w(x,y)$, $\varrho(x,y)$ are unknown complex functions. Evidently, $dw = w_x\cdot dx + w_y dy$, therefore it holds
$$w_x = \varrho A \quad \text{and} \quad w_y = \varrho B. \tag{1.36}$$
We found $\varrho = \frac{w_x}{A} = \frac{w_y}{B}$ (the case $A = 0$ or $B = 0$ is trivial). Then we obtain
$$w_x B - w_y A = 0. \tag{1.37}$$
The condition (1.37) is a homogenous linear partial differential equation respective unknown function w for which there exist standard methods of solution, see [99, 100].

Moreover, if $g_{ij}, A, B \in C^r$ $(r \geq 1)$ then exist $w \in C^r$ and $\varrho \in C^{r-1}$, therefore $f \in C^{r-1}$.

[24] We obtain complex coordinates $z \in \mathbb{C}$ on S for which $S\colon p = p(z)$ and $ds^2 = f(z)\,dz\,d\overline{z}$.

[25] $\Delta = \partial_{uu} + \partial_{vv}$

1.3.4 Gauss, Weingarten and Peterson-Codazzi formulas

For surface $S \in C^2$ Gauss and Weingarten[26] proved

$$\begin{aligned} \boldsymbol{p}_{ij} &= \Gamma^k_{ij}\boldsymbol{p}_k + b_{ij}\cdot \boldsymbol{m} && - \textit{Gauss formula},\\ \boldsymbol{m}_i &= -\,b^k_i\boldsymbol{p}_k && - \textit{Weingarten formula}, \end{aligned} \tag{1.38}$$

where $b^k_i = b_{ij}g^{jk}$, g^{ij} are components of dual tensor of g, i.e. $\|g^{ij}\| = \|g_{ij}\|^{-1}$,

$$\Gamma_{ijk} = \tfrac{1}{2}\,(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}) \quad \text{and} \quad \Gamma^h_{ij} = g^{hk}\Gamma_{ijk} \tag{1.39}$$

are *Christoffel*[27] *symbols* of the first and the second types.

For surface $S \in C^3$ from the integrability condition of PDE's (1.38) with respect to unknown $\boldsymbol{p}_i$ and $\boldsymbol{m}$, which are $\partial_k\boldsymbol{p}_{ij} = \partial_j\boldsymbol{p}_{ik}$ and $\partial_j\boldsymbol{m}_i = \partial_i\boldsymbol{m}_j$, Gauss and Peterson[28], Mainardi[29], Codazzi[30] proved

$$\begin{gathered} R_{1212} = b_{11}b_{22} - b^2_{12} \;-\; \textit{Gauss formula},\\ \partial_2 b_{i1} + \Gamma^1_{i1}b_{12} + \Gamma^2_{i1}b_{22} = \partial_1 b_{i2} + \Gamma^1_{i2}b_{11} + \Gamma^2_{i2}b_{21} - \textit{Peterson-Codazzi formulas}^{31)}, \end{gathered} \tag{1.40}$$

$$\text{where} \quad R_{hijk} = g_{h\alpha}R^\alpha_{ijk} \text{ and } R^h_{ijk} = \partial_j\Gamma^h_{ik} - \partial_k\Gamma^h_{ij} + \Gamma^\alpha_{ik}\Gamma^h_{\alpha j} - \Gamma^\alpha_{ij}\Gamma^h_{\alpha k} \tag{1.41}$$

are the components of Riemannian tensors of the first and the second kind.

From Gauss and Weingarten formulas (1.38) it elementary follows that surfaces which have identical first and second forms in common domain are isometric, i.e. the same with precision to the location in the space $\mathbb{E}^3$.

Moreover, the formulas (1.38) with (1.40) forms full integrability Cauchy type PDE's respective unknown vector functions $\boldsymbol{p}, \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{m}$ which have only one solution for initial conditions, for example, $\boldsymbol{p}(x_0) = \boldsymbol{o}$, and for which $\boldsymbol{p}_i(x_0)\cdot\boldsymbol{p}_j(x_0) = g_{ij}(x_0)$, $\boldsymbol{p}_{ij}(x_0)\cdot\boldsymbol{m}(x_0) = b_{ij}(x_0)$. Then it follows

Theorem 1.2 (*Bonnet theorem*[32] on the existence and the uniqueness of a surface with given first and second fundamental forms, [36, 380])

Let the following two quadratic forms be given:

$$g_{11}\,dx^{1^2} + 2g_{12}\,dx^1dx^2 + g_{22}\,dx^{2^2} \quad \textit{and} \quad b_{11}\,dx^{1^2} + 2b_{12}\,dx^1dx^2 + b_{22}\,dx^{2^2}$$

the first one of which is positive definite, and let the coefficients of these forms satisfy the Gauss equation and the Peterson-Codazzi equations (1.40)*. Then there exists a surface, which is unique up to motions in space, for which these forms are the first and the second fundamental forms, respectively.*

---

[26] Julius Weingarten, 1836–1910, was a German mathematician.

[27] Elwin Bruno Christoffel, 1829–1900, was a German mathematician and physicist.

[28] Karl Mikhailovich Peterson, 1828–1881, was a Russian mathematician. Peterson gave, in his graduation dissertation (1853), but not published until later, in Derpt University (now Tartu, Estonia). In 1879, the University of Odessa awarded him an honorary degree.

[29] Gaspare Mainardi, 1800–1879, was an Italian mathematician.

[30] Delfino Codazzi, 1824–1873, was an Italian mathematician.

[31] The equations were first discovered by Peterson in 1853 and were rediscovered by Mainardi in 1856 and Codazzi in 1867, see [176].

[32] Pierre Ossian Bonnet, 1819–1892, was a French mathematician. He made some important contributions to the differential geometry of surfaces, including the Gauss-Bonnet theorem.

The question whether does exist a surface for which the positive form ${g_{11}\,dx^1}^2 + 2g_{12}\,dx^1 dx^2 + g_{22}\,{dx^2}^2$ is the first quadratic form, was studied in 1920s by É. Cartan and M. Janet for real analytic functions, and by J.F. Nash[33] for n-dimensional Riemannian spaces, see [852] and [225, 531, 532].

Remark on the Isometric Embedding PDE. Let $p\colon \mathbb{R}^2 \to S \subset \mathbb{R}^3$. Saying that p is an isometric embedding amounts to a system of fully nonlinear 1st-order PDE for $p = \boldsymbol{p}(x^1, x^2) = (p^i(x^1, x^2))$. Namely, if (x^1, x^2) are local coordinates for S and $g = g_{ij}(x^1, x^2)\,dx^i\,dx^j$, then p is an isometric embedding if and only if

$$g_{ij}(x^1, x^2) = \frac{\partial \boldsymbol{p}}{\partial x^i} \cdot \frac{\partial \boldsymbol{p}}{\partial x^j}, \quad i, j = 1, 2. \tag{1.42}$$

This is a system of 3 equations for the 3 unknown components of $\boldsymbol{p} = (p^1, p^2, p^3)$.

1.3.5 Isometry and Inner geometry

Recall that *isometries* are diffeomorphisms[34] $S \to \overline{S}$ which map curves in S to curves of the same length in $\overline{S}$ (i.e. lengths of all curvolinear segments are preserved).

It is obvious that a diffeomorphism $f\colon\ S \to \overline{S}$ is an isometry if and only if for any surface patch σ on S, the patches σ and $f \circ \sigma$ on S and $\overline{S}$, respectively, have the same first fundamental form, i.e. $d\overline{s}^2 = ds^2$ (if we use the "common" coordinates x on the image $f(M)$ as on the preimage M, the metric on the image under the isometry f has the same components $\overline{g}_{ij}(x) = g_{ij}(x)$), [173, p. 101].

The *inner geometry* (also *intrinsic geometry*) of a surface includes all properties or objects of the given surface which are derived only from the first fundamental form (in practice, from components of the metric tensor g and their derivatives), i.e. just those properties which are invariant under isometries.

Since the metric $g = (g_{ij})$ belongs to the intrinsic geometry, components of the dual tensor $g^* = (g^{ij})$ also belongs to the intristic geometry as well a *discriminant tensor* ε and a *structure tensor* F defined by relations[35]

$$\varepsilon_{ij} = \sqrt{g_{11}g_{22} - g_{12}^2} \cdot \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \text{and} \qquad F_i^h = \varepsilon_{ij} \cdot g^{jh}. \tag{1.43}$$

It can be easily proved that the vectors $\boldsymbol{v} = v^h \boldsymbol{p}_h$ and $F\boldsymbol{v} = F_i^h v^i \boldsymbol{p}_h$ are orthogonal, and, moreover, they are of the same length $|F\boldsymbol{v}| = |\boldsymbol{v}|$.

The Christoffel symbols (1.39) of the surfaces $S \in C^2$ belong to the intrinsic geometry, and for surfaces $S \in C^3$ the Riemannian tensors (1.41), and Gaussian curvature K (*Gauss Theorema Egregium*, p. 38) belong to the intrinsic geometry as well.

---

[33] John Forbes Nash, 1928–2015, was an American mathematician who made fundamental contributions to game theory, differential geometry, and the study of partial differential equations.

[34] A diffeomorphism is a homeomorphism which is differentiable together with its inverse, see 2.1.3.

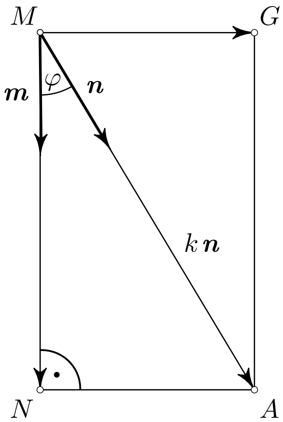
[35] The tensor ε is skew-symmetric and the tensor F defines a complex structure, for which it holds $F^2 = -Id$. In addition, the metric tensor g, the dual tensor g^*, the discriminant tensor ε, and the structure tensor F are covariantly constant.

1.3.6 Normal, geodesic, principal, mean and Gauss curvatures

Let a curve $\ell\colon\ p = p(s)$ in a surface $S \subset \mathbb{E}_3$ be given by its inner coordinates

$$x^i = x^i(s), \qquad i = 1, 2, \tag{1.44}$$

where s is a natural parameter (arc length).



Consider a (non-singular) point $M = p(s)$ of ℓ, the vector $k\,\boldsymbol{n}$ located at M, and denote by A the endpoint of the located vector, i.e. $\overrightarrow{MA} = k\,\boldsymbol{n}$ (k is curvature and $\boldsymbol{n}$ unit principal normal of ℓ, see (1.13)); sometimes, $\overrightarrow{MA}$ is called the *curvature vector* of ℓ at M.

Let $\overrightarrow{MN}$ be the projection of the vector $\overrightarrow{MA}$ onto a normal of the surface S at M, and similarly let $\overrightarrow{MG}$ be the projection of the vector $\overrightarrow{MA}$ onto the tangent plane τ of S at M. Obviously,

$$k\,\boldsymbol{n} = \overrightarrow{MA} = \overrightarrow{MN} + \overrightarrow{MG}; \tag{1.45}$$

$\overrightarrow{MN}$ is called the *normal curvature vector*,
$\overrightarrow{MG}$ the *geodesic curvature vector* of ℓ at M, and their lengths

$$|k_n| = |\overrightarrow{MN}| \quad \text{and} \quad k_g = |\overrightarrow{MG}|, \tag{1.46}$$

respectively are called the *normal curvature* and the *geodesic curvature* of ℓ on S, respectively. From the technical reasons, at a non-flat point, k_n itself is usually considered with a sign: $k_n = |\overrightarrow{MN}| > 0$ (with "plus") if the vectors $\overrightarrow{MN}$ and $\boldsymbol{m}$ have the same direction (i.e., $\overrightarrow{MN}$ is a positive multiple of $\boldsymbol{m}$) while $k_n = -|\overrightarrow{MN}| < 0$ (with "minus") if the vectors $\overrightarrow{MN}$ and $\boldsymbol{m}$ have opposite directions (i.e., $\overrightarrow{MN}$ is a negative multiple of $\boldsymbol{m}$).

Evidently, the *Meusnier formula* $k_n = k \cdot \cos\varphi$ holds, and from (1.46) we get

$$k^2 = k_n^2 + k_g^2. \tag{1.47}$$

Let us pass to coordinate expressions of the above concepts. If a surface S is given by (1.21) and a curve on S is given by its curvolinear coordinates (1.24) then the curve is described by its "vector equation" which is often written as

$$\boldsymbol{p}(s) = \boldsymbol{p}(x^1(s), x^2(s)).$$

We find

$$\boldsymbol{t} = \frac{d\boldsymbol{p}}{ds} = \frac{\partial \boldsymbol{p}}{\partial x^1}\frac{dx^1}{ds} + \frac{\partial \boldsymbol{p}}{\partial x^2}\frac{dx^2}{ds} = \boldsymbol{p}_i\,\frac{dx^i}{ds}$$

and

$$\frac{d\boldsymbol{t}}{ds} = \frac{d^2\boldsymbol{p}}{ds^2} = \boldsymbol{p}_{ij}\,\frac{dx^i}{ds}\frac{dx^j}{ds} + \boldsymbol{p}_i\,\frac{d^2x^i}{ds^2}.$$

Since $\dfrac{d\boldsymbol{t}}{ds} = k\,\boldsymbol{n}$ then from previous and the Gauss equations (1.38) it follows

$$k\,\boldsymbol{n} = \left(\frac{d^2x^h}{ds^2} + \Gamma^h_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds}\right)\boldsymbol{p}_h + b_{ij}\,\frac{dx^i}{ds}\frac{dx^j}{ds}\,\boldsymbol{m}. \tag{1.48}$$

So we have obtained a decomposition of the curvature vector with respect to the basis $\langle \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{m}\rangle$. Comparing (1.43) and (1.48) we conclude

$$\overrightarrow{MN} = b_{ij}\,\frac{dx^i}{ds}\frac{dx^j}{ds}\,\boldsymbol{m}, \quad \text{and} \quad \overrightarrow{MG} = \left(\frac{d^2x^h}{ds^2} + \Gamma^h_{ij}\,\frac{dx^i}{ds}\frac{dx^j}{ds}\right)\boldsymbol{p}_h. \tag{1.49}$$

According to (1.49), the normal curvature (i.e. the length of the vector $\overrightarrow{MN}$ equipped with a sign) is given by the formula

$$k_n = b_{ij}\,\frac{dx^i}{ds}\frac{dx^j}{ds} = \frac{b_{ij}\,dx^i dx^j}{ds^2} = \frac{b_{ij}\,dx^i dx^j}{g_{ij}\,dx^i dx^j} = \frac{II}{I} \tag{1.50}$$

well-known from basic courses of differential geometry. From (1.50) it follows that k_n depends only at the point M and the direction (dx^1, dx^2).

On the other hand, k_n is a curvature of a normal section at the point M on the surface S. A normal section of the surface S at the point M is a plane curve which is an intersection of the a normal plane (at the point M and contains the normal vector) and the surface S.

In our further considerations, the geodesic curvature will play the more important role. The formula for k_g can be obtained e.g. if we calculate scalar quadrat of (1.49). The expression reads

$$k_g^2 = \overrightarrow{MG}^2 = \left(\frac{d^2x^\alpha}{ds^2} + \Gamma^\alpha_{ij}\,\frac{dx^i}{ds}\frac{dx^j}{ds}\right)\left(\frac{d^2x^\beta}{ds^2} + \Gamma^\beta_{pq}\,\frac{dx^p}{ds}\frac{dx^q}{ds}\right)g_{\alpha\beta}. \tag{1.51}$$

Note that the geodesic curvature k_g belongs to the intrinsic geometry of the surface, which is almost obvious from the formula (1.50), and hence it is invariant under an isometric deformation (by a “bending”) of the surface.

On the other hand, depending also on the second fundamental form of the surface (on components b_{ij}), k_n does not belong to the intrinsic geometry of the surface, it can be changed under isometric deformations.

Given a fixed point Q of a surface S and a fixed line t tangent to S at Q (i.e. lying in the tangent space to S at Q), there are infinitely many curves on S passing through Q and having t as its tangent at Q. Curves of this family have, in general, different curvatures k, but it appears that they have the same normal curvature k_n, see the comments of the formula (1.50). And the normal section is uniquely determined by the line t and by the normal of the surface at Q, hence is the same for all curves under consideration.

The maximum and minimum values of the normal curvature k_n at a point M on a surface are called the *principal curvatures* k_1 and k_2, and their corresponding directions are called *principal.* Euler[36)] proved the following formula

$$k_n = k_1\,\cos^2\alpha + k_2\,\sin^2\alpha\,, \tag{1.52}$$

where α is an angle between the direction corresponding to k_1 and k_n.

[36)] Leonhard Euler, 1707-1783, was a greate Swiss mathematician and physicist, who was tutored by Johann I. Bernoulli. He spent most of his adult life in Saint Petersburg, Russia, and in Berlin, then the capital of Prussia. He was the chairman of Imperial Russian Academy of Sciences, and Berlin Academy.

The principal curvatures k_1 and k_2 are the solutions of characteristic equations

$$k_n^2 - 2H\, k_n + K = 0 \tag{1.53}$$

where

$$H = \frac{1}{2}\,(k_1 + k_2) \quad \text{and} \quad K = k_1 \cdot k_2 \tag{1.54}$$

are the *mean* and the *Gaussian curvature*[37)], respectively. As it is known, these curvatures are expressed in the following form

$$H = \frac{g_{11}b_{22} - 2g_{12}b_{12} + g_{22}b_{11}}{2\,(g_{11}g_{22} - g_{12}^2)} \quad \text{and} \quad K = \frac{b_{11}b_{22} - b_{12}^2}{g_{11}g_{22} - g_{12}^2}\ . \tag{1.55}$$

From the Euler formula (1.52) it follows

$$\frac{1}{2}\,(k_n(\alpha) + k_n(\alpha + \pi/2)) = H \quad \text{and} \quad \frac{1}{2\pi}\int_0^{2\pi} k_n(\alpha)\, d\alpha = H, \tag{1.56}$$

* in case $k_1 = k_2$ it holds $k_n = k_1 = k_2$ (these points are called *umbilical points* and, evidently, all directions there are principal), and
* in case $k_1 \neq k_2$ (*non umbilical points*) there exist just two orthogonal principal directions (corresponding to k_1 and k_2).

From (1.40) for $S \in C^3$ it follows the

Gauss Theorema Egregium *The Gaussian curvature K belongs to inner geometry, and it holds*

$$K = \frac{R_{1212}}{g_{11}g_{22} - g_{12}^2}\ . \tag{1.57}$$

Based on the sign of K, it was introduced the points classification: positive – *elliptic*, negative – *hyperbolic*, and zero – *parabolic*, moreover if $b_{ij} = 0$ – *planar*.

The principal directions $d\boldsymbol{p} = \boldsymbol{p}_i\, dx^i$ can be calculated from differential equations

$$\begin{vmatrix} {dx^2}^2 & -dx^1 dx^2 & {dx^1}^2 \\ g_{11} & g_{12} & g_{22} \\ b_{11} & b_{12} & b_{22} \end{vmatrix} = 0. \tag{1.58}$$

Direction $d\boldsymbol{p}$ is principal if and only if *Rodrigues' formula* holds[38)]

$$d\boldsymbol{m} = -k_n\, d\boldsymbol{p}. \tag{1.59}$$

[37)] The Gaussian curvature K is often called *total*, and it was introduced by Euler in 1760, see [71, p. 162].

[38)] Benjamin Olinde Rodrigues, 1795–1851, was a French banker, mathematician, and social reformer. Rodrigues was awarded a doctorate in mathematics on 28 June 1815 by the University of Paris. His dissertation contains the result now called *Rodrigues' formula.*

The *lines of curvature* or *curvature lines* are curves which are always tangent to a principal direction (they are integral curves for the principal direction fields). Monge[39] proved that a curve is line of curvature if and only if the nearly normals of the surface lay in one plane.

The orthogonal coordinate system coordinate lines are curvature lines is called *principal coordinates.* Criterium of principal coordinates is

$$g_{12} = b_{12} = 0. \tag{1.60}$$

An *asymptotic direction* is one in which the normal curvature k_n is zero. An *asymptotic curve* is a curve which is always tangent to an asymptotic direction of the surface (where they exist). It is sometimes called an *asymptotic line*, although it need not be a line.

From (1.50) equations of asymptotic directions and curves follow

$$b_{ij}\,dx^i\,dx^j = 0, \quad \text{i.e.} \quad b_{11}\,{dx^1}^2 + 2b_{12}\,dx^1\,dx^2 + b_{22}\,{dx^2}^2 = 0. \tag{1.61}$$

Asymptotic directions can only occur when the Gaussian curvature is negative (or zero). The direction of the asymptotic direction are the same as the asymptotes of the hyperbola of the Dupin indicatrix[40].

There are two asymptotic directions through every point with negative Gaussian curvature, bisected by the principal directions. If the surface is minimal, the asymptotic directions are orthogonal to one another.

In this case the coordinate system whose coordinate lines are asymptotic lines is called *asymptotic coordinates.* Criterium of these coordinates is $b_{11} = b_{22} = 0$.

A *minimal surface* is a surface that locally minimizes its area, i.e. it is a surface with the smallest area of all surfaces whose border is the given curve. This is equivalent to having zero mean curvature H.[41]

---

[39] Gaspard Monge, 1746-1818, was a French mathematician, the inventor of descriptive geometry (the mathematical basis of technical drawing), and the father of differential geometry. During the French Revolution he served as the Minister of the Marine.

[40] Baron Pierre Charles François Dupin, 1784–1873, was a French mathematician, engineer, economist and politician.

[41] This is equivalent of following equation $(1+\boldsymbol{p}_1^2)\,\boldsymbol{p}_{22} - 2\boldsymbol{p}_1\boldsymbol{p}_2\,\boldsymbol{p}_{12} + (1+\boldsymbol{p}_2^2)\,\boldsymbol{p}_{11} = 0.$ The above partial differential equation was originally found in 1762 by Lagrange[42] [689] and Meusnier[43] discovered in 1776 that it implied a vanishing mean curvature [746].

[42] Joseph Louis Lagrange, 1736–1813, was an Italian and French greatest Enlightenment Era mathematician and astronomer. He made significant contributions to the fields of analysis, number theory, and both classical and celestial mechanics. The author of *Mécanique analytique.*

[43] Jean Baptiste Marie Charles Meusnier de la Place, 1754–1793, was a French mathematician, engineer and Revolutionary general. He is best known for *Meusnier's theorem on the curvature of surfaces.* He also discovered the helicoid. He worked with Lavoisier on the decomposition of water and the evolution of hydrogen.

Evidently, a plane is a minimal surface. Let us state other examples of the minimal surfaces

$$\textit{helicoid}:\ x = u\cos v,\ y = u\sin v,\ z = a\,v, \qquad ds^2 = du^2 + (a^2+u^2)\,dv^2, \quad (1.62)$$

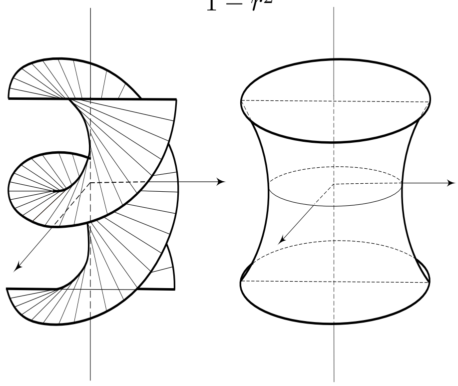
$$\textit{catenoid}:\ x = r\cos\varphi,\ y = r\sin\varphi,\ z = f(r), \quad ds^2 = r^2 d\varphi^2 + \frac{r^2}{1-r^2}\,dr^2, \quad (1.63)$$

where $f(r) = a\ln(\operatorname{th}\frac{r}{2a})$.

The mapping between the helicoid and the catenoid characterized by the relation

$$r = \sqrt{u^2+a^2},\ \varphi = v,$$

is isometric, therefore helicoid and catenoid are (locally) isometric.



Let us note an interesting fact that helicoid and catenoid are the only minimal surfaces that are ruled (also called a scroll) surface and surface of revolution, respectively.

Let the surface S be defined by equations $z = f(x,y)$. Then

$$\boldsymbol{p}_1 = (1,0,f_x),\ \boldsymbol{p}_2 = (0,1,f_y),\ \boldsymbol{p}_{11} = (0,0,f_{xx}),\ \boldsymbol{p}_{12} = (0,0,f_{xy}),\ \boldsymbol{p}_{22} = (0,0,f_{yy})$$

$$\boldsymbol{m} = \frac{1}{\sqrt{1+f_x^2+f_y^2}}\cdot(-f_x,-f_y,1) = \frac{\overrightarrow{\nabla}(z-f)}{\left|\overrightarrow{\nabla}(z-f)\right|}$$

$$g_{11} = 1+f_x^2,\quad g_{12} = f_x f_y,\quad g_{22} = 1+f_y^2,$$

$$b_{11} = \frac{f_{xx}}{\sqrt{1+f_x^2+f_y^2}},\quad b_{12} = \frac{f_{xy}}{\sqrt{1+f_x^2+f_y^2}},\quad b_{22} = \frac{f_{yy}}{\sqrt{1+f_x^2+f_y^2}}$$

$$I = (1+f_x^2)dx^2 + 2f_x f_y dxdy + (1+f_y^2)dy^2\ ,$$

$$II = \frac{1}{\sqrt{1+f_x^2+f_y^2}}\ \left(f_{xx}\,dx^2 + 2f_{xy}\,dxdy + f_{yy}\,dy^2\right),$$

$$S = \iint_D \sqrt{1+f_x^2+f_y^2}\ \ dx\,dy.$$

$$H = \frac{(1+f_y^2)f_{xx} - 2f_x f_y f_{xy} + (1+f_x^2)f_{yy}}{2\,(1+f_x^2+f_y^2)^{3/2}}, \qquad K = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1+f_x^2+f_y^2)^{3/2}}$$

1.4 Geodesics in surfaces

Among all curves passing through Q, we can distinguish those for which the geodesic curvature vanishes. The class of all curves with $k_g = 0$ appears to be an interesting one; we will examine existence (and cardinality) below.

Definition 1.1 A curve on a surface is called a *geodesic* if its geodesic curvature is zero everywhere.

1.4.1 Characterization of geodesics

The sense of the definition is as follows. Given a point of the surface S and a fixed direction (given by a fixed tangent vector), from all curves passing through the point in the given direction we distinguish those for which the curvature (1.46) is a minimal one, since the summand k_n is common for all such curves, it is given by the shape of the surface, and cannot be "deleted". Geometrically speaking, from all curves on the surface, geodesics are most "straight" ones.

Since $k_g = 0$ if and only if $\overrightarrow{MG} = 0$, and $\overrightarrow{MG}$ is a vector projection of $k\boldsymbol{n}$ onto the tangent plane, it follows that $\overrightarrow{MG}$ can vanish only in two cases:

$$(1) \quad k\,\boldsymbol{n} \parallel \boldsymbol{m} \qquad\qquad (2) \quad k\,\boldsymbol{n} = 0, \quad \text{i.e. } k = 0. \tag{1.64}$$

The second possibility indicates that the point is inflex.

Theorem 1.3 *A curve in a surface is a geodesic if and only if for each of its points the following is satisfied: either the principal normal coincides with the normal of the surface at the point (i.e. is perpendicular to the tangent plane), or the point is an inflex one.*

The most familiar examples are the straight lines in Euclidean geometry; on a sphere, the geodesics are the great circles.

Any part of a straight line on a surface is a geodesic according to (2). Consequently, straight lines on surfaces are always geodesics, e.g. straight lines in a (part of a) plane, on quadratic surfaces as a conus, cylinder, hyperboloid of revolution or hyperbolic paraboloid, to mention the most famous examples.

Great circles on a sphere can serve as examples of geodesics satisfying the above condition (1) since their principal normals pass through the center, hence coincide with normals of the sphere. The shortest path from point A to point B on a sphere is given by the shorter piece of the great circle passing through A and B. If A and B are antipodal points (like the North pole and the South pole), then there are infinitely many shortest paths between them. Such an ambivalent situation arises on a cylinder as well.

Recall that k_g belongs to the intrinsic geometry of a surface S, and does not change under isometric deformations (which naturally remains true even in the particular case $k_g = 0$). It means that the concept of geodesic lines belongs to the intrinsic geometry of a surface, and consequently, under isometric deformations, geodesics are transformed again into geodesics (which holds neither for asymptotic curves nor for lines of curvature on surfaces, see p. 39).

It is not difficult to give differential equations of geodesics. If we take into account the formula (1.49) it is clear that $\overrightarrow{MG} = 0$ for a geodesic if and only if

$$\frac{d^2x^h}{ds^2} + \Gamma^h_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds} = 0, \qquad h = 1, 2. \tag{1.65}$$

That is, we obtained a system of two ordinary differential equations of second order of Cauchy type for functions $x^1(s)$, $x^2(s)$ of argument s. The second derivatives of the functions $x^h(s)$ under consideration are expressed, according to (1.65), via the functions themselves (they are composed with Γ^h_{ij}) and their first derivatives.[44)]

Analogously, an n-surface in $\mathbb{E}_{n+1}$ can be defined, [215, p. 16], and a *geodesic* on a n-surface can be introduced as a parametrized curve $c\colon I \to S$ whose acceleration is everywhere orthogonal to S, it has no component of acceleration tangent to the surface, [215, p. 40]. That is, acceleration serves only to keep it in the surface. Roughly, a geodesic is a curve in S which always goes straight ahead in the surface.

1.4.2 Existence and uniqueness of geodesics

Intuitively, it seems that given any point p in an n-surface S and any initial velocity $\lambda_p \in T_pS$ there should be a geodesic in S passing through p with initial velocity λ_p. The following theorem states that this is in fact the case, and that the geodesic satisfying the "initial data" is essentially unique. It is in fact a variant of the Frobenius Theorem. More precisely:

Theorem 1.4 (The existence and uniqueness theorem for geodesics)
Let $S \in C^3$ be an surface in $\mathbb{E}_3$ (or an n-surface in $\mathbb{E}_{n+1}$). Let $p \in S$ be a point and $\lambda_p \in T_pS$ a tangent vector at p. Then there exists a uniquely determined geodesic γ which passes through this point p with the tangent vector λ_p at p.

The geodesic ℓ on I is called the *maximal geodesic* in S passing through p with initial velocity λ_p, I is the maximal open interval in $\mathbb{R}$ (containing 0) on which a geodesic with the given initial data can be defined (in general, I is a proper subset of $\mathbb{R}$).

Proof of this theorem follows from the theory of ordinary differential equations, by noticing that the geodesic equation is a second-order ODEs. Existence and uniqueness then follow from the Picard-Lindelöf theorem for the solutions of ODEs with the prescribed initial data; [90, pp. 32–33].

Let us denote $\lambda^h(s) = dx^h(s)/ds$, then the system (1.65) can be written in the form of first order system of ODEs

$$\frac{dx^h(s)}{ds} = \lambda^h(s), \qquad \frac{d\lambda^h(s)}{ds} = -\Gamma^h_{ij}(x(s))\lambda^i(s)\lambda^j(s). \tag{1.66}$$

That is, according to the existence theorem for differential equations of the given

[44)] Remark that the formula (1.65) for geodesics was discovered by Christoffel in 1868.

type, the system (1.66) admits (locally) a unique solution[45)] with initial data

$$x^h(s_0) = x_0^h, \qquad \lambda^h(s_0) = \lambda_0^h, \quad h = 1, \dots, n. \tag{1.67}$$

But on the geometric language, the initial data[46)] (1.67) mean that for any point $p_0 = (x_0^1, \dots, x_0^n)$, we can find a geodesic which passes through this point and has the prescribed direction given by the vector $\lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$, or even having the prescribed velocity λ_0 in it.

Another speaking, the set of geodesics on a surface behaves (in the well-known sense) similarly as the set of lines in a plane. Which is not surprising since lines in a plane represent a special case of geodesics on a 2-surface. Geodesics form a two-parameter family of curves on a 2-surface.

Note that the theorem on existence of the solution of a system of differential equations which was reffered to has a local character, and hence the existence of a geodesic is guaranteed only in a neighborhood of a point $(x_0^1, \dots, x_0^n)$ of an n-surface, $n \geq 2$. But geodesics can be in fact prolonged on a surface either to infinity, or up to the boundary of the surface, using repeatedly the cited theorem and taking, as an initial point and initial vector, the end of the geodesic segment already constructed and the tangent vector at it. Note that I may not be all of $\mathbb{R}$, as for example for an open disc in $\mathbb{R}^2$.

Let us come back to the system (1.65), $n = 2$. Its examining is rather complicated by the fact that the parameter must have the geometric meaning of arc length, so that the functions $x^1(s), x^2(s)$ must satisfy, besides (1.65), also the additional condition $ds^2 = g_{ij}\, dx^i dx^j$.

Let us get rid of this obstruction. For this purpose, instead of the intrinsic equations (1.24), let us introduce the curve on a surface by an arbitrary parameter t. Then $s = s(t)$, $t = t(s)$, and we find

$$\frac{dx^h}{ds} = \frac{dx^h}{dt}\frac{dt}{ds} = \frac{dx^h}{dt}\frac{1}{s'}, \quad s' = \frac{ds}{dt}, \quad \frac{d^2x^h}{ds^2} = \frac{d^2x^h}{dt^2}\frac{1}{s'} - \frac{dx^h}{dt}\frac{s''}{s'^3}.$$

Now substituting these expressions to (1.65) we get

$$\frac{d^2x^h}{dt^2} + \Gamma_{ij}^h \frac{dx^i}{dt}\frac{dx^j}{dt} = \sigma(t)\frac{dx^h}{dt}. \tag{1.68}$$

Here the function $\sigma(t) = s''/s'$.

On the other hand, for a curve $x = x(t)$ which satisfies (1.68) for some function $\sigma(t) \in C^0$, there exists a parametrization with a parameter, let us say, s, under which the equations take the form (1.65).

It is obvious that the system (1.65) has its advantages in comparison with (1.68). Both systems are equivalent with respect to the object they define, that is, with respect to geodesic lines. We use one or the other representation of geodesics according to the actual mathematical purpose.

[45)] For $\Gamma_{ij}^h \in C^0$ (that is, $S \in C^2$), the Cauchy problem (1.66) and (1.59) has solution, and for $\Gamma_{ij}^h \in C^1$ (that is, $S \in C^3$), the existence of a unique solution is guaranteed.

[46)] Called also Cauchy conditions sometimes.

Now let us rewrite equations for geodesics in another form which appear to be very useful in further considerations. Let us write the equation (1.68) with $h = \alpha$, then once more with $h = \beta$, and eliminate the terms with $\sigma(t)$. We get

$$\frac{d^2x^\alpha}{dt^2}\frac{dx^\beta}{dt} - \frac{d^2x^\beta}{dt^2}\frac{dx^\alpha}{dt} + \left(\Gamma^\alpha_{ij}\frac{dx^\beta}{dt} - \Gamma^\beta_{ij}\frac{dx^\alpha}{dt}\right)\frac{dx^i}{dt}\frac{dx^j}{dt} = 0. \tag{1.69}$$

It can be easily seen that (1.68) and (1.69) are equivalent, and the advantage of (1.69) is that $\sigma(t)$ is excluded.

Remark that if we work with systems similar to those in (1.65), (1.68) and (1.69) (or coming below), we always suppose that differentiability conditions necessary for application of existence theorems are always satisfied.

1.4.3 Semigeodesic coordinates

On arbitrary surface S it is possible to locally introduce a special orthogonal coordinate system, which is called an *semigeodesic coordinate system*, in which the first quadratic form is expressed by

$$ds^2 = {dx^1}^2 + f(x^1, x^2)\, {dx^2}^2, \tag{1.70}$$

where $f(x^1, x^2)$ (> 0) is a function.

We can convince ourselves that x^1-curves are geodesics. By the analysis of these geodesics it easy to see that they are the (locally) shortest curves. In detail, see pp. 148-149.

In this coordinate system non-vanishing components:

$$g_{11} = g^{11} = 1,\ g_{22} = (g^{22})^{-1} = f(x^1, x^2),$$

$$\Gamma^1_{22} = -\tfrac{1}{2}\,\partial_1 f,\quad \Gamma^2_{12} = \Gamma^2_{21} = \partial_1 \ln\sqrt{f},\quad \Gamma^2_{22} = \partial_2 \ln\sqrt{f},\qquad K = -\frac{\partial_{11}(\sqrt{f})}{\sqrt{f}}.$$

1.4.4 Geodesic bifurcations

If the Christoffel symbols are continuous, then geodesics exist for above mentioned. We demonstrate an example of connections whose components are not differentiable, but geodesics have common properties, that is there do not exist bifurcations.

Bifurcation of geodesic is studied in [1041]. There, bifurcation is described as situation where (different) geodesics go through one point and have different tangent vectors. We show bifurcation of geodesics on surfaces of revolution, where two different geodesics go through the *same* point and have the *same* tangent vector.

Let $\mathcal{S}_2$ be a surface of revolution given by the equations:

$$x = r(u)\cos v,\quad y = r(u)\sin v,\quad z = z(u) \tag{1.71}$$

where v is parameter from $(-\pi, \pi)$, $u \in I \subset \mathbb{R}$ and $I = \langle u_1, u_2\rangle$.

In these equations we exclude meridian corresponding to coordinate $v = \pi$. Naturally, we also exclude "poles" which correspond to $r(u) = 0$.

Rotational surface $\mathcal{S}$ given by equations (1.71) has the following metric form

$$ds^2 = \left({r'}^2(u) + {z'}^2(u)\right) du^2 + {r'}^2(u)\, dv^2.$$

Let us choose parameter u as a length parameter of forming curve $(r(u), 0, z(u))$ then ${r'}^2(u) + {z'}^2(u) = 1$. In this case, metric of surface $\mathcal{S}$ is simplified

$$ds^2 = du^2 + f(u)\, dv^2, \tag{1.72}$$

where $f(u) = r^2(u)$, i.e. nonzero components of metric and inverse tensors are $g_{11} = g^{11} = 1$ and $g_{22} = (g^{22})^{-1} = f(u)$. Non-vanishing Christoffel symbols of first kind are $\Gamma_{122} = \Gamma_{212} = \dfrac{1}{2} f'(u)$ and $\Gamma_{221} = -\dfrac{1}{2} f'(u)$ and nonzero Christoffel symbols of second kind are

$$\Gamma^1_{22} = -\frac{1}{2} f'(u) \quad \text{and} \quad \Gamma^2_{12} = \Gamma^2_{21} = \frac{1}{2} \frac{f'(u)}{f(u)}. \tag{1.73}$$

Further, let $u \equiv x^1$ and $v \equiv x^2$. The equations (1.65) of geodesics on surface S can be written in the following form:

$$\ddot{u} = \frac{1}{2} f'(u)\, \dot{v}^2, \qquad \ddot{v} = -\frac{f'(u)}{f(u)}\, \dot{u}\, \dot{v}. \tag{1.74}$$

Because s is parameter of length, tangent vector of these geodesics is unitary, i.e. first integral applies: $\dot{u}^2 + f(u)\,\dot{v}^2 = 1$.

Trivially, we verify that u-coordinate curves ($u = s$, $v = \text{const}$, i.e. *meridian*) are geodesic. In general, the same does not apply for the v-coordinates, v-curves are geodesic if and only if $f'(u) = 0$ (they are also called *gorge circles*).

Further, let us study geodesics, which are none of the mentioned above. Suppose that $v(s) \neq 0$, i.e. $\dot{v}(s) \neq 0$. Then we can rewrite second equation of (1.66) in form $\dfrac{\ddot{v}}{\dot{v}} = -\dfrac{f'(u)}{f(u)}\,\dot{u}$. After modification and integration by s we get:

$$\dot{v} = \frac{C_1}{f(u)}, \quad \dot{u} = \sqrt{1 - \frac{{C_1}^2}{f(u)}}, \quad C_1 \in \mathbb{R} \tag{1.75}$$

Finally, the equations (1.75) determine system of the differential equations of the first order. □

Now we construct example of rotational surface $\mathcal{S}$, where above mentioned bifurcation exists, see Rýparová, Mikeš [920].

Example 1.1 Let $\mathcal{S}$ be a surface of revolution with functions

$$r(u) = \frac{1}{\sqrt{1 - u^{2\alpha}}} \quad \left(\Rightarrow f(u) = \frac{1}{1 - u^{2\alpha}}\right), \quad u \in (-1, 1). \tag{1.76}$$

The function r must to be differentiable so the Christoffel symbols exist and equations of geodesics can be written. On the other hand, the Christoffel symbols can not satisfy the Lipschitz condition and, of course, can not be differentiable (there would be an unique solution and bifurcation would not exist).

Theorem 1.5 *On above mentioned surface of revolution S there exist geodesic bifurcations for $\alpha \in (0.5, 1)$.*

Proof. The statement can be proved by existence of geodesics given by the equations:

$$\begin{aligned} &I. \qquad u = 0, \quad v = s \\ &II. \quad u = \big((1-\alpha)\, s\big)^{\frac{1}{1-\alpha}}, \quad v = s - \frac{\big((1-\alpha)\, s\big)^{\frac{1+\alpha}{1-\alpha}}}{1+\alpha}. \end{aligned} \tag{1.77}$$

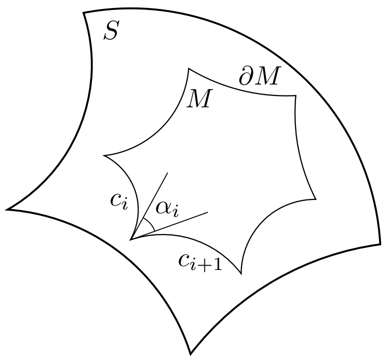
We can verify that curves given by the equations (1.66) are geodesics by direct substitution to fundamental equations (1.66).

These two geodesics go through the same point $(0,0)$ and have the same tangent vector $(0,1)$. The consequence is that through this point in this direction goes infinite number of geodesics and the gorge circle (mentioned above) is one of them. □

Example 1.2 If we set $f(u) = -\dfrac{1}{1-u^{2\alpha}}$ then the metric will be indefinite and the equations (1.77) describe a geodesic bifurcation on a pseudo-Riemannian space. Moreover, in [921] Rýparová and Mikeš construct closed geodesics with bifurcations.

1.4.5 Gauss-Bonnet Theorem

Finally, we present the Gauss-Bonnet Theorem without proof, which is undoubtedly one of the deepest (and even the most beautiful) results differential geometry of surfaces.



Gauss-Bonnet Theorem
Let M be an area of the surface S with boundary ∂M that is constructed with the curves $c_1, \dots, c_r$. These curves intersect under the angle α_i at a common point (see picture). Then

$$\sum_{i=1}^{r} \alpha_i = (r-2)\pi + \iint_M K d\sigma + \sum_{i=1}^{r} \int_{c_i} k_g ds, \tag{1.78}$$

where k_g and K are the geodesic and the Gauss curvature, respectively, ds and $d\sigma$ are the differential of length and area, respectively. [47)]

If M is a geodesic polygon, i.e. the curves c_i are geodesics, then

$$\sum_{i=1}^{r} \alpha_i = (r-2)\pi + \iint_M K d\sigma. \tag{1.79}$$

[47)] The area M does not contain a “hole” of surfaces S (we will not specify this term).

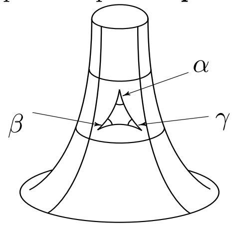
If S is a surface with constant curvature K then $\sum_{i=1}^{r} \alpha_i = (r-2)\pi + K \cdot vol(M)$, where $vol(M)$ is an area of M.

For geodesic triangle on surfaces with constant curvature K we obtain

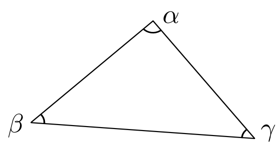
$$\alpha + \beta + \gamma = \pi + K\mathrm{vol}(\Delta). \tag{1.80}$$

For example, surfaces with constant curvature are

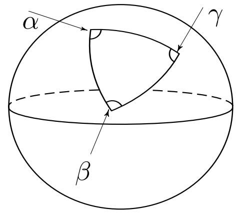
plane:	$\boldsymbol{p} = (u, v, 0)$,	$K = 0$,
sphere:	$\boldsymbol{p} = r \cdot (\cos u \cos v, \cos u \sin v, \sin u)$,	$K = 1/r^2$,
pseudosphere:	$\boldsymbol{p} = r \cdot (\cos u \cos v, \cos u \sin v, \cos u + \ln(\tan(u/2)))$,	$K = -1/r^2$.



Pseudosphere
$\alpha + \beta + \gamma < \pi$



Plane
$\alpha + \beta + \gamma = \pi$



Sphere
$\alpha + \beta + \gamma > \pi$

The above inequalities apply to the geodesic triangles of the surfaces with constant curvature.

It is said that the above inequalities were examined by Gauss as part of the geodetic triangulation of Germany, and by Lobachevsky in the framework of his star observation, but they were within the scope of observational errors.

This was associated with the emergence of the non-euclidean geometry which was founded by Lobachevsky[48], Bolyai[49] and Gauss. The geometry constructed by these authors is now called *hyperbolic* or *Lobachevskian geometry.*

In 1868, Beltrami [351] proved that the surfaces of negative constant curvature are part of the Lobachevsky plain. Poznyak's and Popov's[50] papers [171, 886–891] are dedicated to a detailed study of the Lobachevsky plain, see the monograph by Popov [168].

These results are connected with the Beltrami Theorem, see p. 350, which states that spaces with constant curvature are projective Euclidean.

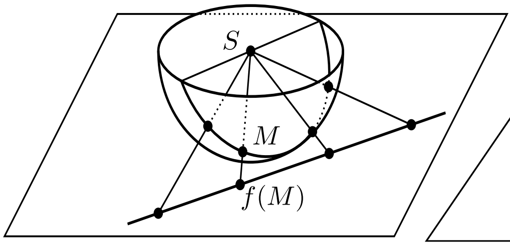
[48] Nikolai Ivanovich Lobachevsky, 1792-1856, a Russian mathematician and physicist, the rector of the Kazan University from 1827 to 1846, who formulated (in 1826) and published (in 1829) basic ideas of non-Euclidean hyperbolic geometry (called also Lobachevsky geometry). He was influenced by Johann Christian Martin Bartels, 1769-1836, a former teacher and friend of Gauss.

[49] János Bolyai, 1802–1860, a Hungarian soldier and mathematician, a son of the mathematician Farkas Bolyai who also studied at Bartels. Between 1820 and 1823 J. Bolyai prepared a treatise on a complete system of non-Euclidean geometry, independently of the results of Lobachevsky. Bolyai's work was published in 1832 as an Appendix to a mathematics textbook by his father. As a matter of interest, let us mention that as a soldier, during his military service, J. Bolyai was for a short time (1832–1833) a member of a garrison also in Olomouc (Czech Republic, germ.: Olmütz), this fact is reminded by his memory desk under his bust at the Army House in Olomouc.

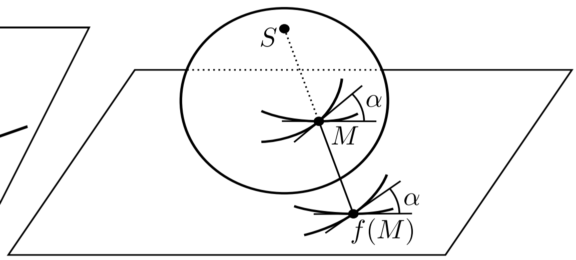
[50] Eduard Genrikhovich Poznyak, 1923-1993, and Andrey Gennadievich Popov, 1962-2014, are Russian mathematicians of Faculty of Physics, Lomonosov Moscow State University.

1.4.6 Gnomonic and stereographic projections

In geometry, the gnomonic and stereographic projections are particular mappings which project a half-sphere or a sphere respectively onto a plane. In the gnomonic projection, the center of projection is in the center of the sphere. In the stereographic projection, the center of projection is on the sphere.



Gnomonic projection



Stereographic projection

It is often stated that the Gnomonic projection is the oldest of all cartographical projections.

It was used in the 6th century BC by Thales of Miletus[51)] to describe star constellations. Thales is considered an author of this projection. Since the 16th century, the gnomonic projection is used to describe the surface of Earth. Currently, it is used for navigational maps and the present name "*gnomonic*" comes from the 19th century. This projection is an important example of the geodesic mappings discussed in Chapters 8-12.

The stereographical projection was used by Hipparchus of Nicaea[52)] in the 2nd century BC, and Claudius Ptolemy[53)] in the 2nd century. The use was mainly for star maps and from the 16th century for the image of Earth's hemispheres. Current name "*stereographical*" comes from the 17th century. This projection is conformal, i.e. it preserves angles. In 1779, the first non-trivial examples of conformal mappings were discovered by Lagrange [120], namely the stereographic projection of a sphere.

Several important cartographic projections, including the Mercator[54)] projection, are conformal maps. It is known, p. 33, that locally any surface can be mapped conformally onto the Euclidean plane. Chapter 6 is dedicated to conformal mappings of Riemannian spaces.

[51)] Thales of Miletus, 624/623-548/545 BC, was a pre-Socratic Greek philosopher, mathematician, and astronomer from Miletus in ancient Greek Ionia.

[52)] Hipparchus of Nicaea,190-120 BC, was a Greek astronomer, geographer, and mathematician. He is considered the founder of trigonometry but is most famous for his incidental discovery of precession of the equinoxes.

[53)] Claudius Ptolemy, 100-170, was a Greco-Roman mathematician, astronomer, geographer and astrologer, lived and worked in Alexandria, Egypt.

[54)] Gerardus Mercator, 1512-1594, was a 16th-century Southern Dutch (current day Belgium) cartographer, geographer and cosmographer. He was renowned for creating the 1569 world map based on a new projection which represented sailing courses of constant bearing (rhumb lines) as straight lines – an innovation that is still employed in nautical charts.

1.4.7 Motivations and applications

In mathematics, a geodesic is a generalization of the notion of a "straight line" to "curved spaces". In presence of a metric, geodesics are defined to be (locally) the shortest paths between points on the space. In the presence of an affine connection, geodesics are defined to be curves whose tangent vectors remain parallel if they are transported along it.

The term "geodesic" comes from geodesy, the science of measuring the size and shape of Earth; in the original sense, a geodesic was the shortest route between two points on the Earth's surface, namely, a segment of a great circle. The term has been generalized to include measurements in much more general mathematical spaces; for example, in graph theory, one might consider a geodesic between two vertices (nodes) of a graph.

Geodesics are commonly seen in the study of Riemannian geometry and more generally metric geometry.

In physics, geodesics describe the motion of point particles; in particular, the path taken by a falling rock, an orbiting satellite, or the shape of a planetary orbit are all described by geodesics in the theory of general relativity. More generally, the topic of sub-Riemannian geometry deals with the paths that objects may take when they are not free, and their movement is constrained in various ways.

Let us mention two theorems from mechanics which clarify natural character of geodesics, and at the same time demonstrate their importance [120].

Theorem 1.6 *On any smooth surface, an elastic band stretched between two points will contract its length, and the resulting shape of the band is a geodesic.*

In fact, let us connect two points of an "absolutely" smooth surface by an elastic band (rubber band, binder). The elasticity forces of such a band are oriented in the direction of the tangent line in a particular point. In the theory of elasticity, one proves that the resulting force of the elasticity forces at any point of the band is involved in the osculating plane (of the curve which represents the band), in which also $k\boldsymbol{n}$ is included. On the other hand, the resistence force of the absolutely smooth surface could be oriented only in direction of the normal of the surface. If the band is balanced then the result of the elasticity forces is in balance with the resistence force of the surface. Therefore the osculating plane of the band must involve the normal of the surface. But in this case, the normal of the surface and the principal normal of the band are parallel (since they belong to the same plane, pass through the same point, and are perpendicular to the tangent of the band curve). That is, the principal normal of the band coincides with the normal of a surface at any point which means, according to Theorem 1.3 that the band takes the form of a geodesic.

There is another interesting mechanical interpretation of geodesics, [173, p. 172]:

Theorem 1.7 *A particle moving on the surface, and subject to no forces except a force acting perpendicular to the surface that keeps the particle on the surface, would move along a geodesic.*

In fact, the only force acting on the mass particle is the resistence of the surface, which has always direction of the normal of the surface. According to the main theorem of dynamics, the resistence of the surface is proportional to the acceleration of the motion, and acceleration is always parallel to the osculating plane. That is, the osculating plane involves the normal of the surface. We get the same situation as in Theorem 1.6, i.e. the trajectory of motion is a geodesic.

In the calculus of variations, one proves that the shortest path between two points in a curved space can be found by writing the equation for the length of a curve, and then minimizing this length using standard techniques of calculus and differential equations. We will postpone details of this view-point to the next chapter.

Equivalently, a different quantity may be defined, termed the energy of the curve; minimizing the energy leads to the same equations for a geodesic. Intuitively, one can understand this second formulation by noting that an elastic band stretched between two points will contract its length, and in so doing will minimize its energy; the resulting shape of the band is a geodesic, as already mentioned.

2 TOPOLOGICAL SPACES

2.1 From metric spaces to abstract topological spaces

We intend to work here with spaces that are more general than the Euclidean space or affine spaces known from algebra and elementary geometry, although they look locally like Euclidean spaces and have various applications in mechanics, theoretical physics etc. Together with a distinguished class of spaces, mathematics is interested also in the family of mappings that preserve the typical properties of spaces from the class under consideration. To describe our field of interest we need also concepts from metric spaces, topology and the theory of continuous functions. Let us recall some basic notions and notation.

As well known from linear algebra, the real n-dimensional vector space $\mathbb{R}^n$ is a family of all n-tuples $x = (x^1, \ldots, x^n)$ of real numbers that are added and multiplied by reals component-wise in a familiar way, which defines on $\mathbb{R}^n$ a linear structure of a finite-dimensional vector space. Besides the linear structure the vector n-space $\mathbb{R}^n$ carries a natural inner product, the *dot product* $x \cdot y$ and the induced *norm* $\|x\|$:

$$x \cdot y = x^1y^1 + \ldots + x^ny^n \quad \text{and} \quad \|x\| = \sqrt{x \cdot x} = \sqrt{\,{x^1}^2 + \ldots + {x^n}^2}\ .$$

By the *Euclidean space* $\mathbb{E}^n$ we usually mean just $\mathbb{R}^n$ endowed with this dot product. The dot product defines naturally a *metric* $d(x, y) = \|x - y\|$, and the metric induces the *metric topology* on $\mathbb{R}^n$, or on $\mathbb{E}^n$: a subset $O \subset \mathbb{R}^n$ is open if and only if for any point $x \in O$, there exists an open ball $B(x, r)$ with center x and radius $r > 0$ which lies entirely in O; it can be seen that this metric topology coincides with the product topology of the Cartesian product $\mathbb{R} \times \ldots \times \mathbb{R}$ where the reals $\mathbb{R}$ are taken with the natural norm. Note that this metric is *translation invariant* in the following sense: $d(x + z, y + z) = d(x, y)$ holds for all x, y, z from $\mathbb{R}^n$; note that this property is common to *all* metric spaces arising from norms.

Definition 2.1 A *metric space* is a set X together with a function d from $X \times X$ to the non negative real numbers, such that for each $x, y, z \in X$:
(1) $d(x, y) = 0$ if and only if $x = y$,
(2) $d(x, y) = d(y, x)$,
(3) $d(x, z) \leq d(x, y) + d(y, z)$;
the function d is called a *metric* on X.

A *pseudometric* d satisfies (2) and (3) from the definition of metric but (1) is substituted by a weaker condition, namely $d(x,y) = 0$ if $x = y$ (that is, even distinct points might have zero distance).

Given a point x in a metric space X and a real number $r > 0$, the *open ball (= disc) of radius r about x (with the center x)* is a subset

$$B(x,r) = \{y \in X : \ d(x,y) < r\}.$$

We call a subset $O \subset X$ *open* in metric space X if for each point $x \in O$ there exists an open ball about x in X which is entirely in O.

If $A \subset X$ we introduce a *diameter* of A: $\operatorname{diam} A = \sup\{d(x,y) : \ x, y \in A\}$.

In a metric space X, a subset $A \subset X$ is *bounded* if it has finite diameter.

2.1.1 A couple of examples

Example 2.1 On at least two-element set, the so-called *discrete metric* is defined by $d(x,y) = 1$ for $x \neq y$, $d(x,y) = 0$ for $x = y$.

Example 2.2 A normed real vector space V with the norm $\|\cdot\|$ determines a natural metric d defined by $d(u,v) = \|u - v\|$.

Example 2.3 Let $X = C^0\langle a,b\rangle$ be the set of all real functions continuous on a closed interval $\langle a,b\rangle$. We can define a metric on X directly by

$$d(f,g) = \sup\{|f(x) - g(x)| : \ x \in \langle a,b\rangle\} \quad \text{for } f,g \in X.$$

This metric comes from a norm. On X, the linear operations $f \mapsto cf$ and $(f,g) \mapsto f + g$ are defined pointwise which turn X into an infinite-dimensional vector (linear) space. If we define $\|f\| = \max_{x\in\langle a,b\rangle} |f(x)|$ for $f \in X$ we have a norm on X that gives just the metric d. Similarly for the space of continuous complex functions.

There are metrics on vector spaces that do not arise from any norm.

Example 2.4 The set X of all real sequences $\{a_n\}_{n=1}^{\infty}$ is a real infinite-dimensional vector space. The function $\varrho(a,b) = \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot \frac{|a_n - b_n|}{1 + |a_n - b_n|}$ is a metric.

Suppose that the metric ϱ comes from some norm $\|\cdot\|$ in the way described above. Then $\|x\| = \varrho(x,0)$ for each $x \in X$ and $\|ax\| = |a| \cdot \|x\|$, that is, ϱ have to satisfy

$$\varrho(ax,0) = |a|\,\varrho(x,0) \qquad \text{for all } a \in \mathbb{R},\ x \in X.$$

Take $x = (1,0,0,\dots)$, then $ax = (a,0,0,\dots)$ for any $a \in \mathbb{R}$. Now $\varrho(x,0) = \frac{1}{4}$, $\varrho(ax,0) = \frac{1}{2} \cdot \frac{|a|}{1+|a|}$. But this equality does not hold in general: if we take $a = 2$ then $\varrho(2x,0) = \frac{1}{3}$, on the other hand $2\varrho(x,0) = \frac{1}{2}$, a contradiction. Therefore such a norm cannot exist.

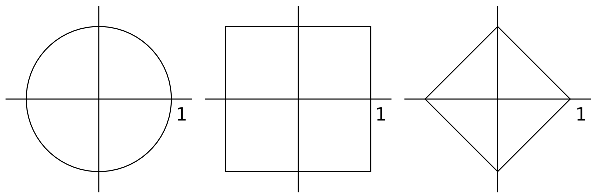
Example 2.5 For two points $P = (x_1, x_2)$ and $Q = (y_1, y_2)$ in $\mathbb{R}^2$, the formulae

$$\begin{aligned} d_1(P,Q) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}, \\ d_2(P,Q) &= \max\{|x_1 - y_1|, |x_2 - y_2|\}, \\ d_3(P,Q) &= |x_1 - y_1| + |x_2 - y_2| \end{aligned}$$

define three metrics which provide the plane with three distinct structures as a metric space. Yet the family of all open sets is the same in all three cases. We can consider them "equivalent". Not only open sets, but of course also all concepts based on open sets are the same: closed sets, as their complements, etc.; also convergence of sequences is the same. It makes us think what is "behind" this fact.

Note that d_3 is sometimes called a *taxicab distance function (metric)*.

The figure shows the ball of radius 1, central at the origin, for each of these three metrics.



2.1.2 Euclidean space

Similar metrics can be introduced in $\mathbb{R}^n$ for arbitrary $n \in \mathbb{N}$; the first metric space is usually called *Euclidean* or *cartesian space*; the three metrics define the same open sets again.

Let us give first a bit of motivation. Let f be a map of the Euclidean space $\mathbb{E}^m$ to $\mathbb{E}^n$. The classical "ε, δ" definition of continuity for f generalizes continuity of a real-valued function of one real variable well-known from the Calculus and goes as follows: f is *continuous at* $x \in \mathbb{E}^m$ if given any $\varepsilon > 0$ there exists $\delta > 0$ such that $\|f(y) - f(x)\| < \varepsilon$ whenever $\|y - x\| < \delta$, $y \in \mathbb{E}^m$.

More geometric speaking, f is *continuous* if for any open ball $D' = B(f(x), \varepsilon)$ in $\mathbb{E}^n$ about $f(x)$ with radius $\varepsilon > 0$ there exists an open ball $D = B(x, \delta)$ in $\mathbb{E}^m$ with center x and radius $\delta > 0$ which is mapped into $B(f(x), \varepsilon)$ under f: $f(D) \subset D'$. The function is *continuous* if it is continuous in each point.

Call a subset U of $\mathbb{E}^m$ a *neighbourhood* of the point $x \in \mathbb{E}^m$ if for some real number $r > 0$ the open ball of radius r and center x lies entirely in U. It is easy to rephrase the above definition of continuity as follows: f is continuous if given any $x \in \mathbb{E}^m$ and any neighbourhood U of the image $f(x)$ in the space $\mathbb{E}^n$, then the inverse image $f^{-1}(U)$ is a neighbourhood of the point x in $\mathbb{E}^m$.

More generally, we can proceed similarly in any metric space. A map $f\colon X \to Y$ of a metric space (X, ϱ) to a metric space (Y, σ) is *continuous* if for any $x \in X$ given $\varepsilon > 0$ there exists $\delta > 0$ such that $\sigma(f(x), f(z)) < \varepsilon$ whenever $\varrho(x, z) < \delta$, $z \in X$. Again, a neighbourhood of the point $x \in X$ is a subset which contains a disc centered at x, and continuity can be rephrased using the concept of neighbourhoods.

2.1.3 Natural topology on metric space

Note that defining neighbourhoods in Euclidean spaces or metric spaces, we use very strongly the distance function. In constructing an "abstract space" we would like to retain the concept of neighbourhood but rid ourselves of *any* dependence, of the definition of the space itself as well as of continuity of maps between abstract spaces, on a distance function. Just this point is crucial: any point of the "space" should be endowed with a family of "neighbourhoods" settled in such a way that a "good" definition of continuity can be expected. Note that Maurice Fréchet, the French mathematician who created the first definition of an abstract topological space, used just this way, namely generating topology by neighbourhoods.

We ask for a set X and for each point $x \in X$ a nonempty collection $\mathcal{U}(x)$ of subsets of X, called *neighbourhoods* of x, that are required to satisfy the following four conditions (axioms):

(a) x lies in each of its neighbourhoods.

(b) The intersection of two neighbourhoods of x is itself a neighbourhood of the point x.

(c) If V is a subset of X which contains U and U is a neighbourhood of x, then V is a neighbourhood of x.

(d) If U is a neighbourhood of x then there exists a neighbourhood O of x such that $O \subset U$ and O is a neighbourhood of z whenever $z \in O$; O is an *interior* of U.

This whole structure can be called a *topological space*, and we say that the assignment of a collection of neighbourhoods satisfying (a) – (d) to each point $x \in X$ gives a *topology* generated by a neighbourhood system on the set X.

We call a subset O of X *open* in this topology if it is open neighbourhood of each of its points. The union of any collection of open sets is open by (c), and the intersection of any *finite* number of open sets is open by axiom (b) (on the other hand, the intersection of an infinite collection of open sets need not be open). The empty set is open, as is the whole space X. Axiom (d) tells us that given a neighbourhood U of a point x, the interior of U is an open set which contains x and which lies in U. To understand better motivation for the last condition we can take the closed ball $\{z \in \mathbb{E}^m : d(x, z) \leq r\}$ as U, then as O we can take the open ball $\{z \in \mathbb{E}^m : d(x, z) < r\}$.

Although the above concept was formulated quite comprehensible and fits well our idea what a space ought to be, unfortunately such a definition is not so practical to work with. It was found out that an equivalent, more manageable, set of axioms can be given. During the time it was discovered that more convenient, especially in proofs, is to start with the idea of open set, then build up a collection of neighbourhoods for each point, and to show that both approaches are equivalent. Then all concepts build up on open sets will be topological notions.

2.1.4 Isometry of metric spaces

As morphisms between metric spaces, we prefer those maps which preserve distances between points. A map of metric spaces $f\colon (X,\varrho) \to (Y,\sigma)$ is an *isometry* if $\varrho(x,y) = \sigma(f(x), f(y))$ for every pair $x,\, y \in X$. Any isometry is a continuous map with continuous inverse (i.e. homeomorphism). All isometries of the given metric space X with map composition operation constitute a group $iso(X)$, the *isometry group* of the space.

Example 2.6 Translations, rotations, symmetries, skew symmetries (and the identity map) are well-known isometries in $\mathbb{E}^2$.

2.1.5 Abstract topological spaces, topology

Let X be a given set. Recall that the family of all subsets in X is called a *potence set* of X and is denoted by $\mathcal{P}(X)$.

Definition 2.2 The system τ of subsets in X, $\tau \subseteq \mathcal{P}(X)$, is called a *topology* on X if the following three axioms hold:
(O1) The empty set $\varnothing$ and the whole space X belong to τ.
(O2) The intersection of two[55] of sets from τ is in τ.
(O3) The union of any family of sets from τ is in τ.

The pair (X,τ) is called a *topological space*.

The sets from τ are called *open*. Given a point $x \in X$ we shall call a subset U of X a *neighbourhood* of x if we can find an open set O such that $x \in O \subset U$. For a fixed $x \in X$, the set of all neighbourhoods of x in the given topology is denoted by $\mathcal{U}(x)$.

We can verify that this definition of neighbourhood makes X into a topological space according to the above "neighbourhood" definition. For each point, at least X is a neighbourhood of x. If U_1, U_2 are neighbourhoods of x and O_1, O_2 are the corresponding open sets satisfying $x \in O_1 \subset U_1$, $x \in O_2 \subset U_2$, then $x \in O_1 \cap O_2 \subset U_1 \cap U_2$ where $O_1 \cap O_2$ is open. Therefore $U_1 \cap U_2$ is a neighbourhood of x, and we have checked axiom (b). To check (a) and (c) is easy. Similarly the converse implication.

Theorem 2.1 *A subset of a topological space is open if and only if it is a neighbourhood of each of its points.*

2.1.6 Examples of topological spaces

Example 2.7 As a well known example of topological space, recall the set of real numbers $X = \mathbb{R}$ with natural topology: a subset O in $\mathbb{R}$ is open when with each point $r \in O$, an open interval $(r-\varepsilon, r+\varepsilon)$ is in O. Notice that this natural topology is just the metric topology corresponding to $d(a,b) = |a-b|$, $a, b \in \mathbb{R}$.

---

[55] And consequently of any finite number.

Example 2.8 For our purpose, particularly the real n-dimensional space $\mathbb{R}^n$ with natural topology is important; this topology arises as the product topology of $\mathbb{R} \times \ldots \times \mathbb{R}$ (n copies), and can be introduced directly as follows. Each point of the space $\mathbb{R}^n$ is an ordered n-tuple $(x^1, x^2, \ldots, x^n)$ of real numbers x^1, x^2, $\ldots$, x^n. Let us consider n open intervals (a^i, b^i) in the reals, $i = 1, 2, \ldots, n$. An *open coordinate parallelepiped*, or *open coordinate box* in $\mathbb{R}^n$ is the set

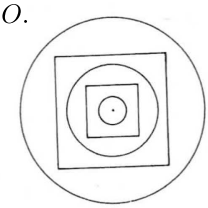
$$\mathcal{K}_n = \{\, x(x^1, x^2, \ldots, x^n) \mid a^i < x^i < b^i,\ i = 1, 2, \ldots, n\}.$$

We consider a subset $O \subseteq \mathbb{R}^n$ open in the "product" topology if for any point $x \in O$ there is an open coordinate box $\mathcal{K}_n$ such that $x \in \mathcal{K}_n \subseteq O$. We can easily check that the family of such subsets O satisfies (O1) – (O3), and hence is a topology in $\mathbb{R}^n$.

The same topology appears to be induced by the Euclidean metric of $\mathbb{R}^n$ as a metric space. In fact, for any pair of points $x = (x^1, x^2, \ldots, x^n)$ and $y = (y^1, y^2, \ldots, y^n)$ in $\mathbb{R}^n$, introduce their Euclidean distance by the formula

$$\varrho(x, y) = \sqrt{(x^1 - y^1)^2 + (x^2 - y^2)^2 + \cdots + (x^n - y^n)^2}.$$

A pair $(\mathbb{R}^n, \varrho)$ is a metric space denoted by $\mathbb{E}^n$ and called *n-dimensional Euclidean space.* As above, we can consider a subset $O \subseteq \mathbb{R}^n$ *open* if for any point $x \in O$ there exists an open ball $B(x, r)$ such that $B(x, r) \subset O$.



The family of all such open sets satisfies the definition of a topology in $\mathbb{R}^n$, and is called *natural*, or *metric* topology (induced by the Euclidean metric). It is easy to prove that the "product" topology of the n-dimensional real space $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ and the natural topology of the n-dimensional Euclidean space $(\mathbb{R}^n, \varrho)$ coincide (hint: to any ball centred at x, a cube centred in x can be inscribed, and vice versa).

Example 2.9 More generally, any metric space (X, d) endowed with its natural metric topology, is a topological space. Recall how the metric d defines a topology. We consider a subset of X open in the metric topology if and only if it is open with respect to the metric d, i.e. a subset $O \subseteq M$ is *open in the metric (natural) topology* when with each of its points, it includes some open ball $B(x, r)$ centered at the point x. We check directly that (O1)– (O3) hold.

Example 2.10 Normed real vector spaces are topological spaces. Any real vector space with a norm $(V, \|\cdot\|)$ has a natural metric $d(x, y) = \|x - y\|$ for $x, y \in V$. This metric turns V to a topological space if we take on V a metric topology corresponding to d, called *topology of norm on V.*

Example 2.11 Real vector spaces with scalar product are topological spaces. If $(V, (\,\cdot\,,\cdot\,))$ is a real vector space with scalar product we define the corresponding norm by $\|x\| = (x, x)^{1/2}$ for each $x \in V$ and consider the *topology of norm on V.*

Example 2.12 Any nonempty set X together with the family of open sets $\{\varnothing, X\}$ is a topological space, both the topology and the space are called *antidiscrete* or *indiscrete.* Every subset $A \subset X$ is open-and-closed [66, p. 31].

Example 2.13 Any nonempty set X with the family of open sets $\tau = \mathcal{P}(X)$ is a topological space, both the topology and the space are called *discrete*. Note that the discrete topology is generated by the discrete metric.

Example 2.14 Let X be any set. If we take as open sets a family of all subsets having finite complement we obtain on X the so-called *topology of finite complements*.

Example 2.15 Let X be an infinite set. If we take, as open sets, a family of all subsets having countable complement we obtain the so-called *topology of countable complements* on X.

Example 2.16 Assume the set $X = C^0\langle a,b\rangle$ of all real functions continuous on $\langle a,b\rangle$. Open sets are just all subsets of X that, with any of its elements, say f_0, contain, for a suitable $\varepsilon > 0$, all $f \in X$ such that $\sup\{|f_0(x) - f(x)|\colon\ x \in \langle a,b\rangle\} < \varepsilon$. This topology is a metric topology induced by the metric from Example 2.3, or, if we consider X as a vector space, induced by the norm $\|f\| = \sup\{f(x)\colon\ x \in \langle a,b\rangle\} = \max\{f(x)\colon\ x \in \langle a,b\rangle\}$. Similarly for continuous complex functions on a closed interval.

Example 2.17 Let X be a set ordered by a relation of ordering $\leq$. The so-called *interval topology* on X consists of subsets which contain, with each of its points, a set of the form $\{z \in X\colon\ a \leq z \leq b\}$, called *interval*, for some $a, b \in X$.

Example 2.18 Let A be a commutative ring with unit. Recall that an ideal $I \subseteq A$ in A, $I \neq A$ is a prime-ideal if $a \cdot b \in I \Longrightarrow a \in I$ or $b \in I$.

The spectrum $Spec(A)$ of the ring A is a set of all prime-ideals in A. On $Spec(A)$ we define a topology as follows. For any fixed ideal $J \subseteq A$ (not necessarily prime) take the set $\mathcal{O}_J = \{P \in\ Spec(A)\colon\ J \not\subseteq P\}$.

Then the system $\tau = \{\mathcal{O}_J\colon\ J$ is an ideal in A distinct from $A\}$ is a topology on $Spec(A)$. Any ideal contains at least zero element, hence $\mathcal{O}_{(0)} = \varnothing$. Further $\mathcal{O}_A = \ Spec(A)$, therefore (O1) holds. Let $\mathcal{O}_I, \mathcal{O}_J$ be from τ,

$$\mathcal{O}_I = \{P \in\ Spec(A)\colon\ I \not\subseteq P\}, \quad \mathcal{O}_J = \{P \in\ Spec(A)\colon\ J \not\subseteq P\}.$$

Then the intersection can be written as $\mathcal{O}_I \cap \mathcal{O}_J = \mathcal{O}_K$ where $K = IJ$ is the product of ideals. For a system of ideals $\{\mathcal{O}_{I_\alpha}\colon\ \alpha \in A\}$ the union of corresponding sets is $\bigcup\limits_{\alpha} \mathcal{O}_{I_\alpha} = \mathcal{O}_L$, where $L = \sum\limits_{\alpha \in A} I_\alpha$ is the sum of ideals.

2.2 Generating of topologies

2.2.1 Closed sets

Complements of open sets play also an important role, and have dual properties.

Definition 2.3 A subset $F{\subset}X$ is called *closed*, if its complement $X \backslash F$ is open.

Note that due to the de Morgan formulae the following can be proved:
(F1) the empty set $\varnothing$ and the set X are closed;
(F2) the intersection of any system of closed sets is a closed set;
(F3) the union of any finite number of closed sets is a closed set.

There is a topology generated (uniquely) by the system of closed sets. If a system $\mathcal{F}$ of subsets of X satisfies (F1) – (F3) then the set of complements $X \setminus F$ of members F from the system $\mathcal{F}$ has the properties (O1) – (O3) of open sets and is called the *topology generated on X by the family of closed sets $\mathcal{F}$*.

Example 2.19 In the classical algebraic geometry the so-called *Zariski topology* was defined for affine and projective varieties. Let $\mathcal{A}_n$ be an affine space. The Zariski topology is defined by specifying its closed sets: the set $\mathcal{F}$ of all algebraic sets (affine varieties) in $\mathcal{A}_n$ satisfies (F1) – (F3). Similarly, the projective Zariski topology in the projective space is given by Zariski-closed sets which are just all projective varieties (zero sets of homogeneous ideals).

2.2.2 Closure operator. Accumulation points

Let A be a subset of X. The intersection of all closed subsets in X containing A is called the *closure of* A and is denoted as $\overline{A}$; i.e. the closure of A is the smallest close set containing A. A point $x \in X$ is from $\overline{A}$ if and only if each neighbourhood of x intersects A.

The concept of a topological space can be considered as an axiomatization of the notion of the "closeness" of a point to a set: a point is close to a set if it belongs to it closure.

On the other hand, a point $x \in X$ is called an *accumulation point* or *limit point* of A if every neighbourhood of x contains at least one point of $A \setminus \{x\}$, i.e. each neighbourhood of x has a common point with A which is different from x. The family of all accumulation points of A in X is the *derived set* of A and is denoted here as A'.

The following properties hold true:

1. *The closure $\overline{A}$ of a subset $A \subset X$ is the union of A and all its accumulation (limit) points; $\overline{A} = A \cup A'$.*
2. *A set is closed if and only if it contains all its accumulation points.*
3. *A set is closed if and only if it is equal to its closure.*

A set whose closure is the whole space is said to be *dense* in the space. For example, the set of all points in $\mathbb{E}^3$ with rational coordinates is dense in $\mathbb{E}^3$.

A set $A \subset X$ is *co-dense* if $X \setminus A$ is dense in X, and *nowhere dense* if $\overline{A}$ is co-dense, i.e. $X \setminus \overline{A}$ is dense in X.

The closure operator $A \mapsto \overline{A}$ in the given topology has the properties:

Theorem 2.2 (K. Kuratowski) *Let A, B be subsets of the topological space (X, τ). Then the following conditions hold:*

$$\text{(1)}\ \ \overline{\varnothing} = \varnothing, \quad \text{(2)}\ \ A \subset \overline{A}, \quad \text{(3)}\ \ \overline{A \cup B} = \overline{A} \cup \overline{B}, \quad \text{(4)}\ \ \overline{\overline{A}} = \overline{A}.$$

It can be verified that the topology is by its closure operator uniquely determined. That is, we have another way how to generate a topology:

Theorem 2.3 *Let (X, τ) be a topological space and let $cl\colon \mathcal{P}(X) \to \mathcal{P}(X)$ be a map satisfying for any $A, B \in \mathcal{P}(X)$*

$$A \subset cl\,A, \quad cl\,\varnothing = \varnothing, \; cl\,X = X, \quad cl(A \cup B) = cl\,A \cup cl\,B, \quad cl(cl\,A) = cl\,A.$$

Then there exists a unique topology on the set X such that $cl\,A = \overline{A}$ holds for all $A \in \mathcal{P}(X)$.

Indeed, if we take all sets $G \in \mathcal{P}(X)$ such that $cl(X \setminus G) = X \setminus G$ we get just the announced topology.

2.2.3 Interior, exterior, boundary

Let $A \subset X$ be a subset in a topological space. Then every point $x \in X$ has exactly one of the following three properties (we speak about *interior*, *exterior* and *boundary* points of A, accordingly; alternative definitions are possible):

(1) there exists a neighbourhood of x which is contained in A;
(2) there exists a neighbourhood of x which is contained in $X \setminus A$;
(3) every neighbourhood of x intersects both the sets A and $X \setminus A$.

The union of all open subsets in the space X that are contained in A is called the *interior of* A and is denoted by $Int\,A$; a point x lies in the interior of A (is an interior point of A) if and only if A is a neighbourhood of x; equivalently, if and only if x has the property (1). The points of $Int\,A$ are *interior points* of A.

Theorem 2.4 *A subset $A \subseteq X$ is open if and only if $A = Int\,A$.*

We can introduce *exterior of* A as $\mathcal{E}xt\,A = Int(X \setminus A)$; a point $x \in \mathcal{E}xt\,A$ if and only if it satisfies (2), and is called an *exterior point of* A. Obviously, $\mathcal{E}xt\,\varnothing = X$. Similarly as above, it is possible to take interior (or exterior) as an axiomatic notion.

Fundamental properties of the interior operator are for every $A, B \in \mathcal{P}(X)$:

$$Int A \subset A,\; Int\,\varnothing = \varnothing,\; Int\,X = X,\; Int(Int A) = Int A,\; Int A \cap Int B = Int(A \cap B).$$

The family of open sets for the topology generated by interior operator consists just from the sets for which $Int\,A = A$.

The set $\delta A = \overline{A} - Int\,A = \overline{A} \cap \overline{(X \setminus A)}$ is called a *boundary* or *frontier* of A. The points from δA are called *boundary (or frontier) points* of A. A point x belongs to δA if and only if it has the property (3).

Among others, the following identities (useful in proofs) can be checked for a subset A of a topological space:

$$Int\,A \subset A \subset \overline{A} = A \cup A' = A \cup \delta A = Int\,A \cup \delta A,$$

$$A \setminus \delta A = Int\,A, \quad X = Int\,A \cup \delta A \cup \mathcal{E}xt\,A \text{ (disjoint union)},$$

$$\delta A = \overline{A} \cap \overline{(X \setminus A)} = \delta(X \setminus A), \qquad \delta\varnothing = \varnothing,$$

$$\varnothing = Int\,A \cap \delta A = \delta(X \setminus A) \cap \mathcal{E}xt\,A = Int\,A \cap \mathcal{E}xt\,A = \delta A \cap \mathcal{E}xt\,A.$$

Theorem 2.5 *A subset A of a topological space satisfies*

$$X \setminus \overline{A} = Int(X \setminus A), \quad X \setminus Int\,A = \overline{(X \setminus A)}.$$

2.2.4 The lattice of topologies. Ordering

On the same set X, more (or even many) topologies can be defined, and they are naturally ordered by inclusion of corresponding systems of open sets.

Let τ and $\tilde{\tau}$ be two topologies, i.e. families of open sets satisfying (O1)–(O3), on the same set X. We say that τ is *finer*, or *bigger* than $\tilde{\tau}$ when the systems satisfy $\tilde{\tau} \subseteq \tau$; in this case, $\tilde{\tau}$ is called *coarser* or *smaller*. Of course there are uncomparable topologies if the underlying set is at least two-element.

From the algebraic point of view, the family of all topologies on the same underlying set, ordered by inclusion, forms a lattice. Indeed, the antidiscrete topology is smaller than any other topology on the same set while discrete topology is the biggest one.

Example 2.20 On a one-element set $X = \{a\}$ there exists a unique topology; the discrete and antidiscrete topology coincide. On a two-element set $X = \{a, b\}$ there are four distinct topologies:

$$\mathcal{O}_0 = \{\varnothing, \{a, b\}\}, \qquad \mathcal{O}_1 = \{\varnothing, \{a\}, \{b\}, \{a, b\}\},$$
$$\mathcal{O}_2 = \{\varnothing, \{a\}, \{a, b\}\}, \qquad \mathcal{O}_3 = \{\varnothing, \{b\}, \{a, b\}\}.$$

Here $\mathcal{O}_0$ is the smallest one, $\mathcal{O}_1$ is the biggest one, $\mathcal{O}_2$ and $\mathcal{O}_3$ are uncomparable. It is interesting to notice how the number of topologies increases with the increasing number of elements of the underlying set X.

2.2.5 Metrization problem

A natural question arises in connection with examples: if we are given a topology is it possible to generate it by some metric, or at least pseudometric? The answer is negative.

A topological space that can be assigned a metric inducing the given topology is called *metrizable*.

The antidiscrete topology on at least two-element set is not metrizable.

The topologies $\mathcal{O}_2$ and $\mathcal{O}_3$ on a two-element set from Example 2.20 are not metrizable (because they are not Hausdorff, which we explain later).

The discrete space is metrizable by the discrete metric.

A great deal of work in general topology was done during examining metrizable spaces, their subspaces and metrizability conditions. Metrization problem was solved in 1951 by a Canadian mathematician R.H. Bing. To be able to formulate necessary and sufficient conditions for a topological space to be metrizable we need to know more about special types of bases and about separation properties which we mention later. Metrization theorems and properties of metrizable spaces are postponed to next sections.

2.2.6 Cover, subcover

Let (X, τ) be a topological space and $A \subseteq X$ a subset. A collection of (open) subsets in X:

$$\mathcal{U} = \{U_\alpha : \ U_\alpha \in \tau, \ \alpha \in \mathcal{J}\}, \qquad \mathcal{J} \text{ is some index set,}$$

is an *(open) cover (or covering)* of A if $A \subset \cup_{\alpha \in \mathcal{J}} U_\alpha$.

Note that the equality holds in the last formula when $X = A$, i.e. $X = \cup_{\alpha \in \mathcal{J}} U_\alpha$.

If $\{U_\alpha \colon U_\alpha \in \tau, \alpha \in \mathcal{J}\}$ is (open) cover of X, under its *(open) subcover* we mean any system $\{U_\alpha \colon U_\alpha \in \tau, \alpha \in \tilde{\mathcal{J}}\}$ where the index set $\tilde{\mathcal{J}} \subseteq \mathcal{J}$.

An (open) cover $\{V_\beta \colon V_\beta \in \tau, \beta \in \mathcal{K}\}$ is a *refinement* of the (open) cover $\{U_\alpha \colon U_\alpha \in \tau, \alpha \in \mathcal{J}\}$ of X if each (open) subset V_β is contained in some U_α.

2.2.7 Bases. Countability Axioms

As we have seen, a topology on a set X can be given by distinguishing a system τ of subsets in X (called "open") satisfying the axioms (O1) – (O3), but it is not the only possibility. We have already mentioned that a topology can be determined, or generated, also in some other way: by the set of closed sets provided they satisfy (F1) – (F3), by the families of neighbourhoods $\mathcal{U}(x)$ of points x from the topological space satisfying (a) – (d), or by the interior or the closure operator. Here we show yet another possibility, namely to give a suitable part of the family of open sets from which all open sets are generated by set union.

Definition 2.4 Let (X, τ) be topological space. A *neighbourhood base of a point* $x \in X$, or a *local base of* $x \in X$ is a subset $\mathcal{B}(x)$ of the set $\mathcal{U}(x)$ of all neighbourhoods of x such that every neighbourhood of x contains a neighbourhood in $\mathcal{B}(x)$.

Example 2.21 Let $X = \mathbb{R}^n$, or more generally, let X be any metric space. The set of open balls with radius $1/n$, $n = 1, 2, \ldots$ around a fixed point x forms a (countable) neighbourhood base of x.

Definition 2.5 A *base of the topology* τ on X is a system $\mathcal{B}$ of subsets in X such that

1. $\mathcal{B} \subset \tau$.
2. For any point $x \in X$ and any neighbourhood U of x there is a subset $V \in \mathcal{B}$ such that $V \subset U$.

We can give another characterization: a subsystem $\mathcal{B}$ of τ forms a base of the topological space (X, τ) if and only if any non-empty set from τ can be expressed as a union of the sets from the system $\mathcal{B}$.

Obviously, a topological space can have many bases. The following theorem characterizes systems of sets that can serve as a base for some topology.

Theorem 2.6 *A family of sets $\mathcal{B}$ is a base of a topology on a set $X = \cup\{B : B \in \mathcal{B}\}$ if and only if for any $A, B \in \mathcal{B}$ and for any $x \in A \cap B$, there exists $C \in \mathcal{B}$ such that $x \in C \subset A \cap B$ holds. The topology is uniquely determined by its base, and it is in fact the coarsest (smallest with respect to set inclusion) topology on X containing the given base $\mathcal{B}$.*

Example 2.22 Let $X = \mathbb{R}$. The set of intervals $\mathcal{B} = \{\langle a, b) \colon a < b, a, b \in \mathbb{R}\}$ is a base for a topology on $\mathbb{R}$, called *Sorgenfrey topology*. The corresponding topological space is a *Sorgenfrey straightline*. Notice that $\bigcup\limits_{n=2}^{\infty} (a + \frac{b-a}{n}, b) = \langle a, b)$.

The Sorgenfrey topology is finer than the usual topology on $\mathbb{R}$.

As mentioned above not every system is a base of a topology. The following weaker concept can be viewed as a compensation, it simplifies considerations concerning possibility of generating a topology by a system of sets.

Definition 2.6 A *subbase of the topology* τ on X is a system $\mathcal{S}$ of subsets in X such that

1. $\mathcal{S} \subset \tau$. 2. Every open set from τ is a union of finite intersections of sets in $\mathcal{S}$.

Every base of topology is also a subbase.

Example 2.23 The system of all intervals (a, ∞) and $(-\infty, a)$, $a \in \mathbb{R}$ is a subbase of the usual topology of $\mathbb{R}$ but not its base.

Theorem 2.7 *Any system $\mathcal{S}$ of sets is a subbase of a uniquely defined topology on the set $X = \bigcup\{S\colon\ S \in \mathcal{S}\}$.*

Indeed, we take all possible finite intersections of sets from $\mathcal{S}$ and use Theorem 2.6 to check that they form a base.

The definition of a topological space is very general. Not many interesting theorems can be proved about all topological spaces. Various classes of topological spaces are studied, ranging from fairly general to more and more special. One type of restrictions is concerned with cardinality of bases.

Recall that a system of sets is *countable* if the system includes at most a countable family of members (i.e. there is a one-one map of elements of the system into natural numbers $\mathbb{N}$).

Definition 2.7 A topological space satisfies the *first countability axiom*, and is called *first countable*, if every point possesses a countable neighbourhood base.

A space satisfies the *second countability axiom*, and is called *second countable*, if it possesses a countable base (of the topology).

A topological space X is called *Lindelöf* if each open cover of X has a countable subcover.

If a space contains a countable dense subset it is called *separable*.

Theorem 2.8 *Any metric (metrizable) space is first countable.*

Theorem 2.9 *If the space is second countable then each its base has a countable subbase.*

The second countability axiom is the strongest one from the list of conditions just mentioned:

Theorem 2.10 *If the topological space is second countable then it is first countable, Lindelöf and separable.*

In the second countable space, all sets containing the fixed point x form obviously a countable neighbourhood base of x. To construct a countable dense subset we choose one element from each member of a fixed countable base (we need the Axiom of Choice); to prove the Lindelöf property is also possible. However in metrizable spaces the following holds:

Theorem 2.11 *If the topological space is metrizable then second countability, Lindelöf property and separability are equivalent.*

Before giving more details it is convenient to introduce new concepts and more terminology.

2.2.8 Sequences in topological spaces, nets

As far as the role and behaviour of convergent sequences is concerned there is a great difference between first coutable spaces and general topological spaces.

Assume a topological space X, a sequence $\{x_n\}_{n\in\mathbb{N}}$ of points from X, i.e. x_n belongs to X for all $n \in \mathbb{N}$, and a fixed point $x \in X$.

We say that the sequence $\{x_n\}$ *converges* to the point x in X, or that x is a *limit point* of the sequence $\{x_n\}$ if for any neighbourhood $U \in \mathcal{U}(x)$ of x there exists a natural number $n \in \mathbb{N}$ such that for all $m \geq n$, $m \in \mathbb{N}$, the point x_m belongs to the neighbourhood U; we use the usual notation $\lim\limits_{n\to\infty} x_n = x$.

We say that $x \in X$ is an *accumulation point* of a sequence $\{x_n\}$ if for any neighbourhood U and for any natural $n \in \mathbb{N}$ there exists $m \geq n$, $m \in \mathbb{N}$ such that $x_m \in U$.

Similarly as in metric spaces (particularly as in real numbers) we can prove:

Theorem 2.12 *If X is a first countable topological space, $A \subset X$, $x \in X$, then the following holds:*

(1) the point x belongs to the closure $\overline{A}$ if and only if there exists a sequence $\{x_n\}$ of points from A, $x_n \in A$, such that $\lim\limits_{n\to\infty} x_n = x$;

(2) the point x is an accumulation point of the sequence $\{x_n\}$ if and only if there exists a subsequence $\{x_{k_n}\}$ of $\{x_n\}$ such that x is its limit point;

(3) the point x is an accumulation point of the set A if and only if there exists a sequence $\{x_n\}$ of points from the set $A \setminus \{x\}$ such that $x = \lim\limits_{n\to\infty} x_n$.

If we omit the assumption on first countability the theorem is no more true (we can construct examples [102]), because convergence of sequences depends not only on the sequence itself but also on the type of ordering of the local base for the point x. To substitute sequences in general topological spaces, we need to find some more general concept which would "work".

Recall that the relation of *directing* on a set D is an ordering of D which satisfies: if d and d' belong to D then there exists $d'' \in D$ such that $d \leq d''$ and $d' \leq d''$, and the pair $(D, \leq)$ is a *directed set*.

Definition 2.8 A *net* $\{x_d \colon d \in (D, \leq)\}$ is an arbitrary function (map) from a non-empty directed set $(D, \leq)$ to the topological space X.

A point $x \in X$ is said to be a *limit point* of a net $\{x_d\colon\ d \in D\}$ in X if for every neighbourhood U of x there exists an element $d_0 \in D$ such that for any $d \in D$ satisfying $d_0 \leq d$, x_d belongs to U; we write $x = \lim\limits_{d\in D} x_d$.

To demonstrate applications of the concept let us mention:

Theorem 2.13 *Let A be a subset of a topological space X. A point $x \in X$ belongs to the closure $\overline{A}$ if and only if there exists a net $\{x_d\colon\ d \in D\}$, $x_d \in A$ for $d \in D$, such that $x = \lim\limits_{d\in D} x_d$.*

Also subnets and accumulation points of a net can be introduced (for more details, [102]).

2.3 Continuous maps

In the family of maps from one topological space to the other, we prefer maps which "preserve topological structure." We start with continuous maps and show that they "pull back" open sets, i.e. preserve topological structure in one direction. Then we pass to one-one onto continuous maps with continuous inverse which peserve topological structure in both directions, and are topological equivalences.

After some experience with generalizing continuity of a real-valued function of real variable(s) to continuity of maps between metric spaces, it might seem quite natural to formulate continuity of maps between topological spaces also in terms of neighbourhoods. Recall that a subset U of (X, τ) is a neighbourhood of a point $x \in X$ if there is an open subset $O \subseteq X$ such that $x \in O \subset U$. If U itself is an open set we speak about an *open neighbourhood*.

2.3.1 Continuous maps of topological spaces

Let (X, τ) and (X', τ') be topological spaces.

Definition 2.9 A map $f\colon X \to X'$ is *continuous* in the point $x \in X$ if for any neighbourhood U' of the point $f(x) \in X'$ there exists a neighbourhood U of $x \in X$ such that $f(U) \subset U'$, i.e. $f(y) \in U'$ for each $y \in U$.

A map f is *continuous* if it is continuous in all points of the set X.

It is convenient to have some criteria for continuity of maps formulated in terms corresponding to various methods of defining or generating topologies. The notion of continuity is particularly easy to formulate in terms of open sets.

Theorem 2.14 *Let (X, τ), (X', τ') be topological spaces and $f\colon X \to X'$ a map of X to X'. The following properties are equivalent:*

(1) $f\colon X \to X'$ is continuous;
(2) inverse images of all open subsets of X' (sets from τ') are open in (X, τ);
(3) inverse images of all closed subsets of X' are closed in (X, τ);
(4) inverse images of all members of a subbase for X' are open in X;
(5) inverse images of all members of a base for X' are open in X;
(6) for every subset $A \subset X$ we have $f(\overline{A}) \subset \overline{f(A)}$;

(7) *for every subset* $B \subset X'$ *we have* $\overline{f^{-1}(B)} \subset f^{-1}(\overline{B})$*;*
(8) *for every point* $x \in X$ *and every net* $\{x_d\colon\ d \in D\}$ *in* X *with* $x = \lim\limits_{d\in D} x_d$*, the net of images* $\{f(x_d)\colon\ d \in D\}$ *in* X' *has a limit point equal* $f(x)$*.*

Further equivalences can be found e.g. in [66, p. 47].

Note that a continuous real-valued function of one real variable is continuous according to our new definition, [66, p. 49]. Important point is that continuity is preserved under maps composition; in the proof, the equality $(gf)^{-1}(A) = f^{-1}(g^{-1}(A))$ is used:

Theorem 2.15 *The composition of two continuos maps is continuos.*

Note that any map $f\colon X \to X'$ is continous whenever the topological space X is discrete, or whenever X' is antidiscrete.

Theorem 2.16 *Let* X, X' *be topological spaces and let* $A \subseteq X$ *have a subspace topology. Suppose* $f\colon X \to X'$ *is continuous. Then the restriction* $f|_A\colon A \to X'$ *is continuous.*

Theorem 2.17 (Glueing Lemma) *Let* X *and* Y *be topological spaces,* $X = A \cup B$*, where* A *and* B *are closed (open) subsets of* X*. Let* $f_1\colon A \to Y$ *and* $f_2\colon B \to Y$ *be continuous maps such that* $f_1(x) = f_2(x)$ *for all* $x \in A \cap B$*.*

Then the map $g\colon X \to Y$ *defined by* $g(x) = \begin{cases} f_1(x) \text{ for } x \in A, \\ f_2(x) \text{ for } x \in B, \end{cases}$ *is continuous.*

Note that without any assumption on A and B, the theorem is false.

2.3.2 Homeomorphisms

Together with topological spaces, we consider maps ("morphisms") that preserve topological structure:

Definition 2.10 A map $f\colon X \to X'$ is a *homeomorphism* (or a *topological map*) if the following conditions are satisfied:

1. f is one-one and onto map, i.e. there exists an inverse map f^{-1};
2. both the maps f and f^{-1} are continuous.

In this case the spaces X and X' are called *homeomorphic* which is denoted by

$$X \cong X'.$$

Since the identity map $\mathrm{id}_X\colon X \to X$ is a homeomorphism, the composition gf of two homeomorphisms f and g as well as the inverse map f^{-1} are again homeomorphisms the following can be checked:

1. $X \cong X$ – *reflexivity*,
2. $X \cong X' \Rightarrow X' \cong X$ – *symmetry*,
3. $X \cong X'$ and $X' \cong X'' \Rightarrow X \cong X''$ – *transitivity*.

Therefore the binary relation " $\cong$ " on the class of all topological spaces is an equivalence relation. Under homeomorphisms, open (closed) sets are mapped again onto open (closed) sets, hence the map f induces one-one onto correspondence between the topologies of X and X'. The topology of two spaces belonging to the same equivalence class is in a sense the same. That is why we identify homeomorphic spaces.

Theorem 2.18 *Let (X, τ), (X', τ') be topological spaces and $f\colon X \to X'$ a map of X to X'. The following properties are equivalent:*

(1) $f\colon X \to X'$ is a homeomorphism;
(2) G is open in τ if and only if $f(G)$ is open in τ';
(3) F is closed in τ' if and only if $f^{-1}(F)$ is closed in τ;
(4) O is open in τ' if and only if $f^{-1}(O)$ is open in τ;
(5) A is closed in τ if and only if the image $f(A)$ is closed in τ';
(6) U is a neighbourhood of $x \in X$ if and only if $f(U)$ is a neighbourhood of $f(x) \in X'$;
(7) for every subset $A \subset X$ we have $f(\overline{A}) = \overline{f(A)}$;
(8) for every net $\{x_d\colon d \in D\}$ in X,

$$x = \lim_{d \in D} x_d \text{ holds if and only if } f(x) = \lim_{d \in D} f(x_d).$$

Further equivalences can be found e.g. in [66, p. 54]. Continuous maps and homeomorphisms of abstract spaces were first considered by M. Fréchet (1910).

If $f\colon X \to Y$ is a one-one map of topological spaces, and if $f\colon X \to f(X)$ is a homeomorphism when we give $f(X)$ the induced topology from Y, we call f an *embedding* of X to Y.

Recall that a map between topological spaces is *open (closed)* if the image of each open (closed) set is open (closed).

Theorem 2.19 *A one-one onto map of topological spaces is a homeomorphism if and only if it is continuous and open.*

2.3.3 Topological invariants

If a property of a topological space is preserved under homeomorphisms it is called a topological property or a *topological invariant*.

The object of topology is to study topological properties. Roughly, every property defined in terms of open sets and in terms of set theory is a topological invariant.

We should say at once that there is no hope of classifying all topological spaces. However, there are techniques which anable us to decide whether two spaces are homeomorphic or not. Showing that two spaces are homeomorphic is rather a geometrical problem, involving the construction of a specific homeomorphism between given spaces, and the techniques used vary with the problem.

On the other hand, a problem of an entirely different nature is attempting to prove that two spaces are not homeomorphic to one another: in this case we look for topological invariants trying to find a topological property in which the spaces differ. It might be one of the well-known topological properties (some of them will be discussed in the sequel) such as countability, existence of special bases, connectedness, compactness, separation properties, or an algebraic structure, such as a group or ring constructed from the space (e.g. fundamental group, homotopy groups, homology groups), or number (e.g. Euler number defined for the surface, Betti numbers[56)]) etc.

[56)] Eduard Čech, 1893-1960, was a Czech mathematician. His research interests included projective differential geometry and topology. He is especially known for the technique known as Stone-Čech compactification (in topology) and the notion of Čech cohomology. See [397].

2.4 Constructions of new topological spaces from given spaces

2.4.1 Projectively and inductively generated topologies (initial and final)

Let us describe methods of generating topologies based on the concept of a continuous map. The following four constructions are particularly useful: (topological) product, subspace, sum (= disjoint union) and quotient. Note that first two constructions are particular cases of a more general construction of the so-called projectively generated topologies while sum and quotient are particular cases of the so-called inductively generated topologies.

Theorem 2.20 *Let X be a set, $\{(Y_t,\tau_t)\colon\ t\in T\}$ a family of topological spaces and $\{f_t\colon\ t\in T\}$ a system of maps where $f_t\colon X\to Y_t$. In the class of all topologies on X that make all maps f_t continuos there exists a coarsest topology τ. One of its bases consists of all sets of the form $\bigcup_{i=1}^{k} f_{t_i}^{-1}(V_i)$ where V_i is open in Y_{t_i}, $t_1,t_2,\dots,t_k\in T$ for $i=1,2,\dots,k$.*

The topology τ is called the topology *projectively determined, or projectively generated, by the system of maps* $\{f_t\colon\ t\in T\}$, also *initial topology*.

Notice that all sets of the form $f_t^{-1}(V_t)$ where V_t is open in Y_t form a subbase for the initial topology [66, p. 51].

Theorem 2.21 *A map f of a topological space (X,τ) to a topological space (X',τ') whose topology is generated projectively by a family of maps $\{f_t\colon\ t\in T\}$ where f_t is a map of X' to X'_t, is continuous if and only if every composite map f_tf is continuous for $t\in T$.*

Now let us assume the "dual" situation, when all arrows in the considered maps are reversed.

Theorem 2.22 *Let X be a set, $\{(Y_t,\tau_t)\colon\ t\in T\}$ a system of topological spaces and $\{f_t\colon\ t\in T\}$ a system of maps where $f_t\colon Y_t\to X$. In the class of all topologies on X that make all maps f_t continuos there exists a finest topology τ. Open sets of this topology are exactly all sets $G\subseteq X$ satisfying $f_t^{-1}(G)\in\tau_t$ for all $t\in T$.*

The topology τ is called the topology *inductively generated on X by the system of maps* $\{f_t\colon\ t\in T\}$, also *final topology*.

Theorem 2.23 *Let $f\colon(X,\tau)\to(X',\tau')$ be a map of topological spaces and let τ be a topology inductively generated on X by a family of maps $\{f_t\colon\ t\in T\}$ where $f_t\colon Y_t\to X$. Then the map f is continuous if and only if every composition $f\circ f_t$ is continuous for $t\in T$.*

2.4.2 Subspace and product

If $A\subseteq X$, we introduce a *subspace topology* on A induced by the topology τ on X (or *relative topology*) as follows:

$$\tau_A=\{Y\cap U\colon\ U\in\tau\}.$$

The topological space (A,τ_A) is called a *topological subspace* in (X,τ). It can be checked that a subspace topology on a subset $A\subset X$ of a topological space (X,τ) is just the topology projectively generated by the one-element system $\{j\}$ where $j\colon A\to X$ is the "*canonical identical embedding*", $j(y)=y$ for every $y\in A$.