CZECH TECHNICAL UNIVERSITY IN PRAGUE

FACULTY OF MECHANICAL ENGINEERING

MATHEMATICS II

May 21, 2026

Experimental version of the lecture notes for the academic year 2025/2026

Based on: *J. Neustupa: Matematika II. Skriptum Strojní fakulty. Vydavatelství ČVUT, Praha 2015*

Contents

I. Differential calculus of functions of several variables **2**

I.1 Euclidean space $\mathbb{E}_n$. Points and sets in $\mathbb{E}_n$ 2
I.2 Functions of several variables, basic concepts 8
I.3 Limit and continuity 14
I.4 Partial derivatives. Differentiable function. Tangent plane. Differential 21
I.5 Directional derivative. Gradient. Higher-order partial derivatives. Partial derivatives of a composite function. 29
I.6 Extrema of functions of several variables 37
I.7 Implicit functions 44
I.8 Exercises 48

II. Riemann's integral of functions of two and three variables **51**

II.1 Double integral – motivation and definition. Two-dimensional Jordan measure and measurable sets in $\mathbb{E}_2$ 51
II.2 Existence and selected properties of the double integral 56
II.3 Computation of the double integral – Fubini's theorem and transformation to polar coordinates 57
II.4 Some physical applications of the double integral 62
II.5 Triple integral – motivation and definition. Three-dimensional Jordan measure and measurable sets in $\mathbb{E}_3$ 63
II.6 Existence and selected properties of the triple integral 66
II.7 Computation of the triple integral – Fubini's theorem and transformation to cylindrical and spherical coordinates 67
II.8 Some physical applications of the triple integral 72
II.9 * More on the substitution method for computing double and triple integrals 73
II.10 Exercises 76

III. Line integrals **79**

III.1 Simple curves 79
III.2 Line integral of a scalar function (Line integral of the first kind) 83
III.3 Some physical applications of the line integral of a scalar function 86
III.4 Line integral of a vector function (Line integral of the 2nd kind) 88
III.5 Green's theorem 91
III.6 Exercises 94

IV. Surface integrals **97**

IV.1 Simple surfaces 97
IV.2 Surface integral of a scalar function (Surface integral of the 1st kind) 103
IV.3 Some physical applications of the surface integral of a scalar function 107
IV.4 Surface integral of a vector function (Surface integral of the 2nd kind) 108
IV.5 Operators of divergence and curl. Gauss–Ostrogradsky (divergence) theorem and Stokes' theorem 112
IV.6 Exercises 117

V. Potential and solenoidal vector fields **121**
V.1 Path-independence of the line integral of a vector function. Potential field – definition and basic properties 121
V.2 Potential fields in $\mathbb{E}_2$ 126
V.3 Potential fields in $\mathbb{E}_3$ 131
V.4 * Solenoidal fields 134
V.5 Exercises 140

Introduction

This text is a study aid for the course Mathematics II, taught at the Faculty of Mechanical Engineering of CTU in the summer semester in the first year of study. It follows the lecture notes Mathematics I (see [5]). The text also contains some parts which are not usually required in the examination in the course Mathematics II; nevertheless, they are closely related to the content of the lecture notes. These parts, either chapters or sections, are marked with an asterisk. If students do not use them immediately, they may return to them later in case of need or interest.

The content of the course Mathematics II, and also the content of this text, is in particular the differential and integral calculus of functions of several variables. This material has many applications in engineering and in the natural sciences. The concepts with which we deal here have a multidimensional character. (When studying Mathematics II you will understand precisely what we mean by this.) Their good and active mastery therefore requires not only knowledge of the basic definitions, theorems, formulas and a certain computational skill, but also good spatial imagination. In explaining such concepts, precisely because of the frequent possibility of their geometric interpretation, for example in three-dimensional space, the teacher's personality has an irreplaceable role. Despite every effort, a written text with two-dimensional figures cannot always replace the teacher's exposition, and therefore we recommend to students that they understand these lecture notes as a supplement to lectures and tutorials, where the material will be explained in detail.

In order to facilitate the understanding of many concepts, before their precise definitions there are placed sections in which we try to describe the motivation for introducing the concepts and to explain why these concepts are defined in the given way. This concerns especially various types of integrals, curves and surfaces. In the sections which have a motivating character, we often leave the exact manner of exposition and allow ourselves on an intuitive level to work even with terms such as, for example, "an infinitely small positive number dt", "an infinitely short segment of a curve", etc., which contradict the present-day conception of the set of real numbers or the geometry of curves. Several further concepts (orientation of a closed curve in the plane, orientation of a surface with respect to the orientation of its boundary, simply connected domain) we introduce in such a way that we rely on the reader's imagination and prefer simplicity and clarity at the expense of mathematical precision. In the other parts of the notes we try, at an acceptable but also understandable level, to keep the logical correctness of the exposition. Keep this in mind when reading the notes and distinguish the above-mentioned two levels of logical rigor in different parts of the text.

The notes contain a number of well-known and important theorems of applied mathematics, such as Green's theorem, the Gauss–Ostrogradsky theorem, Stokes' theorem, etc. The statements of these theorems are expressed by means of certain integral formulas. Students, and also the authors of various professional articles, sometimes incorrectly reduce these theorems merely to the mentioned formulas and forget that an integral part of the theorems are also their assumptions. It would be naive to suppose that the formulas are valid at any time. The opposite is true – these formulas are valid only under certain conditions. The assumptions of the theorems are a concise and simple specification of these conditions and are just as important as the statements of the theorems. Therefore do not forget them.

In the notes some examples are solved as models. Other, unsolved examples are included at the end of each thematic unit. Many further examples, solved and unsolved, can be found in the lecture notes of E. Brožíková and M. Kittlerová [1].

CHAPTER I.

Differential calculus of functions of several variables

I.1 Euclidean space $\mathbb{E}_n$. Points and sets in $\mathbb{E}_n$ 2
I.2 Functions of several variables, basic concepts 8
I.3 Limit and continuity 14
I.4 Partial derivatives. Differentiable function. Tangent plane. Differential 21
I.5 Directional derivative. Gradient. Higher-order partial derivatives. Partial derivatives of a composite function. 29
I.6 Extrema of functions of several variables 37
I.7 Implicit functions 44
I.8 Exercises 48

Functions of several variables in mathematics and its applications describe situations where one quantity (the so-called dependent variable) depends on several other quantities (the so-called independent variables). In various places in this text you will find many concrete examples of such functions.

You will see that the theory of functions of several variables differs at some points from the theory of functions of one variable, which you know from the course Mathematics I. However, between the theory of functions of two, three, or even more variables there is no very large difference. In this text we shall deal with functions which depend in general on n variables. If the exposition seems too abstract to you, imagine that $n = 2$. In order to keep the text understandable, many concepts will first be explained on the example of a function of two variables and only then we will show their generalization to functions of n variables.

We shall usually denote variables by $x_1, x_2, \ldots, x_n$, sometimes also by $t_1, t_2, \ldots, t_k$. In the concrete case when $n = 2$, we shall usually denote the variables by x_1, x_2 or x, y. Similarly, for $n = 3$ we shall usually denote the variables by x_1, x_2, x_3 or x, y, z. You see that variables in general form ordered n-tuples of real numbers. They can therefore be considered as elements of the n-dimensional Euclidean space $\mathbb{E}_n$. A function of n variables in general need not be defined on the whole space $\mathbb{E}_n$, but only on some set in $\mathbb{E}_n$. Properties of a function of n variables are closely related to the type of the set on which the function is defined. In the next chapter we therefore recall the definition of the space $\mathbb{E}_n$ (known from Mathematics I) and then we become acquainted with various types of sets in $\mathbb{E}_n$.

I.1 Euclidean space $\mathbb{E}_n$. Points and sets in $\mathbb{E}_n$

I.1.1. n-dimensional Euclidean space $\mathbb{E}_n$

Assume that n is a natural number. The set of all ordered n-tuples of real numbers we denote by $\mathbb{R}^n$ and call the *n-dimensional arithmetic space*. Elements of $\mathbb{R}^n$ are called points. We denote points by capital letters and their coordinates (ordered n-tuples) we write in square brackets. For example $A = [a_1; \ldots; a_n]$, $X = [x_1; \ldots; x_n]$, etc.

If we define in $\mathbb{R}^n$ the distance of any two points $X = [x_1; \ldots; x_n]$ and $Y = [y_1; \ldots; y_n]$ by the formula

$$\|Y - X\| = \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2}, \tag{I.1.1}$$

then $\mathbb{R}^n$ becomes the so-called *n-dimensional Euclidean space*. We denote this space by $\mathbb{E}_n$.

The distance of points X and Y is sometimes also denoted by $|Y-X|$, $|X-Y|$, $d(X,Y)$ or $\rho(X,Y)$. In this text, however, we will prefer the notation $\|Y-X\|$ or $\|X-Y\|$. (It is the same.)

The origin of the coordinate system in $\mathbb{E}_n$, i.e. the point whose all coordinates are zero, we will denote by O.

From secondary school and from the course Mathematics I you know that you can imagine $\mathbb{E}_1$ as a line, $\mathbb{E}_2$ as a plane and $\mathbb{E}_3$ as the space in which we live. (It is only an approximate idea; astronomers have long ago found that our universe in reality is not $\mathbb{E}_3$.)

By a vector in $\mathbb{E}_n$ we shall always mean the so-called free vector in $\mathbb{E}_n$. (See the lecture notes [5].) In this text we shall usually denote vectors by lowercase bold letters. (Another possibility is to put an arrow above the vector's symbol.) We shall write coordinates of a vector in round brackets. Thus, for example, we may write: $\vec{a}=(a_1;\dots;a_n)$, $\vec{t}=(t_1;\dots;t_n)$, etc.

By the sum of a point $X=[x_1;\dots;x_n]$ and a vector $\vec{a}=(a_1;\dots;a_n)$ we mean the point $X+\vec{a}=[x_1+a_1;\dots;x_n+a_n]$. Conversely, we consider the difference of two points $Y=[y_1;\dots;y_n]$ and $X=[x_1;\dots;x_n]$ as the vector $Y-X=(y_1-x_1;\dots;y_n-x_n)$.

The length of the vector $\vec{a}=(a_1;\dots;a_n)$ is the number

$$\|\vec{a}\| = \sqrt{a_1^2+\dots+a_n^2}. \tag{I.1.2}$$

(Thus the distance $\|Y-X\|$ of two points X, Y can be understood as the length of the vector $Y-X$.)

Draw for yourselves figures in $\mathbb{E}_2$ for all concepts defined in this chapter!

I.1.2. Neighborhood of a point in $\mathbb{E}_n$. Punctured neighborhood of a point in $\mathbb{E}_n$.

A neighborhood of the point A of radius R means the set $\{X\in\mathbb{E}_n;\ \|X-A\|<R\}$ (i.e. the set of all points $X\in\mathbb{E}_n$ whose distance from A is less than R). We denote such a neighborhood by $\mathcal{U}_R(A)$. ($\mathcal{U}$ because the notion "neighborhood of a point" arose in Germany and in German a neighborhood is called "die Umgebung".) If the radius of the neighborhood is not important, we denote it only by $\mathcal{U}(A)$.

In the special case when $n=1$ and A is the point with coordinate a on the real axis (i.e. $A=[a]$), the neighborhood $\mathcal{U}_R(A)$ is the open interval $(a-R;a+R)$. Similarly, if $n=2$, then the neighborhood $\mathcal{U}_R(A)$ is the open disk with center at the point A and radius R.

If we remove from the neighborhood $\mathcal{U}_R(A)$ its center A, we obtain the set $\mathcal{U}_R(A)\setminus\{A\}=\{X\in\mathbb{E}_n;\ 0<\|X-A\|<R\}$. We call this set the *punctured neighborhood* of the point A of radius R and denote it by $\mathcal{P}_R(A)$. If the radius of the punctured neighborhood is not important, we may denote it only by $\mathcal{P}(A)$.

Instead of "punctured neighborhood" one sometimes uses the names "deleted neighborhood" or "reduced neighborhood"; however, in this text we will keep the name "punctured neighborhood".

I.1.3. Interior point of a set. Boundary point of a set

Let $M\subset\mathbb{E}_n$ and $A\in M$. We call the point A an *interior point* of the set M if there exists a neighborhood $\mathcal{U}(A)$ such that $\mathcal{U}(A)\subset M$.

The neighborhood $\mathcal{U}(A)$ in the definition of an interior point may be arbitrarily small. What matters, however, is that some such neighborhood exists.

We call the point A a *boundary point* of the set M if in every neighborhood of it there is at least one point of the set M and at least one point which does not belong to M.

In the definition of a boundary point the word "every" is very important. That also means "arbitrarily large", or rather "arbitrarily small". A boundary point of the set M may itself belong to M, but it need not.

I.1.4. Interior of a set. Open set

Let $M \subset \mathbb{E}_n$. The set of all interior points of M is called the *interior* of the set M. We denote it by M°.

A set $M \subset \mathbb{E}_n$ is called *open* if $M = M^\circ$. (In other words: the set M is open if each of its points is an interior point.)

I.1.5. Boundary of a set

Let $M \subset \mathbb{E}_n$. The set of all boundary points of M is called the *boundary* of M. We denote it by ∂M or Γ_M. (In this text we will prefer ∂M.)

The point A in Fig. 1 is an interior point of the set M, the point B is its boundary point.

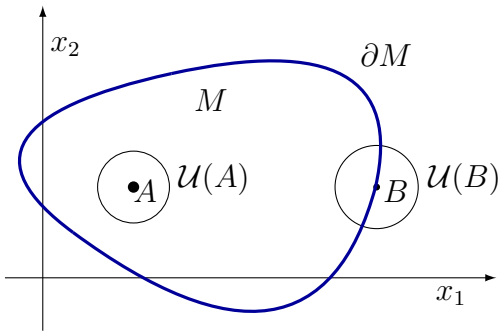


Fig. 1

I.1.6. Closure of a set. Closed set

Let $M \subset \mathbb{E}_n$. The *closure* of M is the union $M \cup \partial M$. We denote the closure of M by $\overline{M}$.

We call the set M *closed* if $M = \overline{M}$.

I.1.7. Examples

1. First assume that $n = 1$ and M is the open interval $(0; 1)$.

   Each point of the interval M is at the same time its interior point. Thus M is, in accordance with Definition I.1.4, an open set. The boundary points are 0 and 1. Therefore the closure of the interval M is the closed interval $\langle 0; 1\rangle$.

2. Now assume that $n = 1$ and M is the set of all natural numbers $\mathbb{N}$. This set M has no interior points, therefore $M^\circ = \emptyset$. Each point of the set M is its boundary point, therefore $\partial M = M$. The closure of M is $\overline{M} = M \cup \partial M = M$. The set M coincides with its closure, hence it is closed.

3. Let $n = 1$ and $M = \mathbb{E}_1$. Each point of M is an interior point, hence M is open. The set M has no boundary points, therefore $\partial M = \emptyset$. The closure of M is $\overline{M} = M \cup \partial M = M \cup \emptyset = M$. The set M equals its closure, therefore it is also closed.

   This is somewhat surprising: the set $M = \mathbb{E}_1$ in $\mathbb{E}_1$ is at the same time open and closed. Similarly one can show that the set $M = \mathbb{E}_n$ in $\mathbb{E}_n$ is at the same time open and closed. One can prove that the only other set in $\mathbb{E}_n$ with this property is the empty set.

4. Now let $n = 2$ and $M = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_2 > 0\}$. Each point of M is its interior point. Therefore $M = M^\circ$ and the set M is open. The boundary of M is the line $x_2 = 0$ (the x_1-axis). The closure of M is $\overline{M} = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_2 \geq 0\}$ (the union of M and the x_1-axis).

5. Let $n = 2$ and $M = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1 \geq 0, \ x_2 > 0\}$. The interior of M is $M^\circ = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1 > 0, \ x_2 > 0\}$ (the open first quadrant). The boundary of M is the union of two rays: $\partial M = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1 \geq 0, \ x_2 = 0\} \cup \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1 = 0, \ x_2 \geq 0\}$. The closure is $\overline{M} = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1 \geq 0, \ x_2 \geq 0\}$ (the closed first quadrant). The set M is neither open nor closed.

6. Let $n = 2$ and $M = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1^2 + x_2^2 \leq 4\}$. The interior of M is $M^\circ = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1^2 + x_2^2 < 4\}$. The boundary is $\partial M = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1^2 + x_2^2 = 4\}$ (the circle with center at the origin and radius 2). The closure is $\overline{M} = M$. The set M is closed.

I.1.8. Some further properties of sets

It is not difficult to prove that the following statements or identities hold:

1. *The union or intersection of finitely many open sets in $\mathbb{E}_n$ is an open set.*
2. *The union or intersection of finitely many closed sets in $\mathbb{E}_n$ is a closed set.*

Further assume that $M \subset \mathbb{E}_n$.

3. *For every point $A \in M$ exactly one of the following two possibilities occurs:*
   a) A is an interior point of M,
   b) or A is a boundary point of M.
4. $\partial M = \partial(\mathbb{E}_n \backslash M) = \overline{M} \cap \overline{\mathbb{E}_n \backslash M}$
5. *∂M is a closed set.*
6. $M^\circ = M \backslash \partial M$
7. *M is open if and only if it contains none of its boundary points (i.e. if $M \cap \partial M = \emptyset$).*
8. *M is closed if and only if it contains all of its boundary points (i.e. if $\partial M \subset M$).*
9. *M is open if and only if its complement $\mathbb{E}_n \backslash M$ is closed.*

I.1.9. Line segment in $\mathbb{E}_n$. Polygonal line in $\mathbb{E}_n$

If A and B are two distinct points in $\mathbb{E}_n$, then the set of all points X of the form $X = A + t\,(B - A)$ (for $t \in \langle 0; 1 \rangle$) is called the *line segment* in $\mathbb{E}_n$ with *endpoints* A and B. We denote it by $\overline{AB}$.

Let A_1, A_2, ..., A_k be k pairwise distinct points in $\mathbb{E}_n$. The union of the segments $\overline{A_1A_2}$, $\overline{A_2A_3}$, ..., $\overline{A_{k-1}A_k}$ is called a *polygonal line* in $\mathbb{E}_n$.

I.1.10. Domain in $\mathbb{E}_n$

A set $D \subset \mathbb{E}_n$ which

a) is open, and

b) any two of its points can be connected by a polygonal line which lies entirely in D,

is called a *domain*.

Property b) is called *connectedness*. Thus a domain in $\mathbb{E}_n$ can also be defined as a set in $\mathbb{E}_n$ which is open and connected.

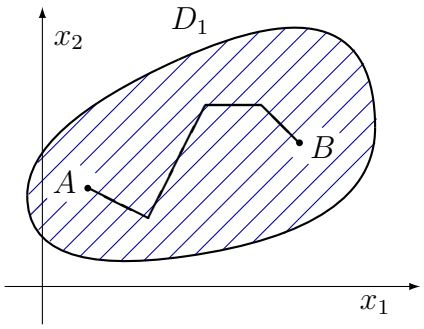


Fig. 2a

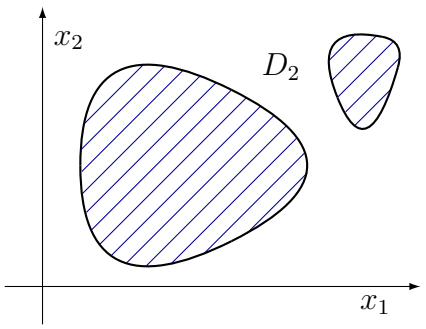


Fig. 2b

I.1.11. Examples

1. The set D_1 in Fig. 2a is connected. The set D_2 in Fig. 2b is not connected. (It consists of two components.)

2. The sets

$$D_3 = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1^2 + x_2^2 < 9\},$$
$$D_4 = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1 < 0\},$$
$$D_5 = (-1; 1) \times (-1; 1)$$

are domains in $\mathbb{E}_2$.

3. The sets

$$D_6 = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1^2 + x_2^2 \leq 9\},$$
$$D_7 = \{X = [x_1; x_2] \in \mathbb{E}_2 : \ x_1 < 0 \text{ or } x_1 > 5\}$$

are not domains in $\mathbb{E}_2$. (D_6 is not open and D_7 is not connected.)

I.1.12. Bounded set in $\mathbb{E}_n$

A set $M \subset \mathbb{E}_n$ is called *bounded* if there exists a neighborhood of the origin $\mathcal{U}_R(O)$ such that $M \subset \mathcal{U}_R(O)$.

(You can surely imagine that in order for a bounded set M to be contained in $\mathcal{U}_R(O)$, the radius R must be "sufficiently large". But that does not matter. What is essential is that some "sufficiently large" radius R and a neighborhood $\mathcal{U}_R(O)$ containing the set M exist.)

A set in $\mathbb{E}_n$ which is not bounded is called *unbounded*.

Find and write down your own examples of bounded and unbounded sets in $\mathbb{E}_n$.

I.1.13. Sequence of points in $\mathbb{E}_n$

Any mapping from the set of natural numbers $\mathbb{N}$ into $\mathbb{E}_n$ is called a *sequence* in $\mathbb{E}_n$. Thus a sequence in $\mathbb{E}_n$ is an assignment which assigns to each natural number k a point X_k in $\mathbb{E}_n$. We denote a sequence by $\{x_1; x_2; x_3, \dots\}$ or $\{X_k\}_{k=1}^{\infty}$ or briefly just $\{X_k\}$.

I.1.14. Limit of a sequence in $\mathbb{E}_n$

A point $A \in \mathbb{E}_n$ is called the *limit* of the sequence $\{X_k\}$ if

(I.1.3) $$\lim_{k \to +\infty} \|X_k - A\| = 0.$$

We use the notation: $\lim\limits_{k \to +\infty} X_k = A$, $\lim X_k = A$ or simply $X_k \longrightarrow A$.

A sequence of points $\{X_k\}$ in $\mathbb{E}_n$ which has a limit in $\mathbb{E}_n$ is called *convergent*. If the limit is the point A, we say that the sequence $\{X_k\}$ *converges* to the point A.

I.1.15. Remark
In other words, we may say that A is the limit of the sequence $\{X_k\}$ if the sequence of distances of X_k from A has limit zero. The sequence of distances of X_k from A is a sequence of real numbers, and the notion of the limit of a sequence of real numbers is known from the course Mathematics I.

From Mathematics I you also know that a sequence of real numbers (which may be regarded as a sequence of points in $\mathbb{E}_1$) can have at most one limit (see [5], Theorem III.1.8). The same statement holds for a sequence of points in $\mathbb{E}_n$:

I.1.16. Theorem
A sequence of points $\{X_k\}$ in $\mathbb{E}_n$ can have at most one limit.

Indeed, if $\{X_k\}$ is a sequence of points in $\mathbb{E}_n$, then clearly it is not possible that there exist two distinct points A and B such that the distances of X_k from A tend to zero as $k \to +\infty$ and at the same time the distances of X_k from B tend to zero as $k \to +\infty$.

I.1.17. Theorem
Let $\{X_k\}$ be a sequence of points in $\mathbb{E}_n$, where $X_k = [x_{1k}; \dots; x_{nk}]$. Let $A = [a_1; \dots; a_n] \in \mathbb{E}_n$. Then

$$\lim_{k\to+\infty} X_k = A \quad \Longleftrightarrow \quad \lim_{k\to+\infty} x_{ik} = a_i \text{ for all } i = 1, 2, \dots, n.$$

This theorem can easily be proved if we use (I.1.3) and carefully expand $\|X_k - A\|$ using (I.1.1).

Theorem I.1.17 says that the limit of a sequence of points $\{X_k\}$ can be computed "coordinatewise". That means: the first coordinate of the limit point A equals the limit of the first coordinates of the points X_k (i.e. $\lim\limits_{k\to+\infty} x_{1k} = a_1$), the second coordinate of the limit point A equals the limit of the second coordinates of the points X_k (i.e. $\lim\limits_{k\to+\infty} x_{2k} = a_2$), etc.

I.1.18. Example Decide whether the sequence of points

$$X_k = \left[\frac{\sin k}{k};\ \frac{k^2 - 7k}{6 - 5k - 2k^2};\ \frac{k+1}{k+5}\right], \quad (k = 1, 2, \dots)$$

is convergent in $\mathbb{E}_3$. If it is, determine its limit.

S o l u t i o n : The considered sequence can be written as $\{X_k\} = \{\,[x_{1k};\ x_{2k};\ x_{3k}]\,\}$, where

$$x_{1k} = \frac{\sin k}{k}, \qquad x_{2k} = \frac{k^2 - 7k}{6 - 5k - 2k^2}, \qquad x_{3k} = \frac{k+1}{k+5}.$$

For the limits of the sequences $\{x_{1k}\}$, $\{x_{2k}\}$ and $\{x_{3k}\}$ we have:

$$\lim_{k\to+\infty} x_{1k} = \lim_{k\to+\infty} \frac{\sin k}{k} = 0, \qquad \lim_{k\to+\infty} x_{2k} = \lim_{k\to+\infty} \frac{k^2 - 7k}{6 - 5k - 2k^2} = -\frac{1}{2},$$

$$\lim_{k\to+\infty} x_{3k} = \lim_{k\to+\infty} \frac{k+1}{k+5} = 1.$$

All sequences $\{x_{1k}\}$, $\{x_{2k}\}$ and $\{x_{3k}\}$ have finite limits and are therefore convergent. By Theorem I.1.17 we conclude: the sequence $\{X_k\}$ is convergent in $\mathbb{E}_3$ and

$$\lim_{k\to+\infty} X_k = \left[0;\, -\tfrac{1}{2};\, 1\right].$$

I.1.19.* Accumulation point of a set. Isolated point of a set
Assume that $M \subset \mathbb{E}_n$. A point $A \in \mathbb{E}_n$ is called an *accumulation point* of the set M if there exists at least one sequence of points $\{X_k\}$ in M, different from A, converging to the point A.

An accumulation point of the set M may itself belong to M, but it need not. The name is derived from the fact that in an arbitrarily small neighborhood of an accumulation point, points of the set M "accumulate", i.e. one always finds infinitely many of them there.

A point $A \in M$ is called an *isolated point* of the set M if A is not an accumulation point of the set M.

I.1.20.* Examples
1. Suppose that $\mathbb{E}_n = \mathbb{E}_1$ and M is the open interval $(0; 1)$.

   Then every point of the closed interval $\langle 0; 1\rangle$ is an accumulation point of M. The set M has no isolated points.

2. Let $\mathbb{E}_n = \mathbb{E}_1$ and $M = \mathbb{N}$ (the set of natural numbers). This set M has no accumulation points. Each of its points is at the same time an isolated point.
3. Let $\mathbb{E}_n = \mathbb{E}_1$ and $M = \{1;\, \frac{1}{2};\, \frac{1}{3};\, \frac{1}{4};\, \dots\}$. Each point of M is an isolated point. The only accumulation point of M is the point 0, which, however, does not belong to M.
4. Let $\mathbb{E}_n = \mathbb{E}_2$ and $M = \{X = [x_1; x_2] \in \mathbb{E}_2 :\ x_1^2 + x_2^2 < 2\}$. ($M$ is the open disk with center at the origin and radius $\sqrt{2}$.) This set M has no isolated points. The accumulation points of M form the set $\overline{M} = \{X = [x_1; x_2] \in \mathbb{E}_2 :\ x_1^2 + x_2^2 \le 2\}$.

I.1.21.* Some further properties of accumulation and isolated points
It is not difficult to prove the following statements:

1. *A point $A \in M$ is an isolated point of the set M if and only if there exists a neighborhood $\mathcal{U}(A)$ in which, besides the point A, there is no other point of the set M.*
2.* *M is closed if and only if M contains all of its accumulation points.*

I.2 Functions of several variables, basic concepts

I.2.1. Functions of several variables – motivation and examples
In mathematical modeling of various phenomena, states and processes, cases very often occur in which one variable quantity depends on a larger number of other variable quantities. Let us give some examples:

– The volume V of a cone whose base is a disk depends on the base radius r and the cone height v as follows: $V = \frac{1}{3}\pi r^2 v$.

– The magnitude of the force by which a material point of mass M, located at the origin of the coordinate system in $\mathbb{E}_3$, attracts another material point of mass m with coordinates $[x_1; x_2; x_3]$ is

$$F = \frac{\kappa M m}{x_1^2 + x_2^2 + x_3^2}.$$

(κ is Newton's gravitational constant.) Thus F depends on M, m, x_1, x_2 and x_3.

– According to a simplified mathematical model (Ch. Cobb and P. Douglas, 1928), the total production P of an economy in one year depends on the amount of labor performed L and the amount of invested capital K as follows: $P = 1.01\, L^{0.75} K^{0.25}$.

– For some functions we know that they exist, but we either do not know their expression by a formula (prescribing the performance of finitely many operations), or this is not possible at all. For example, to each point $X = [x_1; x_2; x_3]$ in a room one can assign a temperature T. Moreover, the temperature depends on time t. Thus T is a function of four variables: x_1, x_2, x_3 and t. An exact expression for T, in a room of complicated shape and with various heat sources, is practically impossible. We can only approximately describe T by a table of temperatures at chosen points X and at chosen instants t. Such a table can be obtained either by measurement or by computation; however, a description of the computation would far exceed the content of the course Mathematics II.

I.2.2. Function of n variables. Domain. Notation of a function

Assume that $n \in \mathbb{N}$ and $M \subset \mathbb{E}_n$, $M \neq \emptyset$. A mapping f from the set M into $\mathbb{R}$ is called a *function of n variables*. (Since the values of f are real numbers, we also speak of a real function of n variables.) The set M is called the *domain (of definition)* of the function f and we denote it by $D(f)$.

A function f, defined on a set $M \subset \mathbb{E}_n$, assigns to each point $X = [x_1; \ldots; x_n] \in M$ a real number, which we denote by $f(X)$ or $f(x_1, \ldots, x_n)$. The quantities $x_1, \ldots, x_n$ are called *independent variables*. If we define f by the equation $y = f(x_1, \ldots, x_n)$, then we say that y (or the function f itself) "depends" on the variables $x_1, \ldots, x_n$.

The same rules for notation apply to functions of several variables as for functions of one variable. On a completely exact level, therefore, the difference between f and $f(X)$ is this: while f denotes the function, $f(X)$ is a number – namely the value of the function f at the point X. Similarly to functions of one variable, however, we will not observe this rule entirely consistently here either, and we will often denote the function both by f and, for example, by $f(X)$, $f(x_1, x_2)$, $f(x_1, \ldots, x_n)$, $f(x, y)$, etc. The first notation (i.e. f) has the advantage that it is exact and short. The other notation (e.g. $f(x_1, x_2)$) has the advantage that it immediately shows on how many and which variables f depends.

We will also often denote functions by letters other than f, for example g, h, F, etc.

Let f be, for example, the function which assigns to each point $X = [x_1; x_2]$ from the set $M = \{[x_1; x_2] \in \mathbb{E}_2 : \ x_2 \geq 1\}$ the number $x_1 + \sqrt{x_2}$. We can write this fact as follows:

1. $f \ : \ y = x_1 + \sqrt{x_2}$ for $x_2 \geq 1$,

2. $f(x_1, x_2) = x_1 + \sqrt{x_2}$; $[x_1; x_2] \in M$,

etc. Keep in mind the following important rule, which you already know from the theory of functions of one variable: if the domain is not given explicitly (i.e. expressly) in the definition of the function, then we consider as the domain the set of all possible values of the independent variables for which the formula defining the function makes sense
Thus, if we define the function g, for example, by the equation

$$g(x, y, z) = x \ln(y + z) \tag{I.2.1}$$

without further specification of the domain, then by the domain of g we mean the set of all points $[x; y; z] \in \mathbb{E}_3$ for which the expression on the right-hand side of (I.2.1) makes sense. It is the set $\{[x; y; z] \in \mathbb{E}_3 : \ y + z > 0\}$. We write:

$$D(g) \ = \ \{[x; y; z] \in \mathbb{E}_3 : \ y + z > 0\}.$$

I.2.3. Range. Graph
Assume that f is a function of n variables defined on a set $M \subset \mathbb{E}_n$.

By the *range* of f we mean the set

$$R(f) = \{y \in \mathbb{R};\ \exists\, X \in M \ : \ y = f(X)\}.$$

The *graph* of f is the set

$$G(f) = \{[X; y] \in \mathbb{E}_{n+1};\ X \in M \text{ and } y = f(X)\}.$$

From secondary school and from the course Mathematics I you know that the graph of a "reasonable" function of one variable is a curve in $\mathbb{E}_2$. (By a "reasonable" function we mean, for example, a continuous function; however, we will not discuss this in detail here.)

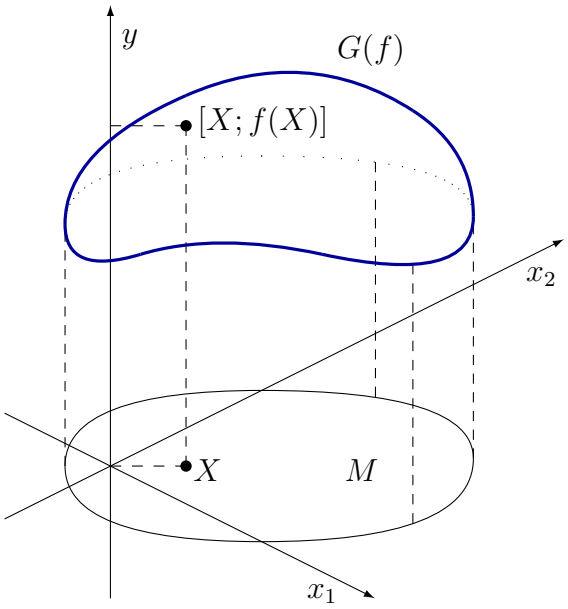


Fig. 3

Similarly, the graph of a "reasonable" function of two variables is a surface in $\mathbb{E}_3$. (See Fig. 3.) Analogously, the graph of a "reasonable" function of three variables is a set in $\mathbb{E}_4$. However, hardly anyone can imagine that.

I.2.4. Level curves – Isolines
Besides the graph there is another possibility of visually representing a function of two variables. This possibility is the construction of a system of so-called level curves.

If $y = f(x_1, x_2)$ is a function of two variables and k is a chosen real number, then the set of points $\{[x_1; x_2] \in \mathbb{E}_2 : \ f(x_1, x_2) = k\}$ is called a *level curve (isoline)* of the function f. The equation $f(x_1, x_2) = k$ is called the equation of the considered level curve.

A function of two variables may have infinitely many level curves: different values of k correspond to different level curves. A level curve may also be the empty set if $k \notin R(f)$. In general, a level curve need not be a curve: for example, for the constant function $y = k$ on a

set M, the level curve with equation $f(x_1, x_2) = k$ is the set identical with M. The term "level curve" became established because for most continuous and non-constant functions f, the level curves are indeed what we imagine under the notion of a "curve". The term "contour line" is also often used.

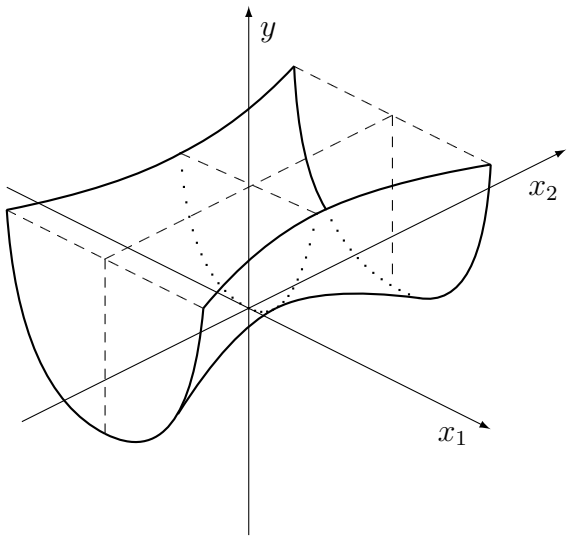


Fig. 4a

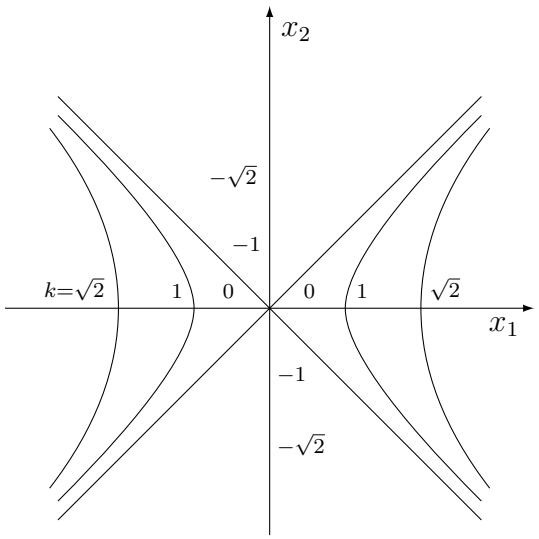


Fig. 4b

I.2.5. Examples of level curves

a) Level curves are commonly used when drawing maps: for example, curves connecting points of equal altitude are level curves of a function $y = f(x_1, x_2)$ giving the altitude of the place with coordinates x_1, x_2. (The term *contour line (isoline)* is used.) Or curves connecting points of equal temperature are level curves of the function $y = T(x_1, x_2)$ giving the temperature at the point $[x_1; x_2]$. (These level curves are called *isotherms*.) Curves connecting points of equal pressure are level curves of the function $y = P(x_1, x_2)$ giving the pressure at the point $[x_1; x_2]$. (These level curves are called *isobars*.)

b) Circles with equations $x_1^2 + x_2^2 = k$ (for various $k > 0$) are level curves of the function $y = x_1^2 + x_2^2$. (The graph of this function is a rotational paraboloid in $\mathbb{E}_3$ – see [5].)

c) In Fig. 4a you see a part of the graph of the function $y = x_1^2 - x_2^2$, and in Fig. 4b you see the corresponding system of level curves. The level curves are hyperbolas with equations $x_1^2 - x_2^2 = k$ for various $k \in \mathbb{R}$.

I.2.6. Level surfaces – Isosurfaces

If f is a function of three variables and k is a real number, then the set of points $\{[x_1; x_2; x_3] \in \mathbb{E}_3 : f(x_1, x_2, x_3) = k\}$ is called a *level surface (isosurface)* of the function f.

A level surface is a concept similar to a level curve, only "one dimension higher". The considered function is a function of three variables, not just two, and a level surface is a set in $\mathbb{E}_3$, not merely in $\mathbb{E}_2$. A level surface can in general be a very diverse set; however, for most continuous and non-constant functions of three variables its shape corresponds to what we imagine under the notion of a "surface".

For example, the sphere $x^2 + y^2 + z^2 = 4$ is a level surface of the function $f(x, y, z) = x^2 + y^2 + z^2$.

I.2.7. Operations with functions

Assume that f and g are functions of n variables with domains $D(f),\ D(g) \subset \mathbb{E}_n$. By the *sum* of f and g we mean the function h whose domain is $D(h) = D(f) \cap D(g)$ and which assigns to each $X \in D(h)$ the value $h(X) = f(X) + g(X)$. In short we write: $h = f + g$. Analogously we define the *difference* and the *product* of two functions.

Similarly we can define the quotient of f and g. In the domain of the quotient, however, no points may occur where the denominator is zero. Thus the definition of the quotient is: By the *quotient* of f and g we mean the function h whose domain is the set $D(h) = D(f) \cap \{X \in D(g);\ g(X) \neq 0\}$ and which assigns to each $X \in D(h)$ the number $h(X) = f(X)/g(X)$. In short we write: $h = f/g$.

You see that operations between functions of several variables are defined in the same way as in the case of functions of one variable. The same holds for the further concepts that follow in this chapter.

I.2.8. Composite function

Assume that $y = f(x_1, \ldots, x_n)$ is a function of n variables $x_1, \ldots, x_n$. If we substitute for these variables other functions depending, for example, on k variables $t_1, \ldots, t_k$:

$$x_1 = \varphi_1(t_1, \ldots, t_k),\ \ldots,\ x_n = \varphi(t_1, \ldots, t_k),$$

we obtain the function

$$y = f(\varphi_1(t_1, \ldots, t_k),\ \ldots,\ \varphi_n(t_1, \ldots, t_k)). \tag{I.2.2}$$

We call this function a *composite function*. f is the *outer function* and $\varphi_1, \ldots, \varphi_n$ are the *inner functions*.

It is clear that for the composite function (I.2.2) to be defined on some set $G \subset \mathbb{E}_k$, all functions $\varphi_1, \ldots, \varphi_n$ must be defined on G and

$$\big[\varphi_1(t_1, \ldots, t_k);\ \ldots;\ \varphi_n(t_1, \ldots, t_k)\big] \in D(f) \ \text{ for all } [t_1; \ldots; t_k] \in G.$$

I.2.9. Example

Assume that $y = f(x_1, x_2)$ is a function of two variables defined on $\mathbb{E}_2$. We can regard x_1 and x_2 as Cartesian coordinates of the point $X = [x_1; x_2]$ in $\mathbb{E}_2$. If r, φ are polar coordinates of the same point X, then $x_1 = r\cos\varphi$ and $x_2 = r\sin\varphi$. (See Section II.3.5.) Substituting these expressions for x_1 and x_2 into f, we obtain the composite function $y = f(r\cos\varphi,\ r\sin\varphi)$.

I.2.10. Bounded function

Let f be a function of n variables and $M \subset D(f)$. We say that f is *bounded above on the set M* if there exists a real number K such that $\forall\, X \in M\ :\ f(X) \leq K$.

Analogously, f is *bounded below on the set M* if there exists a real number L such that $\forall\, X \in M\ :\ f(X) \geq L$.

We say that f is *bounded on M* if it is bounded both below and above on M.

If f is bounded above on its entire domain $D(f)$, we simply say it is *bounded above*. Analogously we define *bounded below* and *bounded*.

I.2.11. Example

The function $f:\ y = x_1^2 + x_2^2$ is bounded below, but not bounded above, on $\mathbb{E}_2$. The graph of f is a paraboloid.

I.2.12. Vector-valued function. Vector field

In mechanics and other disciplines we often work with functions that assign not a real number but a vector to each point of some set. For example, the velocity of a fluid flowing in a region M is a vector at each point $X = [x_1; x_2, x_3] \in M$, not a scalar. Another example is the deformation of a body originally occupying a region $M \subset \mathbb{E}_3$: to each point $X = [x_1; x_2, x_3] \in M$ we can assign a vector describing the displacement of X. Functions whose values are vectors are called vector-valued functions. The precise definition is:

Assume that $M \subset \mathbb{E}_n$. A mapping that assigns to each point $X \in M$ a vector in $\mathbb{E}_n$ is called a *vector-valued function*.

If a vector-valued function $\vec{f}$ is defined on $M \subset \mathbb{E}_n$, then there exist n scalar functions f_1, $\dots$, f_n such that

$$\vec{f}(X) \;=\; \big(f_1(X); \dots; f_n(X)\big) \qquad \text{for every } X \in M.$$

We write this briefly as:

$$\vec{f} = (f_1; \dots; f_n).$$

The functions $f_1, \dots, f_n$ are called the *component functions* of $\vec{f}$. If $\widehat{e}_1, \dots, \widehat{e}_n$ are the unit vectors oriented along the axes $x_1, \dots, x_n$, we can also write

$$\vec{f}(X) \;=\; f_1(X)\,\widehat{e}_1 + \cdots + f_n(X)\,\widehat{e}_n \qquad \text{for every } X \in M$$

or

$$\vec{f} = f_1\widehat{e}_1 + \cdots + f_n\widehat{e}_n.$$

In particular, when $n = 3$, instead of $\widehat{e}_1, \widehat{e}_2, \widehat{e}_3$ we often use $\widehat{\imath}, \widehat{\jmath}, \widehat{k}$ (similarly as we use x, y, z instead of x_1, x_2, x_3). The vectors $\widehat{\imath}, \widehat{\jmath}, \widehat{k}$ are unit vectors oriented along the axes x, y, z. A vector-valued function $\vec{f}$ with component functions U, V, W can then be written as

$$\vec{f} = (U; V; W) \quad \text{or} \quad \vec{f} = U\,\widehat{\imath} + V\,\widehat{\jmath} + W\,\widehat{k}.$$

If $n = 2$ and $\vec{f}$ is a planar vector-valued function with components U, V, we analogously write:

$$\vec{f} = (U; V) \quad \text{or} \quad \vec{f} = U\,\widehat{\imath} + V\,\widehat{\jmath}.$$

Further examples of vector-valued functions are the intensity of the gravitational field, the electrostatic or magnetic field intensity, etc. In all these cases, to each point X of a set $M \subset \mathbb{E}_3$ we can assign a vector – e.g. the force exerted by the gravitational (or electrostatic) field on a unit mass (or unit charge) placed at X, etc. In physics, the term "field" is often used instead of "function": one speaks of the velocity field of a flowing fluid, a gravitational field, an electromagnetic field. Therefore, in mathematics we also often use the term *vector field* instead of "vector-valued function".

In some cases, especially in physical applications, it is useful to emphasize the difference between a vector-valued function (vector field) and an ordinary function whose values are real numbers (scalars). Therefore an "ordinary" function of n variables is sometimes called a *scalar function* or a *scalar field*. Examples include the temperature field in a body occupying $M \subset \mathbb{E}_3$, the illumination intensity on a planar plate covering a set $M \subset \mathbb{E}_2$, etc.

In this text we will deal mainly with scalar functions. To avoid repeating this constantly, we adopt the following convention: whenever we use the word "function" without any adjective, we mean a scalar function. We will encounter vector-valued functions and their integrals mainly in Chapters III.4, III.5, IV.4 and IV.5.

I.3 Limit and continuity

I.3.1. Limit of a function – motivation

Let us recall the notation introduced in [5]: $\mathbb{R}^*$ is the so-called extended set of real numbers. It is the union of the set of real numbers $\mathbb{R}$ and the two-element set $\{-\infty; +\infty\}$. More details about $\mathbb{R}^*$ can be found in [5] .

Assume that f is a function of n variables, $A \in \mathbb{E}_n$ and $a \in \mathbb{R}^*$. When we say that "$f(X)$ approaches a" as "X approaches A", you can certainly imagine what this means. From Mathematics I you know that mathematics can describe such an idea precisely using the concept of "limit".

In the following sections we define two types of limits. In the first case we assume that X approaches A "from all directions" (limit at the point A, Section I.3.2). In the second case we assume that X approaches A but at the same time X belongs to some set M (limit at A with respect to M, Section I.3.8). Thus X may approach A for example along a curve or from a set whose boundary contains A, etc.

I.3.2. Limit of a function

Assume that f is a function of n variables, $A \in \mathbb{E}_n$, $a \in \mathbb{R}^*$, and there exists a punctured neighborhood $\mathcal{P}(A)$ that is entirely contained in the domain of f. If for every sequence of points $\{X_k\}$ in $\mathcal{P}(A)$ the implication

$$X_k \to A \implies f(X_k) \to a,$$

holds, then the number a is called the *limit of f as X approaches A*. We write:

$$\lim_{X \to A} f(X) = a \quad \text{or} \quad \lim_{x_1 \to a_1, \ldots, x_n \to a_n} f(x_1, \ldots, x_n) = a.$$

Recall that $X_k \to A$ is just a shorter notation for $\lim\limits_{k \to +\infty} X_k = A$.
Similarly, $f(X_k) \to a$ means that $\lim\limits_{k \to +\infty} f(X_k) = a$.

You see that for f to have a limit as $X \to A$, it must be defined on some punctured neighborhood of A. The function then has limit a as $X \to A$ if, roughly speaking, as X approaches A from any direction, the values $f(X)$ approach a. The function may or may not be defined at A itself. The existence and value of $\lim_{X \to A} f(X)$ depend exclusively on the behaviour of f in a punctured neighborhood of A, not at the point A itself. This is the same as for functions of one variable.

We defined the limit of a function of n variables at $A \in \mathbb{E}_n$ using two limits of sequences: $\lim X_k$ and $\lim f(X_k)$. Note that this is not the only possible definition. Historically, several definitions arose; all are equivalent.

From the fact that a sequence of real numbers $\{f(X_k)\}$ can have at most one limit a (as $k \to +\infty$), we easily obtain:

I.3.3. Theorem – Uniqueness of a limit

A function f can have at most one limit at a point $A \in \mathbb{E}_n$.

That means only two cases are possible:

1) $\lim\limits_{X\to A} f(X)$ does not exist (i.e. f has no limit at A),
2) or $\lim\limits_{X\to A} f(X)$ exists and has a unique value a.

I.3.4. Example

$f(x,y) = \dfrac{\sin(x^2+y^2)}{x^2+y^2}$ We investigate $\lim\limits_{[x;y]\to[0;0]} f(x,y)$.

The function f is defined on all of $\mathbb{E}_2$ except the point $[0;0]$. $f(x,y)$ depends on x and y only through the expression x^2+y^2. As $[x;y]$ approaches $[0;0]$, the quantity x^2+y^2 approaches 0. Let $z = x^2+y^2$. Then $[x;y] \to [0;0]$ implies $z \to 0$. Therefore

$$\lim_{[x;y]\to[0;0]} f(x,y) = \lim_{[x;y]\to[0;0]} \frac{\sin(x^2+y^2)}{x^2+y^2} = \lim_{z\to 0} \frac{\sin z}{z} = 1.$$

I.3.5. Example

$g(x,y) = \dfrac{x^2-y^2}{x^2+y^2}$ We investigate $\lim\limits_{[x;y]\to[0;0]} g(x,y)$.

The function g is defined on $\mathbb{E}_2 \setminus \{[0;0]\}$. First imagine that $[x;y]$ approaches $[0;0]$ along the x-axis. This means $y = 0$, $x \neq 0$ and $x \to 0$. Then

$$g(x,y) = g(x,0) = \frac{x^2}{x^2} = 1.$$

Thus $g(x,y) \to 1$ when $[x;y] \to [0;0]$ along the x-axis. Now imagine that $[x;y]$ approaches $[0;0]$ along the y-axis. This means $x = 0$, $y \neq 0$ and $y \to 0$. Then

$$g(x,y) = g(0,y) = \frac{-y^2}{y^2} = -1.$$

Thus $g(x,y) \to -1$ when $[x;y] \to [0;0]$ along the y-axis. Therefore the values $g(x,y)$ can approach at least two different numbers (1 and -1) depending on the direction of approach. If the limit existed, it would have to equal both 1 and -1, contradicting Theorem I.3.3. Hence the limit

$$\lim_{[x;y]\to[0;0]} g(x,y)$$

does not exist.

I.3.6. Remark

The method from Example I.3.5 can be generalized: If

1) $g(X) \to a_1$ as X approaches A along some curve C_1, and
2) $g(X) \to a_2$ (with $a_2 \neq a_1$) as X approaches A along another curve C_2,

then $\lim\limits_{X\to A} g(X)$ does not exist.

I.3.7. Example

$g(x_1, x_2) = \dfrac{\sin(x_1 + 2x_2)}{x_1 + 2x_2}$ We investigate $\lim\limits_{[x_1;x_2]\to[0;0]} g(x_1, x_2)$.

The function g is defined for $x_1 + 2x_2 \neq 0$, i.e. at all points of $\mathbb{E}_2$ except the line $x_1 + 2x_2 = 0$. This line passes through $[0; 0]$. Hence g is not defined on any (entire) punctured neighborhood of $[0; 0]$, because every punctured neighborhood of $[0; 0]$ intersects the line $x_1 + 2x_2 = 0$. Therefore the limit

$$\lim_{[x_1;x_2]\to[0;0]} g(x_1, x_2)$$

does not exist.

I.3.8.* Limit of a function with respect to a set

Assume that f is a function of n variables, $a \in \mathbb{R}^*$, M is a set in $\mathbb{E}_n$, A is an accumulation point of M, and there exists a punctured neighborhood $\mathcal{P}(A)$ such that $\mathcal{P}(A) \cap M \subset D(f)$. If for every sequence of points $\{X_k\}$ in $\mathcal{P}(A) \cap M$ the implication

$$X_k \to A \implies f(X_k) \to a,$$

holds, then a is called the *limit of f as X approaches A with respect to the set M*. We write:

$$\lim_{\substack{X\to A \\ X\in M}} f(X) = a \quad \text{or} \quad \lim_{\substack{x_1\to a_1,\,\ldots,\,x_n\to a_n \\ [x_1;\ldots;x_n]\in M}} f(x_1, \ldots, x_n) = a.$$

A typical situation motivating this concept is when A lies on the boundary of M and the independent variable X does not approach A from all directions, but only from within M. It is clear that for f to have a limit at A with respect to M, it need not be defined on an entire punctured neighborhood $\mathcal{P}(A)$; it suffices that it be defined on $\mathcal{P}(A)\cap M$. To make the definition meaningful, we also require that there exists at least one sequence of points in M, different from A, converging to A; this is guaranteed if A is an accumulation point of M.

You can see that the limit as $X \to A$ defined in Section I.3.2 is just a limit with respect to any punctured neighborhood of A.

For limits of functions of several variables at $A \in \mathbb{E}_n$ with respect to a set $M \subset \mathbb{E}_n$, the same theorem as I.3.3 holds:

I.3.9.* Theorem

A function f can have at most one limit at $A \in \mathbb{E}_n$ with respect to a set $M \subset \mathbb{E}_n$.

I.3.10.* Example

Let us return to the limit from Example I.3.7. If we compute it as a limit with respect to the set of $(x_1; x_2)$ for which $x_1 + 2x_2 \neq 0$ (i.e. $[x_1; x_2] \in M = \{[x_1; x_2] \in \mathbb{E}_2 : \ x_1 + 2x_2 \neq 0\}$), we can set $u = x_1 + 2x_2$ and obtain:

$$\lim_{\substack{[x_1;x_2]\to[0;0] \\ [x_1;x_2]\in M}} \frac{\sin(x_1 + 2x_2)}{x_1 + 2x_2} = \lim_{u\to 0} \frac{\sin u}{u} = 1.$$

Note also that instead of $\lim\limits_{\substack{[x_1;x_2]\to[0;0] \\ [x_1;x_2]\in M}}$ we may write under "lim" the condition $\lim\limits_{\substack{[x_1;x_2]\to[0;0] \\ x_1+2x_2\neq 0}}$.

I.3.11.* Example

We investigate the limit $\lim\limits_{\substack{[x;y]\to[2;1] \\ x-2y\geq 0}} \left(2x - y + \sqrt{x - 2y}\right)$.

The function is defined on the half-plane $M = \{[x;y];\ x - 2y \geq 0\}$ whose boundary contains the point $[2;1]$. If $[x;y]$ approaches $[2;1]$ while remaining in M, then clearly $x \to 2$ and $y \to 1$, and

$$\lim_{\substack{[x;y]\to[2;1]\\ x-2y\geq 0}} \left(2x - y + \sqrt{x-2y}\right) = 2\cdot 2 - 1 + \sqrt{2 - 2\cdot 1} = 3.$$

I.3.12. Difference, sum, product and quotient of limits

From the theory of functions of one variable we know that under simple assumptions, the limit of a sum (difference, etc.) equals the sum (difference, etc.) of the limits. The same statements hold for limits of functions of several variables. We state them here again, for simplicity only for the limit defined in Section I.3.2, i.e. the limit as $X \to A$.

Assume that $\lim_{X\to A} f(X) = a$ *and* $\lim_{X\to A} g(X) = b$. *Then*

$$\lim_{X\to A} \big(\, f(X) \pm g(X) \,\big) = a \pm b,$$

$$\lim_{X\to A} f(X)\cdot g(X) = a\cdot b,$$

$$\lim_{X\to A} \frac{f(X)}{g(X)} = \frac{a}{b}$$

provided the expressions on the right-hand sides are meaningful.

The assumption that the expressions on the right-hand sides are meaningful means that, for example, expressions such as $(+\infty) - (+\infty)$, $0\cdot(+\infty)$, $5/0$, etc. must not occur. The same statements hold for the limit as $X \to A$ with respect to a set.

From differential calculus of one variable you also know limits as $x \to \pm\infty$. In the multi-dimensional space $\mathbb{E}_n$, however, it does not make reasonable sense to define two directions "to $+\infty$" and "to $-\infty$" as in $\mathbb{E}_1$. Nevertheless, one can study the behaviour (limit) of a function as X goes unboundedly far from the origin (i.e. $\|X - O\| \to +\infty$) either in any direction or only within some unbounded set. We will not deal with such limits in this text.

I.3.13. Continuity of a function – motivation

In Fig. 5a and Fig. 5b you see the graphs of two functions of two variables on a set $M \subset \mathbb{E}_2$. While the graph of f (Fig. 5a) is an "unbroken surface" in $\mathbb{E}_3$ "above" the set M, the same cannot be said about the graph of g (Fig. 5b). Mathematics describes such behaviour using the notion of "continuity": we can say that f is continuous at the point A and on the set M, while neither is true for g.

In the following sections you will find precise definitions of continuity at a point A, continuity at A with respect to a set M, and continuity on a set M, along with further comments.

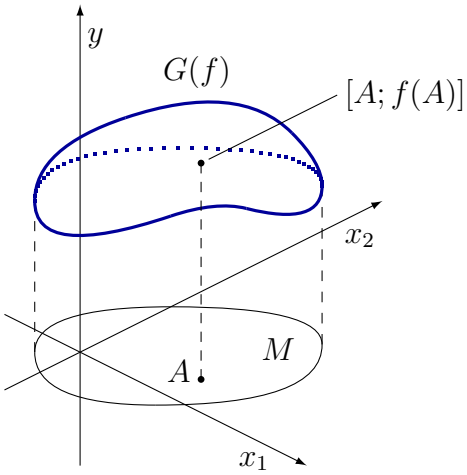


Fig. 5a

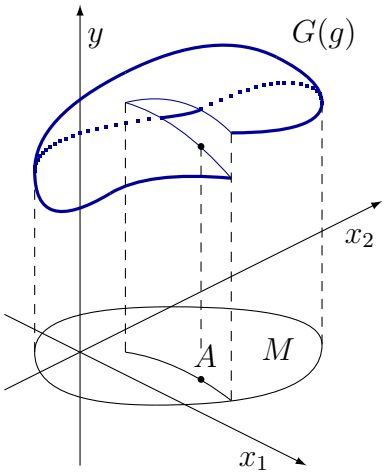


Fig. 5b

I.3.14. Continuity of a function at a point A
Assume that f is a function of n variables, $A \in D(f)$, and there exists a neighborhood $\mathcal{U}(A)$ entirely contained in $D(f)$. We say that f is *continuous at A* if for every sequence of points $\{X_k\}$ in $\mathcal{U}(A)$ the implication

$$X_k \to A \implies f(X_k) \to f(A) \qquad \text{holds.}$$

Comparing this definition with Definition I.3.2 (limit of a function), you see the simple statement: *f is continuous at A if and only if*

$$\lim_{X \to A} f(X) = f(A).$$

Thus the definition of continuity of a function of n variables at $A \in \mathbb{E}_n$ extends the definition of continuity of a one-variable function at $a \in \mathbb{E}_1$ known from Mathematics I. The meaning remains the same: roughly speaking, as X approaches A from any direction, the values $f(X)$ must approach the value of f at the limit point A, i.e. $f(A)$.

I.3.15. Continuity of a function at A with respect to a set M

Assume that f is a function of n variables, $M \subset \mathbb{E}_n$, and $A \in D(f) \cap M$. We say that f is *continuous at A with respect to M* if for every sequence of points $\{X_k\}$ in M we have

$$X_k \to A \implies f(X_k) \to f(A).$$

The meaning is clear: continuity of f at A with respect to M depends only on the behaviour of f at points of M. If X approaches A while remaining in M, then $f(X)$ must approach $f(A)$. How f behaves outside M, and even whether it is defined there, is irrelevant.

A typical situation where it makes sense to speak of continuity at A with respect to M is when A lies on the boundary of M. If A is an interior point of M, then continuity at A (in the sense of Definition I.3.14) and continuity at A with respect to M coincide.

An example from one-variable theory is right-continuity or left-continuity at a point $a \in \mathbb{E}_1$. Right-continuity at a is continuity with respect to the right neighborhood of a, and left-continuity is continuity with respect to the left neighborhood.

I.3.16. Continuity of a function on a set M

Assume that f is a function of n variables and M is a set in $\mathbb{E}_n$ contained in $D(f)$. We say that f is *continuous on the set M* if it is continuous at every point $X \in M$ with respect to M.

A simple example is continuity of a one-variable function on an interval.

From Definition I.3.16 the following immediately follows:

If f (of n variables) is continuous on M and $M_0 \subset M$, then f is also continuous on M_0.

I.3.17.* Remark

From the definition of continuity at A with respect to M it is clear that if A is an isolated point of M, then any function whose domain contains A is continuous at A with respect to M. Hence, if a set M consists only of isolated points, then every function defined on it is automatically continuous on it.

A practical question is how to recognize whether a function is continuous at a point A, continuous at A with respect to M, or continuous on a set M. The following theorems are very useful:

I.3.18. Theorem (continuity of sum, difference, product, absolute value and quotient)

a) Assume that the functions f and g (of n variables) are continuous at a point $A \in \mathbb{E}_n$. Then the functions $f+g$, $f-g$, $f \cdot g$ and $|f|$ are also continuous at A. If $g(A) \neq 0$, then the quotient f/g is also continuous at A.

(The above part also holds for continuity at A with respect to a set M.)

b) If f and g (of n variables) are continuous on a set $M \subset \mathbb{E}_n$, then $f+g$, $f-g$, $f \cdot g$ and $|f|$ are continuous on M as well. If in addition $g(X) \neq 0$ for all $X \in M$, then the quotient f/g is also continuous on M.

By mathematical induction, it is easy to show that the same statements hold not only for the sum of two functions but for the sum of any finite number of functions. The same applies to differences, products, etc.

The next theorem concerns continuity of a composite function. Recall that when forming a composite function using $y = f(x_1, \dots, x_n)$, we substitute for each variable $x_1; \dots; x_n$ a new function depending on, for example, k variables $t_1, \dots, t_k$: $x_1 = \varphi_1(t_1, \dots, t_k)$, $\dots$, $x_n = \varphi_n(t_1, \dots, t_k)$. The idea is simple: if all inner functions $\varphi_1, \dots, \varphi_n$ are continuous at $A = [a_1; \dots; a_k] \in \mathbb{E}_k$, map A to a point $B \in \mathbb{E}_n$, and the outer function f is continuous at B, then the composite function is continuous at A. Similarly on sets. The precise statement is formally a bit complicated:

I.3.19. Theorem (continuity of a composite function).

a) Assume that the function

$$y = f(x_1, \dots, x_n) \tag{I.3.1}$$

is continuous at the point $B = [b_1; \dots; b_n]$, the functions

$$x_1 = \varphi_1(t_1, \dots, t_k), \ \dots \ , \ x_n = \varphi_n(t_1, \dots, t_k) \tag{I.3.2}$$

are all continuous at the point $A = [a_1; \dots; a_k]$ and $B = [\varphi_1(A); \dots; \varphi_n(A)]$. Then the composite function

(I.3.3) $$y = f\big(\varphi_1(t_1, \dots, t_k), \dots, \varphi_n(t_1, \dots, t_k)\big)$$

is continuous at the point A.

b) Assume that the outer function (I.3.1) is continuous on a set $M \subset \mathbb{E}_n$, all inner functions (I.3.2) are continuous on a set $G \subset \mathbb{E}_k$, and the mapping

(I.3.4) $$\begin{aligned} x_1 &= \varphi_1(t_1, \dots, t_k), \\ &\vdots \\ x_n &= \varphi_n(t_1, \dots, t_k) \end{aligned}$$

maps G into M. Then the composite function (I.3.3) is continuous on G.

The assumption that (I.3.4) maps G into M means that for every $T = [t_1; \dots; t_k] \in G$, the point $X = [x_1; \dots; x_n]$ whose coordinates are given by (I.3.4) belongs to M.

Now we can establish continuity of many concrete functions. We give an example:

I.3.20. Example

The function

$$h(x, y, z) = \ln(x + y) + \cos(y - z)$$

is continuous in the half-space $G = \{[x; y; z] \in \mathbb{E}_3 : x + y > 0\}$ (which is its domain).

Clearly G is the domain of h. The condition $x+y > 0$ follows because $x+y$ is the argument of the logarithm, and the logarithm is defined only for positive arguments.

The function $h(x, y, z)$ is the sum of two composite functions: $\ln(x + y)$ and $\cos(y - z)$. The outer function $\ln$ is continuous on $(0; +\infty)$. The inner function $x + y$ is continuous on G and maps G into $(0; +\infty)$. Hence $\ln(x + y)$ is continuous on G. The outer function $\cos$ is continuous on $\mathbb{E}_1$. The inner function $y - z$ is continuous on G and maps G into $\mathbb{E}_1$, so $\cos(y - z)$ is continuous on G. Therefore h, as a sum of two continuous functions on G, is also continuous on G.

At the end of this chapter we state a theorem analogous to the well-known one from the theory of functions of one variable:

I.3.21. Theorem (intermediate value property)

Assume that f (of n variables) is continuous on a set $M \subset \mathbb{E}_n$, A and B are two points in M, and L is a polygonal line connecting A and B lying entirely in M. Then for every number y between $f(A)$ and $f(B)$ (including possible equality) there exists a point $X \in L$ such that $y = f(X)$.

This theorem is also called, just like the analogous theorem for functions of one variable, Darboux's theorem. Its meaning is simple. Imagine that you move along a polygonal line L from A to B. The values of the function f at the points you pass through change continuously. Hence the function f cannot skip any chosen value y between $f(A)$ and $f(B)$. Therefore, somewhere on the polygonal line L you must encounter a point X at which $y = f(X)$.

I.4 Partial derivatives. Differentiable function. Tangent plane. Differential

I.4.1. Partial derivatives – motivation

From differential calculus of functions of one variable you know the notion of a derivative. Surely you recall that $f'(a)$ is the slope of the tangent to the graph of the function f at the point $[a, f(a)]$. The value $f'(a)$ expresses the rate at which the values $f(x)$ change as the variable x passes through the point a. Suppose now that f is a function of n variables and $A = [a_1; \ldots; a_n]$ is a point in the domain of f. If we are interested in the rate of change of the values $f(X)$ as the moving point X passes through A, then we must first determine the direction in which X passes through A. In $\mathbb{E}_n$, unlike $\mathbb{E}_1$, there are many more possibilities. If we choose, for instance, the direction parallel to the axis x_1, then the moving point X has the first coordinate x_1 variable, while all the other coordinates are constant and equal to those of the point A: $X = [x_1, a_2, \ldots, a_n]$. (Draw a picture in $\mathbb{E}_2$.) We can then regard $f(X) = f(x_1, a_2, \ldots, a_n)$ as a function of one variable x_1 and differentiate it, as such, with respect to x_1 at the point $x_1 = a_1$. The corresponding derivative equals the limit

$$\lim_{h\to 0} \frac{f(a_1 + h, a_2, \ldots, a_n) - f(a_1, a_2, \ldots, a_n)}{h}, \tag{I.4.1}$$

provided this limit exists and is finite. (Compare with definition III.4.3 in the textbook [5].) We call the derivative under consideration the partial derivative of f with respect to x_1 at the point A. This derivative expresses the rate of change of the values of f when the moving point X passes through A in the direction consistent with the orientation of the axis x_1. The exact definition of a partial derivative is contained in the following paragraph I.4.2.

We can write the limit (I.4.1) in a shorter way as well: Let us return to the notation $A = [a_1; \ldots; a_n]$. If $\widehat{e}_1$ is the unit vector oriented consistently with the axis x_1 (see the notation in paragraph I.2.13), then $[a_1 + h; a_2; \ldots; a_n] = A + h\widehat{e}_1$. Hence the limit (I.4.1) is identical with the limit

$$\lim_{h\to 0} \frac{f(A + h\widehat{e}_1) - f(A)}{h}. \tag{I.4.2}$$

Now we can proceed to the definition of the partial derivative with respect to a general variable x_i.

I.4.2. Partial derivative at a point A
Suppose that $y = f(x_1, \ldots, x_n)$ is a function of n variables, $A = [a_1; \ldots; a_n] \in D(f)$, i is one of the numbers $1, \ldots, n$ and $\widehat{e}_i$ is the unit vector in $\mathbb{E}_n$ oriented consistently with the coordinate axis x_i. If the finite limit

$$\lim_{h\to 0} \frac{f(A + h\widehat{e}_i) - f(A)}{h}, \tag{I.4.3}$$

exists, then we call its value the *partial derivative* of the function f *with respect to the variable x_i at the point A* and denote it by

$$\frac{\partial f}{\partial x_i}(A), \quad \frac{\partial}{\partial x_i} f(A) \quad \text{or} \quad \frac{\partial y}{\partial x_i}(A).$$

(Instead of A we may also list the individual coordinates $a_1, \ldots, a_n$ in the parentheses.)

I.4.3. Remark

You see that the limit (I.4.3) is a generalization of the limit (I.4.2) (which is nothing but a slightly different notation for the limit (I.4.1)). If the original notation (I.4.1) is closer to you than (I.4.2), you may start from (I.4.1) in definition I.4.2 and, instead of (I.4.3), define the partial derivative with respect to x_i by the limit

$$\lim_{h\to 0}\frac{1}{h}\Big(f(a_1,\ldots,a_{i-1},a_i+h,a_{i+1},\ldots,a_n)-f(a_1;\ldots;a_n)\Big). \tag{I.4.4}$$

It does not matter which limit (i.e., (I.4.3) or (I.4.4)) we use in the definition, because both express the same thing. In any case, if the partial derivative of f with respect to x_i at the point A exists, then

$$\begin{aligned}\frac{\partial f}{\partial x_i}(A) &= \lim_{h\to 0}\frac{f(A+h\widehat{e_i})-f(A)}{h}= \\ &= \lim_{h\to 0}\frac{1}{h}\Big(f(a_1,\ldots,a_{i-1},a_i+h,a_{i+1},\ldots,a_n)-f(a_1;\ldots;a_n)\Big).\end{aligned}$$

I.4.4. Partial derivative as a function

Suppose that $y = f(x_1,\ldots,x_n)$ is a function of n variables. The function that assigns to each point $X = [x_1;\ldots;x_n] \in D(f)$ at which the partial derivative of f with respect to x_i exists the value of this partial derivative is called the *partial derivative of f with respect to x_i*. We denote it by

$$\frac{\partial f}{\partial x_i},\quad \frac{\partial}{\partial x_i}f \quad \text{or} \quad \frac{\partial y}{\partial x_i}.$$

For its domain we have the inclusion

$$D\left(\frac{\partial f}{\partial x_i}\right)\subseteq D(f).$$

I.4.5. Computing a partial derivative

From definition I.4.2 it is clear that the partial derivative of f with respect to x_i is obtained by viewing f as a function of one variable, namely x_i, and differentiating it as such. The other variables are considered parameters or constants. In computing a partial derivative we will follow exactly the same rules. The computation of a partial derivative is then the same as the computation of the derivative of a function of one variable.

I.4.6. Example

$f(x_1,x_2)=2x_1\,x_2^2+3x_1^5\,x_2^7-3x_1^4$ Then:

$$\frac{\partial f}{\partial x_1}=2x_2^2+15x_1^4\,x_2^7-12x_1^3,\quad \frac{\partial f}{\partial x_2}=4x_1\,x_2+21x_1^5\,x_2^6 \quad \text{for } [x_1;x_2]\in\mathbb{E}_2.$$

I.4.7. Example

$f(x,y,z)=2x^2y^3-5xz+\mathrm{e}^{x^2+y^2+z^2}$ Then:

$$\frac{\partial f}{\partial x}(x;y;z) = 4xy^3 - 5z + 2x\,\mathrm{e}^{x^2+y^2+z^2}, \quad \frac{\partial f}{\partial y}(x;y;z) = 6x^2y^2 + 2y\,\mathrm{e}^{x^2+y^2+z^2},$$
$$\frac{\partial f}{\partial z}(x;y;z) = -5x + 2z\,\mathrm{e}^{x^2+y^2+z^2} \quad \text{for } [x;y;z] \in \mathbb{E}_3.$$

I.4.8. Partial derivatives and continuity

From the theory of functions of one variable you know that the existence of the derivative of $y = f(x)$ at a point a implies continuity of f at a. (See the textbook [5].) The question is whether a similar statement holds also for functions of several variables. More precisely: Does the existence of partial derivatives of a function of several variables at a point $A \in \mathbb{E}_n$ imply continuity of this function at A?
The answer is: NO. A function of several variables may have all partial derivatives at A and still fail to be continuous at A. We show this by a simple example:

Example

Let f be a function of two variables defined by

$$f(x,y) = \begin{cases} 0 & \text{for } x \neq 0 \text{ and } y \neq 0, \\ 1 & \text{for } x = 0 \text{ or } y = 0. \end{cases}$$

The function f has the value 0 on the whole plane $\mathbb{E}_2$ except on the two axes x and y, where its values are equal to 1. From the definition of the partial derivative in paragraph I.4.2 and also from its expression in remark I.4.3 it follows easily that

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0} \frac{1}{h}\big(f(h,0) - f(0,0)\big) = \lim_{h\to 0} \frac{1}{h}(1-1) = 0$$

and similarly

$$\frac{\partial f}{\partial y}(0,0) = \lim_{h\to 0} \frac{1}{h}\big(f(0,h) - f(0,0)\big) = \lim_{h\to 0} \frac{1}{h}(1-1) = 0.$$

(You can also use this argument: the function f is constant on both axes x and y. Hence its partial derivatives at $[0;0]$ with respect to both x and y are, as derivatives of constant functions, equal to zero.) Nevertheless, although both partial derivatives of f at $[0;0]$ exist, the function f is not continuous at $[0;0]$.

Continuation

You see that the property "to have partial derivatives" is, in the case of a function of several variables, a weaker property than the property "to have a derivative" for a function of one variable. For a function f of one variable, the existence of the derivative at a ensures continuity of f at a and the existence of a tangent to the graph of f at $[a, f(a)]$. For a function of several variables, the existence of partial derivatives guarantees neither of these. It is therefore natural to ask:

1) *What ensures continuity for a function of several variables?*
2) *What ensures the existence of a tangent plane for a function of several variables and what is its equation?*

   (The graph of f is a set, for example a surface, in $\mathbb{E}_3$ – see Fig. 3. Instead of a tangent line we therefore have to speak of a tangent plane.)

In the following paragraphs you will find answers to these questions. You will see that the property of a function that ensures both continuity and the existence of a tangent plane is the so-called differentiability. From Mathematics I you know that a function $y = f(x)$ that has a derivative at a is sometimes said to be differentiable at a. For a function of several variables, differentiability at a point A means something more than just the existence of partial derivatives at A. In paragraphs I.4.9 and I.4.10 we explain what it means for a function of several variables to be differentiable at A. Since the notion of a differentiable function is closely related to the notion of a tangent plane, we will explain the meaning of both notions simultaneously.

I.4.9. Tangent plane, differentiable function – motivation

For simplicity, imagine that $y = f(x_1, x_2)$ is a function of two variables whose graph is a surface σ, and $A = [a_1; a_2]$ is a point in $D(f)$. In this paragraph we will denote the coordinates in $\mathbb{E}_3$ by x_1, x_2 and y. (x_1 and x_2 are independent variables and y is the dependent variable.) Let P be the point in $\mathbb{E}_3$ with coordinates $x_1 = a_1$, $x_2 = a_2$ and $y = f(a_1, a_2)$. (That is, $P = [a_1;\, a_2;\, f(a_1, a_2)] = [A; f(A)]$.) The point P lies on the surface σ.

The natural requirements for a tangent plane to the surface σ at the point P are:

a) The tangent plane passes through the point P.

b) The tangent plane is not parallel to the y-axis.

c) The surface σ "clings" to the tangent plane at P.

Requirement b) comes from the analogy with functions of one variable, where the derivative ensures the existence of a tangent line that is not parallel to the y-axis. There are infinitely many planes in $\mathbb{E}_3$ that satisfy requirements a) and b). They are the graphs of all possible linear functions $y = L(X)$, where $X = [x_1; x_2]$,

$$L(X) = f(A) + k_1 \cdot (x_1 - a_1) + k_2 \cdot (x_2 - a_2) \tag{I.4.5}$$

and k_1 and k_2 are real coefficients.

Before we clarify what we mean precisely by requirement c), we explain the meaning of this requirement by an example: Imagine that you have a magnifying glass with variable magnification. As you look through it at the point P and increase the magnification more and more, the surface σ should differ less and less in the field of view from the tangent plane. At a hypothetical infinite magnification, the surface σ should coincide in the field of view with the tangent plane. Let us describe this idea mathematically. Points on the tangent plane have coordinates $[X;\, L(X)]$. Points on the surface σ have coordinates $[X;\, f(X)]$. These points differ in the third coordinate. The difference of the third coordinates is $f(X) - L(X)$. Increasing the magnification means that only points in a shrinking neighborhood of P are visible, i.e., points whose first two coordinates satisfy $x_1 \to a_1$ and $x_2 \to a_2$. Then $X \to A$, i.e. $\|X - A\| \to 0$. The requirement that the surface σ should increasingly coincide with the tangent plane in the field of view means that $f(X) - L(X)$ should also tend to zero, and even "faster" than $\|X - A\|$. In mathematical form:

$$\lim_{X \to A} \frac{f(X) - L(X)}{\|X - A\|} = 0. \tag{I.4.6}$$

The situation is drawn in Fig. 6a. In Fig. 6b you see a two-dimensional section of Fig. 6a by the plane passing through the points A and X and parallel to the y-axis. (The y-axis is drawn dashed in the background.)

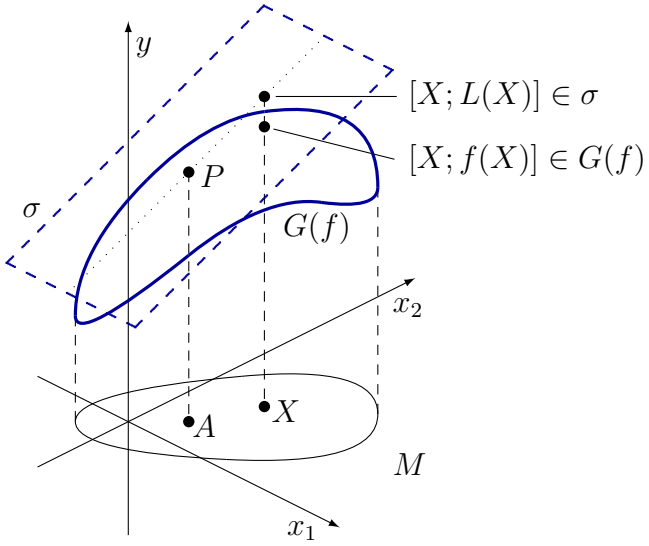


Fig. 6a

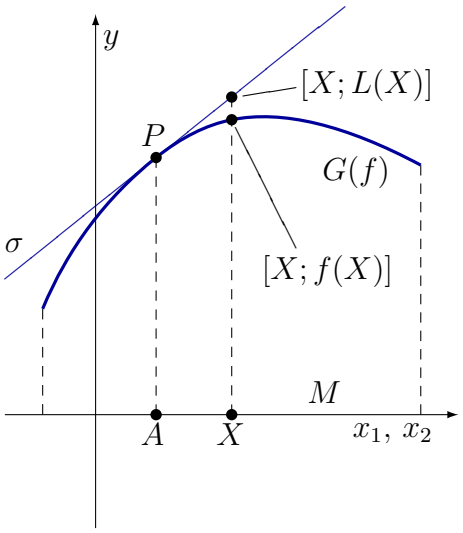


Fig. 6b

Now we can summarize everything: The plane $y = L(X)$ (where the linear function L is given by equation (I.4.5)) satisfies requirements a), b) and c), and we can therefore call it the tangent plane to the surface σ at the point P if condition (I.4.6) is satisfied. One can prove that if condition (I.4.6) is satisfied, then the coefficients k_1 and k_2 are uniquely determined, and thus the tangent plane is also uniquely determined. You can find the proof in paragraph I.4.12. We will return to the tangent plane in paragraph I.4.13, where we will give its precise definition and equation.

The above exposition can now be easily generalized to a function of n variables. In this case $A = [a_1; \ldots; a_n]$, $X = [x_1; \ldots; x_n]$ and the linear function L has the form

$$L(X) = f(A) + k_1 \cdot (x_1 - a_1) + \ldots + k_n \cdot (x_n - a_n). \tag{I.4.7}$$

$k_1, \ldots, k_n$ are real coefficients. Instead of saying that there exists a tangent plane to the surface which is the graph of the function $y = f(X)$ at the point $P = [A; f(A)]$, in mathematical analysis one more often uses the phrase that "the function f is differentiable at the point A." A precise and concise definition is given in the following paragraph.

I.4.10. Differentiable function

Suppose that $y = f(X) = f(x_1, \ldots, x_n)$ is a function of n variables and $A = [a_1; \ldots; a_n] \in \mathbb{E}_n$. We say that f is *differentiable at the point A* if there exists a linear function (I.4.7) such that

$$\lim_{X \to A} \frac{f(X) - L(X)}{\|X - A\|} = 0. \tag{I.4.8}$$

We say that f is *differentiable on a set $M \subset \mathbb{E}_n$* if it is differentiable at every point of M.

In paragraph I.4.15 you will find a simple sufficient condition that makes it possible to recognize that f is differentiable at the point A or on an open set $M \subset \mathbb{E}_n$. Differentiability of f at A has important consequences. The first two will be introduced in paragraphs I.4.11 and I.4.12.

I.4.11. Theorem

If the function f (of n variables) is differentiable at a point $A \in \mathbb{E}_n$, then it is continuous at A.

P r o o f :
From differentiability of f at A it follows that there exists a linear function L of the form (I.4.7) satisfying condition (I.4.8). Continuity of f at A can now be proved as follows:

$$\lim_{X\to A} f(X) = \lim_{X\to A} \big(f(A) + (f(X) - L(X)) + (L(X) - f(A)) \big) =$$

$$= \lim_{X\to A} \left(f(A) + \frac{f(X) - L(X)}{\|X - A\|} \|X - A\| + (L(X) - f(A)) \right) = f(A).$$

I.4.12. Theorem

If the function f (of n variables) is differentiable at a point $A \in \mathbb{E}_n$, then at A it has partial derivatives with respect to all variables.

P r o o f :
Let us compute, for instance, the partial derivative of f with respect to x_1 at A. Choose $X = A + h\widehat{e}_1 = [a_1 + h; a_2; \ldots; a_n]$. Assume first that $h > 0$. Then $\|X - A\| = h$ and $L(X) = f(A) + k_1 \cdot h$. From condition (I.4.8) we obtain

$$0 = \lim_{X\to A} \frac{f(X) - L(X)}{\|X - A\|} = \lim_{h\to 0+} \left(\frac{f(A + h\widehat{e}_1) - f(A)}{h} - k_1 \right).$$

We obtain the same equality also in the case $h < 0$ and $h \to 0_-$. Hence

$$\frac{\partial f}{\partial x_1}(A) = \lim_{h\to 0} \frac{f(A + h\widehat{e}_1) - f(A)}{h} = k_1 .$$

Similarly, the partial derivatives of f with respect to the other variables $x_2, \ldots, x_n$ at A also exist and are successively equal to the coefficients $k_2, \ldots, k_n$.

Now let us return to the tangent plane to the graph of a function f of two variables at the point $P = [A; f(A)] \in \mathbb{E}_3$. Suppose that f is differentiable at $A = [a_1; a_2]$. From paragraph I.4.9 it follows that the tangent plane is the graph of one of the infinitely many linear functions of the form (I.4.5) (namely the one that satisfies condition (I.4.6)). In the proof of Theorem I.4.12 we further derived that in the linear function (I.4.5) satisfying (I.4.6), the coefficients k_1 and k_2 are equal to the partial derivatives of f with respect to x_1 and x_2 at A. We can therefore state the definition:

I.4.13. Tangent plane

Suppose that the function $y = f(x_1, x_2)$, of two variables, is differentiable at the point $A = [a_1; a_2] \in \mathbb{E}_2$. The *tangent plane* to the graph of f at the point $P = [A; f(A)] \in \mathbb{E}_3$ is the plane with equation

(I.4.9) $$y = f(A) + \frac{\partial f}{\partial x_1}(A) \cdot (x_1 - a_1) + \frac{\partial f}{\partial x_2}(A) \cdot (x_2 - a_2).$$

(The coordinates in $\mathbb{E}_3$ are denoted here by x_1, x_2 and y, just as in paragraph I.4.9.)

We can easily generalize the definition to a function of n variables:
Suppose that the function $y = f(x_1, \ldots, x_n)$, of n variables, is differentiable at the point $A = [a_1; \ldots; a_n] \in \mathbb{E}_n$. The *tangent plane* to the graph of f at the point $P = [A; f(A)] \in \mathbb{E}_{n+1}$ is the set of points in $\mathbb{E}_{n+1}$ satisfying the equation

$$y = f(A) + \frac{\partial f}{\partial x_1}(A)\cdot(x_1 - a_1) + \ldots + \frac{\partial f}{\partial x_n}(A)\cdot(x_n - a_n). \tag{I.4.10}$$

(The coordinates in $\mathbb{E}_{n+1}$ are denoted here by $x_1, \ldots, x_n, y$.)

The tangent plane to the graph of a function of n variables is a plane in the space $\mathbb{E}_{n+1}$. If $n > 2$, then the dimension of $\mathbb{E}_{n+1}$ is greater than 3. Such a space, and a plane in it, is hard to imagine. To express that a plane in this space is "something more" than just a plane in $\mathbb{E}_3$, we often call it a "hyperplane". Thus, for a function of n variables we often speak of a *tangent hyperplane* instead of a tangent plane.

I.4.14. Normal line

In this paragraph we again restrict ourselves to a function of two variables. Suppose that the function $y = f(x_1, x_2)$ is differentiable at the point $A = [a_1; a_2] \in \mathbb{E}_2$. The *normal line* to the graph of f at the point $P = [A; f(A)] \in \mathbb{E}_3$ is the line in $\mathbb{E}_3$ that passes through P and is perpendicular to the tangent plane at this point. A direction vector of the normal line is the vector

$$\vec{n} = \left(\frac{\partial f}{\partial x_1}(A);\ \frac{\partial f}{\partial x_2}(A);\ -1\right).$$

The parametric equations of the normal line are

$$x_1 = a_1 + t\,\frac{\partial f}{\partial x_1}(A), \quad x_2 = a_2 + t\,\frac{\partial f}{\partial x_2}(A), \quad y = f(A) - t; \quad t \in \mathbb{R}. \tag{I.4.11}$$

(The coordinates in $\mathbb{E}_3$ are denoted here by x_1, x_2, y.)

I.4.15. Linear approximation. Differential

Suppose that the function f, of n variables, is differentiable at the point $A = [a_1; \ldots; a_n] \in \mathbb{E}_n$. Among all possible linear functions (i.e., functions whose graph is a plane in $\mathbb{E}_{n+1}$), the best approximation of f in a neighborhood of A is evidently the linear function whose graph is the tangent plane. Such a linear function has an equation identical to the equation of the tangent plane, i.e., to (I.4.10).

For $X = [x_1; \ldots; x_n]$ "close" to A we therefore have approximately

$$f(X) \doteq f(A) + \frac{\partial f}{\partial x_1}(A)\cdot(x_1 - a_1) + \ldots + \frac{\partial f}{\partial x_n}(A)\cdot(x_n - a_n). \tag{I.4.12}$$

We often write this approximate expression for $f(X)$ in the form

$$f(X) \doteq f(A) + \mathrm{d}f[A], \tag{I.4.13}$$

where

$$\mathrm{d}f[A] = \frac{\partial f}{\partial x_1}(A)\cdot(x_1 - a_1) + \ldots + \frac{\partial f}{\partial x_n}(A)\cdot(x_n - a_n). \tag{I.4.14}$$

The expression $\mathrm{d}f$ is called the *(total) differential* of f at the point A. You see that the differential depends both on the position of A and on the differences $x_1 - a_1, \ldots, x_n - a_n$. Therefore it is more accurate to write, for example, $\mathrm{d}f[A](x_1 - a_1, \ldots, x_n - a_n)$ instead of just $\mathrm{d}f[A]$.

If we denote $dx_1 = x_1 - a_1, \ \ldots, \ dx_n = x_n - a_n$, then $X = [x_1; \ldots; x_n] = [a_1 + dx_1; \ldots; x_n + dx_n]$ and

(I.4.15) $$f(a_1 + dx_1, \ldots, a_n + dx_n) \doteq f(a_1, \ldots, a_n) + \mathrm{d}f[A](dx_1, \ldots, dx_n),$$

where

(I.4.16) $$\mathrm{d}f[A](dx_1, \ldots, dx_n) = \frac{\partial f}{\partial x_1}(A) \cdot dx_1 + \ldots + \frac{\partial f}{\partial x_n}(A) \cdot dx_n.$$

To emphasize that the differential of a function of several variables is, unlike the differential of a function of one variable, a differential "with respect to all variables", we also often call it the *total differential*.

I.4.16. How do we recognize that a function is differentiable?
In paragraphs I.4.9–I.4.15 you saw that differentiability of f at A is an important property with many consequences. The question, however, is how to recognize easily whether a given function f is differentiable at a given point A. Verifying condition (I.4.8) from the definition of a differentiable function would not be very simple. The answer is provided by the following theorem. Part a) concerns differentiability of f at A, part b) gives a sufficient condition for differentiability of f on an open set $M \subset \mathbb{E}_n$.

Theorem – Sufficient condition of differentiability

Suppose that f is a function of n variables.

a) If f has continuous partial derivatives with respect to all variables at a point $A \in \mathbb{E}_n$, then it is differentiable at A.

b) If M is an open set in $\mathbb{E}_n$ and f has continuous partial derivatives with respect to all variables on M, then it is differentiable on M.

I.4.17. Example

$f(x_1, x_2) = 2 \ln x_1 - x_2^2$

Is f differentiable at the point $A = [1; 3]$? If yes, write the equation of the tangent plane to the graph of f at the point $P = [1;\ 3;\ f(1, 3)]$. Write what the differential of f at A looks like. Using the differential, approximately compute the value $f(1.2,\ 3.2)$.

S o l u t i o n :
The function f has partial derivatives

$$\frac{\partial f}{\partial x_1} = \frac{2}{x_1}, \qquad \frac{\partial f}{\partial x_2} = -2x_2\,.$$

Both partial derivatives are continuous functions at the point A. Hence (by Theorem I.4.16) the function f is differentiable at A. The values of f and its partial derivatives at A are:

$$f(1,3) = -9, \quad \frac{\partial f}{\partial x_1}(1,3) = 2, \quad \frac{\partial f}{\partial x_2}(1,3) = -6.$$

We obtain the equation of the tangent plane at $P = [1; 3; -9]$ by substituting into (I.4.9):

$$y = -9 + 2\,(x_1 - 1) - 6\,(x_2 - 3).$$

(The coordinates in $\mathbb{E}_3$ are denoted by x_1, x_2 and y.) The differential is obtained by substituting into (I.4.15):

$$\mathrm{d}f[1;3](dx_1, dx_2) \;=\; 2\,dx_1 \;-\; 6\,dx_2\,.$$

We compute the approximate value of $f(1.1,\ 3.2)$ using (I.4.15) with $dx_1 = 0.1$ and $dx_2 = 0.2$:

$$f(1.1,\ 3.2) \doteq f(1,3) \;+\; \mathrm{d}f[1;3](dx_1, dx_2) \;=\; -9 + 2\cdot 0.1 - 6\cdot 0.2 \;=\; -10.$$

I.4.18. Summary of important theorems

Suppose that f is a function of n variables and $A \in \mathbb{E}_n$. Theorem I.4.16 says that continuity of the partial derivatives of f at A implies differentiability of f at A. Differentiability then implies continuity of f at A (Theorem I.4.11) and a whole range of other properties. The following scheme symbolically summarizes these and some other important theorems. Items 4 and 5 will be discussed only in Chapter I.5; however, for completeness we include them here as well.

The function f has continuous partial derivatives with respect to all variables at the point A.

$\Downarrow$ (Theorem I.4.16.)

f is differentiable at A. $\Longrightarrow$

1. *f is continuous at A.* (Theorem I.4.11.)
2. *There exists a tangent plane to the graph of f at A. Its equation is (I.4.9) or more generally (I.4.10).*
3. *There exists a differential of f at A. The differential is given by (I.4.16).*
4. *The directional derivative of f in the direction of a nonzero vector $\vec{u}$ exists and can be computed using formula (I.5.2).* (Theorem I.5.5.)
5. *The gradient of f at A has a geometric and physical meaning, described in paragraph I.5.7.*

I.5 Directional derivative. Gradient. Higher-order partial derivatives. Partial derivatives of a composite function.

I.5.1. Directional derivative

In paragraphs I.4.1 and I.4.2 the partial derivative of a function f of n variables at a point $A \in \mathbb{E}_n$ was explained and defined. Recall that the partial derivative of f with respect to x_i at A expresses the rate of change of the values $f(X)$ when the "moving" point X passes through A in a direction consistent with the orientation of the axis x_i. This direction is defined by the unit vector $\widehat{e}_i$. The natural need to express the rate at which the values $f(X)$ change when X passes through A in a direction defined by some general unit vector $\widehat{u}$ (which need not coincide with any of the vectors $\widehat{e}_1, \ldots, \widehat{e}_n$) leads to the following definition:

Directional derivative

Suppose that f is a function of n variables, $A \in D(f)$ and $\widehat{u}$ is a unit vector in $\mathbb{E}_n$. If the finite limit

$$\lim_{h\to 0} \frac{f(A+h\widehat{u}) - f(A)}{h}, \tag{I.5.1}$$

exists, then we call its value the *derivative of f in the direction $\widehat{u}$ at the point A* (also called the *directional derivative of f at A with respect to $\widehat{u}$*) and denote it by

$$\frac{\partial f}{\partial \widehat{u}}(A).$$

I.5.2. Remark

We can easily extend the definition of the directional derivative of f to the case where the direction is defined by a general nonzero vector that need not be a unit vector: If $\vec{u}$ is a nonzero vector in $\mathbb{E}_n$, then the unit vector in the direction of $\vec{u}$ is $\widehat{u} = \vec{u}/\|\vec{u}\|$. We define the directional derivative of f at A in the direction of $\vec{u}$ by

$$\frac{\partial f}{\partial \vec{u}}(A) \;=\; \frac{\partial f}{\partial \widehat{u}}(A).$$

(The derivative on the right-hand side is already known from paragraph I.5.1.)

In paragraph I.5.5 you will see that a sufficient condition for the existence of the directional derivative of f at A is differentiability of f at A. In that paragraph you will also learn how to compute the directional derivative easily. The formula we will use contains the so-called gradient of a function. We therefore introduce this notion now.

I.5.3. Gradient

Suppose that f is a function of n variables $x_1, \ldots, x_n$ and A is a point in $\mathbb{E}_n$ at which f has partial derivatives with respect to all variables. The *gradient of f at the point A* is the vector denoted by $\operatorname{grad} f(A)$ and satisfying:

$$\operatorname{grad} f(A) \;=\; \left(\frac{\partial f}{\partial x_1}(A); \ldots; \frac{\partial f}{\partial x_n}(A)\right),$$

or, in accordance with the notation introduced in paragraph I.2.12,

$$\operatorname{grad} f(A) \;=\; \frac{\partial f}{\partial x_1}(A)\,\widehat{e}_1 \;+\; \ldots \;+\; \frac{\partial f}{\partial x_n}(A)\,\widehat{e}_n\,.$$

The vector-valued function that assigns to each point $X \in D(f)$ at which f has partial derivatives with respect to all variables the vector $\operatorname{grad} f(X)$ is called simply the *gradient* of f. We denote it by $\operatorname{grad} f$. It is clear that $D(\operatorname{grad} f) \subseteq D(f)$.

I.5.4. Example

$f(x,y,z) = 3xy + \mathrm{e}^{-(x^2+y^2+z^2)}$ Compute $\operatorname{grad} f$.

The function f has partial derivatives:

$$\frac{\partial f}{\partial x} = 3y - 2x\,\mathrm{e}^{-(x^2+y^2+z^2)}, \quad \frac{\partial f}{\partial y} = 3x - 2y\,\mathrm{e}^{-(x^2+y^2+z^2)}, \quad \frac{\partial f}{\partial z} = -2z\,\mathrm{e}^{-(x^2+y^2+z^2)}.$$

The gradient of f is the vector-valued function

$$\operatorname{grad} f = \left(3y - 2x\,\mathrm{e}^{-(x^2+y^2+z^2)},\ 3x - 2y\,\mathrm{e}^{-(x^2+y^2+z^2)},\ -2z\,\mathrm{e}^{-(x^2+y^2+z^2)}\right)$$

The gradient of f at a chosen point A, for example $A = [0;1;-1]$, is the vector

$$\operatorname{grad} f(0,1,-1) = \left(3;\ -2\mathrm{e}^{-2};\ 2\mathrm{e}^{-2}\right).$$

I.5.5. Existence and computation of the directional derivative of f in the direction $\vec{u}$

Assuming that f is differentiable at A and $\vec{u} = (u_1, \ldots, u_n)$ is a nonzero vector in $\mathbb{E}_n$, it can be proved that the directional derivative of f at A in the direction of $\vec{u}$ exists and

$$\frac{\partial f}{\partial \vec{u}}(A) = \frac{1}{\|\vec{u}\|} \cdot \left(\frac{\partial f}{\partial x_1}(A) \cdot u_1 + \ldots + \frac{\partial f}{\partial x_n}(A) \cdot u_n\right).$$

This formula can be written more simply using the gradient of f at A. The result is so important that we state it as a separate theorem:

Theorem – calculation of directional derivative

Suppose that the function f (of n variables) is differentiable at a point $A \in \mathbb{E}_n$. Suppose further that $\vec{u}$ is a nonzero vector in $\mathbb{E}_n$. Then the directional derivative of f at A in the direction of $\vec{u}$ exists and can be computed by the formula

(I.5.2) $$\frac{\partial f}{\partial \vec{u}}(A) = \frac{\operatorname{grad} f(A) \cdot \vec{u}}{\|\vec{u}\|} \qquad \left(= \operatorname{grad} f(A) \cdot \widehat{u}, \quad \text{where } \widehat{u} = \frac{\vec{u}}{\|\vec{u}\|}\right).$$

I.5.6. Example

Compute the directional derivative of the function $f(x,y) = 3x^2 + 5xy + y^2$ at the point $A = [1;-1]$ in the direction defined by the vector $\vec{u} = (3;2)$ (if the derivative exists).

S o l u t i o n :
The partial derivatives of f are:

$$\frac{\partial f}{\partial x} = 6x + 5y, \qquad \frac{\partial f}{\partial y} = 5x + 2y.$$

Both partial derivatives are continuous on all of $\mathbb{E}_2$, hence (by Theorem I.4.16) f is differentiable on all of $\mathbb{E}_2$. For us it is important that it is differentiable at A. Therefore the desired directional derivative exists. The gradient of f at A is the vector

$$\operatorname{grad} f(1,-1) = \left(\frac{\partial f}{\partial x}(1,-1);\ \frac{\partial f}{\partial y}(1,-1)\right) = \left(6x + 5y;\ 5x + 2y\right)\Big|_{[x;y]=[1;-1]} = (1;3).$$

Using formula (I.5.2) we obtain:

$$\frac{\partial f}{\partial \vec{u}}(1,-1) = \frac{(1;3)\cdot\vec{u}}{\|\vec{u}\|} = \frac{(1;3)\cdot(3;2)}{\sqrt{3^2+2^2}} = \frac{9}{\sqrt{13}}.$$

I.5.7. Geometric and physical meaning of the gradient

Suppose that f (of n variables) is differentiable at A. Suppose further that $\operatorname{grad} f(A)$ is a nonzero vector. Ask: *In which direction does f have the largest directional derivative?* If α denotes the angle between $\operatorname{grad} f(A)$ and $\vec{u}$, then from (I.5.2) it follows:

$$\frac{\partial f}{\partial \vec{u}}(A) = \frac{\|\operatorname{grad} f(A)\|\,\|\vec{u}\|\,\cos\alpha}{\|\vec{u}\|} = \|\operatorname{grad} f(A)\|\,\cos\alpha.$$

The expression on the right-hand side attains its maximum possible value when $\alpha = 0$ (then $\cos\alpha = 1$). Hence we conclude: The function f has the largest directional derivative at A in the direction defined by the vector $\operatorname{grad} f(A)$.

In other words: the vector $\operatorname{grad} f(A)$ gives the direction of the maximal increase of f at A. It also follows that the vector $\operatorname{grad} f(A)$ is perpendicular to the level curve (for $n=2$) or to the level surface (for $n=3$) passing through A.

Let us show one possible physical interpretation of the derived result: Suppose that $T(x_1,x_2,x_3)$ is a temperature field in a body $V \subset \mathbb{E}_3$ and $A \in V$. The vector $\operatorname{grad} T(A)$ gives the direction in which the temperature increases most rapidly at A. Conversely, $-\operatorname{grad} T(A)$ gives the direction of the greatest decrease of the temperature at A. This has a very important consequence in thermomechanics: according to the so-called Fourier law, heat flows precisely in the direction of the greatest temperature decrease, and the so-called heat flux is therefore the vector equal to $-k\operatorname{grad} T$. (The coefficient k depends on the material of the body and is called the *thermal conductivity coefficient*.)

I.5.8. Tangent and normal to a level curve. Example

What are the equations of the tangent and the normal to the curve in $\mathbb{E}_2$ defined by $x^2+3y^2-2xy+3x-2y=1$ at the point $A=[-1;1]$?

Solution:
The set of points $[x;y]\in\mathbb{E}_2$ satisfying the equation $x^2+3y^2-2xy+3x-2y=1$ is a level curve of the function $f(x,y)=x^2+3y^2-2xy+3x-2y$.

Let us solve the problem first at a general level. Suppose that the function $f(x_1,x_2)$ is differentiable at the point $A=[a_1;a_2]$, $\operatorname{grad} f(A)$ is a nonzero vector, and the level curve $f(x_1,x_2)=k$ passes through A, i.e. $f(a_1,a_2)=k$. The vector $\operatorname{grad} f(A)$ is perpendicular to the level curve (= contour line) of the function f at the point A. (See Section I.5.7.) Therefore it is also perpendicular to the line that is tangent to the level curve of f at A. A point $X=[x_1;x_2]$ lies on the tangent line to the level curve exactly when the vector $X-A$ is perpendicular to $\operatorname{grad} f(A)$, i.e. exactly when $\operatorname{grad} f(A)\cdot(X-A)=0$. Expanding this scalar product we obtain the equation of the tangent line:

$$\frac{\partial f}{\partial x_1}(A)\cdot(x_1-a_1) + \frac{\partial f}{\partial x_2}(A)\cdot(x_2-a_2) = 0. \tag{I.5.3}$$

The normal line to the level curve at A is the line that passes through A and is perpendicular there to the level curve, i.e. perpendicular to the tangent. Its equation (as follows from analytic geometry in $\mathbb{E}_2$) is:

(I.5.4) $$\frac{\partial f}{\partial x_2}(A)\cdot(x_1-a_1)\ -\ \frac{\partial f}{\partial x_1}(A)\cdot(x_2-a_2)\ =\ 0.$$

(The direction vector of the normal is $\operatorname{grad} f(A)$.)

Now let us return to the given function $f(x,y)=x^2+3y^2-2xy+3x-2y$ and to the notation using variables x, y instead of x_1, x_2. The function f has partial derivatives at the point $[-1;1]$:

$$\frac{\partial f}{\partial x}(-1,1)=(2x-2y+3)\Big|_{\substack{x=-1\\y=1}}=-1,\quad \frac{\partial f}{\partial y}(-1,1)=(6y-2x-2)\Big|_{\substack{x=-1\\y=1}}=6.$$

Substituting into the general equation (I.5.3) we obtain the equation of the tangent line to the level curve $x^2+3y^2-2xy+3x-2y=1$ at the point $[-1;1]$:

$$-(x+1)\ +\ 6\,(y-1)\ =\ 0.$$

Substituting into equation (I.5.4) we obtain the equation of the normal line to the same level curve at the point $[-1;1]$:

$$6\,(x+1)\ +\ (y-1)\ =\ 0.$$

I.5.9. Tangent plane and normal to a level surface

In Section I.4.13 we became familiar with the equation of the tangent plane to a surface in $\mathbb{E}_3$ that is the graph of a function of two variables. But what is the equation of the tangent plane and the normal to a surface in $\mathbb{E}_3$ given by the general equation $f(x_1,x_2,x_3)=k$, at the point $A=[a_1;a_2;a_3]$? (In accordance with the terminology introduced in Section I.2.6, this is a level surface of the function f.) We assume that the coordinates of A satisfy the equation, that f is differentiable at A, and that its gradient at A is a nonzero vector.

An analogous problem, only "one dimension lower," was solved in Section I.5.8. Using similar reasoning we can also derive here that the tangent plane has the equation

(I.5.5) $$\frac{\partial f}{\partial x_1}(A)\cdot(x_1-a_1)\ +\ \frac{\partial f}{\partial x_2}(A)\cdot(x_2-a_2)\ +\ \frac{\partial f}{\partial x_3}(A)\cdot(x_3-a_3)\ =\ 0.$$

The normal at A is the line in $\mathbb{E}_3$ that passes through A and is perpendicular to the given surface, and thus also to the tangent plane (I.5.5). Its direction vector is therefore $\operatorname{grad} f(A)$. The parametric equation of the normal line is:

(I.5.6.) $$X\ =\ A\ +\ t\cdot\operatorname{grad} f(A);\quad t\in\mathbb{R}.$$

Written in coordinates:

$$x_1=a_1+t\cdot\frac{\partial f}{\partial x_1}(A),\ \ x_2=a_2+t\cdot\frac{\partial f}{\partial x_2}(A),\ \ x_3=a_1+t\cdot\frac{\partial f}{\partial x_3}(A),\quad t\in\mathbb{R}.$$

I.5.10. Higher-order partial derivatives
Suppose that f is a function of n variables $x_1, \ldots, x_n$ and $i \in \{1; \ldots; n\}$. Recall that the function $\partial f/\partial x_i$ is the partial derivative of f with respect to x_i.

If $j \in \{1; \ldots; n\}$, then the partial derivative with respect to x_j of the function $\partial f/\partial x_i$ is called the *second-order partial derivative or second partial derivative*) of the function f *with respect to x_j and x_i*. We denote it by

$$\frac{\partial^2 f}{\partial x_j\, \partial x_i} \quad \left(\text{or} \quad \frac{\partial^2 f}{\partial x_i^2} \quad \text{if } j = i\right).$$

For the domains we have the inclusion $\quad D\left(\dfrac{\partial^2 f}{\partial x_j\, \partial x_i}\right) \subseteq D\left(\dfrac{\partial f}{\partial x_i}\right) \subseteq D(f).$

Third-, fourth-, and higher-order partial derivatives are defined analogously.

I.5.11. Example

$f(x, y, z) = 5x^2 + 2y^3 + xy^2 \cos z$

Let us compute the partial derivatives $\dfrac{\partial^2 f}{\partial y\, \partial x}$, $\dfrac{\partial^2 f}{\partial z\, \partial x}$ and $\dfrac{\partial^2 f}{\partial x\, \partial y}$.

Solution: $\dfrac{\partial f}{\partial x} = 10x + y^2 \cos z$ and $\dfrac{\partial f}{\partial y} = 6y^2 + 2xy \cos z$.

Therefore:

$$\frac{\partial^2 f}{\partial y\, \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = 2y \cos z, \quad \frac{\partial^2 f}{\partial z\, \partial x} = \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial x}\right) = -y^2 \sin z,$$

$$\frac{\partial^2 f}{\partial x\, \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = 2y \cos z.$$

I.5.12. Interchanging the order of partial derivatives

In Example I.5.11 it turns out that

$$\frac{\partial^2 f}{\partial y\, \partial x} = \frac{\partial^2 f}{\partial x\, \partial y}. \tag{I.5.7}$$

Thus the second partial derivative of f with respect to x and y came out the same whether we first differentiated f with respect to x and then with respect to y, or vice versa. It is natural to ask whether this is a coincidence or whether it must always happen. The correct answer is somewhere in between: equality (I.5.7) is not true in general–see the following example. On the other hand, for "sufficiently reasonable" functions f the equality (I.5.7) does hold–see Theorem I.5.13.

Example

Here we show an example of a function f for which equality (I.5.7) does not hold.
Let $f(x, y) = |x| + y$. It is clear that

$$\frac{\partial f}{\partial y} = 1 \text{ for } [x; y] \in \mathbb{E}_2 \quad \text{and} \quad \frac{\partial f}{\partial x} = \begin{cases} 1 & \text{for } [x; y] \in \mathbb{E}_2,\ x > 0, \\ -1 & \text{for } [x; y] \in \mathbb{E}_2,\ x < 0. \end{cases}$$

The partial derivative of f with respect to x does not exist at points of the form $[0, y]$, i.e. on the y-axis. Hence

$$\frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = 0 \quad \text{for } [x;y] \in \mathbb{E}_2\,,$$

$$\frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = 0 \quad \text{for } [x;y] \in \mathbb{E}_2,\ x \neq 0.$$

You can see that, for example at the point $O = [0;0]$, the derivative $\partial^2 f/\partial x\,\partial y$ exists and equals zero, while the derivative $\partial^2 f/\partial y\,\partial x$ does not exist.

The following theorem shows that for a wide class of functions f, the equality (I.5.7) nevertheless holds. The theorem was discovered and proved by the French mathematician A. Clairaut (1713–1765).

I.5.13. Theorem (on interchanging the order of partial derivatives)

Suppose that f is a function of n variables $x_1, \ldots, x_n$ and $i, j \in \{1;\ldots;n\}$. If both second partial derivatives $\partial^2 f/\partial x_i\,\partial x_j$ and $\partial^2 f/\partial x_j\,\partial x_i$ exist at the point $X = [x_1;\ldots;x_n]$ and at least one of them is continuous at X, then

(I.5.8) $$\frac{\partial^2 f}{\partial x_i\,\partial x_j}(X) = \frac{\partial^2 f}{\partial x_j\,\partial x_i}(X).$$

I.5.14. Partial derivatives of a composite function–motivation

Suppose that in a region $D \subset \mathbb{E}_3$ the function $p(x_1, x_2, x_3)$ gives the pressure at the point $[x_1; x_2; x_3]$. A probe moves in D, and its position at time t is $[x_1(t); x_2(t); x_3(t)]$. What is the change (i.e. the time derivative) of the pressure measured by an instrument placed in the probe?

Assume that all functions are differentiable at all points under consideration. ($p(x_1, x_2, x_3)$ is differentiable at all points of D and $x_1(t)$, $x_2(t)$, $x_3(t)$ are differentiable at all times t under consideration.)

Let p_s denote the pressure measured by the instrument in the probe. p_s depends on the position of the probe, and that position depends on time t through the functions $x_1(t)$, $x_2(t)$, $x_3(t)$. Hence p_s is a function of t: $p_s(t) = p(x_1(t), x_2(t), x_3(t))$. We are to express the derivative of $p_s(t)$ with respect to t.

If all functions are explicitly known, this is a simple problem. For example, if

(I.5.9) $$p(x_1, x_2, x_3) = \mathrm{e}^{-(x_1^2+x_2^2+x_3^2)} + 100 - x_3,\ x_1(t) = x_2(t) = ct,\ x_3(t) = 2ct$$

(where c is a constant), then by substituting $x_1(t)$, $x_2(t)$ and $x_3(t)$ into $p(x_1, x_2, x_3)$ we obtain an explicit expression for $p_s(t)$:

$$p_s(t) = p(x_1(t), x_2(t), x_3(t)) = \mathrm{e}^{-6c^2t^2} + 100 - 2ct.$$

It follows that

(I.5.10) $$\frac{dp_s}{dt}(t) = -12c^2t\,\mathrm{e}^{-6c^2t^2} - 2c.$$

If we do not know explicit formulas for the functions $p(x_1, x_2, x_3)$, $x_1(t)$, $x_2(t)$ and $x_3(t)$, then the question is how we can express, at least on a general level, the derivative of the composite function $p_s(t) = p(x_1(t), x_2(t), x_3(t))$. One can prove that the formula holds:

$$\frac{dp_s}{dt}(t) = \sum_{i=1}^{3} \frac{\partial p}{\partial x_i}\big(x_1(t), x_2(t), x_3(t)\big) \cdot \frac{dx_i}{dt}(t). \tag{I.5.11}$$

In words: *We obtain the derivative of the composite function with respect to t by differentiating the outer function successively with respect to all variables, multiplying each such derivative by the derivative (with respect to t) of the corresponding variable, and summing all the resulting terms.* Verify for yourself that if all functions $p(x_1, x_2, x_3)$, $x_1(t)$, $x_2(t)$ and $x_3(t)$ have the specific form (I.5.9), then using formula (I.5.11) yields the derivative of $p_s(t)$ as in (I.5.10).

A generalization of formula (I.5.11) is given by the following theorem:

I.5.15. Theorem (on the partial derivative of a composite function)

Suppose that $y = f(x_1, \ldots, x_n)$ is a function of n variables and

$$x_1 = \varphi_1(t_1, \ldots, t_k), \ \ldots, \ x_n = \varphi_n(t_1, \ldots, t_k) \tag{I.5.12}$$

are functions of k variables. The derivative of the composite function

$$y = f(\varphi_1(t_1, \ldots, t_k), \ \ldots, \ \varphi_n(t_1, \ldots, t_k))$$

with respect to the variable t_j (where $j \in \{1; \ldots; k\}$) can be expressed by the formula

$$\frac{\partial y}{\partial t_j} = \frac{\partial f}{\partial x_1} \cdot \frac{\partial \varphi_1}{\partial t_j} + \ldots + \frac{\partial f}{\partial x_n} \cdot \frac{\partial \varphi_n}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot \frac{\partial \varphi_i}{\partial t_j}. \tag{I.5.13}$$

The formula holds under the assumption that the outer function f is differentiable at the point $X = [x_1; \ldots; x_n]$, the inner functions $\varphi_1, \ldots, \varphi_n$ are differentiable at the point $T = [t_1; \ldots; t_k]$, and the coordinates of X and T are linked by equations (I.5.12). The partial derivatives of f are considered at X, and the partial derivatives of $\varphi_1, \ldots, \varphi_n$ are considered at T.

Formula (I.5.13) can also be written in this way:

$$\frac{\partial y}{\partial t_j} = \frac{\partial y}{\partial x_1} \cdot \frac{\partial x_1}{\partial t_j} + \ldots + \frac{\partial y}{\partial x_n} \cdot \frac{\partial x_n}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial y}{\partial x_i} \cdot \frac{\partial x_i}{\partial t_j} \tag{I.5.14}$$

(for $j \in \{1; \ldots; k\}$).

I.5.16. Example

Suppose that f is a differentiable function of two variables in $\mathbb{E}_2$ and $g(s,t) = f(s^2 - t^2, t^2 - s^2)$ for $[s;t] \in \mathbb{E}_2$. Show that the function g satisfies in $\mathbb{E}_2$ the equation

$$t\,\frac{\partial g}{\partial s} + s\,\frac{\partial g}{\partial t} = 0.$$

Let $x = s^2 - t^2$ and $y = t^2 - s^2$. Then $g(s,t) = f(x,y)$. By Theorem I.5.15 we have:

$$\frac{\partial g}{\partial s} = \frac{\partial f}{\partial x}\cdot\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\cdot\frac{\partial y}{\partial s} = \frac{\partial f}{\partial x}\cdot(2s) + \frac{\partial f}{\partial y}\cdot(-2s),$$
$$\frac{\partial g}{\partial t} = \frac{\partial f}{\partial x}\cdot\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\cdot\frac{\partial y}{\partial t} = \frac{\partial f}{\partial x}\cdot(-2t) + \frac{\partial f}{\partial y}\cdot(2t).$$

From here it follows that

$$t\,\frac{\partial g}{\partial s} + s\,\frac{\partial g}{\partial t} = \left(2st\,\frac{\partial f}{\partial x} - 2st\,\frac{\partial f}{\partial y}\right) + \left(-2st\,\frac{\partial f}{\partial x} + 2st\,\frac{\partial f}{\partial y}\right) = 0.$$

I.6 Extrema of functions of several variables

I.6.1. Local extrema

Let f be a function of n variables and $A \in D(f)$. We say that f has a *local maximum* at A if there exists a punctured neighborhood $\mathcal{P}(A) \subset D(f)$ such that $\forall X \in \mathcal{P}(A) : f(X) \leq f(A)$.

Similarly, f has a *local minimum* at A if there exists a punctured neighborhood $\mathcal{P}(A)$ such that $\forall X \in \mathcal{P}(A) : f(X) \geq f(A)$.

Local maxima and local minima are collectively called *local extrema*.

If we replace the inequalities in the definition of local extrema by strict inequalities, we obtain the definition of so-called *strict local extrema*:

We say that f has a *strict local maximum* (respectively a *strict local minimum*) at A if there exists a punctured neighborhood $\mathcal{P}(A) \subset D(f)$ such that $\forall X \in \mathcal{P}(A) : f(X) < f(A)$ (respectively $\forall X \in \mathcal{P}(A) : f(X) > f(A)$).

I.6.2. Remark

Can you imagine what the graph of a function of two variables looks like in a neighborhood of a point where the function has a strict local maximum? For example, the graph is a surface in $\mathbb{E}_3$ resembling the Earth's surface near the summit of some mountain. This summit may be a local maximum, but not necessarily the overall, absolute maximum, because a little further away there may be higher mountains.

I.6.3. Necessary condition for a local extremum – discussion

Suppose first, for simplicity, that $y = f(x_1, x_2)$ is a differentiable function of two variables. Its graph is a surface in $\mathbb{E}_3$. If you develop the picture described in the preceding remark I.6.2 a bit, you will not be surprised by the finding that f can have a local extremum only at those points where the tangent plane is horizontal (i.e. parallel to the plane x_1, x_2). The words "can have" and "only" are important: they say that at points other than those where the tangent plane is horizontal, a local extremum cannot exist. At the same time, they do not claim that at every point where the tangent plane is horizontal, the function must have a local extremum. That would not be true: you surely know a place in the mountains called a "saddle." At such a place the tangent plane to the ground surface is horizontal, but there is neither a local maximum nor a local minimum there.

The equation of the tangent plane to the graph of $y = f(x_1, x_2)$ at $A = [a_1; a_2]$ is equation (I.4.9). The tangent plane is horizontal if both partial derivatives of f with respect to x_1 and x_2 at A are zero. In other words: the tangent plane is horizontal if the gradient of f at A is the zero vector.

We have reached the conclusion: a differentiable function f can have (not must, but can!) local extrema only at points where all its partial derivatives are zero, i.e. where its gradient is the zero vector. In the following theorem this important observation is formulated again.

I.6.4. Theorem (necessary condition for a local extremum)
Let f, a function of n variables, be differentiable at the point A. If f has a local extremum at A, then

(I.6.1) $$\operatorname{grad} f(A) = \vec{0}.$$

I.6.5. Remark

In other words: *a differentiable function f can have a local extremum at A only if (I.6.1) holds.*

From Mathematics I you know the necessary condition for a function $y = f(x)$, which has a derivative at a, to be able to have a local extremum at a. The necessary condition is $f'(a) = 0$. Condition (I.6.1) is completely analogous. However, f is now a function of n variables and condition (I.6.1) expresses that

(I.6.2) $$\frac{\partial f}{\partial x_1}(A) = 0, \quad \dots, \quad \frac{\partial f}{\partial x_n}(A) = 0.$$

I.6.6. Critical (stationary) point
If we also include in our considerations functions that are not differentiable at A, we can state: *A function f (of n variables) can have a local extremum at the point $A \in \mathbb{E}_n$ only if*
i) *f is differentiable at A and* $\operatorname{grad} f(A) = \vec{0}$,
ii) *or f is not differentiable at A.*
A point A satisfying condition i) or condition ii) is called a *critical point* of f. (In the literature the term *stationary point* is also often used.)

I.6.7. Sufficient condition for a local extremum – discussion and notation

You surely remember that for a function of one variable the equality $f'(a) = 0$ is only a necessary condition, not a sufficient one, for the function to have a local extremum at a. (See [5].) Similarly, for a function f of n variables the equality (I.6.1) (or the equalities (I.6.2)) is not sufficient for the existence of a local extremum at A. This can be documented by the example of the already mentioned saddle–see Section I.6.3. The question arises: What is a sufficient condition for a local extremum? We will show the answer in the simpler case of a function of two variables: $y = f(x_1, x_2)$. Let us use once again the analogy with a function of one variable: in the case $y = f(x)$ it is sufficient for a local extremum that the conditions $f'(a) = 0$ and $f''(a) > 0$ (strict local minimum) or $f'(a) = 0$ and $f''(a) < 0$ (strict local maximum) hold simultaneously. (See [5].) For a function of two variables the situation is more complicated because the function f has more second partial derivatives at A. We can form from them a matrix which we denote by $\mathcal{M}$:

$$\mathcal{M} = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(A), & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2}(A) \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1}(A), & \dfrac{\partial^2 f}{\partial x_2^2}(A) \end{pmatrix}.$$

Let us further denote

$$\Delta_1(A) = \frac{\partial^2 f}{\partial x_1^2}(A), \qquad \Delta_2(A) = \det \mathcal{M} = \begin{vmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(A), & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2}(A) \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1}(A), & \dfrac{\partial^2 f}{\partial x_2^2}(A) \end{vmatrix}.$$

One can prove the following theorem:

I.6.8. Theorem (sufficient conditions for local extrema)
Suppose that $y = f(x_1, x_2)$ is a function of two variables that has first and second partial derivatives continuous at A and $\operatorname{grad} f(A) = \vec{0}$. Then:

a) If $\Delta_1(A) > 0$ and $\Delta_2(A) > 0$, the function f has a strict local minimum at A.

b) If $\Delta_1(A) < 0$ and $\Delta_2(A) > 0$, the function f has a strict local maximum at A.

c) If $\Delta_2(A) < 0$, then f has no local extremum at A.

I.6.9.* Remark

The inequalities $\Delta_1(A) > 0$ and $\Delta_2(A) > 0$ ensure that the matrix $\mathcal{M}$ is called *positive definite*, i.e. for every nonzero vector $\vec{z}$ in $\mathbb{E}_2$, which we view as a column (a 2×1 matrix), we have: $(\mathcal{M} \cdot \vec{z}) \cdot \vec{z} > 0$. (The second dot denotes the scalar product of the column vectors $\mathcal{M} \cdot \vec{z}$ and $\vec{z}$.) Statement a) of Theorem I.6.8 can thus be formulated as follows: *If A is a critical point of the function $y = f(x_1, x_2)$ and the matrix $\mathcal{M}$ is positive definite, then f has a strict local minimum at A.*

Without further explanation, let us only state that statement b) of Theorem I.6.8 can also be formulated as follows: *If A is a critical point of the function $y = f(x_1, x_2)$ and the matrix $\mathcal{M}$ is negative definite, then f has a strict local maximum at A.*

The inequality $\Delta_2(A) < 0$, together with the assumption that A is a critical point of f, ensures that the graph of f in a neighborhood of A forms a so-called saddle. (See also Section I.6.3.) Thus f has no local extremum at A.

If $\Delta_2(A) = 0$ or $\Delta_2(A) > 0$ and $\Delta_1(A) = 0$, then by Theorem I.6.8 we cannot decide whether f does or does not have a local extremum at A and, if so, what type of extremum it is. (In such a case the theorem says nothing.)

I.6.10. Procedure for finding local extrema of a function f of two variables.

1. *Find all points at which f is differentiable and where both partial derivatives are equal to zero.*

   (Differentiability can be decided using Theorem I.4.16.)

2. *Compute the second partial derivatives of f and determine the values of Δ_1 and Δ_2 at all points found. Then decide on the existence and type of local extremum using Theorem I.6.8.*

The procedure in step 2 is possible only if the assumptions of Theorem I.6.8 are satisfied. If A is a critical point of f and the assumptions of Theorem I.6.8 at A (or none of the conditions a), b), c) of this theorem) are not satisfied, then deciding the existence and type of a local extremum is in general a more complicated problem that can be resolved only by a more detailed analysis of the specific function f.

I.6.11. Example

Let us find the local extrema of the function $f(x, y) = x^4 + y^4 - 4xy + 1$.

Solution:
The first partial derivatives of f are:

$$\frac{\partial f}{\partial x} = 4x^3 - 4y, \qquad \frac{\partial f}{\partial y} = 4y^3 - 4x.$$

Both partial derivatives are continuous in $\mathbb{E}_2$, hence f is differentiable in $\mathbb{E}_2$. Setting the partial derivatives equal to zero, we obtain a system of two equations in the two unknowns x and y:

$$4x^3 - 4y = 0, \qquad 4y^3 - 4x = 0.$$

Solving the first equation for y ($y = x^3$) and substituting into the second, we get: $4x^9 - 4x = 0$, i.e. $x^9 - x = 0$. We can factor $x^9 - x$:

$$x^9 - x = x\,(x^8 - 1) = x\,(x^4 + 1)\,(x^4 - 1) = x\,(x^4 + 1)\,(x^2 + 1)\,(x^2 - 1)$$

$$= x\,(x^4 + 1)\,(x^2 + 1)\,(x + 1)\,(x - 1).$$

Hence the equation $x^9 - x = 0$ has the real roots: $x_1 = 0$, $x_2 = -1$, $x_3 = 1$. The corresponding values of y are: $y_1 = 0$, $y_2 = -1$, $y_3 = 1$. Therefore f has three critical points: $[0; 0]$, $[-1; -1]$ and $[1; 1]$. The function f can have local extrema only at these points. To determine at which of them there is actually a local extremum and whether it is a local maximum or a local minimum, we use Theorem I.6.8. The second partial derivatives of f are:

$$\frac{\partial^2 f}{\partial x^2} = 12x^2, \qquad \frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial^2 f}{\partial y\,\partial x} = -4, \qquad \frac{\partial^2 f}{\partial y^2} = 12y^2.$$

Now we easily compute that $\Delta_1(0, 0) = 0$, $\Delta_2(0, 0) = -16$, $\Delta_1(-1, -1) = 12$, $\Delta_2(-1, -1) = 128$, $\Delta_1(1, 1) = 12$ and $\Delta_2(1, 1) = 128$. The points $[-1; -1]$ and $[1; 1]$ satisfy the assumptions in part a) of Theorem I.6.8 (both Δ_1 and Δ_2 are positive). Therefore f has a sharp local minimum at both of these points. The point $[0; 0]$ satisfies the assumption in part c) of Theorem I.6.8. Hence at this point the function f has neither a local maximum nor a local minimum (it is a "saddle point").

I.6.12. Maximum and minimum of a function on a set
Let f be a function of n variables and let $M \subset D(f)$. We say that f attains at a point $A \in M$ its *maximum on the set M* if $\forall\, X \in M : f(X) \leq f(A)$. We write: $f(A) = \max\limits_{X \in M} f(X)$ or simply $f(A) = \max\limits_{M} f$, respectively $f(A) = \max_M f$.
We denote the maximum of f over its entire domain briefly by $\max f$.
Analogously, we can define the *minimum* of f *on the set M*. We denote it by $\min\limits_{X \in M} f(X)$ or simply $\min\limits_{M} f$, respectively $\min_M f$. The minimum of f over its entire domain $D(f)$ is denoted briefly by $\min f$.
We collectively call the maximum and minimum of f on M the *extrema* of f *on the set M*. To distinguish these extrema from local extrema, we also often use the names *absolute extrema* or *global extrema*.

Similarly as for functions of one variable, for a function of several variables it may happen that one of the extrema (or even both) does not exist.

I.6.13. Example

The function $f : z = x^2 - y^2$ is neither bounded below nor bounded above in $\mathbb{E}_2$. Therefore it attains neither a maximum nor a minimum at any point of $\mathbb{E}_2$.

In the open disk $K = \{[x;y] \in \mathbb{E}_2 : x^2 + y^2 < 1\}$ the function f is bounded, because for $[x;y] \in K$ we have:

$$|f(x,y)| = |x^2 - y^2| \leq x^2 + y^2 < 1. \tag{I.6.3}$$

Nevertheless, neither the maximum nor the minimum of f on K exists. Let us show this, for example, for the maximum: Assume that $\max_K f$ exists. We will show that this assumption leads to a contradiction. By (I.6.3) the value of the maximum must be less than 1. ($\max_K f$ is the greatest function value of f on K, and if all function values are less than 1, then the maximum must also be less than 1.) Choose a point $[x;y] \in K$ on the x-axis. Then its y-coordinate is zero, i.e. $[x;y] = [x;0]$. Suppose the chosen point approaches $[1;0]$, which lies on the boundary of K. Then $x \to 1$ and $f(x,0) = x^2 \to 1$. Hence f takes values in K arbitrarily close to 1. Therefore its maximum on K cannot be less than 1. But this contradicts $\max_K f < 1$.

In the closed disk $\overline{K} = \{[x;y] \in \mathbb{E}_2 : x^2 + y^2 \leq 1\}$ both the maximum and the minimum of f do exist: $\max_{\overline{K}} f = f(1,0) = f(-1,0) = 1$ and $\min_{\overline{K}} f = f(0,1) = f(0,-1) = -1$.

Investigating extrema of a multivariable function on a given set (i.e. their existence, location, and values) is among the most important classes of problems studied in differential calculus of several variables. Finding the location of extrema of f on a given set is treated in Sections I.6.15–I.6.16. The following theorem contains sufficient conditions for the existence of an extremum. Its importance lies in the fact that if the conditions are satisfied, then when searching for extrema we cannot end up searching for something that does not exist.

I.6.14. Theorem (existence of absolute extrema of a function on a given set)
If $M \subset \mathbb{E}_n$ is nonempty, bounded, and closed, and if f is continuous on M, then there exist points X_1 and X_2 in M such that $f(X_1) = \max_M f$ and $f(X_2) = \min_M f$.

In other words: *a continuous function on a nonempty set M that is bounded and closed attains a maximum and a minimum.*

I.6.15. Procedure for finding absolute extrema

Suppose that we have already verified (for example using Theorem I.6.14) that both the maximum and minimum of a function f of n variables on a set $M \in \mathbb{E}_n$ exist, and we are solving the problem of how to find these extrema. Similarly as for a function of one variable, f can attain its absolute extrema on M

a) at points $X \in M^\circ$ where f is differentiable and all partial derivatives are zero,

b) or at points $X \in M^\circ$ where f is not differentiable,

c) or at points $X \in \partial M$.

(M° is the interior and ∂M is the boundary of the set M. Points satisfying conditions a) and b) are critical points of f in M°.) This leads to the procedure for finding the absolute extrema of f on M:

1. *Find the points $X \in M^\circ$ where f is differentiable and has zero partial derivatives.*

   (We recall again that differentiability can be decided using Theorem I.4.16.)

2. *Find the points $X \in M^\circ$ where f is not differentiable.*

   (That is, points $X \in M^\circ$ where the tangent plane to the graph of f does not exist.)

3. *Investigate at which points $X \in \partial M$ the function f can attain its absolute extrema on the boundary of M.*

   (A possible procedure is shown in Example I.6.16.)

4. *Finally, compute the values of f at all the points obtained. The largest value equals $\max_M f$ and the smallest value equals $\min_M f$.*

I.6.16. Example

Investigate the absolute extrema of the function

$$f(x,y) = x^2 - 2xy + 2y$$

on the rectangle $M = \langle 0;3\rangle \times \langle 0;2\rangle = \{[x;y] \in \mathbb{E}_2 :\ 0 \le x \le 3,\ 0 \le y \le 2\}$.
(By investigating extrema we mean investigating their existence, location, and value.)

Solution:
The function f is continuous on the nonempty, bounded, closed set M. Therefore, by Theorem I.6.14, $\max_M f$ and $\min_M f$ exist. The partial derivatives of f are:

$$\frac{\partial f}{\partial x} = 2x - 2y, \qquad \frac{\partial f}{\partial y} = -2x + 2.$$

Both partial derivatives are continuous functions on all of $\mathbb{E}_2$, hence also on M°. Therefore f is (by Theorem I.4.11) differentiable in M°.

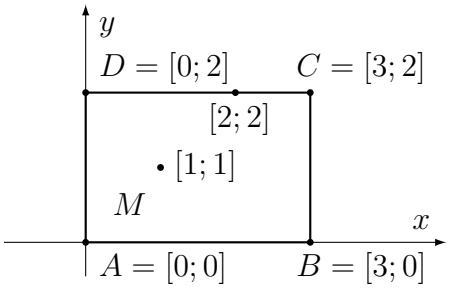
Setting both partial derivatives equal to zero, we obtain a system of two equations in the unknowns x and y:

$$2x - 2y = 0, \qquad -2x + 2 = 0.$$

This system has the unique solution $x = 1$, $y = 1$. The point $[1;1]$ is the only critical point of f. This point lies in M°, so it is of interest.

Now we investigate the extrema of f on the boundary of the rectangle M. The boundary ∂M is the union of four segments: $\overline{AB}$, $\overline{BC}$, $\overline{CD}$ and $\overline{DA}$. (See Fig. 7.)

Fig. 7



On $\overline{AB}$ we have $y = 0$ and $f(x, 0) = x^2$ for $0 \le x \le 3$. The function $f(x, 0)$ of the variable x is increasing, so it has a minimum (equal to 0) at $x = 0$ and a maximum (equal to 9) at $x = 3$. Thus we note that $\max_{\overline{AB}} f = f(3, 0) = 9$ and $\min_{\overline{AB}} f = f(0, 0) = 0$.

On $\overline{BC}$ we have $x = 3$ and $f(3, y) = 9 - 4y$ for $0 \le y \le 2$. The function $f(3, y)$ of the variable y is decreasing, so it has a maximum (equal to 9) at $y = 0$ and a minimum (equal to 5) at $y = 2$. Thus we find that $\max_{\overline{BC}} f = f(3, 0) = 9$ and $\min_{\overline{BC}} f = f(3, 2) = 5$.

On $\overline{CD}$ we have $y = 2$ and $f(x, 2) = x^2 - 4x + 4$ for $0 \le x \le 3$. By a method you know from Mathematics I, we easily find that $\max_{\overline{CD}} f = f(0, 2) = 4$ and $\min_{\overline{CD}} f = f(2, 2) = 0$. (This is also clear because $f(x, 2) = (x - 2)^2$.)

On $\overline{DA}$ we have $x = 0$ and $f(0, y) = 2y$ for $0 \le y \le 2$. Again we easily find that $\max_{\overline{DA}} f = f(0, 2) = 4$ and $\min_{\overline{DA}} f = f(0, 0) = 0$.

Therefore the largest of all values of f on ∂M is $9 = f(3, 0)$ and the smallest is $0 = f(0, 0) = f(2, 2)$. Comparing these values with the value of f at the critical point $[1; 1]$ ($f(1, 1) = 1$), we see that the largest value remains 9 and the smallest value remains 0. We can therefore conclude:

$$\max_M f = f(3, 0) = 9 \quad \text{and} \quad \min_M f = f(0, 0) = f(2, 2) = 0.$$

I.6.17.* Constrained extrema. Lagrange multiplier

Imagine that you are looking for extrema of the function $f(x, y) = x^2 + 2y^2$ on the unit circle $x^2 + y^2 = 1$. We will return to this problem after we show, on a more general level, how it can be solved.

The general problem is: We seek a maximum or minimum of the function $y = f(x_1, \dots, x_n)$ under the condition that

$$g(x_1, \dots, x_n) = 0. \tag{I.6.4}$$

Condition (I.6.4) is called a *constraint* and the extrema of the function f on the set of points satisfying condition (I.6.4) are called *constrained extrema*.

For simplicity, assume that both f and g are differentiable functions in a region $D \subset \mathbb{E}_n$. Form the function

$$L(x_1, \dots, x_n, \lambda) = f(x_1, \dots, x_n) + \lambda\, g(x_1, \dots, x_n). \tag{I.6.5}$$

The variable λ is called the *Lagrange multiplier*. It can be proved that:

Under condition (I.6.4), the function f can have extrema only at those points $[x_1; \dots; x_n] \in D$ for which there exists a value $\lambda \in \mathbb{R}$ such that $[x_1; \dots; x_n, \lambda]$ is a critical point of L.

The critical points of the function L are (given the assumed differentiability of both f and g) those points at which all partial derivatives of L with respect to $x_1, \dots, x_n$ and λ are equal to zero. That is, the points at which the equations

$$\frac{\partial L}{\partial x_1}(x_1, \dots, x_n, \lambda) = \frac{\partial f}{\partial x_1}(x_1, \dots, x_n) + \lambda \frac{\partial g}{\partial x_1}(x_1, \dots, x_n) = 0,$$

(I.6.6)
$$\vdots$$
$$\frac{\partial L}{\partial x_n}(x_1,\dots,x_n,\lambda) = \frac{\partial f}{\partial x_n}(x_1,\dots,x_n) + \lambda\,\frac{\partial g}{\partial x_n}(x_1,\dots,x_n) = 0,$$
$$\frac{\partial L}{\partial \lambda}(x_1,\dots,x_n,\lambda) = g(x_1,\dots,x_n) = 0.$$

This is a system of $n+1$ equations for $n+1$ unknowns: $x_1, \dots, x_n$ and λ. Solving this system yields the coordinates of the points at which the function f may have a constrained extremum. Which of these points actually contains a constrained extremum, and whether it is a maximum or a minimum, is decided according to the nature of the specific problem.

Now let us return to the problem stated at the beginning of this section. First: the function $f(x,y) = x^2 + 2y^2$ is continuous on the unit circle $x^2 + y^2 = 1$, which is a nonempty, closed, and bounded set in $\mathbb{E}_2$. By Theorem I.6.14, constrained absolute extrema of f on the mentioned unit circle exist. The (Lagrange) function L has the form:

$$L(x,y,z,\lambda) = x^2 + 2y^2 + \lambda\,(x^2+y^2-1)$$

and the system of equations (I.6.6) is

$$2x + \lambda\cdot 2x = 0, \quad 4y + \lambda\cdot 2y = 0, \quad x^2+y^2-1 = 0.$$

All solutions of this system are: a) $x = 0$, $y = \pm 1$, $\lambda = -2$, b) $x = \pm 1$, $y = 0$, $\lambda = -1$. Thus the function f can have its constrained extrema on the circle $x^2 + y^2 = 1$ only at the points $[0;1]$, $[0;-1]$, $[1;0]$ and $[-1;0]$. By computing and comparing the values of f at these points, we find that

$$\max_{x^2+y^2=1} f(x,y) = f(0,\pm 1) = 2 \quad \text{and} \quad \min_{x^2+y^2=1} f(x,y) = f(\pm 1,0) = 1.$$

In a similar way one can also solve problems of constrained extrema that contain more than one constraint condition of type (I.6.4). An exposition of the corresponding method can be found, for example, in book [2].

I.7 Implicit functions

I.7.1. Implicit function – motivation

We are used to the fact that a function of one variable $y = f(x)$ is given by an equation, for example $y = 3x^2+1$. But sometimes it happens that the function $y = f(x)$ is the result of some computations, and instead of an explicit (i.e., stated, obvious) expression of y as a function of x, we obtain an equation that looks, for example, like this:

(I.7.1)
$$2y - 3x^2y + 2x^3 + 4 = 0$$

or

(I.7.2)
$$y^5 - 7xy^2 + 2y - 28x^2 - 11 = 0.$$

It is natural to ask:

Does equation (I.7.1) or (I.7.2) uniquely define some function $y = f(x)$? If so, what derivative does this function have?

We will return to equations (I.7.1) and (I.7.2) later. For now, however, let us treat this problem on a general level, where instead of equation (I.7.1) or (I.7.2) we have the equation

(I.7.3)
$$F(x,y) = 0,$$

where F is a function of two variables. Assume that F is continuous. The question remains the same:

(Q) *Does equation (I.7.3) uniquely define some function $y = f(x)$? If so, what derivative does this function have?*

If the answer to the first question is yes, then we say that equation (I.7.3) defines the function $y = f(x)$ *implicitly*. (Implicitly means not explicitly, hidden.) We then call the function $y = f(x)$ simply an *implicit function*.

Question (Q) can be interpreted geometrically: the points $[x; y] \in \mathbb{E}_2$ satisfying equation (I.7.3) are points where the graph of the function $z = F(x, y)$ (a surface in $\mathbb{E}_3$) intersects the $x; y$-plane. Thus the question is: *Is this intersection a curve in the $x; y$-plane that is the graph of some function $y = f(x)$?*

If it is possible to solve equation (I.7.3) for y and obtain a unique expression of y in terms of x, then the problem is simple. An example is equation (I.7.1): from equation (I.7.1) you can easily compute that

$$y = \frac{2x^3 + 4}{3x^2 - 2} \quad \text{for} \quad x \neq \pm\sqrt{\tfrac{2}{3}}. \tag{I.7.4}$$

Equation (I.7.4) gives the function $y = f(x)$ explicitly, and its properties can already be studied by known methods. However, equation (I.7.2) represents a more complicated problem, because y cannot be expressed from it.

On the general level, where equation (I.7.3) may be arbitrarily complicated and y cannot be computed from it, a complete answer to question (Q) does not exist. Nevertheless, mathematical analysis has a theorem that allows one to obtain at least a "locally valid" answer. Roughly speaking, this theorem says that if equation (I.7.3) holds at a point $[x_0; y_0]$ and some other assumptions are satisfied, then in some neighborhood of x_0 there is a function $y = f(x)$ whose graph near $[x_0; y_0]$ coincides with the intersection of the graph of $F(x, y)$ with the $x; y$-plane. In other words, if we substitute $y = f(x)$ into (I.7.3), then equation (I.7.3) (i.e. $F(x, f(x)) = 0$) holds for x in a neighborhood of x_0. The theorem does not say what the function $y = f(x)$ looks like explicitly, nor how large the mentioned neighborhoods of x_0 and $[x_0; y_0]$ are. On the other hand, it does say what derivative the function $y = f(x)$ has in a neighborhood of x_0:

I.7.2. Theorem – implicit functions – 2D version

Assume that

a) *the function $F(x, y)$ has both partial derivatives continuous in some neighborhood of the point $[x_0; y_0]$,*

b) $F(x_0, y_0) = 0$,

c) $\dfrac{\partial F}{\partial y}(x_0, y_0) \neq 0$.

Then there exist numbers $\delta > 0$ and $\epsilon > 0$ and a unique function $y = f(x)$, defined on the interval $(x_0 - \delta; x_0 + \delta)$, such that:

1. $y_0 = f(x_0)$,
2. $\forall\, x \in (x_0 - \delta; x_0 + \delta) \,:\, y = f(x) \in (y_0 - \epsilon; y_0 + \epsilon)$ *and* $F(x, f(x)) = 0$,
3. *the function f is continuous and has a continuous derivative f' on the interval $(x_0 - \delta; x_0 + \delta)$,*
4. $\forall\, x \in (x_0 - \delta; x_0 + \delta) \,:\, f'(x) = -\dfrac{\partial F}{\partial x}(x, f(x)) \Big/ \dfrac{\partial F}{\partial y}(x, f(x))$.

I.7.3. Remark

In order not to complicate the theorem too much, let us add only in this remark that under the more general assumption that the function F has continuous partial derivatives up to order k in some neighborhood of the point $[x_0; y_0]$, it can be proved that the implicitly defined function (which we briefly call the "implicit function") $y = f(x)$ also has continuous derivatives up to order k in a neighborhood of x_0.

The formula in statement 4 of Theorem I.7.2 can be derived easily. We view y in equation (I.7.3) as a function of the variable x and differentiate the equation with respect to x. We differentiate the left-hand side of equation (I.7.3) using the rule for differentiating a composite function (Theorem I.5.14). We obtain:

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \cdot y' = 0. \tag{I.7.5}$$

Solving for y' and noting that $y = f(x)$ and $y' = f'(x)$, we get the desired formula. Using the simplified notation

$$F_x = \frac{\partial F}{\partial x}, \qquad F_y = \frac{\partial F}{\partial y},$$

we can write the formula in the form

$$y' = -\frac{F_x}{F_y}, \tag{I.7.6}$$

where y' is considered at the point x and F_x and F_y at the point $[x; f(x)]$.

In a similar way, one can also obtain an expression for the second derivative of the function $y = f(x)$. Assume that the function F has continuous second partial derivatives. Differentiate equation (I.7.5) once more with respect to x. Denoting, for simplicity, the partial derivatives of F again only by indices, we obtain:

$$F_{xx} + F_{yx}\, y' + F_{xy}\, y' + F_{yy}\, y'^2 + F_y\, y'' = 0.$$

Using the equality $F_{xy} = F_{yx}$ (see Theorem I.5.12) and substituting the expression for y' from (I.7.6), we get:

$$y'' = f''(x) = -\frac{F_{xx}\, F_y^2 - 2F_{xy}\, F_x\, F_y + F_{yy}\, F_x^2}{F_y^3},$$

where y'' is considered at the point x and all derivatives of F at the point $[x; f(x)]$. You do not need to memorize this formula. It is, however, good to remember the procedure that leads to it. That is, we differentiate equation (I.7.3) twice with respect to x and then solve for y''.

I.7.4. Example

Let us return to equation (I.7.2). The function on the left-hand side is

$$F(x, y) = y^5 - 7xy^2 + 2y - 28x^2 - 11.$$

By substituting into equation (I.7.2), you can verify that the equation is satisfied, for example, at the point $[2; 3]$. That is, $F(2, 3) = 0$. Since F is a polynomial in two variables, it is a function that is continuous and has continuous partial derivatives of all orders on all of $\mathbb{E}_2$. The partial derivative of F with respect to y at the point $[2; 3]$ is:

$$\frac{\partial F}{\partial y}(2, 3) = (5y^4 - 14xy + 2)\Big|_{\substack{x=2\\y=3}} = 323 \neq 0.$$

By Theorem I.7.2 there exists a unique function $y = f(x)$, defined in a neighborhood of the point $x_0 = 2$, such that for all x in this neighborhood it satisfies equation (I.7.2). This function is implicitly defined by equation (I.7.2) in a neighborhood of $x_0 = 2$. We do not know the explicit form of this function; Theorem I.7.2 only provides information about the existence of such a function. Moreover, using Theorem I.7.2 we can compute the derivative of this function at the point $x_0 = 2$:

$$y'(2) = f'(2) = -\frac{F_x}{F_y}\Big|_{\substack{x=2\\y=3}} = -\frac{-7y^2 - 56x}{5y^4 - 14xy + 2}\Big|_{\substack{x=2\\y=3}} = -\frac{-166}{323} \doteq 0.51\,.$$

Now compute, yourself, using the procedure described in section I.7.3, the value of $y''(2)$.

I.7.5. Generalization of the implicit function theorem – 3D version
Instead of equation (I.7.3) we may consider the equation

$$F(x, y, z) = 0 \tag{I.7.7}$$

and ask under what conditions this equation implicitly defines a function $z = f(x, y)$. The following theorem gives a "local" answer:

Theorem

Assume that

a) *the function $F(x, y, z)$ has continuous partial derivatives in some neighborhood of the point $[x_0; y_0; z_0]$,*

b) $F(x_0, y_0, z_0) = 0$,

c) $\dfrac{\partial F}{\partial z}(x_0, y_0, z_0) \neq 0$.

Then there exist numbers $\delta > 0$ and $\epsilon > 0$ and a unique function $z = f(x, y)$, defined in a neighborhood $\mathcal{U}_\delta(x_0, y_0)$ of the point $[x_0; y_0]$, such that:

1. $z_0 = f(x_0, y_0)$,
2. $\forall\, [x; y] \in \mathcal{U}_\delta(x_0, y_0) \,:\, z = f(x, y) \in (z_0 - \epsilon; z_0 + \epsilon)$ *and* $F(x, y, f(x, y)) = 0$,
3. *the function f is continuous and has continuous partial derivatives with respect to x and with respect to y in a neighborhood $\mathcal{U}_\delta(x_0, y_0)$,*
4. $\forall\, [x; y] \in \mathcal{U}_\delta(x_0, y_0) \,:\quad \dfrac{\partial f}{\partial x} = -\dfrac{\partial F}{\partial x}\Big/\dfrac{\partial F}{\partial z}, \qquad \dfrac{\partial f}{\partial y} = -\dfrac{\partial F}{\partial y}\Big/\dfrac{\partial F}{\partial z},$

   where the partial derivatives of the function f are considered at the point $[x; y]$ and the partial derivatives of the function F at the point $[x; y; f(x, y)]$.

Now try, using the procedure described in Section I.7.3, to compute what the second partial derivatives $\partial^2 f/\partial x^2$, $\partial^2 f/\partial x\,\partial y$ and $\partial^2 f/\partial y^2$ will look like.

I.8 Exercises

1. Determine the domains and ranges of the given functions. *(See Section I.2.2.)*

a) $f(x,y) = \mathrm{e}^{16-x^2-y^2}$ b) $f(x,y) = \dfrac{1}{x(y+3)}$ c) $f(x,y) = \ln(\mathrm{e}^2 + x^2 + y^2)$

d) $f(x,y) = \sqrt[4]{y-x}$ e) $f(x,y) = \sqrt{y-x^2}$ f) $f(x,y) = \cos(3x^2 - 2y + 5)$

g) $f(x,y,z) = \sqrt{x^2+y^2+z^2-1}$ h) $f(x,y,z) = yz \ln x$

i) $f(x,y,z) = \dfrac{1}{x^2+y^2+z^2}$ j) $f(x,y,z) = \arctan(x+y+z)$

2. Compute the following limits, if they exist. *(See Sections I.3.2–I.3.12.)*

a) $\lim\limits_{[x;y]\to[0;0]} \dfrac{3x^2-y^2+5}{x^2+y^2+2}$ b) $\lim\limits_{[x;y]\to[0;0]} \dfrac{x^4}{x^4+y^2}$ c) $\lim\limits_{[x;y]\to[0;0]} \dfrac{\mathrm{e}^y \sin x}{x}$

d)* $\lim\limits_{\substack{[x;y]\to[0;0]\\ x\neq 0}} \dfrac{\mathrm{e}^y \sin x}{x}$ e)* $\lim\limits_{\substack{[x;y]\to[2;2]\\ x+y\neq 4}} \dfrac{x+y-4}{\sqrt{x+y}-2}$ f)* $\lim\limits_{\substack{[x;y]\to[0;0]\\ x\neq y}} \dfrac{x+y}{x-y}$

3. On which sets $M \subset \mathbb{E}_2$ are the following functions continuous? *(See Sections I.3.14–I.3.20.)*

a) $f(x,y) = \dfrac{x+y}{x-y}$ b) $f(x,y) = \dfrac{x^2+y^4+1}{x^2+x-12}$ c) $f(x,y) = \dfrac{1}{x^2-2y}$

d) $f(x,y) = \ln \dfrac{y}{x}$ e) $f(x,y) = \cos(x^2+xy)$ f) $f(x,y) = \mathrm{e}^{1/(x+y)}$

4. On which sets $M \subset \mathbb{E}_3$ are the following functions continuous? *(See Sections I.3.14–I.3.20.)*

a) $f(x,y,z) = \dfrac{1}{x^2+z^2-4}$ b) $f(x,y,z) = \ln(xyz)$ c) $f(x,y,z) = \mathrm{e}^z \sin(x+y)$

d) $f(x,y,z) = \dfrac{x+y}{x-y}$ e) $f(x,y,z) = \ln \dfrac{1}{xyz}$ f) $f(x,y,z) = \dfrac{1}{|xy|+|z|}$

g) $f(x,y,z) = \dfrac{1}{\ln\sqrt{x^2+y^2+z^2}}$ h) $f(x,y,z) = \dfrac{y+4}{x^2y-xy+4x^2-4x}$

5. Compute the partial derivatives with respect to x and with respect to y. *(See Sections I.4.1–I.4.7.)*

a) $f(x,y) = x^2 - 7xy$ b) $f(x,y) = (x+2)^2 y$ c) $f(x,y) = x^2(3y-5)^7$

d) $f(x,y) = x \sin(xy)$ e) $f(x,y) = \ln(x^2 y)$ f) $f(x,y) = \dfrac{2x}{x - \sin y}$

g) $f(x,y) = \dfrac{x+y}{x-y}$ h) $f(x,y) = \ln(x^2-2y)$ i) $f(x,y) = \sqrt{x^2+y^2}$

j) $f(x,y) = \mathrm{e}^x \ln y$ k) $f(x,y) = \dfrac{1}{\tan(y/x)}$ l) $f(x,y) = y\,\mathrm{e}^{x^2 y}$

6. Compute the partial derivatives with respect to x, y, and z. *(See Sections I.4.1–I.4.7.)*

a) $f(x,y,z) = \dfrac{x^5y^2}{z^3}$ b) $f(x,y,z) = x - \sqrt{y^2+z^2}$

c) $f(x,y,z) = \arctan(x+y+z)$ d) $f(x,y,z) = xy + yz + zx$

e) $f(x,y,z) = \sqrt{x^2+y^2+z^2}$ f) $f(x,y,z) = \dfrac{1}{\sqrt{(x^2+y^2+z^2)}}$

g) $f(x,y,z) = x^2 \sin^2 y \cos z^2$ h) $f(x,y,z) = \dfrac{x^2}{\sqrt{y^2+z^2}}$

7. Compute the second-order partial derivatives of the following functions. *(See Sections I.5.9–I.5.12.)*

a) $f(x,y) = x^2y + \cos y + y \sin x$ b) $f(x,y) = x\,\mathrm{e}^y + y + x^5y^4 - 13$

c) $f(x,y) = \mathrm{e}^{x+3y} + x \ln y + y \ln x + 3$ d) $f(x,y) = y + x^2y + 4y^3x - \ln(y^2+x)$

e) $f(x,y) = y^2 + y(\sin x - x^4)$ f) $f(x,y) = x^2 + 5xy + \sin(xy) + x\,\mathrm{e}^{y^2/2}$

8. Compute the gradient of the function f at the point A and the directional derivative $\partial f/\partial \vec{s}(A)$. *(See Sections I.5.3–I.5.6.)*

a) $f(x,y) = x^2 + 2xy - 3y^2, \quad A = [1;1], \quad \vec{s} = (3;4)$

b) $f(x,y,z) = x^2 + 2y^2 - 3z^3 - 17, \quad A = [1;1;1], \quad \vec{s} = (1;1;1)$

c) $f(x,y,z) = \cos(xy) + \mathrm{e}^{yz} + \ln(zx), \quad A = [1;0;0.5], \quad \vec{s} = (1;2;2)$

9. Write the equation of the tangent plane and the normal line at the point $[x_0; y_0; f(x_0,y_0)]$ to the surface that is the graph of the function $z = f(x,y)$. *(See Section I.4.13.)*

a) $f(x,y) = 2x + y^2, \ x_0 = 1, \ y_0 = 2$ b) $f(x,y) = \sin(x-y), \ x_0 = 0, \ y_0 = \pi$

c) $f(x,y) = \sqrt{x} - y, \ x_0 = 4, \ y_0 = 1$ d) $f(x,y) = \ln(x+y), \ x_0 = 1, \ y_0 = 0$

10. Write the equation of the tangent plane and the normal line at the point A to the surface given by the equation $f(x,y,z) = 0$. *(See Section I.5.10.)*

a) $f(x,y,z) = x^2 + y^2 + z^2 - 3 = 0, \quad A = [1;1;1]$

b) $f(x,y,z) = \cos(\pi x) - x^2y + \mathrm{e}^{xz} + yz - 4 = 0, \quad A = [0;1;2]$

11. Investigate the local extrema of the given functions. *(See Sections I.6.1–I.6.11.)*

a) $f(x,y) = 2xy - 5x^2 - 2y^2 + 4x + 4y$ b) $f(x,y) = x^2 + xy + 3x + 2y + 5$

c) $f(x,y) = 5xy - 7x^2 + 3x - 6y + 2$ d) $f(x,y) = x^2 - 4xy + y^2 + 6y + 2$

e) $f(x,y) = 2x^2 + 3xy + 4y^2 - 5x + 2y$ f) $f(x,y) = x^2 - y^2 - 2x + 4y + 6$

g) $f(x,y) = x^3 + y^3 + 3x^2 - 3y^2 - 8$ h) $f(x,y) = 2x^3 + 2y^3 - 9x^2 + 3y^2 - 12y$

i) $f(x,y) = 4xy - x^4 - y^4 - 11$ j) $f(x,y) = x^4 + y^4 + 4xy + 7$

12. Investigate the existence, location, and value of absolute extrema of the given functions on the given sets. *(See Sections I.6.12–I.6.16.)*

a) $f(x,y) = 2x^2 - 4x + y^2 - 4y + 2, \quad M = \{[x;y];\ x \geq 0,\ y \leq 2,\ y \geq 2x\}$

b) $f(x,y) = x^2 - xy + y^2 + 7, \quad M = \{[x;y];\ x \geq 0,\ y \leq 4,\ y \geq x\}$

c) $f(x,y) = x^2 + xy + y^2 - 6x + 2, \quad M = \{[x;y];\ 0 \leq x \leq 5,\ -3 \leq y \leq 3\}$

d) $f(x,y) = x^2 + xy + y^2 - 6x, \quad M = \{[x;y];\ 0 \leq x \leq 5,\ -3 \leq y \leq 0\}$

e) $f(x,y) = 48xy - 32x^3 - 24y^2, \quad M = \{[x;y];\ 0 \leq x \leq 1,\ 0 \leq y \leq 1\}$

f) $f(x,y) = x^2 - y^2, \quad M = \{[x;y];\ x \geq -1,\ y \geq -1,\ x + 2y \leq 2\}$

13. Show that the equation $F(x,y) = 0$ implicitly defines a function $y = f(x)$ in a neighborhood of the point $[x_0;y_0] \in \mathbb{E}_2$. Compute the derivative of this function at the point x_0 and write the equation of the tangent line to its graph at the point $[x_0;y_0]$.
(See Sections I.7.1–I.7.4.)

a) $F(x,y) = x^3 + y^3 - 6xy + 4, \quad [x_0;y_0] = [1;1]$

b) $F(x,y) = x^2 - 3xy + 5xy^2 - 6y^2 - 32, \quad [x_0;y_0] = [2;3]$

c) $F(x,y) = x\,\mathrm{e}^{y/x} - \mathrm{e}^5, \quad [x_0;y_0] = [1;5]$

d) $F(x,y) = \arctan(2x + y), \quad [x_0;y_0] = [0;0]$

14. Show that the equation $F(x,y,z) = 0$ implicitly defines a function $z = f(x,y)$ in a neighborhood of the point $[x_0;y_0;z_0] \in \mathbb{E}_3$. Compute its partial derivatives at the point $[x_0;y_0]$ and write the equation of the tangent plane to its graph at the point $[x_0;y_0;z_0]$.
(See Section I.7.5.)

a) $F(x,y,z) = z^3 - xy + yz + y^3 - 2 = 0, \quad [x_0;y_0;z_0] = [1;1;1]$

b) $F(x,y,z) = x^2 - 2y^2 + z^2 - 4x + 2z - 5 = 0, \quad [x_0;y_0;z_0] = \left[-1;\ \sqrt{\tfrac{3}{2}};\ 1\right]$

c) $F(x,y,z) = xz^2 - x^2y + y^2z + 2x - y = 0, \quad [x_0;y_0;z_0] = [0;1;1]$

d) $F(x,y,z) = \sin(x+y) + \sin(y+z) + \sin(x+z) = 0, \quad [x_0;y_0;z_0] = [\pi;\pi;\pi]$

CHAPTER II.

Riemann's integral of functions of two and three variables

II.1 Double integral – motivation and definition. Two-dimensional Jordan measure and measurable sets in $\mathbb{E}_2$ 51
II.2 Existence and selected properties of the double integral 56
II.3 Computation of the double integral – Fubini's theorem and transformation to polar coordinates 57
II.4 Some physical applications of the double integral 62
II.5 Triple integral – motivation and definition. Three-dimensional Jordan measure and measurable sets in $\mathbb{E}_3$ 63
II.6 Existence and selected properties of the triple integral 66
II.7 Computation of the triple integral – Fubini's theorem and transformation to cylindrical and spherical coordinates 67
II.8 Some physical applications of the triple integral 72
II.9 * More on the substitution method for computing double and triple integrals 73
II.10 Exercises 76

II.1 Double integral – motivation and definition. Two-dimensional Jordan measure and measurable sets in $\mathbb{E}_2$

II.1.1. Partition of a rectangle and its norm

Let $\langle a\,;b\rangle$ be a closed interval on the x-axis and $\langle c\,;d\rangle$ a closed interval on the y-axis. Then the Cartesian product $O = \langle a\,;b\rangle \times \langle c\,;d\rangle$ is a rectangle in $\mathbb{E}_2$ whose sides are parallel to the coordinate axes x and y.

The rectangle O can be partitioned by a rectangular grid of line segments parallel to the x-axis or the y-axis into n subrectangles $O_1, \ldots, O_n$. The system of these rectangles is called a *partition* of the rectangle O. If we call the described partition D, then the maximum of the side lengths of all subrectangles $O_1, \ldots, O_n$ is denoted by $\|D\|$ and is called the *norm* of the partition D. If the side lengths of the subrectangles $O_1, \ldots, O_n$ are $\Delta x_1, \Delta y_1, \ldots, \Delta x_n, \Delta y_n$ (see Fig. 8), then clearly:

$$\|D\| \;=\; \max\{\Delta x_1;\, \Delta y_1;\, \ldots\,;\, \Delta x_n;\, \Delta y_n\}.$$

II.1.2. Physical motivation

Imagine that a thin plate has the shape of the rectangle $O = \langle a\,;b\rangle \times \langle c\,;d\rangle$ in the xy-plane. The plate need not be homogeneous, and its areal density (i.e., mass per unit area–denote it by $\rho(x,y)$) may vary depending on the position of the point $[x;y]$. We want to compute the mass m of the plate.

If the plate is homogeneous and its areal density ρ is therefore constant, the problem is simple: m equals the product of ρ and the area of the plate, i.e. $m = \rho \cdot (b-a) \cdot (d-c)$.

In the general case, when ρ need not be constant, we can partition the rectangle O by a partition D into many smaller rectangles $O_1, \ldots, O_n$. Let Δx_i and Δy_i be the side lengths of the i-th rectangle O_i (for $i = 1, \ldots, n$). If the partition D is "sufficiently fine" (i.e., the norm $\|D\|$ is "sufficiently small"), then we do not make a large error if we replace the function ρ on each of the rectangles $O_1, \ldots, O_n$ by a constant. A "suitable" constant on O_i is the value $\rho(Z_i)$,

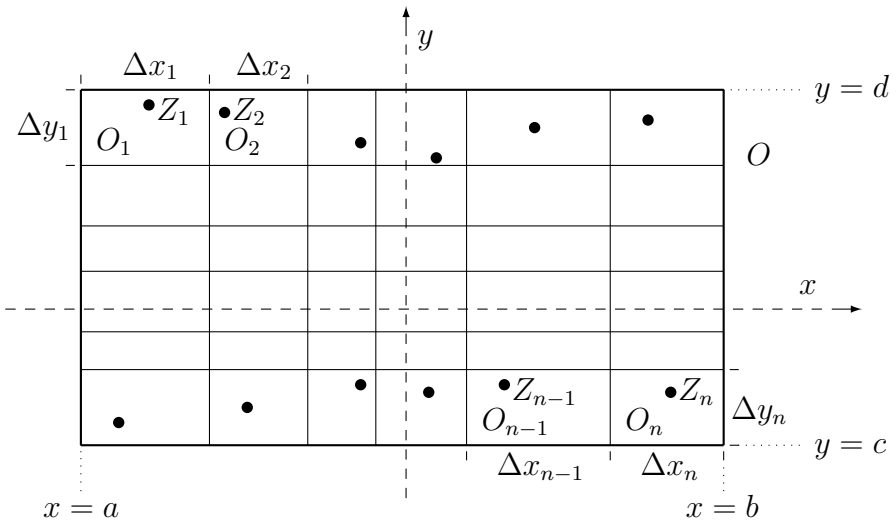


Fig. 8

where Z_i is some chosen point (tag) in O_i. (See Fig. 8.) The mass m_i of the part of the plate covering the rectangle O_i is then approximately equal to the product of $\rho(Z_i)$ and the area of O_i: $m_i \doteq \rho(Z_i) \cdot \Delta x_i \, \Delta y_i$. The mass of the whole plate O can be expressed approximately as:

(II.1.1) $$m = \sum_{i=1}^{n} m_i \doteq \sum_{i=1}^{n} \rho(Z_i) \cdot \Delta x_i \, \Delta y_i \, .$$

It is natural to expect that if the function ρ is "reasonable enough," for example continuous, then the sum on the right-hand side will approach the exact value of the mass of the plate O as we divide O into an ever larger number of smaller parts $O_1, \ldots, O_n$ (i.e. $n \to +\infty$), while the lengths of their sides tend to zero (i.e. $\|D\| \to 0$).

II.1.3. Geometric motivation

Imagine that $z = f(x, y)$ is a nonnegative function on a set $M \subset \mathbb{E}_2$. We want to define and compute the volume V of the set Q bounded above by the graph of f and bounded below by the xy-plane. (See Fig. 9a.) A precise description of Q is: $Q = \{[x; y; z] \in \mathbb{E}_3 : \ [x; y] \in M, \ 0 \leq z \leq f(x, y)\}$. Assume further, for simplicity, that M is the rectangle $O = \langle a \, ; b \rangle \times \langle c \, ; d \rangle$. (Fig. 9b.)

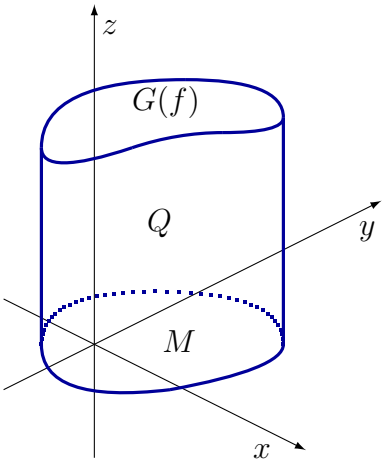


Fig. 9a

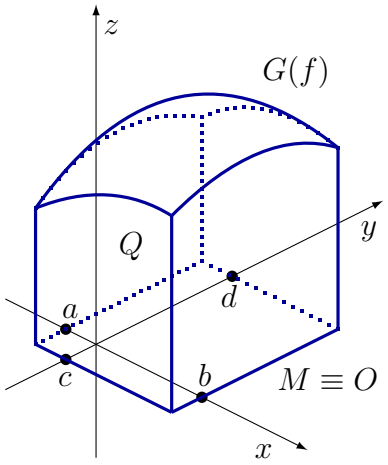


Fig. 9b

If the function f is constant on O, then the set Q is a rectangular box whose edge lengths are $b-a$, $d-c$, and f. The volume of Q is therefore: $V = f \cdot (b-a) \cdot (d-c)$.

In the general case, when f need not be constant, we can partition the rectangle O by a partition D into n smaller rectangles $O_1, \ldots, O_n$. If the rectangles $O_1, \ldots, O_n$ are "very small," then when computing the volume V_i of the set "above" the rectangle O_i we do not make a large error if we choose a point Z_i in O_i and replace the function f on O_i by the constant $f(Z_i)$. Then approximately: $V_i \doteq f(Z_i) \cdot \Delta x_i \, \Delta y_i$, where Δx_i and Δy_i are the side lengths of the rectangle O_i. The volume of the entire set Q between the graph of f and the xy-plane can be expressed approximately as:

$$V = \sum_{i=1}^{n} V_i \doteq \sum_{i=1}^{n} f(Z_i) \cdot \Delta x_i \, \Delta y_i \,. \tag{II.1.2}$$

It is natural to regard the limit of the sums on the right-hand side for $n \to +\infty$ and $\Delta x_i \to 0$, $\Delta y_i \to 0$ (i.e. for $\|D\| \to 0$) as the exact value of the volume V of the set Q, if this limit exists.

Both sums (II.1.1) and (II.1.2) are the same, and both problems described in sections II.1.2 and II.1.3 lead to the limit of the sums (II.1.1) and (II.1.2) as $\|D\| \to 0$. In the following sections we will give a precise definition of this limit. We can already note here that if the limit exists, we will call its value the *double integral* of the function ρ (or f) over the rectangle O.

II.1.4. Riemann sums and their limit

Suppose that $f(x,y)$ is a bounded function on the rectangle $O = \langle a\,;b\rangle \times \langle c\,;d\rangle$ in $\mathbb{E}_2$. Let D be a partition of O into subrectangles $O_1, \ldots, O_n$ whose side lengths are $\Delta x_1, \Delta y_1, \ldots, \Delta x_n, \Delta y_n$. Let $\mathcal{V}$ denote the system of chosen points $Z_i \in O_i$ $(i = 1, 2, \ldots, n)$. (See Fig. 9.) The *Riemann sum* of the function f over the rectangle O corresponding to the partition D and the system of points $\mathcal{V}$ is defined as

$$s(f, D, \mathcal{V}) = \sum_{i=1}^{n} f(Z_i) \cdot \Delta x_i \, \Delta y_i \,. \tag{II.1.3}$$

We say that a number S is the *limit of the Riemann sums* $s(f, D, \mathcal{V})$ *as* $\|D\| \to 0+$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that for every partition D of the rectangle O and for every chosen system $\mathcal{V}$ we have:

$$\|D\| < \delta \quad \Longrightarrow \quad |s(f, D, \mathcal{V}) - S| < \epsilon.$$

We write:

$$\lim_{\|D\| \to 0+} s(f, D, \mathcal{V}) = S. \tag{II.1.4}$$

II.1.5. Double integral over a rectangle

If the limit (II.1.4) exists, then we call the number S the *double integral* of the function f *over the rectangle* O. We usually denote the integral by

$$\iint_O f(x,y)\, dx\, dy \qquad \text{or} \qquad \iint_O f \, dx\, dy.$$

If the limit (II.1.4) exists, we say that "the double integral $\iint_O f \, dx\, dy$ exists" or that "the function f is *integrable* on the rectangle O."

II.1.6. Double integral over a general bounded set in $\mathbb{E}_2$

Suppose that M is a bounded set in $\mathbb{E}_2$ and $f(x,y)$ is a bounded function on M. Let $O =$

$\langle a\,;b\rangle \times \langle c\,;d\rangle$ be a rectangle containing the set M. Define

$$f^*(x;y) = \begin{cases} f(x,y) & \text{for } [x;y] \in M, \\ 0 & \text{for } [x;y] \in O\setminus M. \end{cases}$$

Then by the *double integral* of the function f *over the set* M we mean the double integral of the function f^* over the rectangle O (if this integral exists). We denote the double integral of f over M by

$$\iint_M f(x,y)\,dx\,dy \qquad \text{or} \qquad \iint_M f\,dx\,dy.$$

You see that the double integral of the function f over a general bounded set M is defined by means of the double integral of the function f^* (i.e., f extended by zero outside M) over the rectangle O. The double integral over a rectangle is already known from section II.1.5. One can show that neither the existence nor the value of the double integral $\iint_O f^*\,dx\,dy$ depends on which of the infinitely many rectangles in $\mathbb{E}_2$ containing M we choose and denote by O. This is because the extended function f^* is zero outside the set M.

If the integral $\iint_O f^*\,dx\,dy$ exists, then we also say that "the double integral $\iint_M f\,dx\,dy$ exists" or that "the function f is *integrable* on the set M."

We call the set M the *domain of integration* and the integrated function f is often called the *integrand*.

II.1.7. Physical and geometric meaning of the double integral

From section II.1.2 and from the definition of the double integral, one of the many physical meanings of the double integral is clear: If M is a bounded set (a thin plate) in $\mathbb{E}_2$ and $\rho(x,y)$ is the areal density of the plate (in kg/m^2), then $\iint_M \rho\,dx\,dy$ expresses the *total mass* of the plate M. You will encounter further applications of the double integral in physics in Chapter II.4.

From section II.1.3 and from the definition of the double integral it also follows that if $f(x,y)$ is a nonnegative and integrable function on a set $M \subset \mathbb{E}_2$, then the integral $\iint_M f\,dx\,dy$ can be used to define the *volume* of the set Q between the xy-plane and the graph of the function f over M. (See Fig. 9a.) In Chapters II.5–II.7 you will see that this volume, as well as the volume of more general bodies in $\mathbb{E}_3$, can be defined and computed also using the so-called triple integral.

II.1.8. Historical approach

In the past, the double integral used to be introduced (not defined, because it is not a precise definition) in the following way: For simplicity, let us imagine that M is again a bounded set (a thin plate) in $\mathbb{E}_2$ and $\rho(x,y)$ is the surface density of the plate M. As in Section II.1.2, we want to compute the total mass of this plate. Let us divide M into infinitely many infinitely small rectangles of the form $\langle x;x+dx\rangle \times \langle y;y+dy\rangle$. On each of these rectangles, ρ can be considered as a constant function with value $\rho(x,y)$. (The rectangle $\langle x;x+dx\rangle \times \langle y;y+dy\rangle$ in fact coincides with the point $[x;y]$ because it is "infinitely small".) The mass of the rectangle $\langle x;x+dx\rangle \times \langle y;y+dy\rangle$ is therefore $dm = \rho(x,y)\cdot dx\,dy$. By summing all masses dm over the whole set M, we obtain the total mass m of the considered plate. This sum is denoted by $\iint_M \rho\,dx\,dy$ and is called the double integral of the function ρ over the set M.

However, this approach is imprecise, because it relies on the notions of an "infinitely small rectangle" and a "sum of infinitely many infinitely small numbers" dm. The idea of the exis-

tence of an "infinitely small rectangle" contradicts the modern conception of geometry in $\mathbb{E}_2$, and likewise the notion of a "sum of infinitely many infinitely small numbers" is completely unclear without a precise definition. The development of mathematics and the needs of applications of the double integral therefore forced the later emergence of a precise definition and theory of the double integral. Nevertheless, the idea that an integral is a sum of infinitely many infinitely small numbers $dm = \rho(x,y) \cdot dx\,dy$ is still used, for example in mechanics when deriving various formulas (see Chapter II.4).

II.1.9. A measurable set in $\mathbb{E}_2$ and its Jordan measure

The notion of a bounded set in $\mathbb{E}_2$ is too general for practical purposes. For example, it can be shown that there exist bounded sets $M \in \mathbb{E}_2$ of such complicated shape that even the constant function $f(x,y) = 1$ is not integrable on them. (An example is the set of all points of the square $\langle 0;1\rangle \times \langle 0;1\rangle$ whose both coordinates are rational numbers.) In order to distinguish these "artificial" sets, which are of little importance in technical applications, from "reasonable" sets, we introduce the concept of so-called measurable sets.

Assume that M is a bounded set in $\mathbb{E}_2$. We say that this set is *measurable* (in the Jordan sense) if the double integral of the constant function $f(x,y) = 1$ over M exists. In this case, we call the number

$$\mu_2(M) \;=\; \iint_M 1\; dx\,dy$$

the *two-dimensional Jordan measure* of the set M.

$\mu_2(M)$ has a clear geometric interpretation – it defines and provides a method for computing the *area* (or simply the *area measure*) of the set M. Once you learn how to compute double integrals, you will be able to convince yourself that if the set M has a simple shape (square, rectangle, triangle, circle, etc.), then $\mu_2(M)$ agrees with the formulas for areas that you know from elementary school.

Simple bounded sets (square, rectangle, triangle, circle, sector of a circle, etc.) are measurable. It can be added that practically all other bounded sets that you will encounter in technical applications are also measurable. The criterion by which we can recognize this is given in Section II.1.12. First, however, we will mention some measurable sets for which we can show that their measure is zero.

II.1.10. Examples of sets whose two-dimensional measure is equal to zero.

a) The empty set and all sets consisting of a finite number of points or line segments.

b) Graphs of continuous functions of one variable $y = \varphi(x)$ or $x = \psi(y)$ on bounded closed intervals.

c) So-called simple smooth curves or simple piecewise smooth curves in $\mathbb{E}_2$ (see Chapter III.1).

The following theorem states some fairly obvious properties of sets in $\mathbb{E}_2$ whose two-dimensional Jordan measure is zero (in short: sets of measure zero).

II.1.11. Theorem

a) *The union of finitely many sets of measure zero is a set of measure zero.*

b) *If N is a set of measure zero and $M \subset N$, then M is also a set of measure zero.*

II.1.12. Theorem. (A necessary and sufficient condition for a set in $\mathbb{E}_2$ to be measurable.)
A set $M \subset \mathbb{E}_2$ is measurable (in the Jordan sense) if and only if it is bounded and $\mu_2(\partial M) = 0$.

II.2 Existence and selected properties of the double integral

We recall that the statements "f is an integrable function on the set M" and "the double integral $\iint_M f\,dx\,dy$ exists" have the same meaning. It would be naive to think that the double integral of any function f over any set M always exists (in other words, that for any function f on any set M the construction described in Sections II.1.4–II.1.6 always leads to a unique value of the limit of the Riemann sums (II.1.3)). First of all, one must keep in mind that the construction of the double integral requires the integration domain M to be a bounded set and also requires the function f to be bounded on M. However, boundedness of M and f by itself still does not guarantee the existence of the integral. This is because the notions of a bounded set and a bounded function are still too general and include many sets and functions of very "strange" shapes. We do not wish to deal with such cases in this text. In the following section, we present a simple sufficient condition for the existence of the double integral, which is usually entirely sufficient in technical applications.

II.2.1. Theorem. (A sufficient condition for the existence of the double integral.)
Let M be a measurable set in $\mathbb{E}_2$ and let f be a bounded and continuous function on M. Then the double integral $\iint_M f\,dx\,dy$ exists.

From Theorem I.6.14 it follows that a continuous function on a bounded closed set is bounded. Theorem II.2.1 can therefore be modified. One can require something more of the set M, namely closedness, and in exchange one can omit the assumption that the function f is bounded on M, since this is guaranteed by Theorem I.6.14. Formulate and write down the corresponding modified theorem yourself.

Theorem II.2.1 remains valid if we replace the assumption of continuity of f on the set M by the weaker assumption that f is continuous on M "except on a set of measure zero" (which allows discontinuities of f, but only on a set of measure zero in M).

II.2.2. Some important properties of the double integral.

a) **Linearity of the double integral**

If the functions f and g are integrable on the set $M \subset \mathbb{E}_2$ and $\alpha \in \mathbb{R}$, then the functions $f+g$ and αf are also integrable on the set M and the following holds:

$$\iint_M (f+g)\,dx\,dy = \iint_M f\,dx\,dy + \iint_M g\,dx\,dy,$$

$$\iint_M \alpha \cdot f\,dx\,dy = \alpha \cdot \iint_M f\,dx\,dy.$$

b) **Additivity of the double integral with respect to the domain of integration**

If M_1 and M_2 are measurable sets whose intersection has measure zero (i.e., M_1 and M_2 overlap at most on a set of measure zero) and f is integrable on M_1 and on M_2, then f is also integrable on the union $M_1 \cup M_2$ and

$$\iint_{M_1} f\,dx\,dy + \iint_{M_2} f\,dx\,dy = \iint_{M_1 \cup M_2} f\,dx\,dy.$$

c) *If the function f is integrable on the set $M \in \mathbb{E}_2$ and the function g is bounded and differs from f at most on a set of measure zero, then g is also integrable on M and*

$$\iint_M g \, dx\, dy \;=\; \iint_M f \, dx\, dy.$$

d) *If f is a bounded function on a set $M \in \mathbb{E}_2$ of measure zero, then f is integrable on M and*

$$\iint_M f \, dx\, dy \;=\; 0.$$

e) *If f and g are integrable functions on the set $M \subset \mathbb{E}_2$ such that $f(x,y) \geq g(x,y)$ for all points $[x;y] \in M$, then*

$$\iint_M f \, dx\, dy \;\geq\; \iint_M g \, dx\, dy.$$

In particular, if $f(x,y) \geq 0$ for all points $[x;y] \in M$, then

$$\iint_M f \, dx\, dy \;\geq\; 0.$$

From statement c) it follows that if we change the values of the function f on M on a set of measure zero in M (while keeping the modified function bounded), then the value of the integral $\iint_M f \, dx\, dy$ does not change. When computing the integral, we may even omit a set of measure zero (i.e., instead of integrating over M, we may integrate over $M \setminus N$, where $\mu_2(N) = 0$), and the value of the integral will be the same. In other words: the behavior of the integrand on sets of measure zero (provided the function remains bounded) has no influence on either the existence or the value of the double integral. This also answers a frequently asked question: "Is it important whether we include or do not include the boundary of the domain of integration?" (For example: "Is there a difference between integrating over the closed rectangle $\langle a\,;b\rangle \times \langle c\,;d\rangle$ or over the open rectangle $(a;b)\times(c;d)$?") The answer is: NO. If the boundary of the domain of integration has measure zero and the integrand is bounded there, then it does not matter whether we include it in the domain of integration or not.

II.3 Computation of the double integral – Fubini's theorem and transformation to polar coordinates

Fubini's theorem reduces the computation of a double integral to the computation of two simple (that is: one-dimensional) integrals. It can be used when integrating over so-called elementary domains of integration. The theorem was first published in 1907 by Guido Fubini (1879–1943).

II.3.1. Elementary domain of integration in $\mathbb{E}_2$

a) Let $y = \phi_1(x)$ and $y = \phi_2(x)$ be continuous functions on the interval $\langle a\,;b\rangle$ and let $\phi_1(x) \le \phi_2(x)$ for all $x \in \langle a\,;b\rangle$. Then the set

$$M = \big\{[x;y] \in \mathbb{E}_2 :\ a \le x \le b,\ \phi_1(x) \le y \le \phi_2(x)\big\}$$

is called an *elementary domain of integration with respect to the x-axis.* (See Fig. 10a.)

b) Let $x = \psi_1(y)$ and $x = \psi_2(y)$ be continuous functions on the interval $\langle c\,;d\rangle$ and let $\psi_1(y) \le \psi_2(y)$ for all $y \in \langle c\,;d\rangle$. Then the set

$$M = \big\{[x;y] \in \mathbb{E}_2 :\ c \le y \le d,\ \psi_1(y) \le x \le \psi_2(y)\big\}$$

is called an *elementary domain of integration with respect to the y-axis.* (See Fig. 10b.)

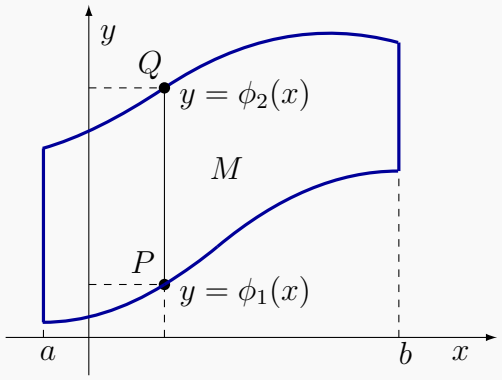


Fig. 10a

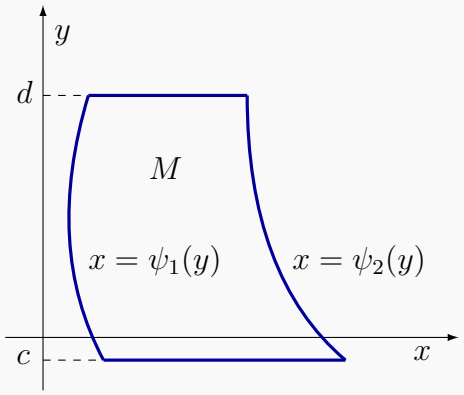


Fig. 10b

Elementary domains of integration are measurable sets in $\mathbb{E}_2$. We now explain the idea of computing the double integral of a function $f(x,y)$ over an elementary domain of integration with respect to the x-axis. (See Fig. 10a.) Imagine that the domain of integration is cut into infinitely many infinitely thin vertical strips. One such strip is the segment $\overline{PQ}$ in Fig. 10a. First, we integrate $f(x,y)$ over each such segment as a function of the variable y – we obtain $F(x) = \int_{\phi_1(x)}^{\phi_2(x)} f(x,y)\,dy$. The result depends on x, since the position of the segment $\overline{PQ}$ depends on x. Then we integrate $F(x)$ as a function of x from a to b. In this way we obtain formula (II.3.1) in Section II.3.2.

In the following theorem we formulate precisely the assumptions under which the integral may be computed in the described way. The theorem also treats the case when the domain of integration is an elementary domain with respect to the y-axis.

II.3.2. Fubini's theorem for the double integral.

a) *Let M be an elementary domain of integration with respect to the x-axis, described in Section II.3.1. Let the function $f(x,y)$ be continuous on M. Then*

(II.3.1) $$\iint_M f(x,y)\,dx\,dy \;=\; \int_a^b \left(\int_{\phi_1(x)}^{\phi_2(x)} f(x,y)\,dy \right) dx.$$

b) *Let M be an elementary domain of integration with respect to the y-axis, described in Section II.3.1. Let the function $f(x,y)$ be continuous on M. Then*

(II.3.2) $$\iint_M f(x,y)\,dx\,dy \;=\; \int_c^d \left(\int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\,dx \right) dy.$$

II.3.3. Remark

The integrals on the right-hand sides of (II.3.1) and (II.3.2), in contrast to the double integrals on the left-hand sides, are often called "iterated integrals".

If M is the rectangle $\langle a\,;b\rangle \times \langle c\,;d\rangle$ and the function f is continuous on M, then from both parts of Theorem II.3.2 we easily obtain the formula:

$$\iint_M f(x,y)\,dx\,dy \;=\; \int_a^b \left(\int_c^d f(x,y)\,dy \right) dx \;=\; \int_c^d \left(\int_a^b f(x,y)\,dx \right) dy.$$

II.3.4. Example

Let us compute the integral $\iint_M (2x+3y+5)\,dx\,dy$, where M is the region bounded by the curves $y = 2x$, $y = 1/x$ and $x = \sqrt{2}$.

S o l u t i o n :
The given curves divide the xy-plane into various parts (draw a picture!), but only one of them is bounded, and that is precisely M. The set M can be described as the set of points $[x;y] \in \mathbb{E}_2$ such that $\sqrt{2}/2 \le x \le \sqrt{2}$ and $1/x \le y \le 2x$.

M is an elementary region of integration with respect to the x-axis. The function f is continuous on M. Using Fubini's theorem II.3.2 we obtain:

$$\iint_M (2x+3y+5)\,dx\,dy \;=\; \int_{\sqrt{2}/2}^{\sqrt{2}} \left(\int_{1/x}^{2x} (2x+3y+5)\,dy \right) dx =$$

$$= \int_{\sqrt{2}/2}^{\sqrt{2}} \left[2xy + \tfrac{3}{2}y^2 + 5y \right]_{y=1/x}^{y=2x} dx \;=\; \int_{\sqrt{2}/2}^{\sqrt{2}} \left(4x^2 + 6x^2 + 10x - 2 - \frac{3}{2x^2} - \frac{5}{x} \right) dx =$$

$$= \left[\frac{10x^3}{3} + 5x^2 - 2x + \frac{3}{2x} - 5\,\ln x \right]_{\sqrt{2}/2}^{\sqrt{2}} \;=\; \frac{79}{12}\sqrt{2} + 7.5 - 5\,\ln 2.$$

From the theory of one-dimensional integrals you may remember that an effective method of integration is substitution. We can use this method also when computing a double integral. When using it, we usually say that we transform the integral to different coordinates. The most commonly used coordinates in $\mathbb{E}_2$ other than Cartesian coordinates are the so-called polar (or generalized polar) coordinates.

II.3.5. Polar coordinates in $\mathbb{E}_2$

The position of a point $X \in \mathbb{E}_2$ can be uniquely determined by its *polar coordinates* r, φ, which have the following geometric meaning: r is the distance of X from the origin O and φ is the angle between the positive part of the x-axis and the segment $\overline{OX}$ (measured from the x-axis toward $\overline{OX}$). (See Fig. 11.) The relationship between the Cartesian coordinates x, y and the polar coordinates r, φ is given by the equations

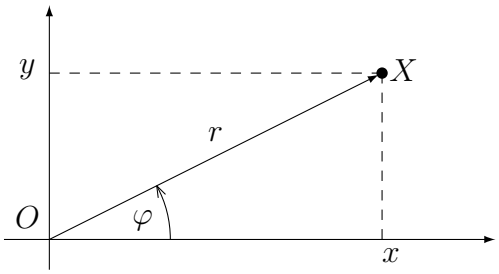


Fig. 11

$$x = r\cos\varphi, \quad y = r\sin\varphi. \tag{II.3.3}$$

From the geometric meaning of polar coordinates it follows that $r \geq 0$ and φ may be taken from any interval whose length does not exceed 2π (for example from the interval $\langle 0; 2\pi \rangle$).

II.3.6. Transformation of a double integral to polar coordinates

Imagine that we are computing the integral $\iint_M f(x,y)\,dx\,dy$. Using equations (II.3.3) we can replace x, y by the expressions $r\cos\varphi$ and $r\sin\varphi$. In addition, however, it is necessary to

a) describe the region of integration M in polar coordinates and replace the old description (in Cartesian coordinates) by a new description (in polar coordinates),

b) substitute an appropriate expression in the double integral for $dx\,dy$ (similarly as when making the substitution $x = g(t)$ in a one-dimensional integral, where we substitute $dx = g'(t)\,dt$).

The set M corresponds to some set M' in polar coordinates. (You may think of M' as the description of M in polar coordinates.) In the optimal case, each point $[x;y] \in M$ corresponds to exactly one point $[r;\varphi] \in M'$ (such that $x = r\cos\varphi$ and $y = r\sin\varphi$). Sets of measure zero, however, play no role, and therefore the one-to-one correspondence between points $[x;y] \in M$ and points $[r;\varphi] \in M'$ may be violated on a set of measure zero.

When transforming to polar coordinates, the expression $dx\,dy$ must be replaced by:

$$dx\,dy = r\,dr\,d\varphi. \tag{II.3.4}$$

The factor r on the right-hand side is the so-called "Jacobian" (short for "Jacobi determinant"). You will learn more about it in Chapter II.9.

Transforming a double integral to polar coordinates makes sense if it leads to a simplification of the integrand (see Example II.3.8) or to a simplification of the description of the region of integration (see Example II.3.7).

II.3.7. Example

Let us compute the integral $\iint_M (x+y)\,dx\,dy$, where $M = \{[x;y] \in \mathbb{E}_2 :\ x > 0,\ y > 0,\ x^2 + y^2 < 4\}$.

Solution:

M is the intersection of the disk (centered at the origin with radius 2) with the first quadrant. It corresponds to the region $M' = \{[r;\varphi] \in \mathbb{E}_2 :\ 0 < r < 2,\ 0 < \varphi < \pi/2\}$ in polar coordinates $[r;\varphi]$. Using equations (II.3.3) and (II.3.4) we obtain:

$$\iint_M (x+y)\,dx\,dy \;=\; \iint_{M'} (r\,\cos\varphi + r\,\sin\varphi)\,r\,dr\,d\varphi =$$

$$= {}^{1)} \int_0^2 \left(\int_0^{\pi/2} r^2\,(\cos\varphi + \sin\varphi)\,d\varphi \right) dr \;=\; \int_0^2 r^2 \big[\sin\varphi - \cos\varphi\big]_{\varphi=0}^{\varphi=\pi/2}\,dr =$$

$$= \int_0^2 2r^2\,dr \;=\; \frac{16}{3}.$$

[1]) We used Fubini's theorem.

II.3.8. Example

Let us compute the integral $\iint_M (x^2+y^2)^{-1/2}\,dx\,dy$, where M is the triangle with vertices $[0;0]$, $[1;0]$, $[1;1]$.

Solution:
M can be defined as the set of points $[x;y] \in \mathbb{E}_2$ such that $0 < x < 1$ and $0 < y < x$. Transforming these inequalities to polar coordinates we obtain:

(II.3.5) $$0 < r\,\cos\varphi < 1, \quad 0 < r\,\sin\varphi < r\,\cos\varphi.$$

From the second inequality it follows that $0 < \sin\varphi < \cos\varphi$, which means $0 < \varphi < \pi/4$. From the first inequality in (II.3.5) it follows that $0 < r < 1/\cos\varphi$. Thus M corresponds to the set $M' = \{[r;\varphi] \in \mathbb{E}_2 : 0 < \varphi < \pi/4,\ 0 \le r \le 1/\cos\varphi\}$ in polar coordinates. Using transformations (II.3.3), (II.3.4) and then applying Fubini's theorem we obtain:

$$\iint_M \frac{1}{\sqrt{x^2+y^2}}\,dx\,dy \;=\; \iint_{M'} \frac{1}{r}\,r\,dr\,d\varphi \;=\; \int_0^{\pi/4} \left(\int_0^{1/\cos\varphi} dr \right) d\varphi =$$

$$= \int_0^{\pi/4} \frac{1}{\cos\varphi}\,d\varphi \;=\; \int_0^{\pi/4} \frac{\cos\varphi}{1-\sin^2\varphi}\,d\varphi \;=\; {}^{2)} \int_0^{\sqrt{2}/2} \frac{dt}{1-t^2} =$$

$$= \frac{1}{2} \int_0^{\sqrt{2}/2} \left(\frac{1}{1-t} + \frac{1}{1+t} \right) dt \;=\; \tfrac{1}{2} \big[-\ln(1-t) + \ln(1+t) \big]_0^{\sqrt{2}/2} \;=\; \frac{1}{2}\,\ln\frac{2+\sqrt{2}}{2-\sqrt{2}}.$$

[2]) We used the substitution $\sin\varphi = t$ and the resulting equality $\cos\varphi\,d\varphi = dt$.

II.3.9. Generalized polar coordinates in $\mathbb{E}_2$

We will again denote these coordinates by r, φ. They have a similar geometric meaning as polar coordinates, but their origin need not coincide with the origin of Cartesian coordinates, and they are also not "isotropic", i.e., the rate of change of r in the direction of the x-axis and in the direction of the y-axis may differ. The relations between Cartesian coordinates x, y and generalized polar coordinates r, φ are:

(II.3.6) $$x \;=\; x_0 + ar\,\cos\varphi, \quad y \;=\; y_0 + br\,\sin\varphi,$$

where $[x_0;y_0]$ is a chosen point in $\mathbb{E}_2$ and a, b are chosen positive constants.

It can be proved that when transforming a double integral to generalized polar coordinates, the expression $dx\,dy$ must be transformed as follows:

(II.3.7) $$dx\,dy \;=\; rab\,dr\,d\varphi.$$

The factor rab on the right-hand side is again called the "Jacobian", and the derivation of its form can be found in Chapter II.9.

By transforming a double integral to generalized polar coordinates we usually achieve a simplification of the integrand if it contains expressions of the form $(x-x_0)^2/a^2 + (y-y_0)^2/b^2$.

The description of the region of integration becomes simpler if, for example, it is the entire interior or only a sector of the ellipse $(x-x_0)^2/a^2 + (y-y_0)^2/b^2 = 1$.

II.3.10. Example

Let us compute the integral $\iint_M x\,dx\,dy$, where $M = \{[x;y] \in \mathbb{E}_2 :\ (x-2)^2 + (y-1)/4^2 \le 1\}$.

S o l u t i o n :
It is clear that if we use the transformation

$$x = 2 + r\,\cos\,\varphi, \quad y = 1 + 2r\,\sin\,\varphi, \tag{II.3.8}$$

then the points $[x;y]$ fill the region M precisely when the points $[r;\varphi]$ fill the region $M' = \{[r;\varphi] \in \mathbb{E}_2 :\ 0 \le r \le 1,\ 0 \le \varphi < 2\pi\}$. Using transformation (II.3.8), the identity $dx\,dy = 2r\,dr\,d\varphi$ (following from (II.3.7)), and then applying Fubini's theorem, we obtain:

$$\iint_M x\,dx\,dy = \iint_{M'} (2 + r\,\cos\,\varphi)\,2r\,dr\,d\varphi = \int_0^{2\pi}\left(\int_0^1 (4r + 2r^2\,\cos\,\varphi)\,dr\right)d\varphi =$$
$$= \int_0^{2\pi}\left[2r^2 + \tfrac{2}{3}r^3\,\cos\,\varphi\right]_{r=0}^{1} d\varphi = \int_0^{2\pi} (2 + \tfrac{2}{3}\,\cos\,\varphi)\,d\varphi = 4\pi.$$

II.3.11. Remark

You have surely noticed that the transformation (II.3.8) in Example II.3.10 is not a one-to-one mapping from the set M' onto the set M. The one-to-one correspondence between points $[x;y]$ and points $[r;\varphi]$ is broken on the subset $M_0' = \{[r;\varphi] \in \mathbb{E}_2 :\ r = 0,\ 0 \le \varphi < 2\pi\}$ of the set M'. This is obvious, because transformation (II.3.8) assigns to all points lying in M_0' a single point $[2;1]$ in M. The point $[2;1]$ in M therefore has infinitely many preimages in M'–namely all points $[r;\varphi] \in M_0'$. However, since $\mu_2(M_0') = 0$, this has no effect on either the existence or the value of the integral.

II.4 Some physical applications of the double integral

Suppose that a thin material plate is placed in the xy-plane so that it exactly covers a measurable set M. The plate need not be homogeneous. This means that the surface density of the plate (i.e., mass per unit area) need not be constant (for example because the thickness of the plate changes, or because of the material from which it is made). Let us denote this surface density by $\rho(x,y)$ and assume that $\rho(x,y)$ is given in $\mathrm{kg}\cdot\mathrm{m}^{-2}$. The double integral makes it possible to define and compute some basic mechanical characteristics of the plate:

mass of the plate ..... $m = \iint_M \rho(x,y)\,dx\,dy$ [kg],

static moment

with respect to the x-axis $m_x = \iint_M y\cdot\rho(x,y)\,dx\,dy$ [kg · m],

with respect to the y-axis $m_y = \iint_M x\cdot\rho(x,y)\,dx\,dy$ [kg · m],

center of mass coordinates . $x_T = \dfrac{m_y}{m}, \quad y_T = \dfrac{m_x}{m}$ $[\mathrm{m}]$,

moment of inertia

with respect to the x-axis $J_x = \iint_M y^2 \cdot \rho(x,y)\,dx\,dy$ $[\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the y-axis $J_y = \iint_M x^2 \cdot \rho(x,y)\,dx\,dy$ $[\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the origin $J_0 = \iint_M (x^2+y^2) \cdot \rho(x,y)\,dx\,dy$ $[\mathrm{kg}\cdot\mathrm{m}^2]$.

Propose yourself a formula by which we can define, and also compute, the moment of inertia with respect to a general line in $\mathbb{E}_2$ given by the equation $ax + by + c = 0$.

II.5 Triple integral – motivation and definition. Three-dimensional Jordan measure and measurable sets in $\mathbb{E}_3$

The theory of the triple integral is almost the same as the theory of the double integral. The main difference is that we integrate functions of three variables and the regions of integration are subsets of $\mathbb{E}_3$. Therefore we present the theory of the triple integral only briefly and do not explain many details.

On the other hand, one extra dimension means that the integrands and especially the regions of integration can be substantially more diverse. Although the methods of integration are again based mainly on Fubini's theorem and transforming the integral to other coordinates, these methods are usually technically more complicated than in the case of the double integral.

II.5.1. Partition of a rectangular box and its norm

Let $\langle a\,;b\rangle$, $\langle c\,;d\rangle$ and $\langle r\,;s\rangle$ be closed intervals on the coordinate axes x, y and z, respectively. Then the Cartesian product $K = \langle a\,;b\rangle \times \langle c\,;d\rangle \times \langle r\,;s\rangle$ is a rectangular box in $\mathbb{E}_3$ whose edges are parallel to the coordinate axes x, y or z.

The box K can be partitioned by a grid of planes parallel to the planes xy, xz or yz into n sub-boxes $K_1, \ldots, K_n$. We call the system of these boxes a *partition* of the box K. If we denote the described partition by D, then the *norm* of the partition D is defined as the length of the longest of all edges of the boxes $K_1, \ldots, K_n$. We denote the norm of the partition D by $\|D\|$. If the edge lengths of the sub-boxes $K_1, \ldots, K_n$ are $\Delta x_1, \Delta y_1, \Delta z_1, \ldots, \Delta x_n, \Delta y_n, \Delta z_n$, then we can express $\|D\|$ as:

$$\|D\| \;=\; \max\{\Delta x_1;\, \Delta y_1;\, \Delta z_1;\, \ldots\,;\, \Delta x_n;\, \Delta y_n;\, \Delta z_n\}.$$

II.5.2. Physical motivation

Suppose we want to compute the mass m of a body having the shape of the box $K = \langle a\,;b\rangle \times \langle c\,;d\rangle \times \langle r\,;s\rangle$ and in which the matter is distributed with density $\rho(x,y,z)$. The density may be variable, so the box need not be homogeneous.

In the special case when the density ρ is constant, we have $m = \rho\cdot(b-a)\cdot(d-c)\cdot(s-r)$.

In the general case, when the density need not be constant, we can partition K by a partition D into n sub-boxes $K_1, \ldots, K_n$. If the sub-boxes are "very small", then we do not make a large error by considering the density ρ to be constant on each of them, successively equal to $\rho(Z_1)$, $\ldots$, $\rho(Z_n)$, where $Z_1, \ldots, Z_n$ are chosen points in $K_1, \ldots, K_n$. The mass of the box K_i (for

$i = 1, \ldots, n$) is then approximately $m_i \doteq \rho(Z_i) \cdot \Delta x_i \, \Delta y_i \, \Delta z_i$, where Δx_i, Δy_i and Δz_i are the edge lengths of the box K_i. The mass of the whole body K can be expressed approximately as:

$$m = \sum_{i=1}^{n} m_i \doteq \sum_{i=1}^{n} \rho(Z_i) \cdot \Delta x_i \, \Delta y_i \, \Delta z_i. \tag{II.5.1}$$

If the function ρ has "reasonable properties", it is natural to expect that we obtain the exact value of m as the limit of the sums on the right-hand side as $n \to +\infty$ and the numbers Δx_i, Δy_i, Δz_i tend to zero (i.e., as $\|D\| \to 0$).

II.5.3. Riemann sums and their limit

Suppose that $f(x, y, z)$ is a bounded function on the box $K = \langle a\,;b\rangle \times \langle c\,;d\rangle \times \langle r\,;s\rangle$. Let D be a partition of the box K into sub-boxes $K_1, \ldots, K_n$, described in Section II.5.1. Denote the edge lengths of the boxes $K_1, \ldots, K_n$ by $\Delta x_1, \Delta y_1, \Delta z_1, \ldots, \Delta x_n, \Delta y_n, \Delta z_n$. Let $\mathcal{V}$ be a system of chosen points $Z_i \in K_i$ $(i = 1, \ldots, n)$. The *Riemann sum* of the function f on the box K corresponding to the partition D and the system $\mathcal{V}$ of chosen points $Z_1, \ldots, Z_n$ is the sum

$$s(f, D, \mathcal{V}) = \sum_{i=1}^{n} f(Z_i) \cdot \Delta x_i \, \Delta y_i \, \Delta z_i. \tag{II.5.2}$$

We say that a number S is the *limit of the Riemann sums* $s(f, D, \mathcal{V})$ *as* $\|D\| \to 0+$ if for every chosen $\epsilon > 0$ there exists $\delta > 0$ such that for every partition D of the box K and for every chosen system $\mathcal{V}$ we have:

$$\|D\| < \delta \quad \Longrightarrow \quad |s(f, D, \mathcal{V}) - S| < \epsilon.$$

We write:

$$\lim_{\|D\| \to 0+} s(f, D, \mathcal{V}) = S. \tag{II.5.3}$$

II.5.4. Triple integral over a rectangular box

If the limit (II.5.3) exists, then we call its value S the *triple integral* of the function f over the box K. This integral is usually denoted by

$$\iiint_K f(x, y) \, dx \, dy \, dz \qquad \text{or} \qquad \iiint_K f \, dx \, dy \, dz.$$

If the limit (II.5.3) exists, we say that "the triple integral $\iiint_K f \, dx \, dy \, dz$ exists" or that "the function f is *integrable* on the box K".

II.5.5. Triple integral over a general bounded set in $\mathbb{E}_3$

Suppose that M is a bounded set in $\mathbb{E}_3$ and $f(x, y, z)$ is a bounded function on M. Let $K = \langle a\,;b\rangle \times \langle c\,;d\rangle \times \langle r\,;s\rangle$ be a box that contains the set M. Define

$$f^*(x, y, z) = \begin{cases} f(x, y, z) & \text{for } [x; y; z] \in M, \\ 0 & \text{for } [x; y; z] \in K \setminus M. \end{cases}$$

Then by the *triple integral* of the function f *over the set* M we mean the triple integral of the function f^* over the box K (provided this integral exists). We denote the triple integral of f over M by

$$\iiint_M f(x, y) \, dx \, dy \, dz \qquad \text{or} \qquad \iiint_M f \, dx \, dy \, dz.$$

It can be shown that neither the existence nor the value of the triple integral $\iiint_K f^*\,dx\,dy\,dz$ depends on the particular choice of the box K containing M. (This is because the extended function f^* is zero outside the set M.)

In the literature, the term *volume integral* is sometimes used instead of "triple integral".

If the integral $\iiint_K f^*\,dx\,dy\,dz$ exists, then we also say that "the triple integral $\iiint_M f\,dx\,dy\,dz$ exists" or that "the function f is *integrable* on the set M".

We call the set M the *region of integration* (or *integration domain*) and we call the integrated function f the *integrand*. A simple sufficient condition for the existence of the integral can be found in Section II.6.1.

II.5.6. Physical meaning of the triple integral

From Section II.5.2 and the definition of the triple integral, the physical meaning of the triple integral is clear: If M is a bounded set (a body) in $\mathbb{E}_3$ and $\rho(x,y,z)$ is the density of the body M, then $\iiint_M \rho\,dx\,dy\,dz$ expresses the *total mass* of the body M. We will become familiar with further applications of the triple integral in physics in Chapter II.8.

II.5.7. Historical approach

The triple integral used to be introduced in the past, similarly to the double integral, by means of the idea of an infinitely small cuboid with edge lengths dx, dy, dz, etc. This approach was explained in Section II.1.8.

II.5.8. A measurable set in $\mathbb{E}_3$ and its Jordan measure

In $\mathbb{E}_3$ there exist sets of such complicated shapes that even the integral of the constant function $f(x,y,z) = 1$ does not exist on them. (An example is the set of all points of the cube $\langle 0;1\rangle \times \langle 0;1\rangle \times \langle 0;1\rangle$ whose some coordinate is a rational number.) In order to distinguish these "artificial" sets, which have little significance in engineering applications, from "reasonable" sets, we introduce the notion of a "measurable set".

Assume that M is a bounded set in $\mathbb{E}_3$. We say that the set M is *measurable* (in the sense of Jordan) if the constant function $f(x,y,z) = 1$ is integrable on M. In this case we call the value

$$\mu_3(M) \;=\; \iiint_M 1\,dx\,dy\,dz$$

the three-dimensional Jordan measure of the set M.

$\mu_3(M)$ has a clear geometric meaning: It defines and provides a method for computing the *volume* of the set M.

II.5.9. Examples of sets whose three-dimensional Jordan measure equals zero.

a) Sets consisting of a finite number of points or curves. (What we mean by a curve is explained precisely in Sections III.1.1, III.1.2 and III.1.8.)

b) Surfaces that are graphs of continuous functions of two variables $z = \varphi(x,y)$ or $y = \psi(x,z)$ or $x = \eta(y,z)$, defined on bounded closed sets in $\mathbb{E}_2$.

c) So-called simple smooth surfaces and simple piecewise smooth surfaces (see Chapter IV.1).

In this chapter and in Chapters II.6–II.8, dealing with the triple integral, by the phrase "a set of measure zero" we will mean a set in $\mathbb{E}_3$ whose three-dimensional Jordan measure equals zero.

II.5.10. Theorem

a) *The union of finitely many sets of measure zero is a set of measure zero.*

b) *If N is a set of measure zero and $M \subset N$, then M is also a set of measure zero.*

II.5.11. Theorem. (A necessary and sufficient condition for a set in $\mathbb{E}_3$ to be measurable.) *Set $M \subset \mathbb{E}_3$ is measurable (in the sense of Jordan) if and only if it is bounded and $\mu_3(\partial M) = 0$.*

From Theorem II.5.11 it follows, for example, that the so-called elementary domains of integration in $\mathbb{E}_3$, which you will learn in Section II.7.1 (see also Fig. 12), are measurable sets. In engineering applications you will practically never encounter bounded sets other than measurable ones.

II.6 Existence and selected properties of the triple integral

We recall that the two statements "the function f is integrable on the set M" and "the triple integral $\iiint_M f\,dx\,dy\,dz$ exists" say exactly the same thing.

II.6.1. Theorem. (A sufficient condition for the existence of the triple integral.) *Let M be a measurable set in $\mathbb{E}_3$ and let f be a bounded and continuous function on M. Then the triple integral $\iiint_M f\,dx\,dy\,dz$ exists.*

Just as with the analogous theorem II.2.1, Theorem II.6.1 can also be modified. (One can additionally require that the set M be closed, and in return omit the assumption that f is bounded.) Formulate and write down the corresponding modified theorem yourself.

Theorem II.6.1 will also remain valid if we replace the assumption of continuity of f on M by the weaker assumption that f is continuous on M "except on a set of measure zero" (which allows discontinuities of f, but only on a set of measure zero in M).

II.6.2. Some important properties of the triple integral.

a) **Linearity of the triple integral**

If the functions f and g are integrable on a set $M \subset \mathbb{E}_3$ and $\alpha \in \mathbb{R}$, then

$$\iiint_M (f+g)\,dx\,dy\,dz = \iiint_M f\,dx\,dy\,dz + \iiint_M g\,dx\,dy\,dz,$$

$$\iiint_M \alpha \cdot f\,dx\,dy\,dz = \alpha \cdot \iiint_M f\,dx\,dy\,dz.$$

b) **Additivity of the triple integral with respect to the domain of integration.**

If M_1 and M_2 are measurable sets in $\mathbb{E}_3$ that overlap at most on a set of measure zero (i.e. $\mu_3(M_1 \cap M_2) = 0$) and f is integrable on M_1 as well as on M_2, then

$$\iiint_{M_1} f\,dx\,dy\,dz + \iiint_{M_2} f\,dx\,dy\,dz = \iiint_{M_1 \cup M_2} f\,dx\,dy\,dz.$$

c) *If the function f is integrable on a set $M \in \mathbb{E}_3$ and the bounded function g differs from f at most on a set of measure zero in M, then g is also integrable on M and*

$$\iiint_M g\,dx\,dy\,dz = \iiint_M f\,dx\,dy\,dz.$$

d) *If f is a bounded function on a set $M \in \mathbb{E}_3$ of measure zero, then f is integrable on M and*

$$\iiint_M f \, dx\, dy\, dz = 0.$$

e) *If f and g are integrable functions on a set $M \subset \mathbb{E}_3$ such that $f(x,y,z) \geq g(x,y,z)$ for all points $[x;y;z] \in M$, then*

$$\iiint_M f \, dx\, dy\, dz \geq \iiint_M g \, dx\, dy\, dz.$$

In particular, if $f(x,y,z) \geq 0$ for all points $[x;y;z] \in M$, then

$$\iiint_M f \, dx\, dy\, dz \geq 0.$$

From item c) it is clear that if the integrand is bounded, then its behavior on a set of measure zero affects neither the existence nor the value of the triple integral. Therefore we can deal with the triple integral of a function f over a set M even in the case when f is defined only on $M - N$, where N is a set of measure zero.

II.7 Computation of the triple integral – Fubini's theorem and transformation to cylindrical and spherical coordinates

Fubini's theorem for the triple integral reduces the computation of a triple integral to the computation of one double integral and one simple (i.e. one-dimensional) integral. It can be used on so-called elementary domains of integration in $\mathbb{E}_3$.

II.7.1. Elementary domain of integration in $\mathbb{E}_3$

Let M_{xy} be a measurable closed set in $\mathbb{E}_2$ and let $z = \phi_1(x,y)$ and $z = \phi_2(x,y)$ be continuous functions on M_{xy} such that $\phi_1(x,y) \leq \phi_2(x,y)$ for all $[x;y] \in M_{xy}$. Then the set

$$M = \{[x;y;z] \in \mathbb{E}_3 : [x;y] \in M_{xy}, \phi_1(x,y) \leq z \leq \phi_2(x,y)\}$$

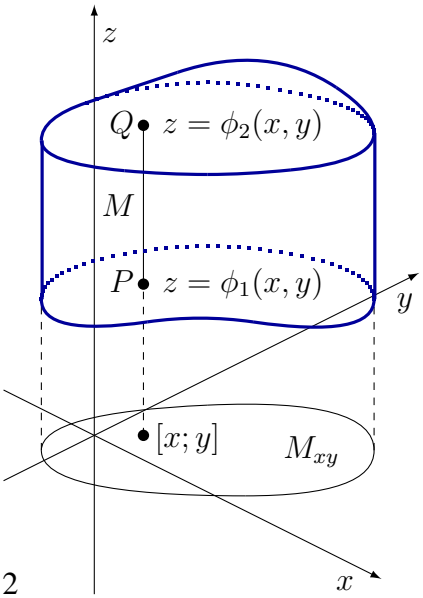


Fig. 12

is called an *elementary domain of integration with respect to the xy-plane*. (See Fig. 12.)

Analogously, we can also define an elementary domain of integration with respect to the xz-plane and an elementary domain with respect to the yz-plane. As an exercise, write down the definitions of these elementary integration domains yourself.

Elementary integration domains are measurable sets in $\mathbb{E}_3$. Integration of a function $f(x,y,z)$ over an elementary domain of integration with respect to the xy-plane is based on the following idea: We split the domain of integration M into infinitely many vertical segments. One of them

is the segment $\overline{PQ}$ in Fig. 12. First we integrate the function f along each such segment as a function of one variable z – we obtain $F(x,y) = \int_{\phi_1(x,y)}^{\phi_2(x,y)} f(x,y,z)\,dz$. The result depends on x and y because the position of the segment $\overline{PQ}$ depends on x and y. Then we integrate $F(x,y)$ as a function of two variables x and y over the set M_{xy}. In this way we obtain formula (II.7.1).

II.7.2. Fubini's theorem for the triple integral.
Let M be an elementary domain of integration with respect to the xy-plane, defined in Section II.7.1. Let $f(x,y,z)$ be a continuous function on M. Then

$$\text{(II.7.1)} \qquad \iiint_M f(x,y,z)\,dx\,dy\,dz = \iint_{M_{xy}} \left(\int_{\phi_1(x,y)}^{\phi_2(x,y)} f(x,y,z)\,dz \right) dx\,dy.$$

Formulate analogous theorems yourself for integration over elementary domains of integration with respect to the xz and yz planes.

II.7.3. Example

Let us compute the integral $\iiint_M (z+2)\,dx\,dy\,dz$, where M is the set in $\mathbb{E}_3$ bounded by the surfaces $x^2+y^2=2$, $z=-2-x$ and $z=2+y$.

S o l u t i o n :
The given surfaces split $\mathbb{E}_3$ into several sets, but only one of them is bounded, namely M. The set M is the part of the cylinder $x^2+y^2 \le 2$, bounded from below by the plane $z=-2-x$ and from above by the plane $z=2+y$. The set M is an elementary domain of integration with respect to the xy-plane, with $M_{xy} = \{[x;y] \in \mathbb{E}_2 : \ x^2+y^2 \le 2\}$ and $\phi_1(x,y) = -2-x$, $\phi_2(x,y) = 2+y$. Using Theorem II.7.2 we obtain:

$$\iiint_M (z+2)\,dx\,dy\,dz = \iint_{M_{xy}} \left(\int_{-2-x}^{2+y} (z+2)\,dz \right) dx\,dy =$$

$$= \iint_{x^2+y^2\le 2} \left[z^2/2 + 2z \right]_{-2-x}^{2+y} dx\,dy = \iint_{x^2+y^2\le 2} \left(\tfrac{1}{2}\,y^2 - \tfrac{1}{2}\,x^2 + 4y + 8 \right) dx\,dy =$$

$$= {}^{1)} \int_0^{\sqrt{2}} \left(\int_0^{2\pi} \left(\tfrac{1}{2}\,r^2 \sin^2\varphi - \tfrac{1}{2}\,r^2 \cos^2\varphi + 4r \sin\varphi + 8 \right) r\,d\varphi \right) dr =$$

$$= \int_0^{\sqrt{2}} 16\pi\,r\,dr = 16\pi.$$

[1]) We transform the double integral over M_{xy} to polar coordinates.

II.7.4. Remark

Fubini's theorem II.7.2 splits a triple integral into two integrals – an outer double integral and an inner simple integral. Sometimes it is useful to perform the decomposition in the opposite way, i.e. into an outer simple integral and an inner double integral. We omit the corresponding theory because it is analogous to the contents of Sections II.7.1 and II.7.2, and we show the procedure only in an example.

II.7.5. Example

Let us compute the volume V of the oblique cone

$$K = \{[x;y;z] \in \mathbb{E}_3 : \ 0 < z < 5, \ (x-2z)^2 + y^2 < z^2\}.$$

S o l u t i o n :
For the volume V of the cone K we have:

$$V = \iiint_K dx\,dy\,dz = \int_0^5 \left(\iint_{(x-2z)^2+y^2<z^2} dx\,dy \right) dz =$$

$$= {}^{2)} \int_0^5 \left(\int_0^z \left(\int_0^{2\pi} r\,d\varphi \right) dr \right) dz = \int_0^5 \pi\, z^2\, dz = \frac{125\,\pi}{3}.$$

[2]) We transform the inner double integral from Cartesian coordinates x, y to generalized polar coordinates r, φ using the relations $x = 2z + r\,\cos\,\varphi$, $y = r\,\sin\,\varphi$, $dx\,dy = r\,d\varphi\,dr$.

II.7.6. Cylindrical coordinates in $\mathbb{E}_3$
Cylindrical coordinates of a point $X = [x; y; z] \in \mathbb{E}_3$ are r, φ, w. They have the following geometric meaning: r, φ are the polar coordinates of the point $[x; y]$ in the xy-plane and $w = z$. Thus, between Cartesian coordinates and cylindrical coordinates of the point X the following relations hold:

(II.7.2) $$x = r\,\cos\,\varphi, \quad y = r\,\sin\,\varphi, \quad z = w.$$

When transforming a triple integral to cylindrical coordinates, we must also replace the expression $dx\,dy\,dz$. It can be shown that the correct replacement is

(II.7.3) $$dx\,dy\,dz = r\,dr\,d\varphi\,dw.$$

(You will find more details in Chapter II.9.)

From the geometric meaning of r, φ and w it is clear that $r \geq 0$, φ can be chosen from any interval of length 2π and $w \in \mathbb{R}$. Transforming a triple integral to cylindrical coordinates makes sense when it simplifies either the integrand or the description of the integration domain. The integrand usually simplifies when it depends on x and y through the expression x^2+y^2 (which becomes r^2 under the transformation). The description of the domain of integration is usually simpler in cylindrical coordinates if the domain is a cylinder or some simple part of a cylinder.

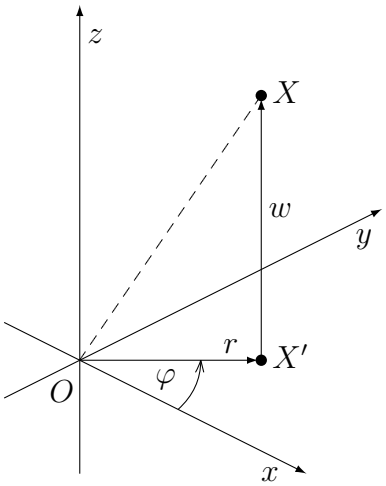


Fig. 13

By transforming the triple integral $\iiint_M f\ dx\,dy\,dz$ we obtain an integral over a domain M' in the variables r, φ, w. In the optimal case, the equations (II.7.2) define a one-to-one mapping from M' onto M. However, since the behavior of the integrand on a set of measure zero has no effect on the existence or the value of the integral (if the function is bounded), the one-to-one correspondence between points $[r, \varphi, w] \in M'$ and points $[x; y; z] \in M$ may be violated on a set of measure zero. This also holds for transformations of a triple integral to other coordinate systems (spherical, generalized cylindrical, etc.), so we will not explicitly repeat it again.

II.7.7. Example

Let us compute the volume of the solid M, bounded by the cylindrical surface $x^2+(y-1)^2=1$, the plane $z=0$ and the paraboloid $z=x^2+y^2$.

S o l u t i o n :
The volume of M is identical to the three-dimensional Jordan measure of the region M (see Section II.5.8). We may therefore denote it by $\mu_3(M)$. It is defined by the integral $\iiint_M dx\,dy\,dz$. We compute this integral using a transformation to cylindrical coordinates.

M is the set of points $[x;y;z]\in\mathbb{E}_3$ such that $0\le z\le x^2+y^2$ and $x^2+(y-1)^2\le 1$. The last inequality can be written as: $x^2+y^2-2y\le 0$. Substituting for x and y from equations (II.7.2), we obtain:

$$r^2-2r\,\sin\varphi\le 0,$$
$$r\le 2\,\sin\varphi.$$

The inequalities $0\le z\le x^2+y^2$ correspond to the inequalities

$$0\le w\le r^2.$$

These inequalities show the limits in which we must integrate with respect to r and with respect to w. The limits for integration with respect to φ are clear if we imagine what the orthogonal projection of M onto the plane xy looks like. The projection is the circle M_{xy} with center $[0,1]$ and radius 1. The angle formed by all rays emanating from the origin and pointing into M_{xy} is the angle between the extreme values $\varphi=0$ and $\varphi=\pi$. Therefore, the desired volume is:

$$\mu_3(M)=\iiint_M 1\,dx\,dy\,dz=\int_0^{\pi}\left(\int_0^{2\sin\varphi}\left(\int_0^{r^2} r\,dw\right)dr\right)d\varphi=\int_0^{\pi}4\,\sin^4\varphi\,d\varphi=\frac{3}{2}\,\pi.$$

The integral $\int_0^{\pi}\big(\int_0^{2\sin\varphi}\big(\int_0^{r^2} r\,dw\big)dr\big)d\varphi$ is, unlike the triple integral, called an "iterated triple integral." To make it immediately clear with respect to which variable we integrate and over which limits, we often write the iterated triple integral in the following form:

$$\int_0^{\pi}\left(\int_0^{2\sin\varphi}\left(\int_0^{r^2} r\,dw\right)dr\right)d\varphi=\int_0^{\pi}d\varphi\int_0^{2\sin\varphi}dr\int_0^{r^2} r\,dw.$$

II.7.8. Spherical coordinates in $\mathbb{E}_3$

The spherical coordinates of a point $X=[x;y;z]$ in $\mathbb{E}_3$ are r, φ and ϑ. They have the following geometric meaning: r is the distance of X from the origin O. φ is the angle between the segment $\overline{OX'}$ (where X' is the orthogonal projection of X onto the plane xy) and the positive part of the x-axis (measured from the x-axis). ϑ is the angle between the segment $\overline{OX'}$ and the segment $\overline{OX}$ (measured from $\overline{OX'}$). This geometric interpretation of spherical coordinates easily leads to the following equations:

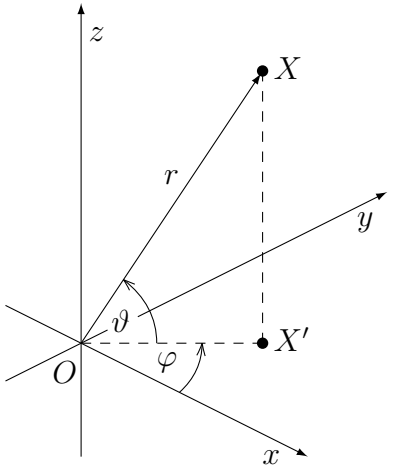


Fig. 14

$$\begin{aligned} x &= r\,\cos\vartheta\,\cos\varphi,\\ y &= r\,\cos\vartheta\,\sin\varphi,\\ z &= r\,\sin\vartheta. \end{aligned}\qquad\text{(II.7.4)}$$

When transforming a triple integral from Cartesian coordinates x, y, z to spherical coordinates r, φ, ϑ, it is also necessary to transform the expression $dx\,dy\,dz$. It can be proven that the correct replacement of this expression is:

(II.7.5) $$dx\,dy\,dz \;=\; r^2\,\cos\,\vartheta\;dr\,d\varphi\,d\vartheta.$$

(In Chapter II.9 you will learn more about why this is so.)

From the geometric meaning of r, φ and ϑ it follows that $r \geq 0$, φ may be chosen from any interval of length 2π (for example $\langle 0; 2\pi\rangle$), and ϑ may be chosen, for example, from the interval $\langle -\pi/2, \pi/2\rangle$. Transforming a triple integral to spherical coordinates usually simplifies the integrand if the function depends on x, y, z through the expression $x^2+y^2+z^2$, which under the transformation becomes r^2. The description of the domain of integration is usually simplified when the domain is a ball centered at the origin or some sector of it.

II.7.9. Example

Let us compute the volume of the solid M which has the shape of the upper part of the ball $x^2+y^2+z^2 \leq 1$, cut off by the conical surface $z^2 = \frac{1}{3}\,(x^2+y^2)$.

S o l u t i o n :
The considered ball has a very simple description in spherical coordinates r, φ, ϑ: $r \leq 1$. Substituting from equations (II.7.4) into the equation of the conical surface we obtain:

$$\begin{aligned} r^2\,\sin^2\vartheta \;&=\; \tfrac{1}{3}\,r^2\,\cos^2\vartheta\,\big(\cos^2\varphi+\sin^2\varphi\big),\\ \sin^2\vartheta \;&=\; \tfrac{1}{3}\,\cos^2\vartheta,\\ \tan\,\vartheta \;&=\; \pm\sqrt{3}/3, \end{aligned}$$

which means that $\vartheta = \pm\pi/6$. Since M lies above the conical surface (and on it), the coordinate ϑ of points in M satisfies: $\vartheta \in \langle \pi/6,\, \pi/2\rangle$. Finally, the angle φ runs from 0 to 2π. Therefore, for the volume of M we have:

$$\mu_3(M) \;=\; \iiint_M 1\,dx\,dy\,dz \;=\; \int_0^1\left(\int_0^{2\pi}\left(\int_{\pi/6}^{\pi/2} r^2\,\cos\,\vartheta\,d\vartheta\right)d\varphi\right)dr \;=\; \frac{\pi}{3}.$$

II.7.10. Generalized cylindrical coordinates in $\mathbb{E}_3$

We will again denote these coordinates of the point $[x;y;z] \in \mathbb{E}_3$, as in the case of cylindrical coordinates, by r, φ and w. Unlike cylindrical coordinates, however, generalized cylindrical coordinates need not have the same origin O as Cartesian coordinates, and distances in different directions may be measured with different "weights": r and φ represent generalized polar coordinates of the point $[x;y]$ in the plane xy and w is a linear function of the variable z. The relations between generalized cylindrical and Cartesian coordinates are therefore:

(II.7.6) $$x = x_0 + ar\,\cos\,\varphi, \quad y = y_0 + br\,\sin\,\varphi, \quad z = z_0 + cw,$$

where $[x_0; y_0; z_0]$ is a chosen point in $\mathbb{E}_3$ (the origin of the generalized cylindrical coordinate system) and a, b, c are positive parameters.

Analogously to equations (II.7.3) and (II.7.5), when transforming a triple integral to generalized cylindrical coordinates, we must also transform the expression $dx\,dy\,dz$, in the following way:

(II.7.7) $$dx\,dy\,dz \;=\; abc\,r\;dr\,d\varphi\,dw.$$

When transforming a triple integral to generalized cylindrical coordinates, the integrand is usually simplified if it depends on x, y and z through the expression $(x-x_0)^2/a^2+(y-y_0)^2/b^2$,

which under the transformation becomes r^2. The description of the domain of integration is usually simplified if it is the interior of the elliptic cylindrical surface $(x-x_0)^2/a^2+(y-y_0)^2/b^2=R^2$ or some simple part of it.

II.7.11. Generalized spherical coordinates in $\mathbb{E}_3$
We will denote these coordinates of the point $[x;y;z]\in\mathbb{E}_3$ in the same way as spherical coordinates, i.e. r, φ, ϑ. Unlike spherical coordinates, generalized spherical coordinates need not have the same origin as Cartesian coordinates, and distances in different directions may be measured with different weights. The relations between Cartesian and generalized spherical coordinates are:

(II.7.8) $$x=x_0+ar\,\cos\vartheta\,\cos\varphi,\quad y=y_0+br\,\cos\vartheta\,\sin\varphi,\quad z=z_0+cr\,\sin\vartheta,$$

where $[x_0;y_0;z_0]$ is a chosen point in $\mathbb{E}_3$ (the origin of the generalized spherical coordinate system) and a, b, c are positive parameters.

When transforming a triple integral from Cartesian to generalized spherical coordinates, the expression $dx\,dy\,dz$ must be transformed according to the formula:

(II.7.9) $$dx\,dy\,dz\;=\;abc\,r^2\,\cos\vartheta\,dr\,d\varphi\,d\vartheta.$$

Transforming a triple integral to generalized spherical coordinates usually leads to a simplification of the integrand if the function depends on x, y and z through the expression $(x-x_0)^2/a^2+(y-y_0)^2/b^2+(z-z_0)/c^2$, which under the transformation becomes r^2. The description of the domain of integration is usually simplified if the domain is the interior of the ellipsoid $(x-x_0)^2/a^2+(y-y_0)^2/b^2+(z-z_0)^2/c^2=R^2$ or some simple part of it.

II.8 Some physical applications of the triple integral

Assume that a three-dimensional body fills a measurable set M in $\mathbb{E}_3$. The body need not be homogeneous, which means that its density need not be constant. Let us denote the density by $\rho(x,y,z)$ and assume it is given in $\mathrm{kg}\cdot\mathrm{m}^{-3}$. Using the triple integral, the following basic mechanical characteristics of the body can be defined and computed:

mass of the body ..... $m=\iiint_M \rho(x,y,z)\,dx\,dy\,dz$ [kg],

static moment

with respect to the plane xy ..... $m_{xy}=\iiint_M z\cdot\rho(x,y,z)\,dx\,dy\,dz$ [kg · m],

with respect to the plane xz ..... $m_{xz}=\iiint_M y\cdot\rho(x,y,z)\,dx\,dy\,dz$ [kg · m],

with respect to the plane yz ..... $m_{yz}=\iiint_M x\cdot\rho(x,y,z)\,dx\,dy\,dz$ [kg · m],

coordinates of the center of mass $x_T=\dfrac{m_{yz}}{m}$, $y_T=\dfrac{m_{xz}}{m}$ $z_T=\dfrac{m_{xy}}{m}$ [m],

moment of inertia

with respect to the plane xy $\quad J_{xy} = \iiint_M z^2 \cdot \rho(x,y,z)\,dx\,dy\,dz \quad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the plane xz $\quad J_{xz} = \iiint_M y^2 \cdot \rho(x,y,z)\,dx\,dy\,dz \quad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the plane yz $\quad J_{yz} = \iiint_M x^2 \cdot \rho(x,y,z)\,dx\,dy\,dz \quad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the x-axis .. $\quad J_x = \iiint_M (y^2+z^2) \cdot \rho(x,y,z)\,dx\,dy\,dz \quad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the y-axis .. $\quad J_y = \iiint_M (x^2+z^2) \cdot \rho(x,y,z)\,dx\,dy\,dz \quad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the z-axis .. $\quad J_z = \iiint_M (x^2+y^2) \cdot \rho(x,y,z)\,dx\,dy\,dz \quad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the origin .. $\quad J_0 = \iiint_M (x^2+y^2+z^2) \cdot \rho(x,y,z)\,dx\,dy\,dz \quad [\mathrm{kg}\cdot\mathrm{m}^2]$.

Try to propose yourself a formula by which we can define and also compute the moment of inertia with respect to a general line or plane in $\mathbb{E}_3$.

II.9 * More on the substitution method for computing double and triple integrals

The idea on which the general substitution method for computing double and triple integrals is based is the same for both integrals. Therefore, in this chapter we will explain it simultaneously for the double and the triple integral. We will use the following simplified notation:

- $\mathbb{E}_k$ will denote either $\mathbb{E}_2$ (if $k=2$) or $\mathbb{E}_3$ (if $k=3$),
- the integral $\int$ will mean either $\iint$ (if $k=2$) or $\iiint$ (if $k=3$),
- a point $X \in \mathbb{E}_k$ will have the form $[x_1; x_2]$ (for $k=2$) or $[x_1; x_2, x_3]$ (for $k=3$), and
- dX will denote either $dx_1\,dx_2$ (for $k=2$) or $dx_1\,dx_2\,dx_3$ (for $k=3$).

Some commonly used substitutions are already known to us–they are defined by the equations (II.3.3), (II.3.6), (II.7.2), (II.7.4), (II.7.6), and (II.7.8) and are called the transformations to polar, generalized polar, cylindrical, spherical, generalized cylindrical, or generalized spherical coordinates. In this chapter we will deal with the substitution method on a general level.

Assume that $M \subset \mathbb{E}_k$ ($k=2$ or $k=3$) and that we compute the integral $\int_M f(X)\,dX$. If every point $X \in M$ can be expressed in the form $X = \mathcal{F}(Y)$, where the points Y range over some set $M' \subset \mathbb{E}_k$, then the integral can be transformed to another integral in the variable Y over the domain M'. However, it is necessary that both integrals over M and M' exist and that the mapping $\mathcal{F}$ has certain properties. The required properties of the mapping are discussed in the following sections.

II.9.1. A regular mapping and its Jacobian

Let M and M' be regions in $\mathbb{E}_k$. Assume that $\mathcal{F}$ is a mapping from M' to M, defined by the equations

$$x_1 = \phi_1(y_1, y_2), \quad x_2 = \phi_2(y_1, y_2) \quad \text{(for } k = 2\text{)},$$

$$\text{or} \quad x_1 = \phi_1(y_1, y_2, y_3), \quad x_2 = \phi_2(y_1, y_2, y_3), \quad x_3 = \phi_3(y_1, y_2, y_3) \quad \text{(for } k = 3\text{)}.$$

Let J denote the determinant defined by

$$J = \begin{vmatrix} \frac{\partial \phi_1}{\partial y_1}, & \frac{\partial \phi_1}{\partial y_2} \\ \frac{\partial \phi_2}{\partial y_1}, & \frac{\partial \phi_2}{\partial y_2} \end{vmatrix} \text{ (if } k = 2\text{)} \quad \text{or} \quad J = \begin{vmatrix} \frac{\partial \phi_1}{\partial y_1}, & \frac{\partial \phi_1}{\partial y_2}, & \frac{\partial \phi_1}{\partial y_3} \\ \frac{\partial \phi_2}{\partial y_1}, & \frac{\partial \phi_2}{\partial y_2}, & \frac{\partial \phi_2}{\partial y_3} \\ \frac{\partial \phi_3}{\partial y_1}, & \frac{\partial \phi_3}{\partial y_2}, & \frac{\partial \phi_3}{\partial y_3} \end{vmatrix} \text{ (if } k = 3\text{)}.$$

J is called the *Jacobian determinant* or, for short, the *Jacobian* of the mapping $\mathcal{F}$. It is clear that J depends on Y.

The mapping $\mathcal{F}$ is called *regular* if the functions ϕ_i ($i = 1, 2$ or $i = 1, 2, 3$) have continuous partial derivatives in the region M' and $J(Y) \neq 0$ at all points $Y \in M'$.

II.9.2. A one-to-one correspondence

A mapping $\mathcal{F}$ is called a *one-to-one* mapping of the set M' onto the set M if

a) it is injective
(i.e., for all $Y_1, Y_2 \in M'$ it holds: $Y_1 \neq Y_2 \implies \mathcal{F}(Y_1) \neq \mathcal{F}(Y_2)$)

b) and it is onto the set M (i.e., its range covers the whole set M).

II.9.3. Example

Verify that the mapping defined by the equations

(II.9.1) $$x_1 = \phi_1(r, \varphi) = r \cos \varphi, \quad x_2 = \phi_2(r, \varphi) = r \sin \varphi$$

(compare with equations (II.3.3)) is a one-to-one regular mapping of the open rectangle $M' = \{[r; \varphi] \in \mathbb{E}_2 : r \in (0; 2), \varphi \in (0; 2\pi)\}$ onto the region $M = \{[x_1; x_2] \in \mathbb{E}_2 : x_1^2 + x_2^2 < 4\} \setminus \{[x_1; x_2] \in \mathbb{E}_2 : x_1 \in \langle 0; 2), x_2 = 0\}$. ($M$ is an open disk in $\mathbb{E}_2$ centered at the origin, from which the segment $\overline{OP}$ is removed, where O is the origin and $P = [2; 0]$.) See Fig. 15. The one-to-one correspondence between points $[x_1; x_2] \in M$

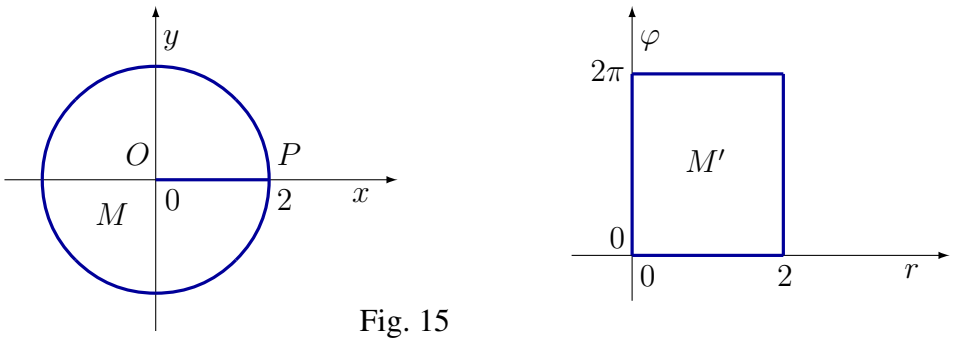


Fig. 15

and $[r; \varphi] \in M'$ is obvious. It is also obvious that the functions ϕ_1 and ϕ_2 have continuous partial derivatives in M'. To verify regularity of the mapping (II.9.1), it therefore suffices to compute its Jacobian:

$$J(r,\varphi) = \begin{vmatrix} \dfrac{\partial\phi_1}{\partial r}, & \dfrac{\partial\phi_1}{\partial\varphi} \\ \dfrac{\partial\phi_2}{\partial r}, & \dfrac{\partial\phi_2}{\partial\varphi} \end{vmatrix} = \begin{vmatrix} \cos\varphi, & -r\sin\varphi \\ \sin\varphi, & r\cos\varphi \end{vmatrix} = r.$$

It is clear that $J(r,\varphi) \neq 0$ in M'. The mapping (II.9.1) is therefore regular in M'.

II.9.4. Theorem (substitution)
Let $X = \mathcal{F}(Y)$ be a one-to-one regular mapping of the region $M' \in \mathbb{E}_k$ onto the region $M \in \mathbb{E}_k$. Then

(II.9.2) $$\int_M f(X)\,dX = \int_{M'} f(\mathcal{F}(Y))\cdot|J(Y)|\,dY,$$

provided that both integrals exist.

II.9.5. Remark

Adding or removing a set of measure zero to/from the domain of integration does not affect the value of the integral. Therefore, Theorem II.9.4 can be generalized:

If the assumptions of Theorem II.9.4 are satisfied and A, respectively A', are sets in $\mathbb{E}_k$ that differ from M, respectively M', only by a set of measure zero, then

(II.9.3) $$\int_A f(X)\,dX = \int_{A'} f(\mathcal{F}(Y))\cdot|J(Y)|\,dY,$$

provided that both integrals exist.

From (II.9.2) and (II.9.3) it is clear that dX in the integral on the left-hand side transforms under substitution into $|J(Y)|\,dY$ in the integral on the right-hand side. In Example II.9.3 we already showed that if $k = 2$ and $Y = [r;\varphi]$ represents polar coordinates, then the Jacobian equals r. From the general equation

(II.9.4) $$dX = |J(Y)|\,dY$$

it therefore follows, as a special case, equation (II.3.4) (i.e., $dx\,dy = r\,dr\,d\varphi$). Similarly, the correctness of the replacements of the expressions $dx\,dy$ and $dx\,dy\,dz$ in transformations to generalized polar, cylindrical, and other coordinate systems (see equations (II.3.7), (II.7.3), (II.7.5), (II.7.7), and (II.7.9)) can be verified by computing the Jacobians of the mappings defined by equations (II.3.6), (II.7.2), (II.7.4), (II.7.6), and (II.7.8).

II.10 Exercises

1. Do the following integrals exist? *(See sections II.1.6, II.2.1, II.5.5 and II.6.1.)*

a) $\iint_M \frac{dx\,dy}{x+y+1}$; $M = \langle 0;1\rangle \times \langle 0;1\rangle$

b) $\iint_M \frac{\sin(x^2+y^2)}{x^2+y^2}\,dx\,dy$; $M = \{[x;y] \in \mathbb{E}_2 :\ x^2+y^2 \le 9\}$

c) $\iint_M \frac{x}{x^2+y^2}\,dx\,dy$; $M = \{[x;y] \in \mathbb{E}_2 :\ x^2+y^2 \le 9\}$

d) $\iint_M \frac{dx\,dy}{(1-xy)^2}$; M is the square $PQRS$, where $P=[1;2]$, $Q=[3;2]$, $R=[3;4]$, $S=[1;4]$

e) $\iiint_M \sqrt{1-\frac{x^2}{4}-\frac{y^2}{9}-\frac{z^2}{16}}\,dx\,dy\,dz$; $M = \left\{[x;y;z] \in \mathbb{E}_3 :\ \frac{x^2}{4}+\frac{y^2}{9}+\frac{z^2}{16} < 1\right\}$

f) $\iiint_M \sqrt{x^2+y^2+z^2}\,dx\,dy\,dz$; $M = \{[x;y;z] \in \mathbb{E}_3 :\ x^2+y^2+z^2 \le 9\}$

2. Compute the area of the region M in the xy-plane bounded by the curves

a) $y = 2x+4,\ y = 4-x^2$,
b) $xy = 1,\ y = x,\ x = 4$,
c) $y = x^2,\ 4y = x^2,\ y = 4$,
d) $y^2 = 4+x,\ x+3y = 0$,
e) $y = \ln x,\ x-y = 1,\ y = -1$,
f) $x^2+y^2 = 2x,\ x^2+y^2 = 4x,\ y = x,\ y = 0$,
g) $y = x^2$ (from the right), $x+y = 2$ (from the left) and $y = 4$ (from above).

(See II.1.9 and II.3.2.)

3. Compute the volume of the given region M in $\mathbb{E}_3$. *(See II.1.7, II.5.8 and II.7.2.)*

a) M is bounded above by the paraboloid $z = x^2+y^2$ and below by the triangle between the lines $y = x,\ x = 0,\ x+y = 2$ in the xy-plane,
b) M is bounded above by the parabolic cylindrical surface $z = x^2$ and below by the region in the xy-plane between the parabola $y = 6-x^2$ and the line $y = x$,
c) M is a region in the half-space $z \ge 0$, bounded by the surfaces $x^2+y^2-z^2 = 0$, $z = 6-x^2-y^2$,
d) M is a region in the half-space $z \ge 0$, bounded by the surfaces $az = x^2+y^2$, $x^2+y^2+x^2 = 2a^2$, $(a > 0)$,
e) M is bounded by the surfaces $y^2 = 4a^2-3ax,\ y^2 = ax,\ z = h,\ z = -h$, $(a > 0,\ h > 0)$,
f) M is bounded by the surfaces $x^2+y^2 = z,\ x^2+y^2 = 1,\ z = 0$.

4. Compute the following integrals. *(See Chapters II.3 and II.7.)*

a) $\iint_M (1+x)\,dx\,dy$; M is a region in $\mathbb{E}_2$ bounded by the curves $y = x^2-4$ and $y^2 = 2x,\ y = -3x$,

b) $\iint_M \frac{dx\,dy}{(x+y)^2}$; $M = \langle 3;4\rangle \times \langle 1;2\rangle$,

c) $\iint_M xy\,dx\,dy$; M is a region in $\mathbb{E}_2$ bounded by the line $y = x-4$ and the parabola $y^2 = 2x$,

d) $\iiint_V (x+y+z)\,dx\,dy\,dz$; $V = \langle 0;1\rangle \times \langle 0;2\rangle \times \langle 0;3\rangle$,

e) $\iiint_V x\,dx\,dy\,dz$; V is a region in $\mathbb{E}_3$ bounded by the planes $x=0,\ y=0,\ z=0,$ $y=2,\ x+z=1$,

f) $\iiint_V xy^2z^3\,dx\,dy\,dz$; V is a region in $\mathbb{E}_3$ bounded by the surfaces $z=xy,\ y=x,$ $x=1,\ z=0$,

g) $\iiint_V x^3y^2z\,dx\,dy\,dz$; $V = \{[x;y;z] \in \mathbb{E}_3 :\ 0 \le x \le 1,\ 0 \le y \le x, 0 \le z \le xy\}$

h) $\iiint_V y\cos(x+z)\,dx\,dy\,dz$; V is a region in $\mathbb{E}_3$ bounded by the surfaces $y=\sqrt{x},$ $y=0,\ z=0,\ x+z=\pi/2$,

i) $\iint_D \sqrt{1-x^2-y^2}\,dx\,dy$; $D = \{[x;y] \in \mathbb{E}_2 :\ x^2+y^2 \le 1\}$,

j) $\iint_D \frac{dx\,dy}{(x^2+y^2)^2}$; D is a region in the xy-plane bounded by the lines $y=x,\ y=2x$ and the circles $x^2+y^2=4x,\ x^2+y^2=8x$,

k) $\iint_D y\,dx\,dy$; D is the upper half of the disk $(x-a)^2+y^2 \le a^2\ \ (a>0)$,

l) $\iint_D x\,dx\,dy$; D is a sector of the disk $x^2+y^2 \le a^2$, containing points $[x;y]$ for which it holds $x \ge 0$ and $-x \le \sqrt{3}\,y \le 1$,

m) $\iiint_V \sqrt{x^2+y^2+z^2}\,dx\,dy\,dz$; V is the ball $x^2+y^2+z^2 \le a^2,\ (a>0)$,

n) $\iiint_D (x+y+z)^2\,dx\,dy\,dz$; D is a region in the half-space $z \ge 0,$ bounded by the paraboloid $z = \frac{1}{2}(x^2+y^2)$ and the spherical surface $x^2+y^2+z^2=3$,

o) $\iiint_V z\,dx\,dy\,dz$; V is a region in $\mathbb{E}_3$ bounded by the surfaces $z=\sqrt{x^2+y^2}$ and $z=1$,

p) $\iiint_D (x^2+y^2)\,dx\,dy\,dz$; D is a region in $\mathbb{E}_3$ bounded by the surfaces $x^2+y^2=2z$ and $z=2$,

q) $\iiint_V \left(\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}\right)dx\,dy\,dz$; V is the interior of the ellipsoid $x^2/a^2+y^2/b^2+z^2/c^2=1,\ (a>0, b>0,\ c>0)$.

5. Determine the coordinates of the center of mass of a homogeneous plate in $\mathbb{E}_2$ bounded by the curves

a) $y=\sin x,\ y=0;\ x \in \langle 0;\pi\rangle$, b) $x^2+y^2=a^2,\ y=0;\ (y \ge 0,\ a>0)$,

c) $y^2=ax,\ x=0,\ y=a;\ (y>0,\ a>0)$, d) $y^2=4x+4,\ y^2=-2x+4$.

(See Chapter II.4.)

6. Compute the moment of inertia about the x-axis of a homogeneous plate in $\mathbb{E}_2$ bounded by the lines $y = x/2,\ y = a,\ x = a\ \ (a > 0)$. The density is $\rho = 1$.
(See Chapter II.4.)

7. Compute the mass of a body in $\mathbb{E}_3$ bounded by the surfaces

a) $x = 0, x = a, y = 0, y = b, z = 0, z = c\ (a > 0,\ b > 0,\ c > 0)$, if the density is $\rho(x, y, z) = x + y + z$,

b) $2x + z = 2a, x + z = a, y^2 = ax, y = 0$ (for $y > 0$), if $a > 0$ and the density is $\rho(x, y, z) = y$,

c) $x^2 + y^2 + z^2 = a^2, x^2 + y^2 + z^2 = 4a^2\ \ (a > 0)$, if the density is $\rho(x, y, z) = 2/\sqrt{x^2 + y^2 + z^2}$.

(See Chapter II.8.)

CHAPTER III.

Line integrals

III.1 Simple curves 79
III.2 Line integral of a scalar function (Line integral of the first kind) 83
III.3 Some physical applications of the line integral of a scalar function 86
III.4 Line integral of a vector function (Line integral of the 2nd kind) 88
III.5 Green's theorem 91
III.6 Exercises 94

III.1 Simple curves

Line integrals are integrals whose domains of integration are curves. You will see that these integrals have broad applications. Before we begin studying line integrals, however, we must clarify what exactly we mean by the term "curve". In the literature one can find definitions of various types of curves. In this text we restrict ourselves to two types: the so-called "simple smooth curve" and the "simple piecewise smooth curve".

III.1.1. Simple smooth curve – motivation and notation

The definition of a simple smooth curve is based on the following idea: A point A moves in $\mathbb{E}_2$ or in $\mathbb{E}_3$ over the time interval $\langle a\,;b\rangle$ and its position at time t is $P(t)$. Assume that

a) at no two different instants $t_1,\ t_2 \in \langle a\,;b\rangle$, $t_1 < t_2$ (with the possible exception of the case $t_1 = a$ and $t_2 = b$) is the point A located at the same place, and

b) the velocity $\dot{P}(t)$ of the motion is bounded, varies continuously, and is nonzero, with the possible exception of the cases $t = a$ or $t = b$.

The trajectory traced by the point A during the time interval $\langle a\,;b\rangle$ will be called a simple smooth curve. The position function P will be called a "parametrization" of the simple smooth curve.

In the precise definition we may forget the picture of motion of a point and simply regard P as a mapping from the interval $\langle a\,;b\rangle$ into $\mathbb{E}_2$ (a curve in $\mathbb{E}_2$) or into $\mathbb{E}_3$ (a curve in $\mathbb{E}_3$). The simple smooth curve will then be the range of the mapping P. Since for each fixed $t \in \langle a\,;b\rangle$ the value $P(t)$ is a point of $\mathbb{E}_k$ ($k = 2$ or $k = 3$), $P(t)$ has two or three coordinates. If we denote them by $\phi(t)$, $\psi(t)$ ($k = 2$) or $\phi(t)$, $\psi(t)$, $\vartheta(t)$ ($k = 3$), we may write:

$$\begin{aligned} P(t) &= \big[\phi(t);\psi(t)\big] && \text{if } k = 2,\\ P(t) &= \big[\phi(t);\psi(t);\vartheta(t)\big] && \text{if } k = 3. \end{aligned}$$

The functions ϕ and ψ (respectively ϕ, ψ and ϑ) are functions of the single variable t, defined on the interval $\langle a\,;b\rangle$. We call them the *coordinate functions* of the mapping P. The variable t is called the *parameter*, and differentiation with respect to t is denoted, as is customary in physics, by a dot. We regard the derivative of P as a vector-valued function. (This is natural, because in physics the derivative of position with respect to time is velocity, and velocity is a vector.) The coordinates of $\dot{P}(t)$ are therefore written in parentheses. We may also use the vectors $\widehat{\imath}$, $\widehat{\jmath}$ and $\widehat{k}$ (unit vectors oriented consistently with the x, y and z axes – see Section I.2.12) and write:

$$\begin{aligned} \dot{P}(t) &= \big(\dot{\phi}(t);\dot{\psi}(t)\big) = \dot{\phi}(t)\,\widehat{\imath} + \dot{\psi}(t)\,\widehat{\jmath} && \text{if } k = 2,\\ \dot{P}(t) &= \big(\dot{\phi}(t);\dot{\psi}(t);\dot{\vartheta}(t)\big) = \dot{\phi}(t)\,\widehat{\imath} + \dot{\psi}(t)\,\widehat{\jmath} + \dot{\vartheta}(t)\,\widehat{k} && \text{if } k = 3. \end{aligned}$$

We consider the mapping P to be continuous if all of its coordinate functions are continuous. Similarly, we say that the mapping P has a continuous derivative if all coordinate functions have a continuous derivative.

If there is no danger of confusion with the notation for the coordinate axes, we may denote the coordinate functions $\phi(t)$, $\psi(t)$, $\vartheta(t)$ of the mapping P also by $x(t)$, $y(t)$, $z(t)$.

III.1.2. Simple smooth curve – definition

Let P be a continuous mapping of the interval $\langle a\,;b\rangle \subset \mathbb{R}$ into $\mathbb{E}_k$ (where $k = 2$ or $k = 3$). Assume that

a) the mapping P is one-to-one on the interval $\langle a\,;b\rangle$, with the possible exception of the case $P(a) = P(b)$, and

b) P has a bounded, continuous, and nonzero derivative $\dot{P}$ on the open interval $(a;b)$.

Then the set of all points $P(t)$ for $t \in \langle a\,;b\rangle$ (i.e. the range of the mapping P) is called a *simple smooth curve* in $\mathbb{E}_k$. The mapping P is called a *parametrization*.

We call the above simple smooth curve *closed* if $P(a) = P(b)$.

We will usually denote simple smooth curves by capital letters, for example C, K, C_1, C_2, etc.

In Fig. 16a and 16b you can see examples of simple smooth curves. On the other hand, the curves in Fig. 16c and 16d are not simple smooth curves. (The curve in Fig. 16c is not simple and the curve in Fig. 16d is not smooth.)

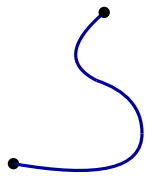


Fig. 16a

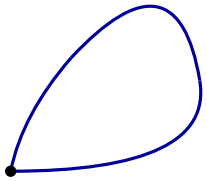


Fig. 16b

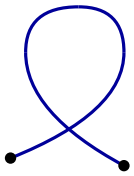


Fig. 16c

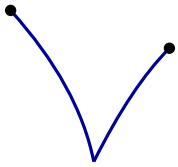


Fig. 16d

Every simple smooth curve has infinitely many parametrizations. This is easy to understand if we return to the idea explained at the beginning of Section III.1.1, according to which a simple smooth curve can be viewed as a trajectory traced, under certain rules, by a material point over some time interval. One concrete trajectory can obviously be traversed, in accordance with the given rules, in one direction, in the opposite direction, over many different time intervals, and with a speed that varies in different ways. Each way of traversing the curve corresponds to one parametrization. Since the curve can be traversed in infinitely many ways, it has infinitely many different parametrizations.

III.1.3. Tangent vector. Oriented simple smooth curve.

Suppose that P is a parametrization, defined on the interval $\langle a\,;b\rangle$, of a simple smooth curve C in $\mathbb{E}_k$ (where $k = 2$ or $k = 3$). Then the vector $\dot{P}(t)$ has, for each $t \in (a;b)$, a tangent direction to the curve C. (Again this is easy to understand – for the same reason that the velocity of the corresponding motion of a point in $\mathbb{E}_2$ or $\mathbb{E}_3$ is tangent to the path of motion.) The vectors $\pm\dot{P}(t)/\|\dot{P}(t)\|$ are also tangent to the curve C and, moreover, have unit length.

We call the simple smooth curve C *oriented by a unit tangent vector* $\widehat{\tau}$ if we choose

a) $\widehat{\tau} = \dfrac{\dot{P}(t)}{\|\dot{P}(t)\|}$ at all points $P(t)$ corresponding to $t \in (a;b)$,

b) or $\widehat{\tau} = -\dfrac{\dot{P}(t)}{\|\dot{P}(t)\|}$ at all points $P(t)$ corresponding to $t \in (a;b)$.

In case a) we say that the simple smooth curve C is oriented *consistently with the parametrization P*. We call $P(a)$ the *initial point* of the curve C (denoted $i.p.\,C$) and $P(b)$ the *terminal point* of the curve C (denoted $t.p.\,C$).

In case b) we say that the simple smooth curve C is oriented *oppositely to the parametrization P*. The names and notation of the points $P(a)$, $P(b)$ are opposite to those in case a).

If the curve C is not closed, then its orientation can also be defined uniquely by choosing one of the points $P(a)$ as the initial point and the other as the terminal point. (For a closed curve we do not have this possibility, because the points $P(a)$ and $P(b)$ coincide.)

In figures, the orientation of a curve is usually indicated by arrows – see Fig. 17a and 17b. The curve C_1 in Fig. 17a is oriented consistently with the parametrization P_1, while the curve in Fig. 17b is oriented oppositely to the parametrization P_2. (For simplicity we assume that both parametrizations are defined on the same interval $\langle a\,;b\rangle$.)

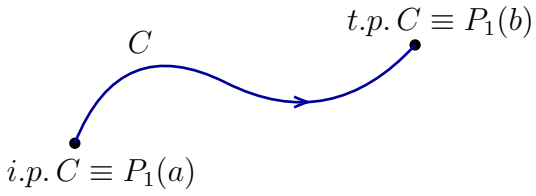


Fig. 17a

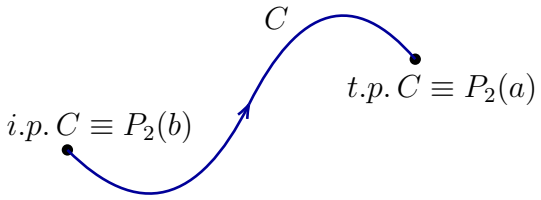


Fig. 17b

The relationship between the orientation of a simple smooth curve C and its parametrization P (defined on the interval $\langle a\,;b\rangle$) can also be illustrated using the already mentioned idea that C is a trajectory traced by a point A over the time interval $\langle a\,;b\rangle$, where $P(t)$ denotes the position of A at time t. If the curve C is oriented consistently with the parametrization P, the point A moves along the curve C in the direction of its orientation. If the curve C is oriented oppositely to the parametrization P, the point A moves along the curve C against the direction of its orientation.

III.1.4. Example

Every line segment in $\mathbb{E}_k$ is a simple smooth curve. For instance, the segment $\overline{AB}$ in $\mathbb{E}_3$, where $A = [1;2;4]$ and $B = [3;-1;7]$, can be parametrized by the map

$$P(t) = A + t \cdot (B - A); \quad t \in \langle 0;1\rangle.$$

Writing it out in coordinates, we obtain the coordinate functions ϕ, ψ, and ϑ of this parametrization:

$$x = \phi(t) = 1 + 2t, \;\; y = \psi(t) = 2 - 3t, \;\; z = \vartheta(t) = 4 + 3t; \quad t \in \langle 0;1\rangle.$$

Verify yourself that the mapping P has all the properties required in Definition III.1.2.

If the segment is oriented so that A is its initial point and B is its terminal point, then the orientation agrees with the parametrization P given above.

III.1.5. Example

The arc of the parabola $y = x^2 + 1$ between the points $[1; 2]$ and $[3; 10]$ (oriented from $[3; 10]$ to $[1; 2]$) is a simple smooth curve in $\mathbb{E}_2$. Verify yourself that one possible parametrization is, for example, the mapping

$$P : \quad x = \phi(t) = t, \;\; y = \psi(t) = t^2 + 1; \quad t \in \langle 1; 3 \rangle.$$

This parametrization corresponds to moving along the curve from its terminal point to its initial point during the time interval $\langle 1; 3 \rangle$. The curve is therefore oriented opposite to the chosen parametrization.

III.1.6. Example

The circular arc $x^2 + y^2 = 9, \; x \geq 0, \; y \geq 0,$ oriented from the point $[3; 0]$ to the point $[0; 3]$, is a simple smooth curve in $\mathbb{E}_2$. Verify yourself that one possible parametrization is, for example, the mapping defined by

$$x = \phi(t) = 3 \cos t, \;\; y = \psi(t) = 3 \sin t; \quad t \in \langle 0; \pi/2 \rangle.$$

The given orientation of the curve agrees with this parametrization. (Verify yourself that the mapping $x = t$, $y = \sqrt{9 - t^2}$; $t \in \langle 0; 3 \rangle$ does not satisfy the requirements imposed on a parametrization.)

III.1.7. Example

The circle $C : \; (x - 3)^2 + y^2 = 4$ in $\mathbb{E}_2$ (oriented counterclockwise) is a closed simple smooth curve. Verify yourself that one possible parametrization is the mapping defined by

$$x = \phi(t) = 3 + 2 \cos t, \;\; y = \psi(t) = 2 \sin t; \quad t \in \langle 0; 2\pi \rangle.$$

The given orientation of the curve C agrees with this parametrization.

III.1.8. Simple piecewise smooth curve

Suppose that $C_1, \ldots, C_n$ are simple smooth curves in $\mathbb{E}_k$ ($k = 2$ or $k = 3$) such that

a) $t.p.\, C_1 = i.p.\, C_2, \;\; t.p.\, C_2 = i.p.\, C_3, \ldots, t.p.\, C_{n-1} = i.p.\, C_n,$

b) apart from the points mentioned in a), and apart from the possible case when $i.p.\, C_1 = t.p.\, C_n,$ no two of the curves $C_1, \ldots, C_n$ have any other common point.

Then the union $C = \cup_{i=1}^{n} C_i$ is called a *simple piecewise smooth curve* in $\mathbb{E}_k$. (We often abbreviate this as "simple p.w. smooth curve".)

The orientation of a simple p.w. smooth curve C is determined by the orientation of its individual smooth parts $C_1, \ldots, C_n$. We set $i.p.\, C = i.p.\, C_1$ (the initial point of C) and $t.p.\, C = t.p.\, C_n$ (the terminal point of C).

The curve that differs from C only by orientation is denoted $-C$.

A simple p.w. smooth curve C is called *closed* if $i.p.\, C = t.p.\, C$.

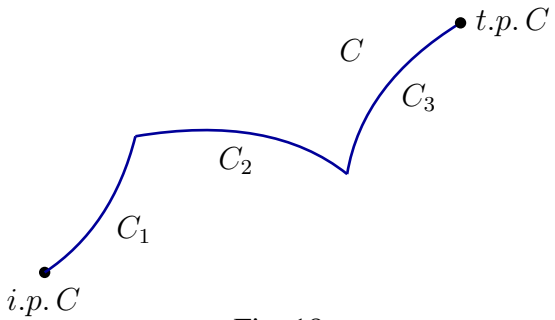


Fig. 18a

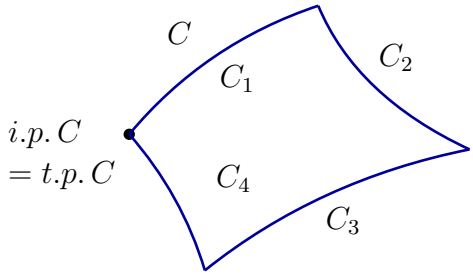


Fig. 18b

In Figs. 18a and 18b you can see examples of simple p.w. smooth curves. The curve in Fig. 18b is closed.

The concept of a simple p.w. smooth curve generalizes the concept of a simple smooth curve. A simple p.w. smooth curve in $\mathbb{E}_k$ ($k = 2$ or $k = 3$) is a bounded measurable set in $\mathbb{E}_k$ whose k-dimensional measure is equal to zero.

III.2 Line integral of a scalar function (Line integral of the first kind)

III.2.1. Physical motivation

Imagine that a string or a wire has the shape of a simple smooth curve C in $\mathbb{E}_k$ ($k = 2$ or $k = 3$). The string need not be homogeneous; its linear density ρ (i.e., mass per unit length) is a function of two variables x, y (if $k = 2$) or of three variables x, y, z (if $k = 3$). We want to compute the total mass of the string.

We could arrive at an expression and a computation of the mass of the curve C by partitioning the curve into many smaller parts, passing to Riemann sums, and considering the limit of the Riemann sums under refinement of the partition. (In an analogous way, in Sections II.1.1 and II.1.7 we derived the mass of a lamina via a double integral, and in Sections II.5.1 and II.5.6 the mass of a solid via a triple integral.) Let us take a different approach. This section is only motivational, so we may allow ourselves to use "archaisms" that are not precisely and clearly defined notions. We have in mind an "infinitesimal" positive number dt and the idea of splitting the curve C into infinitely many "infinitesimally short" segments. (At this point we recommend returning to the textbook [5] and reading Sections V.1.3 and V.1.4.) If P is a parametrization of the curve C defined on the interval $\langle a\,;b\rangle$, then a typical "infinitesimally short" piece of C is the segment with endpoints $P(t)$ and $P(t+dt)$, where $t \in \langle a;b)$ and dt is an "infinitesimal" positive number. The density can be regarded as constant on this segment and equal to $\rho(P(t))$. The mass of the segment is then $dm = \rho(P(t)) \cdot ds$, where ds is the length of the segment. Since the segment is "infinitesimally short", its length can be expressed as: $ds = \|P(t+dt) - P(t)\| = \|\dot{P}(t)\|\,dt$. The total mass of the curve C is obtained by summing the masses of all the infinitesimally short segments into which we have split C:

$$m = \int_C \rho\,ds = \int_a^b \rho(P(t)) \cdot \|\dot{P}(t)\|\,dt.$$

We will now use this expression as the basis for the definition of the so-called line integral of a scalar function. In the definition we will denote the integrand by f instead of ρ. In accordance with the convention from Section I.2.12, instead of "scalar function" we will usually use simply the term "function".

III.2.2. Line integral of a scalar function on a simple smooth curve

Let C be a simple smooth curve in $\mathbb{E}_2$ or in $\mathbb{E}_3$, and let P be its parametrization defined on the interval $\langle a\,;b\rangle$. Let f be a function that is defined and bounded on the curve C. If the Riemann integral $\int_a^b f(P(t))\cdot\|\dot{P}(t)\|\,dt$ exists, then we say that the function f is *integrable* on the curve C. The *line integral* of the scalar function f on the curve C is denoted $\int_C f\,ds$ and is defined by

$$\int_C f\,ds = \int_a^b f(P(t))\cdot\|\dot{P}(t)\|\,dt. \tag{III.2.1}$$

We will return to the computation of the integral on the right in Section III.2.7, and we will also illustrate it with examples in Sections III.2.8 and III.2.9.

The statements "the function f is integrable on the curve C" and "the line integral of f on C exists" will mean exactly the same thing.

III.2.3. Remark – independence of the choice of parametrization

The integrability of the function f on a simple smooth curve C and the line integral of f on C are defined using a parametrization P of the curve C. However, the curve C has infinitely many different parametrizations (see Section III.1.2). This raises the question: Can the integrability of f on C or the value of the line integral of f on C depend on the particular choice of parametrization, or are these notions independent of that choice? The answer is: It can be proved that both the existence and the value of the line integral f on C are independent of the choice of parametrization of the curve C.

III.2.4. Line integral of a scalar function on a simple p.w. smooth curve

Let C be a simple p.w. smooth curve in $\mathbb{E}_2$ or in $\mathbb{E}_3$, composed of simple smooth curves $C_1,\dots,C_n$ (see Section III.1.8). Let f be a scalar function that is defined and bounded on the curve C. If the function f is integrable on each of the curves $C_1,\dots,C_n$, then we say that it is *integrable* on the curve C. The *line integral* of the scalar function f on the curve C is then defined by

$$\int_C f\,ds = \sum_{i=1}^{n}\int_{C_i} f\,ds. \tag{III.2.2}$$

(The line integrals on the right-hand side are line integrals of the function f over simple smooth curves, which are already known from Definition III.2.2.)

Instead of "line integral of a scalar function", the term *line integral of the first kind* is often used.

Instead of $\int_C f\,ds$ we may write, for example, $\int_C f(x,y)\,ds$ (if $C\subset\mathbb{E}_2$ and f is a function of two variables) or $\int_C f(x,y,z)\,ds$ (if $C\subset\mathbb{E}_3$ and f is a function of three variables). Instead of the symbol ds at the end of the integral, one often also uses dl, dr, etc.

III.2.5. Length of a curve

If C is a simple p.w. smooth curve in $\mathbb{E}_2$ or in $\mathbb{E}_3$, then the integral $\int_C ds$ defines the *length* of the curve C. We denote it by $l(C)$.

From (III.2.1) it follows that in the special case when C is a simple smooth curve and P is its parametrization defined on the interval $\langle a\,;b\rangle$, we can express the length of C as

$$l(C) = \int_C 1\,ds = \int_a^b \|\dot{P}(t)\|\,dt. \tag{III.2.3}$$

III.2.6. Some important properties of the line integral of a scalar function

Since the line integral of a scalar function is defined using the one-dimensional Riemann integral, most properties of the two integrals are the same. Let us therefore mention only some of them:

a) **Sufficient condition for existence of the line integral of a scalar function**
If the function f is continuous on a simple p.w. smooth curve C, then it is integrable on C (i.e., the integral $\int_C f\ ds$ exists).

b) **Linearity of the line integral**
If f and g are integrable functions on the curve C and $\alpha \in \mathbb{R}$, then $f+g$ and αf are also integrable on C, and

$$\int_C (f+g)\ ds = \int_C f\ ds + \int_C g\ ds,$$

$$\int_C \alpha \cdot f\ ds = \alpha \cdot \int_C f\ ds.$$

c) *If f is an integrable function on C and the function g differs from f at most at finitely many points, then g is also integrable on C and*

$$\int_C g\ ds = \int_C f\ ds.$$

d) *If f is integrable on C, then it is also integrable on $-C$ and*

$$\int_{-C} f\ ds = \int_C f\ ds.$$

e) *If f and g are integrable on C and $f(X) \geq g(X)$ at all points $X \in C$, then*

$$\int_C f\ ds \geq \int_C g\ ds.$$

In particular, if $f(X) \geq 0$ at all points $X \in M$, then $\int_C f\ ds \geq 0.$

Statement a) can be generalized: *If C is a simple p.w. smooth curve and f is continuous on each of its smooth parts, then f is integrable on C.*

Statement d) says that neither the existence nor the value of the line integral of a scalar function depends on the orientation of the curve.

III.2.7. Computing the line integral of a scalar function

Suppose that $P = [\phi; \psi; \vartheta]$ is a parametrization of a simple smooth curve C, defined on the interval $\langle a\,; b\rangle$. Then the line integral of f on C can be computed using formula (III.2.1). We obtain $f(P(t))$ by substituting into f the variables x, y, z as follows:

(III.2.4) $$x = \phi(t), \quad y = \psi(t), \quad z = \vartheta(t),$$

and for ds we substitute

(III.2.5) $$ds = \|\dot{P}(t)\|\,dt = \big\|\big(\dot{\phi}(t);\, \dot{\psi}(t);\, \dot{\vartheta}(t)\big)\big\|\,dt = \sqrt{\dot{\phi}(t)^2 + \dot{\psi}(t)^2 + \dot{\vartheta}(t)^2}\,dt$$

and integrate with respect to the parameter t over the interval $\langle a\,; b\rangle$ on which the parametrization of C is defined.

To compute the line integral of f on a simple p.w. smooth curve composed of simple smooth curves $C_1, \ldots, C_n$, we first compute the integrals over $C_1, \ldots, C_n$ (for example, using their parametrizations) and then sum the results. (See formula (III.2.2).)

III.2.8. Example

Let C be the union of the segments $C_1 = \overline{OP}$ and $C_2 = \overline{PQ}$, where $O = [0;0;0]$, $P = [1;1;0]$, and $Q = [1;1;1]$. Compute the line integral of the function $f(x,y,z) = x - 3y^2 + z$ over the curve C.

Solution:
The simplest parametrizations of the segments C_1 and C_2 are:

$$\begin{aligned} C_1 &: \quad P_1(t) = O + (P-O)\,t = [t;t;0]; \quad t \in \langle 0;1\rangle, \\ C_2 &: \quad P_2(t) = P + (Q-P)\,t = [1;1;t]; \quad t \in \langle 0;1\rangle. \end{aligned}$$

It follows that $\dot{P}_1(t) = (1;1;0)$, $\dot{P}_2(t) = (0;0;1)$ and $\|\dot{P}_1(t)\| = \sqrt{2}$, $\|\dot{P}_2(t)\| = 1$. Using formulas (III.2.1) and (III.2.2) we obtain:

$$\begin{aligned} \int_C (x - 3y^2 + z)\,ds &= \int_{C_1} (x - 3y^2 + z)\,ds + \int_{C_2} (x - 3y^2 + z)\,ds = \\ &= \int_0^1 (t - 3t^2)\,\sqrt{2}\,dt + \int_0^1 (t-2)\,dt = -\frac{\sqrt{2}+3}{2}. \end{aligned}$$

III.2.9. Example

Compute the line integral $\int_C (x^2 + y)\,ds$, where C is the circle $x^2 + (y-5)^2 = 4$. The orientation of the circle is not specified because it has no effect on the existence or the value of the line integral of a scalar function.

Solution:
C is a closed simple smooth curve, which can be parametrized, for example, by

$$x = \phi(t) = 2\,\cos t, \quad y = \psi(t) = 5 + 2\,\sin t; \quad t \in \langle 0;\, 2\pi\rangle.$$

Compute $\|\dot{P}(t)\|$: $\|\dot{P}(t)\| = \big\|(-2\,\sin t;\, 2\,\cos t)\big\| = \sqrt{4\,\sin^2 t + 4\,\cos^2 t} = 2$.
Using formula (III.2.1) we get:

$$\int_C (x^2 + y)\,ds = \int_0^{2\pi} (4\,\cos^2 t + 5 + 2\,\sin t)\cdot 2\,dt = 28\,\pi.$$

III.3 Some physical applications of the line integral of a scalar function

The line integral of a scalar function has many applications. In this text we restrict ourselves to applications in mechanics.

Suppose that a wire or string has the shape of a curve C in $\mathbb{E}_k$ ($k = 2$ or $k = 3$). Along the curve, mass is distributed with linear density $\rho(x,y)$ (if $k = 2$) or $\rho(x,y,z)$ (if $k = 3$). (ρ is mass per unit length; it is given in $\mathrm{kg}\cdot\mathrm{m}^{-1}$.) Using the line integral of a scalar function we can define and compute some basic mechanical characteristics of the curve C.

I. $k = 2$

total mass	$m = \int_C \rho(x,y)\,ds$	[kg],
static moment		
with respect to the x-axis	$m_x = \int_C y\cdot\rho(x,y)\,ds$	[kg · m],
with respect to the y-axis ..	$m_y = \int_C x\cdot\rho(x,y)\,ds$	[kg · m],
center of mass coordinates	$x_T = \frac{m_y}{m}, \quad y_T = \frac{m_x}{m}$	[m],

moment of inertia

with respect to the x-axis $\quad J_x = \int_C y^2 \cdot \rho(x,y)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the y-axis $\quad J_y = \int_C x^2 \cdot \rho(x,y)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the origin $\quad J_0 = \int_C (x^2+y^2) \cdot \rho(x,y)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$.

II. $k = 3$

total mass . . . . . $m = \int_C \rho(x,y,z)\, ds \qquad [\mathrm{kg}]$,

static moment

with respect to the xy-plane $\quad m_{xy} = \int_C z \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}]$,

with respect to the xz-plane $\quad m_{xz} = \int_C y \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}]$,

with respect to the yz-plane $\quad m_{yz} = \int_C x \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}]$,

center of mass coordinates $\quad x_T = \frac{m_{yz}}{m}, \quad y_T = \frac{m_{xz}}{m} \quad z_T = \frac{m_{xy}}{m} \qquad [\mathrm{m}]$,

moment of inertia

with respect to the xy-plane $\quad J_{xy} = \int_C z^2 \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the xz-plane $\quad J_{xz} = \int_C y^2 \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the yz-plane $\quad J_{yz} = \int_C x^2 \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the x-axis . . $\quad J_x = \int_C (y^2+z^2) \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the y-axis . . $\quad J_y = \int_C (x^2+z^2) \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the z-axis . . $\quad J_z = \int_C (x^2+y^2) \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the origin . . $\quad J_0 = \int_C (x^2+y^2+z^2) \cdot \rho(x,y,z)\, ds \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$.

Propose yourself a formula by which it would be possible to define and compute the moment of inertia with respect to a general line or plane in $\mathbb{E}_3$.

III.4 Line integral of a vector function (Line integral of the 2nd kind)

The notion of a vector function is explained in Section I.2.12. In this chapter we restrict ourselves to vector functions in $\mathbb{E}_2$ and $\mathbb{E}_3$. Vector functions in $\mathbb{E}_3$ will usually be written as:

$$\vec{f}(x,y,z)$$
$$= \big(U(x,y,z);\, V(x,y,z);\, W(x,y,z)\big) \,=\, U(x,y,z)\,\widehat{\imath} + V(x,y,z)\,\widehat{\jmath} + W(x,y,z)\,\widehat{k}$$

or simply
$$\vec{f} \,=\, (U;V;W) \,=\, U\,\widehat{\imath} + V\,\widehat{\jmath} + W\,\widehat{k}.$$

(U, V and W are the coordinate functions of the vector function $\vec{f}$.) We recall that instead of "vector function" we often use the name "vector field".

Vector functions in $\mathbb{E}_2$ will be written similarly, only we will always have one fewer variable and one fewer component.

We say that a vector function $\vec{f}$ in $\mathbb{E}_k$ ($k = 2$ or $k = 3$) is continuous on a set $M \subset \mathbb{E}_k$ if all its coordinate functions are continuous on M.

III.4.1. Physical motivation

Imagine that a body moves along a curve C under the action of a force $\vec{f}$. We want to compute the work W done by the force $\vec{f}$ while acting along the entire trajectory C of the body. Again we use the idea that the curve C can be divided into infinitely many "infinitesimal" segments, similarly as in Section III.2.1. A typical segment is located at $[x;y;z]$, has length ds, and the unit tangent vector to the curve C at the point $[x;y;z]$ is $\widehat{\tau}(x,y,z)$. The segment can be regarded as an "infinitesimally short" line segment, and the work dW done by the force $\vec{f}$ on it is $dW = \vec{f}(x,y,z)\cdot\widehat{\tau}(x,y,z)\;ds$. The total work of the force $\vec{f}$ along the whole curve C is then:

$$W = \int_C \vec{f}(x,y,z)\cdot\widehat{\tau}(x,y,z)\;ds.$$

The product $\vec{f}(x,y,z)\cdot\widehat{\tau}(x;y;z)$ is a scalar. Hence the integral on the right-hand side is a line integral of a scalar function, which is already known from Chapter III.2.

III.4.2. Line integral of a vector function
Let C be a simple piecewise smooth curve in $\mathbb{E}_k$ (where $k = 2$ or $k = 3$) and let $\vec{f}$ be a vector function that is defined and bounded on the curve C. We say that the vector function $\vec{f}$ is *integrable* on the curve C if the scalar function $\vec{f}\cdot\widehat{\tau}$ is integrable on C (in the sense explained in Sections III.2.2 and III.2.4). The integral $\int_C \vec{f}\cdot\widehat{\tau}\;ds$ is called the *line integral* of the vector function $\vec{f}$ along the curve C, and we denote it more briefly by $\int_C \vec{f}\cdot d\vec{s}$.

Instead of "line integral of a vector function" the name *line integral of the 2nd kind* is often used.

You can see that the line integral of a vector function $\vec{f}$ along a curve C is actually nothing new. It is just the line integral of the scalar function $\vec{f}\cdot\widehat{\tau}$, i.e. the projection of the vector function $\vec{f}$ in the tangential direction to the curve C. Remember the formula by which the line integral of a vector function is defined:

(III.4.1)
$$\int_C \vec{f}\cdot d\vec{s} \,=\, \int_C (\vec{f}\cdot\widehat{\tau})\;ds.$$

III.4.3. Physical meaning of the line integral of a vector function.

From Section III.4.1 it is clear that if $\vec{f}$ is a force, then the line integral $\int_C \vec{f} \cdot d\vec{s}$ defines the work done by the force $\vec{f}$ while acting along the path C.

III.4.4. Remark

The unit tangent vector $\widehat{\tau}$ need not exist at those points of a simple piecewise smooth curve C where the individual smooth curves that make up the piecewise smooth curve C meet. However, there are only finitely many such points, and the line integral does not depend on the behavior of the integrand at finitely many points.

III.4.5. Remark – different notations for the line integral of a vector function

The line integral of a vector function can be denoted and written in various ways. It is very important to understand these ways and always to recognize correctly what a given notation means. One way of writing it is already known–namely the integral on the left-hand side of formula (III.4.1). In this section we become acquainted with another possible notation for the line integral of a vector function.

By comparing the two integrals in (III.4.1) we obtain the formal equality $\widehat{\tau}\, ds = d\vec{s}$. The term on the right can be understood as an "infinitesimal" tangent vector to the curve C whose components are dx, dy and dz. Thus we may write:

$$\widehat{\tau}\, ds \;=\; d\vec{s} \;=\; (dx; dy; dz) \;=\; \widehat{\imath}\, dx + \widehat{\jmath}\, dy + \widehat{k}\, dz.$$

If $\vec{f} = (U; V; W)$, we can express the scalar product $\vec{f} \cdot d\vec{s}$ as

$$\vec{f} \cdot d\vec{s} \;=\; (U; V; W) \cdot (dx; dy; dz) \;=\; U\, dx + V\, dy + W\, dz$$

and the line integral of the vector function $\vec{f}$ can be written as:

(III.4.2)
$$\int_C \vec{f} \cdot d\vec{s} \;=\; \int_C \big(U\, dx + V\, dy + W\, dz\big).$$

The line integral of a "planar" vector function along a curve C in $\mathbb{E}_2$ can be written similarly. (On the right-hand side one only needs to omit $W\, dz$.)

III.4.6. Example

In accordance with the notation and writing of the integral (see Section III.4.5), the integral $\int_C (2x^2 + 3y)\, dz$ is the line integral of the vector function $(0;\, 0;\, 2x^2 + 3y)$:

$$\int_C (2x^2 + 3y)\, dz \;=\; \int_C 0 \cdot dx + 0 \cdot dy + (2x^2 + 3y)\, dz \;=\; \int_C \big(0;\, 0;\, 2x^2 + 3y\big) \cdot d\vec{s}.$$

III.4.7. Remark

The line integral of a vector function $\vec{f}$ is defined using the line integral of the scalar function $\vec{f} \cdot \widehat{\tau}$, and therefore the basic properties of the line integral of a vector function and of the line integral of a scalar function are the same. Items a), b) and c) of Section III.2.6 can simply be rewritten with the scalar function $\vec{f} \cdot \widehat{\tau}$ and we obtain valid statements about the line integral of a vector function.

However, there is one big difference between the line integral of a scalar function and the line integral of a vector function: the line integral of a vector function depends on the orientation of the curve. More precisely:

III.4.8. Theorem

If the vector function $\vec{f}$ is integrable on the curve C, then it is integrable also on the curve $-C$ and

$$\int_{-C} \vec{f} \cdot d\vec{s} = -\int_{C} \vec{f} \cdot d\vec{s}.$$

This theorem follows immediately from the definition of the line integral of a vector function. The integral $\int_C \vec{f} \cdot d\vec{s}$ is equal to the integral $\int_C \vec{f} \cdot \widehat{\tau}\ ds$, where $\widehat{\tau}$ is the unit tangent vector. This vector indicates the direction of the orientation of the curve C. If we reverse the orientation of the curve, the unit tangent vector changes its sign, which leads to a change of sign of the whole integral.

III.4.9. Computing the line integral of a vector function

The line integral of a vector function $\vec{f} = (U; V; W)$ along a simple smooth curve C can be computed using a parametrization. Suppose that P is a parametrization of the curve C defined on the interval $\langle a\,; b\rangle$. Suppose further that the curve C is oriented consistently with the parametrization P. The unit tangent vector at each point $P(t)$ of the curve C corresponding to $t \in (a; b)$ (that is, except for the endpoints of the curve C) can be expressed as $\widehat{\tau} = \dot{P}(t)/\|\dot{P}(t)\|$. Using formulas (III.4.1) and (III.2.1) (or (III.2.5)), we obtain:

$$\int_C \vec{f} \cdot d\vec{s} = \int_C \vec{f} \cdot \widehat{\tau}\, ds = \int_a^b \vec{f}(P(t)) \cdot \frac{\dot{P}(t)}{\|\dot{P}(t)\|}\, \|\dot{P}(t)\|\, dt,$$

$$\int_C \vec{f} \cdot d\vec{s} = \int_a^b \vec{f}(P(t)) \cdot \dot{P}(t)\, dt. \tag{III.4.3}$$

We obtain $\vec{f}(P(t))$ again by substituting into $\vec{f}$ (respectively into all its coordinate functions U, V and W) the expressions (III.2.4): $x = \phi(t)$, $y = \psi(t)$, $z = \vartheta(t)$. By computing the scalar product of the vector functions $\vec{f} = (U; V; W)$ and $\dot{P}(t) = \big(\dot{\phi}(t);\, \dot{\psi}(t);\, \dot{\vartheta}(t)\big)$ in the integral on the right-hand side of (III.4.3) we obtain another formula:

$$\int_C \vec{f} \cdot d\vec{s} = \int_a^b \big(U\ \dot{\phi}(t) + V\ \dot{\psi}(t) + W\ \dot{\vartheta}(t)\big)\ dt, \tag{III.4.4}$$

where $U = U\big(\phi(t), \psi(t), \vartheta(t)\big)$, $V = V\big(\phi(t), \psi(t), \vartheta(t)\big)$ and $W = W\big(\phi(t), \psi(t), \vartheta(t)\big)$. Formula (III.4.4) can also be derived formally from the integral form (III.4.2) using the substitutions $x = \phi(t)$, $y = \psi(t)$, $z = \vartheta(t)$ and the resulting replacements for dx, dy and dz: $dx = \dot{\phi}(t)\, dt$, $dy = \dot{\psi}(t)\, dt$, $dz = \dot{\vartheta}(t)\, dt$.

You will see the use of formulas (III.4.3) and (III.4.4) on concrete examples in Sections III.4.10 and III.4.12.

If the curve C is oriented oppositely to the parametrization P, then formulas (III.4.3) and (III.4.4) hold with a minus sign in front of the integral on the right-hand side.

Formula (III.4.3) is also valid for the line integral of a vector function in $\mathbb{E}_2$. Formula (III.4.4) can be easily modified (by omitting the term $W\ \dot{\vartheta}(t)$) so that it is also valid for the line integral of a vector function in $\mathbb{E}_2$.

If C is a simple piecewise smooth curve that is the union of simple smooth curves $C_1, \ldots, C_n$, then the line integral of a vector function along C can be computed by first computing the integrals along the simple smooth curves $C_1, \ldots, C_n$ (for example using parametrizations of these smooth curves) and then summing the results.

Finally, note that the line integral of a vector function can sometimes be computed using Green's theorem, Stokes' theorem (if the curve is closed), or, in the case when $\vec{f}$ is a so-called potential field, also using formula (V.1.3). Details can be found in Sections III.5.4, IV.5.10 and V.1.5.

III.4.10. Example

Determine the work done by the force $\vec{f}(x,y,z) = (y-x^2)\,\widehat{\imath} + (z-y^2)\,\widehat{\jmath} + (x-z^2)\,\widehat{k}$ while acting along the curve $C:\ P(t) = [t;\, t^2;\, t^3];\quad t \in \langle 0;1\rangle$ from the point $[0;0;0]$ to the point $[1;1;1]$.

Solution:
The curve C is given by its parametrization P. Since $i.p.\,C = [0;0;0] = P(0)$ and $t.p.\,C = [1;1;1] = P(1)$, C is oriented consistently with the parametrization P. We easily find that $\dot{P}(t) = (1;\, 2t;\, 3t^2)$. Using formula (III.4.3) we get:

$$\int_C \vec{f}\cdot d\vec{s} = \int_C \big(y-x^2;\, z-y^2;\, x-z^2\big)\cdot d\vec{s} = \int_0^1 \big(0;\, t^3-t^4;\, t-t^6\big)\cdot\big(1;\, 2t;\, 3t^2\big)\,dt =$$

$$= \int_0^1 \big(2t^4 - 2t^5 + 3t^3 - 3t^8\big)\,dt = \Big[\tfrac{2}{5}t^5 - \tfrac{2}{6}t^6 + \tfrac{3}{4}t^4 - \tfrac{3}{9}t^9\Big]_0^1 = \frac{29}{60}.$$

III.4.11. Circulation of a vector field along a closed curve
Let C be a closed curve in $\mathbb{E}_2$ or $\mathbb{E}_3$ and let $\vec{f}$ be a vector field (= vector function) defined on C. The line integral $\int_C \vec{f}\cdot d\vec{s}$ is called the *circulation* of the vector field $\vec{f}$ along the curve C. To emphasize that C is a closed curve, we often use the notation

$$\oint_C \vec{f}\cdot d\vec{s}.$$

III.4.12. Example

Compute the circulation of the vector field

$$\vec{f}(x,y) = -\frac{y}{x^2+y^2}\,\widehat{\imath} + \frac{x}{x^2+y^2}\,\widehat{\jmath}$$

along the circle $C:\ x^2+y^2 = r^2$ $(r>0)$, oriented clockwise.

Solution:
The circle C can be parametrized by

$$x = \phi(t) = r\,\cos t,\quad y = \psi(t) = r\,\sin t;\quad t\in\langle 0;2\pi\rangle. \tag{III.4.5}$$

As t runs through the interval $\langle 0;2\pi\rangle$ from 0 to 2π, the point $[x;y] = [\phi(t);\psi(t)]$ goes around the circle counterclockwise, i.e. opposite to the given orientation of C. Thus the parametrization (III.4.5) and the prescribed orientation of the curve C are inconsistent. Using formula (III.4.4) (with a "$-$" sign on the right-hand side) we get:

$$\oint_C \vec{f}\cdot d\vec{s} = \oint_C \Big(-\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy\Big) =$$

$$= -\int_0^{2\pi}\left(\Big(-\frac{r\,\sin t}{r^2}\Big)(-r\,\sin t) + \frac{r\,\cos t}{r^2}\,r\,\cos t\right)dt = -\int_0^{2\pi} dt = -2\pi.$$

III.5 Green's theorem

In this chapter we show that, under certain conditions, we can compute the circulation of a planar vector field along a closed curve C in $\mathbb{E}_2$ in a different way than by using a parametrization of the curve C. First, however, let us discuss closed curves in $\mathbb{E}_2$ a bit more carefully. We recall

the convention that by "curve" we mean a simple piecewise smooth curve and by "closed curve" a closed simple piecewise smooth curve.

The following theorem states something that is, at first glance, completely intuitive. We do not give its proof. If you were interested in the proof, you would be surprised how complicated it is.

III.5.1. Jordan theorem
Let C be a closed curve in $\mathbb{E}_2$. Then there exist two disjoint regions G_1 and G_2 in $\mathbb{E}_2$ such that C is their common boundary and

a) $\mathbb{E}_2 = G_1 \cup C \cup G_2$ *and*

b) one of the regions G_1, G_2 is bounded and the other is unbounded.

III.5.2. Interior and exterior of a closed curve in $\mathbb{E}_2$. Let C be a closed curve in $\mathbb{E}_2$ and let G_1, G_2 be the regions whose existence is guaranteed by the Jordan theorem. The bounded one of the regions G_1, G_2 is called the *interior* of the curve C and is denoted $\operatorname{Int} C$. The other region, which is unbounded, is called the *exterior* of the curve C and is denoted $\operatorname{Ext} C$.

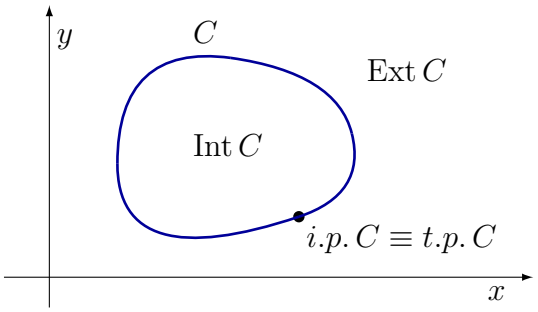


Fig. 19

III.5.3. Positive and negative orientation of a closed curve in $\mathbb{E}_2$
Let C be a closed curve in $\mathbb{E}_2$. We say that the curve C is oriented *positively* if, when traversing C in the direction of its orientation, we have the interior of the curve C on our left. (See Fig. 19.) Otherwise we say that the curve C is oriented *negatively*.

The above definition is simple and intuitive. From a logical standpoint, however, it is not correct. This is obvious. Mathematical notions must be defined precisely and their meaning must not depend on knowing where we have our left and right hands. In other words: how would you explain the definition to an intelligent being (say, from outer space) that has no hands and for which the notions of left and right are foreign? Since a logically correct definition is somewhat complicated, we do not present it in this text, and for our purposes we are satisfied with the above definition III.5.3.

The orientation of a closed curve in $\mathbb{E}_2$ can be defined in yet another simple way, again not completely logically correct: a closed curve C in $\mathbb{E}_2$ is called positively oriented if it is oriented counterclockwise, and negatively oriented if it is oriented clockwise.

The following theorem converts the line integral of a vector function along a closed planar curve into a double integral over the interior of the curve. The theorem was first published by G. Green (1793–1841) in a short book entitled *An Essay on the Application of Mathematical Analysis to Electricity and Magnetism*.

III.5.4. Green's theorem

Suppose that

a) $\vec{f} = (U; V)$ *is a vector function in a region* $G \subset \mathbb{E}_2$ *and the coordinate functions* U, V *have continuous partial derivatives in* G,

b) C *is a positively oriented closed curve in* G *such that* $\operatorname{Int} C \subset G$.

Then:

$$\oint_C \vec{f} \cdot d\vec{s} = \iint_{\operatorname{Int} C} \left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} \right) dx\, dy. \tag{III.5.1}$$

III.5.5. Remark

Using the formal equality $\vec{f} \cdot d\vec{s} = U\, dx + V\, dy$ we can also write formula (III.5.1) in the form

$$\oint_C (U\, dx + V\, dy) = \iint_{\operatorname{Int} C} \left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} \right) dx\, dy. \tag{III.5.2}$$

If the assumptions of Green's theorem are satisfied except that the curve C is negatively oriented, then formulas (III.5.1) and (III.5.2) hold with a "$-$" sign in front of the integrals on the right-hand sides.

III.5.6. Example

Compute the circulation of the vector field $\vec{f} = (-x^2 y; xy^2)$ along the positively oriented circle $x^2 + y^2 = a^2$, where $a > 0$.

Solution:
Let U and V be the coordinate functions of the vector function $\vec{f}$. Then $U(x, y) = -x^2 y$ and $V(x, y) = xy^2$. You can verify yourself that the assumptions of Green's theorem are satisfied. Therefore, by formula (III.5.1) we have:

$$\oint_C \vec{f} \cdot d\vec{s} = \oint_C (-x^2 y\, dx + xy^2\, dy) = \iint_{x^2+y^2<a^2} (x^2 + y^2)\, dx\, dy =$$

$$= {}^{1)} \int_0^{2\pi} \left(\int_0^a r^3\, dr \right) d\varphi = \frac{\pi a^4}{2}.$$

[1]) We transform the double integral into polar coordinates.

III.5.7. Computing the area of $\operatorname{Int} C$ **using a line integral**

If the curve C and the coordinate functions U and V satisfy the assumptions of Green's theorem and if moreover in $\operatorname{Int} C$

$$\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} = 1, \tag{III.5.3}$$

then from formula (III.5.2) it follows that $\oint_C (U\, dx + V\, dy) = \mu_2(\operatorname{Int} C)$.

If we choose, for example, $U = -\frac{1}{2} y$ and $V = \frac{1}{2} x$, then condition (III.5.3) is satisfied and thus we obtain the formula

$$\mu_2(\operatorname{Int} C) = \frac{1}{2} \oint_C (-y\, dx + x\, dy). \tag{III.5.4}$$

III.6 Exercises

1. Decide about the existence of the integral $\int_C ds/(x^2+y^2)$, where C is a circle with center S and radius 1. *(See Sections III.2.2 and III.2.6.)*

a) $S=[0;0]$, b) $S=[1;0]$, c) $S=[0;-2]$.

2. Compute the length of the curve C, which is given by its parametrization.

a) $P(t)=[3t;\,3t^2;\,2t^3],\ t\in\langle 0;1\rangle$,
b) $P(t)=[a\,\cos t;\,a\,\sin t;\,bt],\ t\in\langle 0;2\pi\rangle\ \ (a>0,\ b>0)$.

3. Compute the given line integrals. Which of them are line integrals of scalar functions and which are line integrals of vector functions?

a) $\displaystyle\int_C \frac{ds}{x-y}$; C is a part of the line $y=\frac{1}{2}x-2$ between the points $[0;-2]$ and $[4;0]$,

b) $\displaystyle\int_C y\ ds$; C is an arc of the parabola $y^2=2px$ between the points $[0;0]$ and $[2p;2p]$ $(p>0)$,

c) $\displaystyle\int_C xy\ ds$; C is a part of the ellipse $x^2/a^2+y^2/b^2=1$ in the second quadrant,

d) $\displaystyle\int_C \sqrt{2y}\ ds$; C is a part of the cycloid $x=a(t-\sin t),\ y=a(1-\cos t)$, corresponding to $t\in\langle 0;2\pi\rangle$,

e) $\displaystyle\int_C (x-y)\ ds$; C is the circle $x^2+y^2=2x$,

f) $\displaystyle\int_C \frac{z^2}{x^2+y^2}\ ds$; $C:\ x=a\,\cos t,\ y=a\,\sin t,\ z=at,\ t\in\langle 0;2\pi\rangle\ (a>0)$,

g) $\displaystyle\int_C xyz\ ds$; C is the intersection of the surfaces $x^2+y^2+z^2=R^2$ and $x^2+y^2=R^2/4$ in the first octant and $R>0$,

h) $\displaystyle\int_C (x+y)\ ds$; C is a quarter of the circle $x^2+y^2+z^2=R^2,\ y=x$ in the first octant,

i) $\displaystyle\int_C (x+y)\ dx$; C is the segment from the point $[a;0]$ to the point $[0;b]$,

j) $\displaystyle\int_C (x^2-y^2)\ dx$; C is an arc of the parabola $y=x^2$ from the point $[0;0]$ to the point $[2;4]$,

k) $\displaystyle\int_C \big(-x\,\cos y\,dx+y\,\sin x\,dy\big)$; C is a segment, $i.p.\ C=[0;0]$, $t.p.\ C=[\pi;\,2\pi]$,

l) $\displaystyle\int_C (y;-x)\cdot d\vec{s}$; C is the positively oriented ellipse $x^2/a^2+y^2/b^2=1$,

m) $\displaystyle\int_C \frac{y^2\,dx-x^2\,dy}{x^2+y^2}$; C is an arc of the circle $x^2+y^2=a^2$ $(a>0)$ in the first and second quadrants, from the point $[a;0]$ to the point $[-a;0]$,

n) $\int_C (2a-y;-a+y)\cdot d\vec{s}$; C: $x=a(t-\sin t)$, $y=a(1-\cos t)$, $t\in\langle 0;2\pi\rangle$; C is oriented from the point $[2\pi a;0]$ to the point $[0;0]$,

o) $\int_C \left(y^2\,dx+z^2\,dy+x^2\,dz\right)$; C is the intersection of the sphere $x^2+y^2+z^2=R^2$ with the cylinder $x^2+y^2=Rx$, $(R>0)$, $z\geq 0$, C is positively oriented when viewed from the origin,

p) $\int_C \left(2xy\,\vec{i}-x^2\,\vec{j}\right)\cdot d\vec{s}$; C is the union of segments from $[0;0]$ to $[2;0]$ and from $[2;0]$ to $[2;1]$,

q) $\int_C \left(yz\,\vec{i}+z\sqrt{R^2-y^2}\,\vec{j}+xy\,\vec{k}\right)\cdot d\vec{s}$; C: $x=R\cos t$, $y=R\sin t$, $z=at/(2\pi)$, $(a>0)$, C is oriented from the intersection with the plane $z=0$ to the intersection with the plane $z=a$,

r) $\int_C (1-x^2)y\;dx+x(1+y^2)\;dy$; C is the boundary of the square $\langle 0;2\rangle\times\langle 0;2\rangle$, positively oriented,

s) $\int_C (xy+x+y)\,dx+(xy+x-y)\,dy$; C is the ellipse $9x^2+36x+4y^2=0$, negatively oriented,

t) $\int_C (x+y)\,dx-2x\,dy$; C is the boundary of the triangle with sides on the lines $x=0$, $y=0$, $x+y=5$, negatively oriented,

u) $\int_C (dx/y-dy/x)$; C is the boundary of the triangle with vertices $[1;1]$, $[2;1]$ and $[2;2]$, positively oriented,

v) $\int_C \left(2(x^2+y^2)\,\vec{i}+(x+y)^2\,\vec{j}\right)\cdot d\vec{s}$; C is the boundary of the triangle with vertices $[1;1]$, $[2;2]$, $[1;3]$, positively oriented.

4. Compute the work of the force $\vec{f}$ acting along the curve C. *(See Section III.4.3.)*

a) $\vec{f}=(x-y;x)$, C is the boundary of the square with vertices $[-2;-2]$, $[1;-2]$, $[1;1]$, $[-2;1]$, oriented clockwise,

b) $\vec{f}=(x+y;2x)$, C is the circle $x^2+y^2=4$, positively oriented,

c) $\vec{f}=(y;2)$, C is a closed curve consisting of coordinate semiaxes and a quarter of the ellipse $x=2\cos t$, $y=\sin t$, $t\in\langle 0;2\pi\rangle$ in the first quadrant, positively oriented.

5. C_1 is the segment from $[0;0]$ to $[1;1]$, C_2 is the arc of the parabola $y=x^2$ from $[0;0]$ to $[1;1]$, $I_1=\int_{C_1}(x+y)^2\,dx-(x-y)^2\,dy$, $I_2=\int_{C_2}(x+y)^2\,dx-(x-y)^2\,dy$. Using Green's theorem compute the difference I_1-I_2.

6. Using a line integral, compute the area of the interior of the closed curve consisting of an arc of the cycloid $x=a(t-\sin t)$, $y=a(1-\cos t)$, $t\in\langle 0;2\pi\rangle$ and the segment joining the points $[0;0]$ and $[2\pi a;0]$. *(See Section III.5.7.)*

7. Using a line integral derive a formula for the area of the interior of the ellipse $x^2/a^2+y^2/b^2=1$. *(See Section III.5.7.)*

8. Using a line integral compute the area of the interior of the closed curve whose equation in polar coordinates is $r = a(1 - \cos\varphi)$, where $a > 0$ (the so-called "cardioid"). *(See Section III.5.7.)*

9. Using a line integral compute the area of the interior of the so-called astroid, whose equation is $x^{2/3} + y^{2/3} = a^{2/3}$ $(a > 0)$. (You may use the parametric equations $x = a\cos^3 t$, $y = a\sin^3 t$, $t \in \langle 0; 2\pi\rangle$.) *(See Section III.5.7.)*

10. Compute the circulation of the vector field $\vec{f}$ along the closed curve C. If possible, use Green's theorem. *(See Sections III.4.11 and III.5.4.)*

a) $\vec{f}(x,y) = \big(e^x \sin y - y^2;\, e^x \cos y - 1\big)$, $C = C_1 \cup C_2$, $C_1 = \big\{[x;y] \in \mathbb{E}_2 : x^2 + y^2 + 2x = 0,\ y \le 0\big\}$, $C_2 = \big\{[x;y] \in \mathbb{E}_2 :\ -1 \le x \le 0,\ y = 0\big\}$, C is positively oriented,

b) $\vec{f}(x,y) = (x+y)\,\vec{\imath} + (y-x)\,\vec{\jmath}$, $C = \big\{[x;y] \in \mathbb{E}_2 :\ x^2/a^2 + y^2/b^2 = 1\big\}$, C is negatively oriented,

c) $\vec{f}(x,y) = (x^2; y^2)$, C is the boundary of the triangle with vertices $A = [1;1]$, $B = [2;1]$, $D = [2;3]$, positively oriented.

CHAPTER IV.

Surface integrals

IV.1 Simple surfaces 97
IV.2 Surface integral of a scalar function (Surface integral of the 1st kind) 103
IV.3 Some physical applications of the surface integral of a scalar function 107
IV.4 Surface integral of a vector function (Surface integral of the 2nd kind) 108
IV.5 Operators of divergence and curl.
Gauss–Ostrogradsky (divergence) theorem and Stokes' theorem 112
IV.6 Exercises 117

IV.1 Simple surfaces

In $\mathbb{E}_3$ there exists a large number of objects that can be called surfaces. It is not easy to select from them those surfaces that can be described in a reasonably clear way, that meet the needs of technical applications, and on which a reasonably clear theory of surface integrals can be built. In this text we restrict ourselves to two types of surfaces: the so-called "simple smooth surface" and the "simple piecewise smooth surface".

IV.1.1. Simple smooth surface – motivation and notation

Behind the definition of a simple surface lies the following idea: B is a subset of $\mathbb{E}_2$, bounded by a closed curve Γ. You cut out B from the plane $\mathbb{E}_2$ and move it somewhere into $\mathbb{E}_3$. Elastic deformations are allowed, but they must not destroy smoothness. This means that the cut-out object may be stretched, compressed, and bent in various directions, but it must not be torn nor may two distinct points be glued together. The result is a simple smooth surface in $\mathbb{E}_3$.

This idea can be modeled mathematically. The described procedure moves each point $[u;v] \in B$ to a position $P(u,v) = \big[\phi(u,v);\, \psi(u,v);\, \vartheta(u,v)\big]$ in $\mathbb{E}_3$. Thus P is a mapping from B into $\mathbb{E}_3$. The requirement that the deformation of B be elastic and preserve smoothness leads to the condition that P is a continuous mapping and has continuous partial derivatives on a sufficiently large subset of B. The requirement that two distinct points originally in B cannot be glued together leads to the condition that the mapping P is injective on B. The resulting simple smooth surface is the range of the mapping P, which we shall call a "parametrization" of the simple smooth surface. The functions

$$x = \phi(u,v), \quad y = \psi(u,v), \quad z = \vartheta(u,v)$$

are called the *coordinate functions* of the mapping P. The partial derivatives of P with respect to the variables u, v will be denoted by P_u, P_v and will be regarded as vectors. Thus we may write:

$$P_u(u,v) = \left(\frac{\partial\phi(u,v)}{\partial u}; \frac{\partial\psi(u,v)}{\partial u}; \frac{\partial\vartheta(u,v)}{\partial u}\right) \quad \text{or briefly} \quad P_u = \left(\frac{\partial\phi}{\partial u}; \frac{\partial\psi}{\partial u}; \frac{\partial\vartheta}{\partial u}\right),$$

$$P_v(u,v) = \left(\frac{\partial\phi(u,v)}{\partial v}; \frac{\partial\psi(u,v)}{\partial v}; \frac{\partial\vartheta(u,v)}{\partial v}\right) \quad \text{or briefly} \quad P_v = \left(\frac{\partial\phi}{\partial v}; \frac{\partial\psi}{\partial v}; \frac{\partial\vartheta}{\partial v}\right).$$

The mapping P is considered continuous if all its coordinate functions are continuous. Similarly, we say that P has continuous partial derivatives if all coordinate functions have continuous partial derivatives.

If there is no danger of confusion with the notation of coordinate axes, we may denote the coordinate functions $\phi(u,v)$, $\psi(u,v)$, $\vartheta(u,v)$ also by $x(u,v)$, $y(u,v)$, $z(u,v)$.

The vector product of the vectors P_u and P_v is denoted by $P_u \times P_v$. We recall that

$$P_u \times P_v = \begin{vmatrix} \widehat{i}, & \widehat{j}, & \widehat{k} \\ \dfrac{\partial\phi}{\partial u}, & \dfrac{\partial\psi}{\partial u}, & \dfrac{\partial\vartheta}{\partial u} \\ \dfrac{\partial\phi}{\partial v}, & \dfrac{\partial\psi}{\partial v}, & \dfrac{\partial\vartheta}{\partial v} \end{vmatrix} = \left(\frac{\partial\psi}{\partial u}\frac{\partial\vartheta}{\partial v} - \frac{\partial\vartheta}{\partial u}\frac{\partial\psi}{\partial v}; \ \frac{\partial\vartheta}{\partial u}\frac{\partial\phi}{\partial v} - \frac{\partial\phi}{\partial u}\frac{\partial\vartheta}{\partial v}; \ \frac{\partial\phi}{\partial u}\frac{\partial\psi}{\partial v} - \frac{\partial\psi}{\partial u}\frac{\partial\phi}{\partial v} \right).$$

IV.1.2. Simple smooth surface – definition
Let Γ be a closed simple p.w. smooth curve in $\mathbb{E}_2$ and $B = \Gamma \cup \operatorname{Int}\Gamma$. Let P be a continuous mapping from B into $\mathbb{E}_3$. Assume that

a) P is injective on B,

b) P has continuous and bounded partial derivatives P_u and P_v on $B\backslash K$, where K is at most a finite set of points lying on the boundary Γ of the set B,

c) $P_u \times P_v \neq \vec{0}$ on $B\backslash K$.

The set of all points $P(u,v)$ for $[u;v] \in B$ (i.e., the range of the mapping P) is then called a *simple smooth surface* in $\mathbb{E}_k$. The mapping P is called a *parametrization*.

Simple smooth surfaces will usually be denoted by Greek letters, for example σ, σ_1, σ_2, κ, etc.

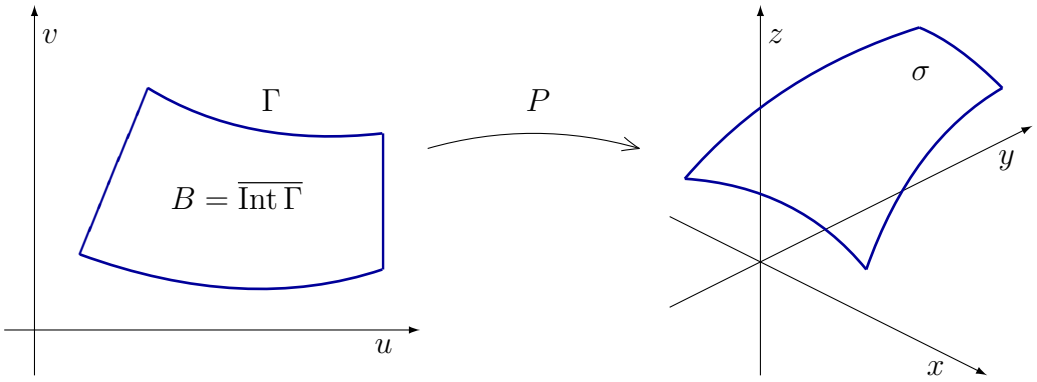


Fig. 20

Similarly to a simple smooth curve, every simple smooth surface also has infinitely many parametrizations. (One particular surface in $\mathbb{E}_3$ can evidently be obtained, while respecting the given rules, in infinitely many ways, i.e., by cutting out a part of $\mathbb{E}_2$, moving it, elastically deforming it, etc.)

A simple p.w. smooth curve in $\mathbb{E}_3$ which is the image of the boundary Γ of the set B under the mapping P is called the *boundary (edge)* of the simple smooth surface.

The definition of a simple smooth surface is not too complicated. However, we pay for this by the fact that some "nice" smooth surfaces are not "covered" by this definition, i.e., these surfaces are not simple smooth surfaces. For example, the spherical surface $x^2 + y^2 + z^2 = R^2$ or the cylindrical surface $x^2 + y^2 = R^2$, $0 \leq z \leq h$ are not simple smooth surfaces. Attempts to modify the above definition so that, for instance, it would also include the spherical surface, lead elsewhere to such an increase of complications that it is not worth it. This is a

consequence of the fact that the three-dimensional Euclidean space $\mathbb{E}_3$ allows a great variety of sets and objects, and as soon as we "loosen" the definition somewhere so that it includes the spherical surface, various strange and complicated objects immediately penetrate among the simple smooth surfaces, which elsewhere bring confusion into the theory or make it immensely harder. Nevertheless, you will soon see that both the spherical surface and a bounded cylindrical surface can be regarded as simple p. w. smooth surfaces, and these surfaces will therefore belong to the class of surfaces that we shall deal with and over which we shall integrate.

IV.1.3. Orientation of a simple smooth surface. Normal vector

Let us assume that P is a parametrization of a simple smooth surface σ, defined on a set $B \subset \mathbb{E}_2$. Let $X = P(u,v)$ for $[u;v] \in B\backslash K$ (see Definition IV.1.2). The vectors $P_u(u,v)$ and $P_v(u,v)$ are tangent to the surface σ at the point X and, due to condition c) in Definition IV.1.2, are linearly independent. Their vector product is perpendicular to both of them, hence it is also perpendicular to the surface σ. Dividing the vector product by its magnitude, we obtain a unit vector which is perpendicular to the surface σ at the point X.

We can *orient* the surface σ by defining on the surface a *normal vector* $\widehat{n}$ (i.e., a unit vector perpendicular to the surface σ which determines the orientation of the surface σ) either by the equation

$$\widehat{n} = \frac{P_u(u,v) \times P_v(u,v)}{\|P_u(u,v) \times P_v(u,v)\|} \quad \text{for all } [u;v] \in B\backslash K, \tag{IV.1.1}$$

or by the equation

$$\widehat{n} = -\frac{P_u(u,v) \times P_v(u,v)}{\|P_u(u,v) \times P_v(u,v)\|} \quad \text{for all } [u;v] \in B\backslash K. \tag{IV.1.2}$$

If the normal vector $\widehat{n}$ is given by equation (IV.1.1), we say that the simple smooth surface σ is oriented *consistently with the parametrization P*. In the opposite case, when the vector $\widehat{n}$ is defined by equation (IV.1.2), we say that the simple smooth surface σ is oriented *inconsistently with the parametrization P*.

The above method of orienting the surface σ guarantees that the normal vector $\widehat{n}$ changes continuously when moving along σ. Thus it cannot happen that $\widehat{n}$ points to different sides of the surface σ at different points.

IV.1.4. Relation between the orientation of a simple smooth surface and its boundary

We say that a simple smooth surface σ is *oriented consistently with its boundary* C if, when traversing the curve C in the direction of its orientation, the side of the surface σ toward which the normal vector $\widehat{n}$ points lies on our left.

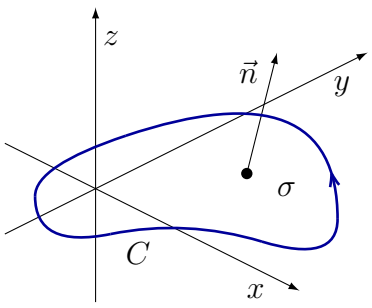


Fig. 21

It is clear that this definition is not logically correct. The reasons are the same as for Definition III.5.3. Nevertheless, the definition is illustrative and at our level it cannot lead to misunderstanding.

IV.1.5. Example

σ is a part of the conical surface $z = \sqrt{x^2+y^2}$, corresponding to $x \geq 0$ and $x^2+y^2 \leq 4$. It is oriented "upward", i.e., the third coordinate of the normal vector is positive. Show that σ is a simple smooth surface, find its parametrization, decide whether σ is oriented consistently with the chosen parametrization, and define the orientation of the boundary of σ so that the orientations of the boundary and the surface σ are consistent.

S o l u t i o n :
For a parametrization of σ we use the equation $z = \sqrt{x^2+y^2}$: We set

$$P: \quad x = \phi(u,v) = u, \quad y = \psi(u,v) = v, \quad z = \vartheta(u,v) = \sqrt{u^2+v^2}$$

for $[u;v] \in B$, where the set B is defined by the conditions $u \geq 0$ and $u^2+v^2 \leq 4$:

$$B = \{[u;v] \in \mathbb{E}_2 : \ u \geq 0, \ u^2+v^2 \leq 4\}.$$

One can verify that the mapping P has all the properties required in Definition IV.1.2. (Here $K = \{[0;0]\}$.) Therefore σ is a simple smooth surface and P is its parametrization. The partial derivatives P_u and P_v are:

$$P_u(u,v) = \left(1;\ 0;\ \frac{u}{\sqrt{u^2+v^2}}\right), \qquad P_v(u,v) = \left(0;\ 1;\ \frac{v}{\sqrt{u^2+v^2}}\right)$$

and their vector product is

$$P_u \times P_v = \left(-\frac{u}{\sqrt{u^2+v^2}};\ -\frac{v}{\sqrt{u^2+v^2}};\ 1\right).$$

You can easily verify that the vector $P_u \times P_v$ has length $\sqrt{2}$. Thus the vector $(P_u \times P_v)/\sqrt{2}$ is unit and perpendicular to σ. Its third coordinate is positive, therefore this vector coincides with the normal vector $\widehat{n}$. Hence the simple smooth surface σ is oriented consistently with the parametrization P.

The orientation of the boundary of the surface σ, which is the closed curve C, consistent with the orientation of the surface σ, is indicated in Fig. 22. For example, the unit tangent vector to C at the point $X = [2;0;2]$ is $\widehat{\tau} = (0;1;0)$.

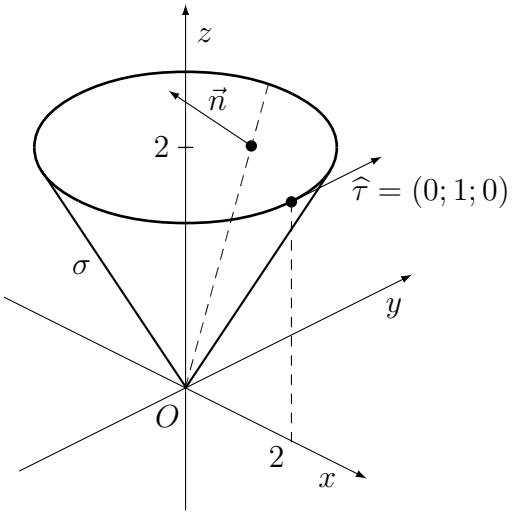


Fig. 22

IV.1.6. Simple piecewise smooth surface composed of two simple smooth surfaces
Suppose that σ_1 and σ_2 are two simple smooth surfaces which are either both oriented consistently with their boundaries C_1 and C_2, or are both oriented inconsistently with their boundaries. Suppose that

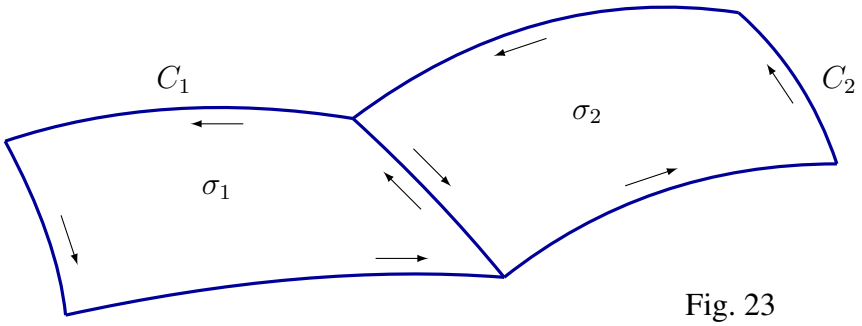


Fig. 23

a) $\sigma_1 \cap \sigma_2 = C_1 \cap C_2$ and this intersection forms one or more simple smooth curves,

b) the orientations of the curves C_1 and C_2 are opposite at all their common points (i.e., on $C_1 \cap C_2$).

The union $\sigma = \sigma_1 \cup \sigma_2$ is then called a *simple piecewise smooth surface* in $\mathbb{E}_3$, composed of two simple smooth surfaces σ_1 and σ_2.

The orientation of the surface σ is defined by the orientation of its parts σ_1 and σ_2.

The boundary of the surface σ is the closure of the set $(C_1 \cup C_2) \setminus (C_1 \cap C_2)$. (See Fig. 23.) It may be either empty, or it may be a closed simple p.w. smooth curve, or it may consist of several (a finite number of) closed simple p.w. smooth curves.

The notion of "consistent (or inconsistent) orientation of a surface and its boundary", defined in Section IV.1.4 for a simple smooth surface, can naturally be extended also to simple p.w. smooth surfaces.

IV.1.7. Simple piecewise smooth surface composed of several simple smooth surfaces
Suppose that σ_1 and σ_2 are simple smooth surfaces from the preceding Section IV.1.6. If, step by step, we attach to their union, while respecting the same rules, further simple smooth surfaces $\sigma_3, \sigma_4, \ldots, \sigma_n$, we obtain a simple smooth surface composed of n simple smooth surfaces σ_1, $\sigma_2, \ldots, \sigma_n$. (See Fig. 24.)

We shall usually shorten the name simple piecewise smooth surface to "simple p.w. smooth surface". If we use in the sequel the term "surface" without any adjective, we shall always mean a simple p.w. smooth surface.

A surface which differs from σ only by its orientation will be denoted by $-\sigma$.

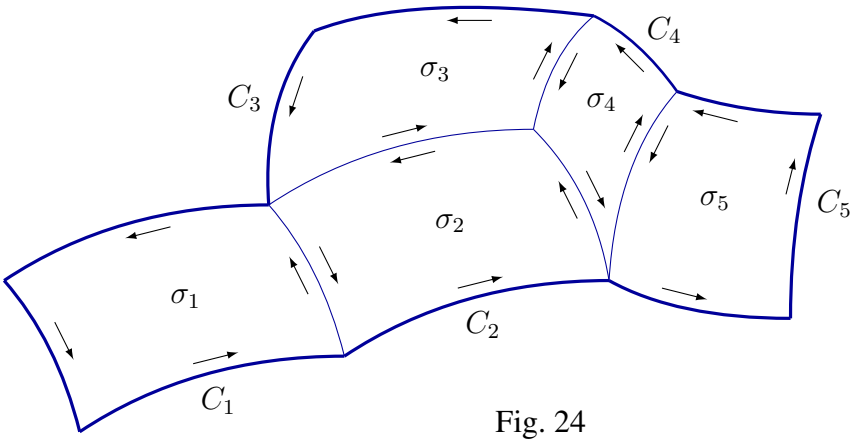


Fig. 24

IV.1.8. Closed simple p.w. smooth surface
A simple p.w. smooth surface whose boundary is the empty set is called *closed*.

IV.1.9. Example

A cylindrical surface composed of two simple smooth surfaces $\sigma_1 : x^2 + y^2 = 1,\ y \geq 0,\ 0 \leq z \leq 1$ and $\sigma_2 : x^2 + y^2 = 1;\ y \leq 0,\ 0 \leq z \leq 1$ is a simple p.w. smooth surface. (See Fig. 25.)

IV.1.10. Example

A conical surface composed of two simple smooth surfaces $\sigma_1 : x^2+y^2 = z^2,\ y \geq 0,\ 0 \leq z \leq 1$ and $\sigma_2 : x^2 + y^2 = z^2;\ y \leq 0,\ 0 \leq z \leq 1$ is a simple p.w. smooth surface. (See Fig. 26.)

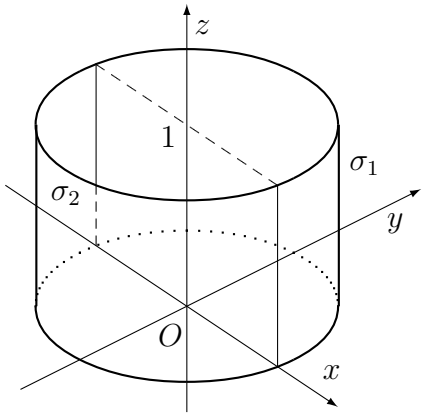


Fig. 25

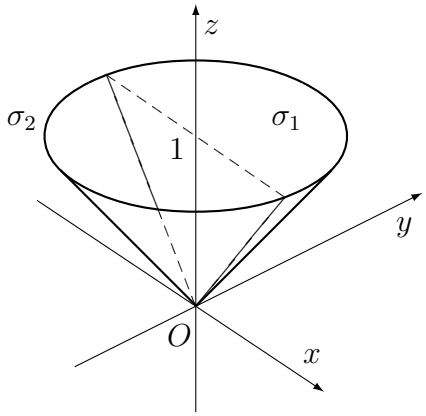


Fig. 26

IV.1.11. Example

A spherical surface composed of two simple smooth surfaces $\sigma_1 : x^2+y^2+z^2 = 4,\ z \geq 0$ and $\sigma_2 : x^2+y^2+z^2 = 4,\ z \leq 0,$ is a closed simple p.w. smooth surface.

Further examples of closed simple p.w. smooth surfaces are the surface of a cube, the surface of a tetrahedron, etc.

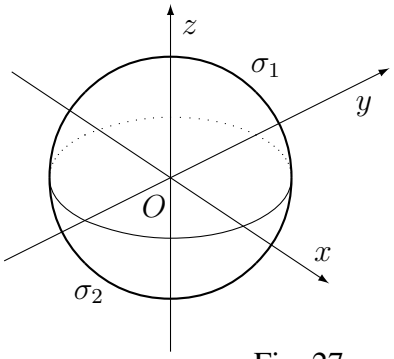


Fig. 27

IV.2 Surface integral of a scalar function (Surface integral of the 1st kind)

IV.2.1. Physical motivation

Let us imagine that a bent metal sheet has the shape of a simple smooth surface σ in $\mathbb{E}_3$ and its surface density (i.e., mass per unit area of the surface) is ρ, which in general is a function of three variables x, y, and z. We want to compute the total mass of the metal sheet.

Assume that P is a parametrization of the surface σ, defined on a set $B \subset \mathbb{E}_2$. Imagine that we partition B into infinitely many "infinitesimal" rectangles of the form $\langle u; u+du\rangle \times \langle v; v+dv\rangle$. The mapping P maps each of these rectangles to an "infinitesimal" parallelogram on the surface σ with vertices $A_1 = P(u,v)$, $A_2 = P(u+du,v) \doteq P(u,v) + P_u(u,v)\cdot du$, $A_3 = P(u+du, v+dv) \doteq P(u,v) + P_u(u,v)\cdot du + P_v(u,v)\cdot dv$ and $A_4 = P(u, v+dv) \doteq P(u,v) + P_v(u,v)\cdot dv$. (See Fig. 28.) Its area is $dp = \|A_2 - A_1\| \cdot \|A_4 - A_1\| \cdot \sin\alpha = \|(A_2 - A_1) \times (A_4 - A_1)\|$. Substituting for the points A_1, A_2, A_4 we obtain: $dp = \|P_u(u,v) \times P_v(u,v)\|\, du\, dv$. Hence the mass of the parallelogram $A_1A_2A_3A_4$ is $dm = \rho(A_1)\cdot dp = \rho(P(u,v))\, \|P_u(u,v) \times P_v(u,v)\|\, du\, dv$. We obtain the total mass of the metal sheet as the sum of the contributions dm over the entire surface σ:

$$m = \iint_\sigma dm = \iint_\sigma \rho(x,y,z)\, dp = \iint_B \rho(P(u,v))\, \|P_u(u,v) \times P_v(u,v)\|\, du\, dv.$$

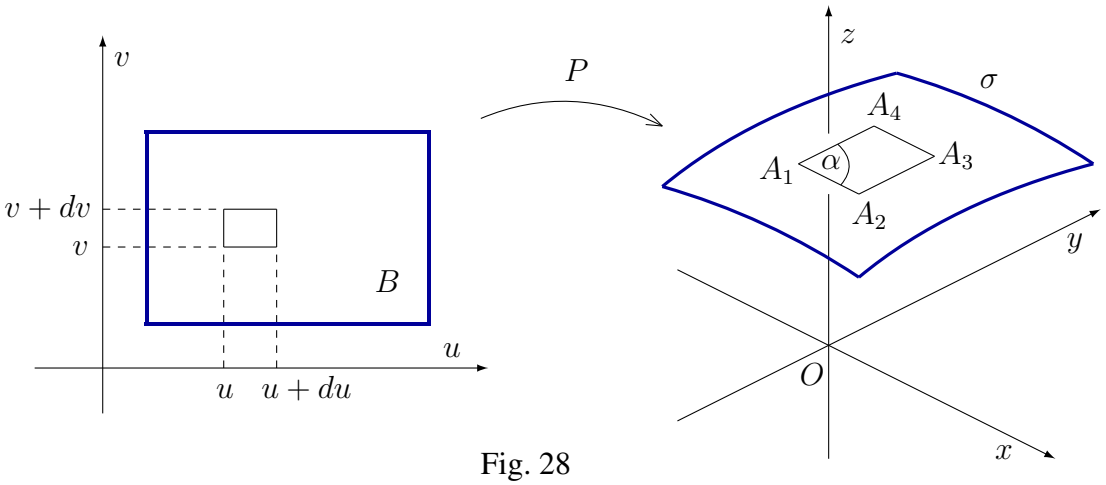


Fig. 28

IV.2.2. Surface integral of a scalar function on a simple smooth surface

Let σ be a simple smooth surface in $\mathbb{E}_3$ and let P be its parametrization, defined on a set $B \subset \mathbb{E}_2$. Let f be a function which is defined and bounded on the surface σ. If the double integral $\iint_B f(P(u,v))\, \|P_u(u,v) \times P_v(u,v)\|\, du\, dv$ exists, then we say that the function f is *integrable* on the surface σ. The *surface integral* of the function f on the surface σ is denoted by $\iint_\sigma f\, dp$ and is defined by the equation

$$\iint_\sigma f\, dp = \iint_B f(P(u,v))\, \|P_u(u,v) \times P_v(u,v)\|\, du\, dv. \tag{IV.2.1}$$

We shall return to the computation of the integral on the right in Section IV.2.7, and we shall also illustrate the computation by examples in Sections IV.2.8–IV.2.10.

The statements "the function f is integrable on the surface σ" and "the surface integral of f over σ exists" have the same meaning.

In the literature, the notation for a surface integral often uses dS or $d\sigma$ instead of dp. In this text we shall use dp.

IV.2.3. Remark – independence of the choice of parametrization
The integrability of the function f on a simple smooth surface σ and the surface integral of f over σ are defined using a parametrization P of the surface σ. However, a simple smooth surface σ has infinitely many different parametrizations (see Section IV.1.2). Nevertheless, one can prove that both the existence and the value of the surface integral of f over σ are independent of the choice of a parametrization of the surface σ.

IV.2.4. Surface integral of a scalar function on a simple p.w. smooth surface
Let σ be a simple p.w. smooth surface in $\mathbb{E}_3$, composed of simple smooth surfaces $\sigma_1, \ldots, \sigma_n$ according to the rules described in Sections IV.1.5 and IV.1.6. Let f be a function which is bounded on the surface σ. If f is an integrable function on each of the surfaces $\sigma_1, \ldots, \sigma_n$, then we say that it is *integrable* on the surface σ. The *surface integral* of the function f over the surface σ is then defined by the equation

$$\iint_\sigma f \, dp = \sum_{i=1}^{n} \iint_{\sigma_i} f \, dp. \tag{IV.2.2}$$

(The surface integrals on the right-hand side are surface integrals over simple smooth surfaces, which are already known from Definition IV.2.2.)

Instead of "surface integral of a scalar function" we often use the name *surface integral of the 1st kind*. If we want to emphasize on which variables the function f depends, we may write $\int_\sigma f(x,y,z)\, dp$ instead of $\iint_\sigma f \, dp$.

IV.2.5. Area of a surface
If σ is a simple p.w. smooth surface, then the number which is the value of the integral $\iint_\sigma dp$ is called the *area* of the surface σ.

In the case when σ is a simple smooth surface and P is its parametrization defined on a set $B \subset \mathbb{E}_2$, the area of the surface σ equals

$$A(\sigma) = \iint_\sigma 1 \, dp = \iint_B \|P_u(u,v) \times P_v(u,v)\| \, du\, dv. \tag{IV.2.3}$$

IV.2.6. Some important properties of the surface integral of a scalar function
The surface integral of a scalar function is defined by means of a double integral. Its basic properties can therefore be derived from the properties of the double integral. Let us mention some of them:

a) **Sufficient condition for the existence of the surface integral of a scalar function**
If f is a continuous function on the surface σ, then it is integrable on σ (i.e., the surface integral $\iint_\sigma f \, dp$ exists).

b) **Linearity of the surface integral**
If the functions f and g are integrable on the surface σ and $\alpha \in \mathbb{R}$, then the functions $f+g$ and αf are also integrable on σ and

$$\iint_\sigma (f+g)\, dp = \iint_\sigma f \, dp + \iint_\sigma g \, dp,$$
$$\iint_\sigma \alpha \cdot f \, dp = \alpha \cdot \iint_\sigma f \, dp.$$

c) *If f is an integrable function on the surface σ and a bounded function g differs from f on σ at most in finitely many points or curves, then g is also integrable on the surface σ and*

$$\iint_{\sigma} g\, dp \;=\; \iint_{\sigma} f\, dp.$$

d) *If the function f is integrable on the surface σ, then it is also integrable on $-\sigma$ and*

$$\iint_{-\sigma} f\, dp \;=\; \iint_{\sigma} f\, dp.$$

Statement a) can be generalized: *If σ is a simple p.w. smooth surface and the function f is continuous on each smooth part of the surface σ, then f is integrable on σ.*

Statement d) shows that neither the existence nor the value of the surface integral of a scalar function depends on the orientation of the surface.

IV.2.7. Computation of the surface integral of a scalar function

The surface integral of a function f over a simple smooth surface σ can be computed using a parametrization $P = [\phi; \psi; \vartheta]$ of the surface σ and formula (IV.2.1). In the function f we substitute, for the variables x, y, z,

(IV.2.4) $$x \;=\; \phi(u,v), \quad y \;=\; \psi(u,v), \quad z \;=\; \vartheta(u,v),$$

for dp we substitute

(IV.2.5) $$dp \;=\; \|P_u(u,v) \times P_v(u,v)\|\, du\, dv$$

and integrate with respect to u and v over the set $B \subset \mathbb{E}_2$ on which the parametrization P is defined.

The surface integral of a function f over a simple p.w. smooth surface composed of simple smooth surfaces $\sigma_1, \ldots, \sigma_n$ is computed by first computing the integrals over the surfaces σ_1, $\ldots$, σ_n and then adding the results. (See formula (IV.2.2).)

IV.2.8. Example

Let us compute the surface integral of the function $f(x,y,z) = x + 2y$ over the simple smooth surface $\sigma \,:\; x + y + z = 1,\; x \geq 0,\; y \geq 0,\; z \geq 0$.

S o l u t i o n :
The surface σ can be parametrized by the mapping

$$P(u,v) \,:\quad x = u,\;\; y = v,\;\; z = 1 - u - v\,; \quad [u;v] \in B,$$

where $B = \{[u;v] \in \mathbb{E}_2 \,:\; 0 \leq u \leq 1,\; 0 \leq v \leq 1 - u\}$. (Verify yourself that the mapping P really is a parametrization, i.e., that it has the properties required in Section IV.1.2.) We easily compute that $P_u = (1;\,0;\,-1)$, $P_v = (0;\,1;\,-1)$, $P_u \times P_v = (1;1;1)$ and $\|P_u \times P_v\| = \sqrt{3}$. Using formula (IV.2.1) we obtain:

$$\iint_{\sigma} (x + 2y)\, dp \;=\; \iint_{B} (u + 2v)\,\sqrt{3}\, du\, dv \;=\; \sqrt{3} \int_0^1 \int_0^{1-u} (u + 2v)\, dv\, du =$$

$$= \sqrt{3} \int_0^1 (1 - u)\, du \;=\; \sqrt{3}/2.$$

IV.2.9. Example

Let us compute the surface integral of the function $g(x,y,z) = xyz$ over the surface σ, which is the surface of the cube $\langle 0;1\rangle \times \langle 0;1\rangle \times \langle 0;1\rangle$.

S o l u t i o n :
The surface of the given cube consists of six squares. Three of them lie in the planes xy, xz, and yz, hence on them $xyz = 0$ and the surface integral of the function g over them is zero. Let us denote the remaining squares by σ_1, σ_2, and σ_3:

$$\begin{aligned}\sigma_1 &: \quad x = 1, \quad 0 \le y \le 1, \quad 0 \le z \le 1,\\ \sigma_2 &: \quad y = 1, \quad 0 \le x \le 1, \quad 0 \le z \le 1,\\ \sigma_3 &: \quad z = 1, \quad 0 \le x \le 1, \quad 0 \le y \le 1.\end{aligned}$$

A natural parametrization of the simple smooth surface σ_1 is:

$$P(u,v) : \quad x = 1, \;\; y = u, \;\; z = v\,; \quad [u;v] \in B = \langle 0;1\rangle \times \langle 0;1\rangle.$$

From this it is clear that $P_u = (0;1;0)$, $P_v = (0;0;1)$, $P_u \times P_v = (1;0;0)$ and $\|P_u \times P_v\| = 1$. Using formula (IV.2.1) we obtain:

$$\iint_{\sigma_1} xyz\; dp \;=\; \int_0^1 \int_0^1 uv\; du\, dv \;=\; \frac{1}{4}.$$

In the same way we can compute that the integrals over the surfaces σ_2 and σ_3 are also equal to $\frac{1}{4}$. (This also follows from the symmetries of the integrand and of the surfaces σ_1, σ_2, and σ_3.) Therefore:

$$\iint_{\sigma} xyz\; dp \;=\; \iint_{\sigma_1} xyz\; dp \;+\; \iint_{\sigma_2} xyz\; dp \;+\; \iint_{\sigma_3} xyz\; dp \;=\; \frac{3}{4}.$$

IV.2.10. Example: computation of a surface integral over a spherical surface
Although the spherical surface $\sigma_R : \; x^2 + y^2 + z^2 = R^2$ is not a simple smooth surface, there exists a mapping P from a closed set $B \subset \mathbb{E}_2$ onto the surface σ_R which has all the properties of a parametrization (listed in Section IV.1.2) with one exception: it is not injective on the whole set B. This mapping arises from the detextbookion of the surface in spherical coordinates; the parameters are the angles φ and ϑ (which we now denote by u and v) and it is defined by the equations:

$$\begin{aligned} x &= \phi(u,v) = R\,\cos u\,\cos v,\\ y &= \psi(u,v) = R\,\sin u\,\cos v,\\ z &= \vartheta(u,v) = R\,\sin v\end{aligned}$$

for $[u;v] \in B = \langle 0;2\pi\rangle \times \langle -\pi/2;\pi/2\rangle$. The mapping P is not injective because of its behavior on the boundary of the set B: for example, it assigns to different points $[u;v] = [u_1;\pi/2]$ and $[u;v] = [u_2;\pi/2]$ (for $u_1 \neq u_2$) on the boundary of B the same point $[x;y;z] = [0;0;R]$ on the surface σ_R. One can verify that inside B the mapping P is injective. The boundary of the set B has two-dimensional measure zero, and the behavior of the integrand (as long as it remains bounded) on a subset of B of measure zero does not affect the value of the double integral over B. Therefore, the mapping P can be used to compute the surface integral over the spherical surface σ_R in the same way as if P were a genuine parametrization. (We often indeed speak of the mapping P, not quite precisely, as a parametrization of the surface σ_R.)

If, for example, $f(x,y,z) = x^2 + y^2$, then using formula (IV.2.1) we obtain

$$\iint_{\sigma_R} f(x,y,z)\; dp \;=\; \iint_B R^2\,\cos^2 v\;\|P_u(u,v) \times P_v(u,v)\|\; du\, dv.$$

The vectors P_u, P_v, $P_u \times P_v$ and the number $\|P_u \times P_v\|$ are:

$$P_u(u,v) = \big(-R\,\sin\,u\,\cos\,v,\ R\,\cos\,u\,\cos\,v,\ 0\big),$$
$$P_v(u,v) = \big(-R\,\cos\,u\,\sin\,v,\ -R\,\sin\,u\,\sin\,v,\ R\,\cos\,v\big),$$
$$P_u(u,v)\times P_v(u,v) = \big(R^2\,\cos\,u\,\cos^2 v,\ R^2\,\sin\,u\,\cos^2 v,\ R^2\,\sin\,v\,\cos\,v\big),$$
$$\|P_u(u,v)\times P_v(u,v)\| = R^2\,\cos\,v.$$

Substituting into the above integral and decomposing the double integral into an iterated integral using Fubini's theorem II.3.2 we obtain:

$$\iint_B R^2\,\cos^2 v\,\|P_u(u,v)\times P_v(u,v)\|\,du\,dv \;=\; \int_0^{2\pi}\left(\int_{-\pi/2}^{\pi/2} R^4\,\cos^3 v\,dv\right)du \;=\; \tfrac{8}{3}\,\pi\,R^4.$$

IV.2.11. Remark on computing a surface integral using a parametrization

The procedure explained in Example IV.2.10 can also be used when computing surface integrals over other simple p.w. smooth surfaces, such as an ellipsoid, a bounded conical surface, a bounded cylindrical surface, etc. This means that in the computation we may use as a parametrization a mapping whose injectivity is violated only on a set of measure zero in its domain (usually on the boundary).

For example, if σ is a bounded conical surface $x^2+y^2=z^2,\ z\in\langle 0;2\rangle$, then the mapping

$$P:\quad x=\phi(u,v)=u\,\cos\,v,\ \ y=\psi(u,v)=u\,\sin\,v,\ \ z=\vartheta(u,v)=u$$

(defined for $u\in\langle 0;2\rangle,\ v\in\langle 0;2\pi\rangle$) has all the required properties of a parametrization of the surface σ (see Section IV.1.2) with the exception that it is injective only in the interior of its domain, i.e., in the open rectangle $(0;2\pi)\times(0;2)$ and not in $\langle 0;2\pi\rangle\times\langle 0;2\rangle$. (For example, to two different points $[u;v]=[u;0]$ and $[u;v]=[u;2\pi]$ (for $u\in\langle 0;2\rangle$) the mapping P assigns the same point $[x;y;z]=[u;0;u]\in\sigma$.) Nevertheless, the mapping P can be used in the computation of the surface integral over σ in the same way as if P were an actual parametrization of σ.

IV.3 Some physical applications of the surface integral of a scalar function

Assume that on a surface σ in $\mathbb{E}_3$ there is distributed a mass with surface density $\rho(x,y,z)$ (given in $\mathrm{kg}\cdot\mathrm{m}^{-2}$). Using the surface integral we can define and compute some basic mechanical characteristics of the surface σ:

total mass ..... $m = \iint_\sigma \rho(x,y,z)\,dp$ [kg],

static moment

with respect to the plane xy ..... $m_{xy} = \iint_\sigma z\cdot\rho(x,y,z)\,dp$ [kg · m],

with respect to the plane xz ..... $m_{xz} = \iint_\sigma y\cdot\rho(x,y,z)\,dp$ [kg · m],

with respect to the plane yz ..... $m_{yz} = \iint_\sigma x\cdot\rho(x,y,z)\,dp$ [kg · m],

coordinates of the center of mass $x_T=\dfrac{m_{yz}}{m},\quad y_T=\dfrac{m_{xz}}{m},\quad z_T=\dfrac{m_{xy}}{m}$ [m],

moment of inertia

with respect to the plane xy $\quad J_{xy} = \iint_\sigma z^2 \cdot \rho(x,y,z)\ dp \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the plane xz $\quad J_{xz} = \iint_\sigma y^2 \cdot \rho(x,y,z)\ dp \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the plane yz $\quad J_{yz} = \iint_\sigma x^2 \cdot \rho(x,y,z)\ dp \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the x-axis .. $\quad J_x = \iint_\sigma (y^2+z^2) \cdot \rho(x,y,z)\ dp \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the y-axis .. $\quad J_y = \iint_\sigma (x^2+z^2) \cdot \rho(x,y,z)\ dp \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the z-axis .. $\quad J_z = \iint_\sigma (x^2+y^2) \cdot \rho(x,y,z)\ dp \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$,

with respect to the origin .. $\quad J_0 = \iint_\sigma (x^2+y^2+z^2) \cdot \rho(x,y,z)\ dp \qquad [\mathrm{kg}\cdot\mathrm{m}^2]$.

Propose yourselves a formula by which it would be possible to define and compute the moment of inertia with respect to an arbitrary line or plane in $\mathbb{E}_3$.

IV.4 Surface integral of a vector function (Surface integral of the 2nd kind)

IV.4.1. Physical motivation

Imagine that σ is a surface in a flow field of an incompressible fluid and we wish to express and compute the flux of the fluid through the surface σ. By flux we mean the volume of fluid that passes through the surface in the chosen direction per unit time. Let us assume that the flow is steady, the fluid velocity is $\vec{v}(x,y,z)$ and $\widehat{n}(x,y,z)$ is the normal vector on the surface σ at the point $[x;y;z]$. (Recall that by a normal vector we mean a unit vector perpendicular to the surface σ, indicating the direction of the orientation of the surface.) The flux of the fluid through an "infinitesimal" part of the surface σ, located at the point $[x;y;z]$ and having area dp, is $\|\vec{v}(x,y,z)\| \cos\alpha\ dp = \vec{v}(x,y,z)\cdot\widehat{n}(x,y,z)\ dp$, where α is the angle between the velocity vector $\vec{v}$ and the normal vector $\widehat{n}$ at the point $[x;y;z]$. The flux of the fluid through the whole surface σ is then $\iint_\sigma \vec{v}(x,y,z)\cdot\widehat{n}(x,y,z)\ dp$.

The same reasoning can be used if we want to express, for example, the flux of a magnetic field through a chosen surface.

IV.4.2. Surface integral of a vector function

Let σ be a simple p.w. smooth surface in $\mathbb{E}_3$ and let $\vec{f}$ be a vector function which is defined and bounded on the surface σ. We say that the vector function $\vec{f}$ is *integrable* on the surface σ if the scalar function $\vec{f}\cdot\widehat{n}$ is integrable on σ (in the sense explained in Sections IV.2.2 and IV.2.4). The integral $\iint_\sigma \vec{f}\cdot\widehat{n}\ dp$ is called the *surface integral of the vector function* $\vec{f}$ over the surface σ and we denote it more briefly by
$\iint_\sigma \vec{f}\cdot d\vec{p}$.

The surface integral of a vector function (= a vector field) $\vec{f}$ expresses the *flux of the vector field* $\vec{f}$ through the surface under consideration. This integral is often called the *surface integral of the 2nd kind*.

From the definition of the surface integral of a vector function $\vec{f}$ it is clear that this integral is in fact the surface integral of the scalar function $\vec{f} \cdot \widehat{n}$, which is the projection of the vector function $\vec{f}$ in the direction perpendicular to the surface σ. The surface integral of a vector function can therefore be expressed by the formula

$$\iint_\sigma \vec{f} \cdot d\vec{p} = \iint_\sigma (\vec{f} \cdot \widehat{n})\, dp. \tag{IV.4.1}$$

IV.4.3. Remark

The normal vector $\widehat{n}$ need not exist at those points of a simple p.w. smooth surface σ where the individual smooth surfaces from which the p.w. smooth surface σ is composed meet. Such points form at most finitely many simple p.w. smooth curves, and neither the existence nor the value of the surface integral depends on the behavior of the integrand (provided it is bounded) on finitely many curves. (Compare with the remarks in Sections II.2.2 and III.4.3.)

IV.4.4. Remark – various notations for the surface integral of a vector function

The surface integral of a vector function can be denoted and written in various ways. It is necessary to understand these ways and always be able to recognize what each notation means precisely. One way of writing it is the integral on the left-hand side of formula (IV.4.1). In this section we shall become familiar with further ways of writing the surface integral of a vector function.

Assume that the vector function $\vec{f}$ has coordinate functions U, V, and W. Then the integrals

$$\iint_\sigma \vec{f} \cdot d\vec{p}, \quad \iint_\sigma \vec{f} \cdot \widehat{n}\, dp, \quad \iint_\sigma (U; V; W) \cdot d\vec{p}, \quad \iint_\sigma (U; V; W) \cdot \widehat{n}\, dp,$$

$$\iint_\sigma (U\,\widehat{\imath} + V\widehat{\jmath} + W\widehat{k}) \cdot d\vec{p}, \quad \iint_\sigma (U\,\widehat{\imath} + V\widehat{\jmath} + W\widehat{k}) \cdot \widehat{n}\, dp$$

all have the same meaning.

Another notation that sometimes appears in the specialist literature is based on expressing the vector $d\vec{p}$ (formally equal to $\widehat{n}\, dp$) in the form

$$d\vec{p} = \big(dy\, dz;\ dx\, dz;\ dx\, dy\big) = \widehat{\imath}\, dy\, dz + \widehat{\jmath}\, dx\, dz + \widehat{k}\, dx\, dy.$$

Substituting into the integral $\iint_\sigma (U; V; W) \cdot d\vec{p}$ and computing the dot product $(U; V; W) \cdot d\vec{p}$ we obtain:

$$\iint_\sigma \vec{f} \cdot d\vec{p} = \iint_\sigma (U; V; W) \cdot d\vec{p} = \iint_\sigma U\, dy\, dz + V\, dx\, dz + W\, dx\, dy.$$

However, this notation for the surface integral of a vector function can rather easily lead to misunderstanding, and therefore we shall not use it in this text.

IV.4.5. Remark

The surface integral of a vector function $\vec{f}$ is defined by means of the surface integral of the scalar function $\vec{f} \cdot \widehat{n}$ and therefore the basic properties of the surface integral of a vector function and the surface integral of a scalar function are the same. We can therefore rewrite items a), b), and c) of Section IV.2.6 with the scalar function $\vec{f} \cdot \widehat{n}$ and obtain valid statements about the surface integral of a vector function.

However, there is one big difference between the surface integral of a scalar function and the surface integral of a vector function: The surface integral of a vector function depends on the orientation of the surface.

More precisely:

IV.4.6. Theorem

If the vector function $\vec{f}$ is integrable on the surface σ, then it is also integrable on the surface $-\sigma$ and

$$\iint_{-\sigma} \vec{f} \cdot d\vec{p} = -\int_{\sigma} \vec{f} \cdot d\vec{p}.$$

This theorem follows from the definition of the surface integral of a vector function. The integral $\iint_\sigma \vec{f} \cdot d\vec{p}$ is equal to the integral $\iint_\sigma \vec{f} \cdot \widehat{n}\ dp$, where $\widehat{n}$ is the normal vector. This vector indicates the direction in which the surface σ is oriented. If we change the orientation of the surface, the normal vector changes its sign, which leads to a change of sign of the whole integral.

IV.4.7. Computation of the surface integral of a vector function

The surface integral of a vector function $\vec{f} = (U; V; W)$ over a simple smooth surface σ can be computed using a parametrization of the surface σ. Assume that P is such a parametrization, defined on a set $B \subset \mathbb{E}_2$. Assume further that the surface σ is oriented consistently with the parametrization P. The normal vector at each point $P(u, v)$ of the surface σ corresponding to $[u; v] \in B \setminus K$ (i.e., except for some points on the boundary of the surface σ) can be expressed as: $\widehat{n} = (P_u \times P_v)/\|P_u \times P_v\|$. (See Section IV.1.3, formula (IV.1.1).) Using formulas (IV.4.1) and (IV.2.1) we obtain:

$$\iint_\sigma \vec{f} \cdot d\vec{p} = \iint_\sigma \vec{f} \cdot \widehat{n}\ dp =$$

$$= \iint_B \vec{f}(P(u,v)) \cdot \frac{P_u(u,v) \times P_v(u,v)}{\|P_u(u,v) \times P_v(u,v)\|}\ \|P_u(u,v) \times P_v(u,v)\|\ du\, dv,$$

$$\iint_\sigma \vec{f} \cdot d\vec{p} = \iint_B \vec{f}(P(u,v)) \cdot \big(P_u(u,v) \times P_v(u,v)\big)\ du\, dv. \tag{IV.4.2}$$

If the surface σ is not oriented in accordance with the parametrization P, then formula (IV.4.2) holds with a "$-$" sign in front of the integral on the right-hand side.

You will see the use of formula (IV.4.2) in a concrete example in Section IV.5.8.

If σ is a simple p.w. smooth surface composed of simple smooth surfaces $\sigma_1, \ldots, \sigma_n$, then we can compute the surface integral of a vector function over σ by first computing the integrals over the simple smooth surfaces $\sigma_1, \ldots, \sigma_n$ (for example using parametrizations of these smooth surfaces), and then summing the results.

When computing the surface integral of a vector function over some simple p.w. smooth surfaces, such as a spherical surface, an ellipsoid, a bounded conical surface, a bounded cylindrical surface, etc., we may also use a mapping which has all the properties of a parametrization except that it is not injective on the whole of its domain B, but only in the interior of B. (See Sections IV.2.10 and IV.2.11, where this procedure is explained and used when computing the surface integral of a scalar function.) With such a mapping we can proceed in the same way as with an actual parametrization, and the integral can then be computed using formula (IV.4.2).

Further possibilities for computing the surface integral of a vector function are, under certain circumstances, provided by the Gauss–Ostrogradsky theorem (if the surface σ is closed) and Stokes' theorem. You will encounter these theorems in Chapter IV.5.

IV.4.8. Example

Let us compute the flux of the vector field $\vec{f}(x,y,z) = yz\,\widehat{\jmath} + z^2\,\widehat{k}$ through the surface σ, which is a part of half of the cylindrical surface $y^2 + z^2 = 4,\ z \geq 0,$ cut off by the planes $x = -1$ and $x = 1$. The surface σ is oriented by a normal vector $\widehat{n}$ which at the point $[0;0;2]$ has the form $\widehat{n} = \widehat{k}$.

S o l u t i o n :
The simple smooth surface σ can be parametrized by the mapping

$$P(u,v)\ :\quad x = u,\ \ y = 2\,\cos\,v,\ \ z = 2\,\sin\,v\,;\quad [u;v] \in B = \langle -1;1\rangle \times \langle 0;\pi\rangle.$$

(Verify yourself that P indeed has all the properties of a parametrization listed in Section IV.1.2.) We easily find that $P_u = (1;\,0;\,0),\ P_v = (0;\,-2\,\sin\,v;\,2\,\cos\,v)$ and $P_u \times P_v = (0;\,-2\,\cos\,v;\,-2\,\sin\,v)$. The unit vector perpendicular to the surface σ at the point $[0;0;2]$ (corresponding to $u = 0$ and $v = \pi/2$), expressed using the parametrization P, is

$$\left.\frac{P_u \times P_v}{\|P_u \times P_v\|}\right|_{[u;v]=[0;\,\pi/2]} = (0;\,0;-1) = -\widehat{k}.$$

It is clear that this vector is opposite to the given normal vector $\widehat{n}$ at the point $[0;0;2]$. Therefore the prescribed orientation of the surface is opposite to the chosen parametrization P. This means that we must use formula (IV.4.2) with a "$-$" sign in front of the integral on the right-hand side:

$$\begin{aligned}\iint_\sigma (yz\,\widehat{\jmath} + z^2\,\widehat{k})\cdot d\vec{p} &= -\iint_B \left(4\,\sin\,v\,\cos\,v\,\widehat{\jmath} + 4\,\sin^2 v\,\widehat{k}\right)\cdot(P_u \times P_v)\,du\,dv = \\ &= -\int_{-1}^{1}\int_0^{\pi} \left(0;\,4\,\sin\,v\,\cos\,v;\,4\,\sin^2 v\right)\cdot\left(0;\,-2\,\cos\,v;\,-2\,\sin\,v\right)\,dv\,du = \\ &= -\int_{-1}^{1}\int_0^{\pi}\left(-8\,\sin\,v\,\cos^2 v - 8\,\sin^3 v\right)\,dv\,du = 32.\end{aligned}$$

IV.5 Operators of divergence and curl. Gauss–Ostrogradsky (divergence) theorem and Stokes' theorem

IV.5.1. The nabla operator

We denote by ∇ a vector operator called the *nabla operator*, whose components are, in order, the partial derivatives with respect to x, y, and z:

$$\nabla = \left(\frac{\partial}{\partial x};\ \frac{\partial}{\partial y};\ \frac{\partial}{\partial z}\right).$$

The nabla operator is used to denote various vector and scalar fields. From Section I.5.3 we know that the gradient of a scalar field φ in a region $D \subset \mathbb{E}_3$ is a vector field whose coordinate functions are the partial derivatives of φ with respect to x, y, and z. Using the nabla operator, the gradient of φ can be written as:

$$\operatorname{grad}\varphi = \nabla\varphi = \left(\frac{\partial\varphi}{\partial x};\ \frac{\partial\varphi}{\partial y};\ \frac{\partial\varphi}{\partial z}\right).$$

IV.5.2. Divergence of a vector field

Let $\vec{f} = (U; V; W)$ be a vector field in a region $D \subset \mathbb{E}_3$. The *divergence* of $\vec{f}$ is the scalar field denoted by $\operatorname{div}\vec{f}$ and defined by the equation

$$\operatorname{div}\vec{f} = \frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} + \frac{\partial W}{\partial z}$$

(at all points $[x; y; z] \in D$ where the right-hand side makes sense). Using the nabla operator we can write the divergence as: $\operatorname{div}\vec{f} = \nabla \cdot \vec{f}$.

IV.5.3. Curl of a vector field

Let $\vec{f} = (U; V; W)$ be a vector field in a region $D \subset \mathbb{E}_3$. The *curl* of $\vec{f}$ is the vector field denoted by $\operatorname{curl}\vec{f}$ and defined by the equation

$$\operatorname{curl}\vec{f} = \left(\frac{\partial W}{\partial y} - \frac{\partial V}{\partial z};\ \frac{\partial U}{\partial z} - \frac{\partial W}{\partial x};\ \frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right)$$

(at all points $[x; y; z] \in D$ where the right-hand side makes sense). Using the nabla operator, the curl can be expressed as:

$$\operatorname{curl}\vec{f} = \nabla \times \vec{f} = \begin{vmatrix} \widehat{i}, & \widehat{j}, & \widehat{k} \\ \dfrac{\partial}{\partial x}, & \dfrac{\partial}{\partial y}, & \dfrac{\partial}{\partial z} \\ U, & V, & W \end{vmatrix}.$$

(The determinant on the right is not a determinant in exactly the sense in which we know it from Mathematics I. Nevertheless, we may view it as a useful scheme providing a recipe for computing the curl.)

IV.5.4. Remark

We see that the gradient maps a scalar field to a vector field, the divergence maps a vector field to a scalar field, and the curl maps a vector field to a vector field.

The gradient, divergence, and curl are important in theories of various types of fields (velocity fields of a flowing fluid, temperature fields, gravitational fields, electromagnetic fields, etc.). Their properties and the relations among them are therefore interesting not only from the standpoint of applied mathematics itself, but also from the standpoint of many other disciplines. A detailed study would go beyond the scope of this text; nevertheless, we will show here two formulas whose validity can easily be verified by computation:

If φ is a scalar field and $\vec{f}$ is a vector field in a region $D \subset \mathbb{E}_3$ and if both fields have second partial derivatives in D, then

$$\operatorname{curl} \operatorname{grad} \varphi = \vec{0}, \tag{IV.5.1}$$

$$\operatorname{div} \operatorname{curl} \vec{f} = 0. \tag{IV.5.2}$$

In Section I.5.7 the geometric meaning of the gradient of a scalar function φ at the point $[x; y; z]$ was explained: $\operatorname{grad} \phi(x, y, z)$ is a vector indicating the direction of the greatest increase of the values of the function ϕ at the point $[x; y; z]$. We shall become familiar with the physical meaning of the operators divergence and curl in Sections IV.5.9 and IV.5.13.

From Section III.5.1 you know the two-dimensional version of Jordan's theorem. The following section is devoted to the three-dimensional version of this theorem. The theorem again asserts something that is, at first sight, clear and understandable. The proof, however, would again be surprisingly long and complicated.

IV.5.5. Jordan's theorem

Let σ be a closed simple p.w. smooth surface in $\mathbb{E}_3$. Then in $\mathbb{E}_3$ there exist two disjoint regions G_1 and G_2 whose common boundary is the surface σ, and furthermore:

a) $\mathbb{E}_3 = G_1 \cup \sigma \cup G_2$,

b) one of the regions G_1, G_2 is bounded and the other is unbounded.

IV.5.6. Interior and exterior of a closed surface

Let σ be a closed simple p.w. smooth surface in $\mathbb{E}_3$ and let G_1, G_2 be the regions whose existence follows from Jordan's theorem. The one of the regions G_1, G_2 that is bounded is called the *interior* of the surface σ and is denoted by $\operatorname{Int} \sigma$. The other region, which is unbounded, is called the *exterior* of the surface σ and is denoted by $\operatorname{Ext} \sigma$.

We say that a closed surface σ is oriented outward (toward its exterior) if its normal vector points, at all points of the surface σ where it exists, into the exterior of σ. Otherwise we say that the closed surface is oriented toward its interior.

The following theorem shows that, under certain conditions, the flux of a vector field through a closed surface equals the integral of the divergence of the vector field over the interior of the surface. The theorem was proved independently by C. F. Gauss (1777–1855) and M. V. Ostrogradsky (1801–1862).

IV.5.7. Gauss–Ostrogradsky (divergence) theorem

Assume that

a) the vector function $\vec{f}$ has continuous partial derivatives in a region $D \subset \mathbb{E}_3$,

b) σ is a closed p. w. smooth surface in D, oriented outward, and such that $\operatorname{Int}\sigma \subset D$.

Then:

(IV.5.3) $$\iint_{\sigma} \vec{f}\cdot d\vec{p} \;=\; \iiint_{\operatorname{Int}\sigma} \operatorname{div}\vec{f}\; dx\,dy\,dz.$$

IV.5.8. Example

Let us compute the flux of the vector field $\vec{f}(x,y,z) = xy\,\widehat{\imath} + yz\,\widehat{\jmath} + xz\,\widehat{k}$ through the surface σ of the cube $\langle 0;1\rangle \times \langle 0;1\rangle \times \langle 0;1\rangle$, oriented outward.

S o l u t i o n :
The coordinate functions of $\vec{f}$ are continuously differentiable in $\mathbb{E}_3$ and σ is a closed simple p. w. smooth surface in $\mathbb{E}_3$, oriented outward. By the Gauss–Ostrogradsky theorem we have:

$$\iint_{\sigma} \vec{f}\cdot d\vec{p} = \iiint_{\operatorname{Int}\sigma} \operatorname{div}\vec{f}\; dx\,dy\,dz = \int_0^1\int_0^1\int_0^1 \left(\frac{\partial(xy)}{\partial x} + \frac{\partial(yz)}{\partial y} + \frac{\partial(xz)}{\partial z}\right) dx\,dy\,dz =$$
$$= \int_0^1\int_0^1\int_0^1 (y+z+x)\; dx\,dy\,dz \;=\; \frac{3}{2}.$$

IV.5.9. Physical meaning of divergence

Assume that the vector field $\vec{f}$ has continuous partial derivatives in a region $D \subset \mathbb{E}_3$ and let A be a point in D. Let σ_r be the spherical surface with center A and radius r, oriented outward. Then

$$\operatorname{div}\vec{f}(A) \;=\; \frac{\operatorname{div}\vec{f}(A)}{\mu_3(\operatorname{Int}\sigma_r)} \iiint_{\operatorname{Int}\sigma_r} dx\,dy\,dz \;=\; \frac{1}{\frac{4}{3}\pi r^3} \iiint_{\operatorname{Int}\sigma_r} \operatorname{div}\vec{f}(A)\; dx\,dy\,dz =$$
$$= \lim_{r\to 0+} \frac{1}{\frac{4}{3}\,\pi r^3} \iiint_{\operatorname{Int}\sigma_r} \operatorname{div}\vec{f}(x,y,z)\; dx\,dy\,dz \;=\; \lim_{r\to 0+} \frac{1}{\frac{4}{3}\,\pi r^3} \iint_{\sigma_r} \vec{f}(x,y,z)\cdot d\vec{p}.$$

If the field $\vec{f}$ has a source at the point A, then the flux of the field $\vec{f}$ through the surface σ_r, i.e., the integral $\iint_{\sigma_r} \vec{f}\cdot d\vec{p}$, is for positive and sufficiently small r also positive. The limit as $r \to 0+$ of this integral, divided by the volume of $\operatorname{Int}\sigma_r$, expresses the intensity of the source at the point A.

Thus, at points where $\operatorname{div}\vec{f} > 0$, the field $\vec{f}$ has sources of positive intensity (so-called springs). Conversely, at points where $\operatorname{div}\vec{f} < 0$, the field $\vec{f}$ has sources of negative intensity (so-called sinks or holes).

This has a number of physical consequences. For example, in the flow field of an incompressible fluid, due to the law of conservation of mass the fluid cannot disappear or be created anywhere, and due to its incompressibility it cannot accumulate or expand anywhere. Therefore the velocity field has no sources of positive or negative intensity. Hence the vector function $\vec{v}$, describing the flow velocity, satisfies at all points of the flow field the so-called *continuity equation*

$$\operatorname{div}\vec{v} = 0,$$

which is one of the fundamental equations of the mechanics of incompressible fluids.

The following theorem shows the relation between the surface integral of the curl of a vector function $\vec{f}$ over a surface whose boundary is a closed curve C and the line integral of the function $\vec{f}$ along the curve C. The theorem was discovered and proved by the English mathematician G. G. Stokes (1819–1903). Note that Stokes' theorem generalizes Green's theorem to the case of a space curve C and a space vector function $\vec{f}$.

IV.5.10. Stokes' theorem

Assume that

a) the vector function $\vec{f}$ has continuous partial derivatives in a region $D \subset \mathbb{E}_3$,

b) σ is a simple p.w. smooth surface in D whose boundary is a simple closed p.w. smooth curve C,

c) the surface σ is oriented consistently with its boundary (edge) C.

Then:

(IV.5.4) $$\iint_\sigma \operatorname{curl} \vec{f} \cdot d\vec{p} \;=\; \oint_C \vec{f} \cdot d\vec{s}.$$

IV.5.11.* Remark

Stokes' theorem can be extended to cases where the surface σ is closed (i.e., its boundary is the empty set) or where the boundary of the surface σ is the union of n closed simple p.w. smooth curves $C_1, \ldots, C_n$, and σ is oriented consistently with its boundary. In the first case we obtain

$$\iint_\sigma \operatorname{curl} \vec{f} \cdot d\vec{p} \;=\; 0$$

and in the second case it holds that

$$\iint_\sigma \operatorname{curl} \vec{f} \cdot d\vec{p} \;=\; \sum_{k=1}^{n} \oint_{C_k} \vec{f} \cdot d\vec{s}.$$

IV.5.12. Example

Let us compute the line integral $\oint_C \vec{f} \cdot d\vec{s}$, where $\vec{f}(x,y,z) = xz\,\widehat{\imath} + xy\,\widehat{\jmath} + 3xz\,\widehat{k}$ and the curve C is the boundary of the surface σ, which is part of the plane $2x + y + z = 2$ in the first octant. C is oriented counterclockwise when viewed from above.

S o l u t i o n :
The coefficients in the equation of the plane whose part is the surface σ form the coordinates of a vector perpendicular to σ: $(2;1;1)$. If you draw a picture, you will see that the surface σ (the triangle) is oriented consistently with its boundary C if the normal vector $\widehat{n}$ has the same direction as the vector $(2;1;1)$. Since the normal vector is unit, we have $\widehat{n} = (2;1;1)/\|(2;1;1)\| = (2;1;1)/\sqrt{6}$.

The coordinate functions of the vector field $\vec{f}$ are continuously differentiable in the whole space $\mathbb{E}_3$ and $\operatorname{curl} \vec{f} = (0;\, x - 3z;\, y)$. Therefore, by Stokes' theorem we have:

$$\oint_C \vec{f} \cdot d\vec{s} \;=\; \iint_\sigma \operatorname{curl} \vec{f} \cdot d\vec{p} \;=\; \iint_\sigma \big(0;\, x - 3z;\, y\big) \cdot d\vec{p}.$$

The surface σ can be parametrized by the mapping

$$P(u,v) \;:\quad x = u,\;\; y = v,\;\; z = 2 - 2u - v\,; \quad [u;v] \in B,$$

where $B = \{[u; v] \in \mathbb{E}_2 : \ 0 \le u \le 1, \ 0 \le v \le 2 - 2u\}$. We easily find that $P_u = (1;\ 0; -2)$, $P_v = (0;\ 1; -1)$ and $P_u \times P_v = (2; 1; 1)$. Since the vector $P_u \times P_v$ is oriented to the same side of the surface σ as the normal vector $\widehat{n}$, the surface σ is oriented consistently with the parametrization P. Using formula (IV.4.2) we obtain:

$$\iint_\sigma \big(0;\ x - 3z;\ y\big) \cdot d\vec{p} = \iint_B \big(0;\ u - 6 + 6u + 3v;\ v\big) \cdot (2; 1; 1)\, du\, dv =$$

$$= \int_0^1 \int_0^{2-2u} (\,7u + 4v - 6\,)\, dv\, du \ = \ -1.$$

IV.5.13. Physical meaning of curl

Assume that the vector function $\vec{f}$ has continuous partial derivatives in a region $D \subset \mathbb{E}_3$, $A \in D$, and $\widehat{n}$ is a unit vector. Let σ_r denote a disk with center at the point A, radius r, and normal vector $\widehat{n}$. Let C_r denote the circle which is the boundary (edge) of the surface σ_r and which is oriented consistently with σ_r. Then:

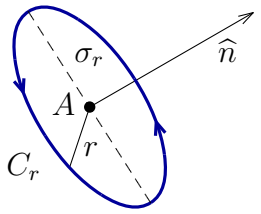


Fig. 29

$$\operatorname{curl} \vec{f}(A) \cdot \widehat{n} \ = \ \frac{\operatorname{curl} \vec{f}(A) \cdot \widehat{n}}{A(\sigma_r)} \iint_{\sigma_r} dp \ = \ \frac{1}{\pi r^2} \iint_{\sigma_r} \operatorname{curl} \vec{f}(A) \cdot \widehat{n}\, dp =$$

$$= \lim_{r \to 0+} \frac{1}{\pi r^2} \iint_{\sigma_r} \operatorname{curl} \vec{f}(x, y, z) \ \cdot \ d\vec{p} = \lim_{r \to 0+} \frac{1}{\pi r^2} \oint_{C_r} \vec{f}(x, y, z) \cdot d\vec{s}.$$

$\operatorname{curl} \vec{f}(A)$ is therefore a vector whose dot product with any unit vector $\widehat{n}$ expresses the degree of vorticity of the field $\vec{f}$ in the plane perpendicular to $\widehat{n}$. (See Fig. 29.) Since the dot product is largest when $\widehat{n}$ has the same direction as $\operatorname{curl} \vec{f}(A)$, the field $\vec{f}$ has the greatest vorticity at the point A in the plane that is perpendicular to the vector $\operatorname{curl} \vec{f}(A)$.

IV.6 Exercises

1. $\sigma = \{[x;y;z] \in \mathbb{E}_3 : \ x^2+y^2+z = 4,\ x \geq 0,\ y \geq 0,\ z \geq 0\}$, σ is oriented so that the normal vector $\widehat{n}$ at each point of σ satisfies the inequality $\widehat{n} \cdot \widehat{\imath} \geq 0$.

a) Verify that the mapping $P(u,v) = [2\cos u;\ 2\sin u;\ v]$ for $[u;v] \in B = \langle -\pi/2; \pi/2\rangle \times \langle 1;4\rangle$ is a parametrization of σ (i.e., that it has all the properties listed in Section IV.1.2). Decide whether σ is oriented consistently with the parametrization P. *(See Sections IV.1.2 and IV.1.3.)*

b) Show that the mapping $Q(u,v) = [\sqrt{4-u^2};\ u;\ v]$ for $[u;v] \in B = \langle -2;2\rangle \times \langle 1;4\rangle$ is not a parametrization of the surface σ. *(See Section IV.1.2.)*

2. σ is the upper hemisphere $\{[x;y;z] \in \mathbb{E}_3 : \ x^2+y^2+z^2 = a^2,\ z \geq 0\}$ $(a>0)$, oriented by the normal vector $\widehat{n} = (n_1;n_2;n_3)$ such that $n_3 \geq 0$. The set B is: $B = \{[u;v] \in \mathbb{E}_2 : \ u^2+v^2 \leq a^2\}$.

a) Show that the mapping

$$P(u,v) = \left[\frac{2a^2u}{a^2+u^2+v^2};\ \frac{2a^2v}{a^2+u^2+v^2};\ \frac{2a^3}{a^2+u^2+v^2} - a\right];\quad [u;v] \in B$$

is a parametrization of the surface σ (i.e., it has all the properties listed in Section IV.1.2). Decide whether σ is oriented consistently with the parametrization P. *(See Sections IV.1.2 and IV.1.3.)*

b) Show that the mapping $Q(u,v) = [u;\ v;\ \sqrt{a^2-u^2-v^2}]$, $[u;v] \in B$, is not a parametrization of the surface σ. *(See Section IV.1.2.)*

3. σ is a simple smooth surface oriented by the normal vector $\widehat{n}$. Find its parametrization, verify the properties listed in Section IV.1.2, and decide whether σ is oriented consistently with the chosen parametrization. *(See Sections IV.1.2 and IV.1.3.)*

a) σ is the triangle with vertices $A = [1;-1;2]$, $B = [2;1;3]$, $C = [-1;2;4]$, $\widehat{n} \cdot \widehat{\jmath} < 0$,

b) $\sigma = \{[x;y;z] \in \mathbb{E}_3 : \ x^2+y^2 = 4,\ x \geq 0,\ 0 \leq z \leq 4\}$, $\widehat{n} = (1;0;0)$ at the point $P = [2;0;2]$,

c) $\sigma = \{[x;y;z] \in \mathbb{E}_3 : \ x^2+y^2 = z,\ y \geq 0,\ z \leq 1\}$, $\widehat{n} = (0;0;-1)$ at the point $P = [0;0;0]$,

d) σ is the parallelogram with vertices $A = [1;1;1]$, $B = [1;4;4]$, $C = [0;5;6]$, $D = [0;2;3]$, $\widehat{n} \cdot \widehat{k} > 0$,

e) σ is the disk in the plane $x = 2$ with center at $[2;-1;3]$ and radius $r = 4$, $\widehat{n} = (-1;0;0)$ at every point of σ,

f) $\sigma = \{[x;y;z] \in \mathbb{E}_3 : \ x^2+y^2+z^2 = 4;\ z \geq \sqrt{2}\}$, $\widehat{n} = (0;0;1)$ at the point $P = [0;0;2]$,

g) $\sigma = \{[x;y;z] \in \mathbb{E}_3 : \ xy - z = 0,\ x^2+y^2 \leq a^2\}$ $(a>0)$, $\widehat{n} \cdot \widehat{k} > 0$.

4. Verify that the set $\sigma = \{X \in \mathbb{E}_3 : \ X = P(u,v),\ [u;v] \in B\}$ is a simple smooth surface in $\mathbb{E}_3$ and that P is its parametrization. *(See Section IV.1.2.)*

a) $P(u,v) = [u;\ 4u^2+9v^2;\ v]$, $B = \{[u;v] \in \mathbb{E}_2 : \ u^2/9 + v^2/4 \leq 1\}$,

b) $P(u,v) = [u;\ v;\ 4-u-v]$, $B = \{[u;v] \in \mathbb{E}_2 : \ u \geq 0,\ v \geq 0,\ u+v \leq 4\}$,

c) $P(u,v) = \big[3\cos u \cos v;\ 3 \sin u \cos v;\ 3 \sin v\big],\ B = \langle 0;\pi\rangle \times \langle 0;\pi/4\rangle$.

5. Verify that σ is a simple p.w. smooth surface. Find parametrizations of the simple smooth surfaces of which σ is composed. *(See Sections IV.1.6, IV.1.7 and IV.1.2.)*

a) $\sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ z = 4 - \sqrt{x^2+y^2},\ z \geq 0\big\}$,

b) $\sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ z = 4 - \sqrt{x^2+y^2},\ 0 \leq z \leq 2\big\}$,

c) $\sigma_1 \cup \sigma_2$ where $\sigma_1 = \big\{[x;y;z] \in \mathbb{E}_3 :\ x^2+y^2 \leq 16,\ z = 0\big\},\ \sigma_2 = \big\{[x;y;z] \in \mathbb{E}_3 :\ z = 4 - \sqrt{x^2+y^2},\ z \geq 0\big\}$,

d) σ is the boundary of the solid $D = \big\{[x;y;z] \in \mathbb{E}_3 :\ x^2+y^2+z^2 \leq 4,\ 4-x^2-y^2 \leq 4z\big\}$.

6. Decide whether the integral $\iint_\sigma f\ dp$ exists. *(See Sections IV.2.2 and IV.2.6.)*

a) $f(x,y,z) = (xy\ \ln|x|)/z,\ \sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ (x-2)^2+y^2+z^2 = 1,\ z \geq 0\big\}$,

b) $f(x,y,z) = (xy\ \ln|x|)/z,\ \sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ z = 1+x^2+y^2,\ z \leq 2\big\}$,

c) $f(x,y,z) = (x^2+y^2+z^2-1)^{-1}$, σ is the spherical surface with center at $S = [0;0;3]$ and radius $r = a$.

7. Compute the areas of the surfaces from Examples 3g, 4c, 5b, 5c. *(See Section IV.2.5.)*

8. Compute the following surface integrals. *(See Section IV.2.7.)*

a) $\displaystyle\iint_\sigma xyz\ dp,\ \sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ y^2+9z^2 = 9,\ 1 \leq x \leq 3,\ y \geq 0,\ z \geq 0\big\}$,

b) $\displaystyle\iint_\sigma xz\ dp$, σ is the triangle with vertices $A = [1;0;0]$, $B = [0;1;0]$ and $C = [0;0;1]$,

c) $\displaystyle\iint_\sigma (x^2+y^2)\ dp,\ \sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ x^2+y^2+z^2 = a^2\big\}\ (a > 0)$,

d) $\displaystyle\iint_\sigma (xy+yz+xz)\ dp,\ \sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ y = \sqrt{x^2+z^2},\ x^2+z^2 \leq 2x\big\}$,

e) $\displaystyle\iint_\sigma (x+y+z)\ dp,\ \sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ x^2+y^2+z^2 = a^2,\ x \leq 0\big\}\ (a > 0)$,

f) $\displaystyle\iint_\sigma \frac{dp}{x^2+y^2+z^2},\ \sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ x^2+y^2 = 9,\ 0 \leq z \leq 3\big\}$,

g) $\displaystyle\iint_\sigma (x^2+y^2)\ dp$, σ is the surface from Example 5d.

9. Determine the coordinates of the center of mass of the surface σ, on which mass is distributed with surface density ρ. *(See Chapter IV.3.)*

a) $\rho(x,y,z) = x,\ \sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ x = \sqrt{y^2+z^2},\ y \geq 0,\ 0 \leq x \leq 2\big\}$,

b) $\rho(x,y,z) = xyz,\ \sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ x^2+z^2 = 4,\ x \geq 0,\ z \geq 0,\ 0 \leq y \leq 3\big\}$.

10. κ is the lateral surface of a right circular cone with base radius $r = a > 0$ and height $h > 0$. The surface density of mass distribution on κ is $\rho = \text{const}$. Compute the moment of inertia of κ with respect to the axis of the cone. *(See Chapter IV.3.)*

11. Compute the flux of the vector field $\vec{f}$ across the surface σ.
(See Sections IV.4.2 and IV.4.7.)

a) $\vec{f}(x,y,z) = y\,\widehat{\imath} - x\,\widehat{\jmath} + z\widehat{k}$, $\sigma = \{\,[x;y;z] \in \mathbb{E}_3 :\ z = 4 - x^2 - y^2/9,\ y \geq 0,\ z \geq 0\,\}$, σ is oriented so that its normal forms an acute angle with the vector $\widehat{k} = (0;0;1)$,

b) $\vec{f}(x,y,z) = (0;0;y)$, σ is the triangle with vertices $A = [0;0;0]$, $B = [5;0;1]$, $C = [1;4;1]$, oriented so that the normal $\widehat{n}$ forms an acute angle with the vector $\widehat{k}$,

c) $\vec{f}(x,y,z) = x\,\widehat{\imath} + y\,\widehat{\jmath} + z\,\widehat{k}$, σ is the cylindrical surface $x^2 + y^2 = 9$, $0 \leq z \leq 4$, oriented by the normal vector $\widehat{n}$ which at the point $[3;2;2]$ equals $-\widehat{\imath}$,

d) $\vec{f}(x,y,z) = (x^3, y^3, z^3)$, σ is the part (corresponding to $z \geq 0$) of the spherical surface with center $S = [0;0;0]$ and radius $r = 2$, oriented so that its normal at the point $[0;0;2]$ is the vector $\widehat{n} = \widehat{k} = (0;0;1)$,

e) $\vec{f}(x,y,z) = x\,\widehat{\imath} - x\,\widehat{\jmath} + y\,\widehat{k}$, σ is the parallelogram $A = [0;0;0]$, $B = [0;3;3]$, $C = [-1;4;5]$, $D = [-1;1;2]$, oriented by the normal vector $\vec{n} = (1;-1;1)$,

f) $\vec{f}(x,y,z) = (x^2 - y^2,\ y^2 - z^2,\ z^2 - x^2)$, $\sigma = \sigma_1 \cup \sigma_2$, $\sigma_1 = \big\{[x;y;0] \in \mathbb{E}_3 :\ x^2 + y^2 \leq 1,\ x \geq 0,\ y \geq 0\big\}$, $\sigma_2 = \big\{[x;0;z] \in \mathbb{E}_3 :\ x^2 + z^2 \leq 1,\ z \geq 0,\ z \geq 0\big\}$, the normal vector on σ_2 is $\widehat{n} = -\widehat{\jmath} = (0;-1;0)$,

g) $\vec{f}(x,y,z) = x\,\widehat{\imath} + y\,\widehat{\jmath} - 2z\,\widehat{k}$, $\sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ y = 9 - \sqrt{x^2 + z^2},\ y \geq 3\big\}$, $\widehat{n} \cdot \widehat{\jmath} \leq 0$,

h) $\vec{f}(x,y,z) = (x^2; y^2; z^2)$, $\sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ y^2/16 + z^2/4 = 1,\ z \geq 0,\ 0 \leq x \leq 3\big\}$, the normal vector at the point $P = [1;0;2]$ is $\widehat{n} = -\widehat{k} = (0;0;-1)$,

i) $\vec{f}(x,y,z) = x\,\widehat{\imath} + y\,\widehat{\jmath} - z\,\widehat{k}$, $\sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ x^2 + y^2 + z^2 = 4,\ x \geq 0\big\}$, the normal vector at the point $P = [2;0;0]$ is $\widehat{n} = \widehat{\imath} = (1;0;0)$,

j) $\vec{f}(x,y,z) = (z;x;y)$, $\sigma = \big\{[x;y;z] \in \mathbb{E}_3 :\ x + z = 2,\ x^2 + y^2 \leq 4\big\}$, the normal vector is $\widehat{n} = (1;0;1)/\sqrt{2}$.

12. Compute the flux of the vector field $\vec{f}$ across the closed surface σ. If possible, use the Gauss–Ostrogradsky (divergence) theorem. *(See Sections IV.4.2, IV.4.7 and IV.5.7.)*

a) $\vec{f}(x,y,z) = x\,\widehat{\imath} + y\,\widehat{\jmath} + z\,\widehat{k}$, σ is the spherical surface with center $S_0 = [x_0; y_0; z_0]$ and radius $a > 0$, oriented outward,

b) $\vec{f}(x,y,z) = y\,\widehat{\imath} + z\,\widehat{\jmath} + x\,\widehat{k}$, σ is the spherical surface with center $S_0 = [x_0; y_0; z_0]$ and radius $a > 0$, oriented outward,

c) $\vec{f}(x,y,z) = (x^2; y^2; z^2)$, σ is the spherical surface with center $S_0 = [x_0; y_0; z_0]$ and radius $a > 0$, oriented outward,

d) $\vec{f}(x,y,z) = (y; 2x; -z)$, σ is the boundary of the solid $D = \big\{[x;y;z] \in \mathbb{E}_3 :\ x^2 + y^2 \leq a^2,\ -a \leq z \leq a\big\}$ $(a > 0)$, σ is oriented inward,

e) $\vec{f}(x,y,z) = (x^2; y^2; z^2)$, σ is the boundary of the solid $D = \big\{[x;y;z] \in \mathbb{E}_3 :\ -2 \leq z \leq 4 - x^2 - y^2,\ x^2 + y^2 \leq 4\big\}$, σ is oriented inward,

f) $\vec{f}(x,y,z) = x\,\widehat{\imath} + y\,\widehat{\jmath} + z\,\widehat{k}$, σ is the boundary of the solid $D = \big\{[x;y;z] \in \mathbb{E}_3 :\ y^2 + z^2 \leq x^2,\ 0 \leq x \leq 3\big\}$, σ is oriented outward,

g) $\vec{f}(x,y,z) = (x^3; z; y)$, σ is the boundary of the solid $D = \big\{[x;y;z] \in \mathbb{E}_3 :\ x^2 + y^2 \leq z \leq 4\big\}$, σ is oriented outward,

h) $\vec{f}(x,y,z) = 2xy\,\widehat{\imath} - y^2\widehat{\jmath} + 2z\,\widehat{k}$, $\sigma = \{[x;y;z] \in \mathbb{E}_3 : x^2/4 + y^2/4 + z^2/9 = 1\}$, σ is oriented inward,

i) $\vec{f}(x,y,z) = x^3\widehat{\imath} + y^3\widehat{\jmath} + z^3\widehat{k}$, σ is the boundary of the solid $D = \{[x;y;z] \in \mathbb{E}_3 : x^2 + y^2 + z^2 \leq a^2,\ y \geq 0\}$ $(a > 0)$, σ is oriented outward,

j) $\vec{f} = (x-1; y+2; 2)$, σ is a closed surface in $\mathbb{E}_3$, oriented outward and such that $\mu_3(\mathrm{Int}\,\sigma) = 5$.

13. Compute the circulation of the vector field $\vec{f}$ along the closed curve C. If possible, use Stokes' theorem. *(See Sections III.4.11 and IV.5.10.)*

a) $\vec{f}(x,y,z) = (y;z;x)$, C is the intersection of the cylindrical surface $x^2 + y^2 = 4$ with the plane $x + z = 0$. C is oriented so that its tangent vector at the point $[2;0;-2]$ is $\widehat{\tau} = \widehat{\jmath} = (0;1;0)$,

b) $\vec{f}(x,y,z) = (xy; yz; zx)$, $C = \{[x;y;z] \in \mathbb{E}_3 : x + z = 1,\ y^2 + z^2 = 1\}$. C is oriented consistently with the surface $\lambda = \{[x;y;z] \in \mathbb{E}_3 : x + z = 1,\ y^2 + z^2 \leq 1\}$ and the orientation of λ is determined by the normal vector $\widehat{n} = (1;0;1)/\sqrt{2}$,

c) $\vec{f}(x,y,z) = (z+1)\widehat{\imath} + (x-y)\widehat{\jmath} + y\,\widehat{k}$, C is the circle given by the intersection of the spherical surface $x^2 + y^2 + z^2 = 2$ and the plane $x + y + z = 0$. When viewed from the point $[0;0;10]$, C is oriented clockwise,

d) $\vec{f}(x,y,z) = \big(-y/(x^2+y^2);\ x/(x^2+y^2);\ 2z\big)$, C is the circle $x^2 + y^2 = a^2$ $(a > 0)$, $z = h$ $(h > 0)$, oriented counterclockwise when viewed from the point $[0;0;2h]$.

CHAPTER V.

Potential and solenoidal vector fields

V.1 Path-independence of the line integral of a vector function. Potential field – definition and basic properties 121
V.2 Potential fields in $\mathbb{E}_2$ 126
V.3 Potential fields in $\mathbb{E}_3$ 131
V.4 * Solenoidal fields 134
V.5 Exercises 140

V.1 Path-independence of the line integral of a vector function. Potential field – definition and basic properties

The work required to move a material body in a gravitational field from point A to point B does not depend on the chosen path (curve) between the points A and B, but only on the positions of A and B. You certainly know this from physics. The same statement holds when moving an electric charge in an electrostatic field. If you realize that the work can be expressed as a line integral of a vector function (equal to the negative of the intensity of the gravitational field multiplied by the mass of the moved body), you can speak about the path-independence of the line integral. Path-independence is an important property possessed by line integrals of some vector fields (= vector functions). In this chapter we will study this property in more detail and at a more general level.

In order not to complicate the text, let us adopt the convention that whenever we use the term curve (without further specification), we mean a simple p.w. smooth curve.

V.1.1. Path-independence of the line integral of a vector function.
Let $\vec{f}$ be a k–dimensional vector field in a domain $D \subset \mathbb{E}_k$ ($k = 2$ or $k = 3$). If for any two curves K_1 and K_2 in D having the same initial and terminal point (i.e. $i.p.\, K_1 = i.p.\, K_2$ and $t.p.\, K_1 = t.p.\, K_2$), the equality

$$\int_{K_1} \vec{f} \cdot d\vec{s} = \int_{K_2} \vec{f} \cdot d\vec{s}, \tag{V.1.1}$$

holds, then we say that *the line integral of the vector function $\vec{f}$ is path-independent in D*.

V.1.2. Theorem
The line integral of the vector function $\vec{f}$ is path-independent in a domain $D \subset \mathbb{E}_k$ (where $k = 2$ or $k = 3$) if and only if the circulation of $\vec{f}$ along any closed curve C in D is zero.

P r o o f:
a) First assume that the line integral of $\vec{f}$ is path-independent in D and that C is an arbitrary closed curve in D. We will prove that the circulation of $\vec{f}$ along C is zero. The curve C can be decomposed into the union of two curves C_1 and C_2 such that $t.p.\,C_1 = i.p.\,C_2$ and $t.p.\,C_2 = i.p.\,C_1$. Setting $K_1 = C_1$ and $K_2 = -C_2$, we obtain two curves in D with the same initial and terminal point. From path-independence of the line integral of $\vec{f}$ in D it follows that $\int_{K_1} \vec{f}\cdot d\vec{s} = \int_{K_2} \vec{f}\cdot d\vec{s}$. This means that

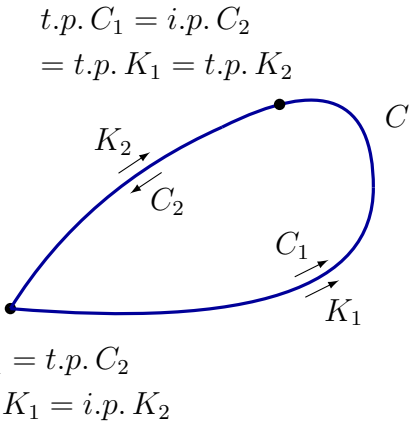


Fig. 30

$$\oint_C \vec{f}\cdot d\vec{s} = \int_{C_1} \vec{f}\cdot d\vec{s} + \int_{C_2} \vec{f}\cdot d\vec{s} = \int_{K_1} \vec{f}\cdot d\vec{s} - \int_{K_2} \vec{f}\cdot d\vec{s} = 0.$$

b) Now conversely assume that the circulation of $\vec{f}$ along any closed curve in D is zero and prove that the line integral of $\vec{f}$ is path-independent in D. Let C_1 and C_2 be any two curves in D such that $i.p.\,C_1 = i.p.\,C_2$ and $t.p.\,C_1 = t.p.\,C_2$. For simplicity, assume that the curves C_1 and C_2 have no other common points, i.e. they do not intersect at any other points. (See Fig. 31a.) Then the union $C = C_1 \cup (-C_2)$ is a closed curve in D and by assumption the circulation of $\vec{f}$ along C is zero. Hence:

$$\int_{C_1} \vec{f}\cdot d\vec{s} = \oint_{C_1\cup(-C_2)} \vec{f}\cdot d\vec{s} + \int_{C_2} \vec{f}\cdot d\vec{s} = \int_{C_2} \vec{f}\cdot d\vec{s}.$$

A similar approach can be used in the case when the curves C_1 and C_2 have more common points than just the initial and terminal points. (See Fig. 31b.) Try to think through and write down yourself how the proof could be completed in this case.

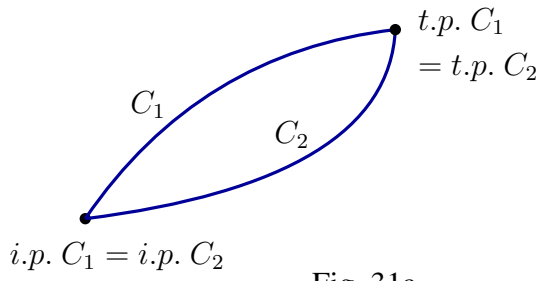


Fig. 31a

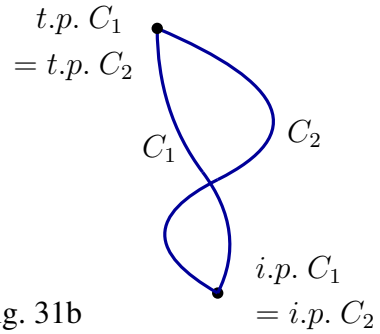


Fig. 31b

V.1.3. Potential vector field
A vector field $\vec{f}$ in a domain $D \subset \mathbb{E}_k$ ($k = 2$ or $k = 3$) is called a *potential field* in D if there exists a scalar field (= a scalar function) φ in D such that

(V.1.2) $$\vec{f} = \operatorname{grad}\varphi$$

in D. The scalar function φ is called the *potential* of the vector field $\vec{f}$ in D.

V.1.4. Some properties of the potential

Recall that

$$\operatorname{grad}\varphi = \left(\frac{\partial\varphi}{\partial x};\ \frac{\partial\varphi}{\partial y}\right)\ (\text{for } k=2), \qquad \operatorname{grad}\varphi = \left(\frac{\partial\varphi}{\partial x};\ \frac{\partial\varphi}{\partial y};\ \frac{\partial\varphi}{\partial z}\right)\ (\text{for } k=3).$$

It is clear that if $\vec{f}$ is a potential field in a domain D and D' is a domain contained in D, then $\vec{f}$ is also a potential field in D'.

The potential φ of a potential field $\vec{f}$ has some properties similar to those of an antiderivative of a function of one variable. For example:

If $\vec{f}$ is a potential field in a domain $D \subset \mathbb{E}_k$, then its potential φ is determined uniquely up to an additive constant.

This means that:

a) *$\varphi + c$ (where c is any real constant) is also a potential of $\vec{f}$ in D.*

b) *Any other potential ζ of the field $\vec{f}$ in D differs from φ by at most an additive constant.* In other words: *if ζ is also a potential of $\vec{f}$ in D, then there exists a constant c such that $\zeta = \varphi + c$ in D.*

Statement a) is very simple: From the equation $\operatorname{grad}\varphi = \vec{f}$ in D it follows immediately that $\operatorname{grad}(\varphi + c) = \vec{f}$ and thus $\varphi + c$ is also a potential of $\vec{f}$ in D.

Statement b) is also quite clear: If φ and ζ are two potentials of $\vec{f}$ in D, then $\vec{f} = \operatorname{grad}\varphi$ and $\vec{f} = \operatorname{grad}\zeta$ in D. Hence $\operatorname{grad}\zeta - \operatorname{grad}\varphi = \operatorname{grad}(\zeta - \varphi) = \vec{0}$. The only function whose gradient is zero in a domain D is a constant function. This means that there exists a constant c such that $\zeta - \varphi = c$. Then of course $\zeta = \varphi + c$ in D.

Another analogy between the potential φ of a potential field $\vec{f}$ and an antiderivative of a function of one variable is the similarity between formula (V.1.3) (see the following theorem) and the Newton–Leibniz formula (V.4.1) (see the textbook [5]).

V.1.5. Theorem

If $\vec{f}$ is a potential and continuous vector field in a domain D, φ is a potential of $\vec{f}$ in D and C is a curve in D, then

(V.1.3) $$\int_C \vec{f}\cdot d\vec{s} = \varphi(t.p.\,C) - \varphi(i.p.\,C).$$

P r o o f:

Since φ is a potential of $\vec{f}$ in D, $\vec{f}$ equals $\operatorname{grad}\varphi$ in D. First assume that C is a simple smooth curve in D and P is its parametrization, defined on the interval $\langle a\,;b\rangle$, and that the curve C is oriented consistently with the parametrization P. Let $x(t)$, $y(t)$ and $z(t)$ be the coordinate functions of the map P. Then

$$\int_C \vec{f}\cdot d\vec{s} = \int_C \operatorname{grad}\varphi\cdot d\vec{s} = \int_a^b \operatorname{grad}\varphi(P(t))\cdot\dot{P}(t)\,dt = \int_a^b \Big(\frac{\partial\varphi}{\partial x}\big(x(t),y(t),z(t)\big);$$
$$\frac{\partial\varphi}{\partial y}\big(x(t),y(t),z(t)\big);\ \frac{\partial\varphi}{\partial z}\big(x(t),y(t),z(t)\big)\Big)\cdot\big(\dot{x}(t);\dot{y}(t);\dot{z}(t)\big)\,dt =$$

$$= \int_a^b \Big(\frac{\partial\varphi}{\partial x}\big(x(t),y(t),z(t)\big)\,\dot{x}(t) + \frac{\partial\varphi}{\partial y}\big(x(t),y(t),z(t)\big)\,\dot{y}(t) +$$

$$+ \frac{\partial\varphi}{\partial z}\big(x(t),y(t),z(t)\big)\,\dot{z}(t)\Big)\,dt = \int_a^b \frac{d}{dt}\,\varphi\big(x(t),y(t),z(t)\big)\,dt =$$

$$= \varphi\big(x(b),y(b),z(b)\big) - \varphi\big(x(a),y(a),z(a)\big) = \varphi(t.p.\,C) - \varphi(i.p.\,C).$$

The same expression for the integral $\int_C \vec{f}\cdot d\vec{s}$ can also be derived in the case when C is a simple p.w. smooth curve. (Try to think this through and write it down yourself.)

Since the value of the integral $\int_C \vec{f}\cdot d\vec{s}$ of the potential field $\vec{f}$ depends only on the position of the initial and terminal point of the curve C, we often write this integral as

$$\int_C \vec{f}\cdot d\vec{s} = \int_A^B \vec{f}\cdot d\vec{s},$$

where $A = i.p.\,C$ and $B = t.p.\,C$.

The following theorem is very important. It relates path-independence of the line integral to the fact that the vector field is potential.

V.1.6. Theorem
Assume that $\vec{f}$ is a continuous vector field in a domain $D \subset \mathbb{E}_k$ ($k = 2$ or $k = 3$). Then the following two statements are equivalent:

a) $\vec{f}$ is a potential field in D.

b) The line integral of $\vec{f}$ is path-independent in D.

P r o o f:
The implication a) $\Longrightarrow$ b) is an immediate consequence of formula (V.1.3).

Now let us prove the converse implication, i.e. b) $\Longrightarrow$ a). Assume therefore that condition b) holds. Let U, V, and W be the coordinate functions of $\vec{f}$. Choose a point $A \in D$. The point A is chosen arbitrarily, but from this moment on we will regard it as fixed. Let $X = [x;y;z]$ be another point in D. Define

$$\varphi(x,y,z) = \int_A^X \vec{f}\cdot d\vec{s} = \int_C \vec{f}\cdot d\vec{s}, \tag{V.1.4}$$

where C is a curve in D whose initial point is A and whose terminal point is X. From path-independence of the line integral of the field $\vec{f}$ in D it follows that the value $\varphi(x,y,z)$ does not depend on the particular choice of the curve C connecting the points A and $X = [x;y;z]$. We will show that $\operatorname{grad}\varphi = \vec{f}$ in D. This means that

$$\frac{\partial\varphi}{\partial x}(X) = U(X), \quad \frac{\partial\varphi}{\partial y}(X) = V(X) \quad \text{and} \quad \frac{\partial\varphi}{\partial z}(X) = W(X). \tag{V.1.5}$$

Let us prove, for example, the first of these equalities. Using the definition of the partial derivative of φ with respect to x at the point $X = [x;y;z]$ we obtain:

$$\frac{\partial\varphi}{\partial x}(X) = \frac{\partial\varphi}{\partial x}(x;y;z) = \lim_{h\to 0}\frac{\varphi(x+h,y,z)-\varphi(x,y,z)}{h}.$$

The value $\varphi(x+h,y,z)$ can be expressed as the line integral of $\vec{f}$ along a simple p.w. smooth curve which is the union of C and the segment $\overline{XX'}$ leading from the point X to the point $X'=[x+h;y;z]$. The unit tangent vector on $\overline{XX'}$ is $\widehat{\tau}=(1;0;0)$. Hence:

$$\frac{\partial\varphi}{\partial x}(X) = \lim_{h\to 0}\frac{1}{h}\left(\int_{C\cup\overline{XX'}}\vec{f}\cdot d\vec{s}-\int_C \vec{f}\cdot d\vec{s}\right) = \lim_{h\to 0}\frac{1}{h}\int_{\overline{XX'}}\vec{f}\cdot d\vec{s} =$$

$$= \lim_{h\to 0}\frac{1}{h}\int_{\overline{XX'}}(U;V;W)\cdot(1;0;0)\,ds = \lim_{h\to 0}\frac{1}{h}\int_{\overline{XX'}}U\,ds = U(X).$$

Equalities (V.1.5) show that $\vec{f}=\operatorname{grad}\varphi$ in D and therefore the vector field $\vec{f}$ is a potential field in D.

V.1.7. Corollary
An immediate consequence of Theorem V.1.2 and Theorem V.1.6 is the following theorem:

Theorem
A vector field $\vec{f}$, continuous in a domain D, is potential in D if and only if its circulation along every closed curve in D is zero.

A potential field is often called an **irrotational field**. From the stated theorem it is clear why.

V.1.8. Remark
If $\vec{f}$ is a potential field in a domain D and $\vec{f}$ has the physical meaning of a force, then from Theorem V.1.7 it follows that the work performed by the force $\vec{f}$ along any closed curve in D is zero. In other words, acting by the force $\vec{f}$ along any closed curve in the domain D no energy is lost or gained. This fact is the basis for another, often used name for a potential field: a **conservative field**.

We have presented the proofs of Theorems V.1.5 and V.1.6 because they are very instructive and their ideas can be used in many other situations. Formula (V.1.3) provides a simple guide for computing the line integral of a potential vector field when its potential is known. The proof of Theorem V.1.6 then shows one possible way to construct the potential of a vector field. We will use this method in a specific case in Section V.2.10. In Sections V.2.10 and V.3.10 we will also explain another method for finding the potential of a vector field $\vec{f}$, provided we know that it is a potential field.

We have learned that a potential vector field $\vec{f}$ in a domain D has interesting and useful properties – in particular, that its line integral is path-independent in D (Theorem V.1.6) and that the line integral can moreover be computed using formula (V.1.3) (Theorem V.1.5). Therefore it is very important to be able to recognize whether a given vector field $\vec{f}$ in a domain D is or is not potential. In the following sections we will deal with this question. We will distinguish between the two-dimensional case (Chapter V.2) and the three-dimensional case (Chapter V.3).

V.2 Potential fields in $\mathbb{E}_2$

V.2.1. Theorem. (Potential fields in $\mathbb{E}_2$ – necessary condition.)
Assume that $\vec{f} = (U; V)$ is a potential field in a domain $D \subset \mathbb{E}_2$ and its coordinate functions U, V have continuous partial derivatives in D. Then

$$\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} = 0 \quad \text{in } D. \tag{V.2.1}$$

P r o o f:
If φ is a potential of $\vec{f}$ in D, then $\vec{f} = \big(\partial\varphi/\partial x;\ \partial\varphi/\partial y\big)$. Thus

$$U = \frac{\partial \varphi}{\partial x} \quad \text{and} \quad V = \frac{\partial \varphi}{\partial y}$$

in D. From this form of the functions U and V and from the information about continuity of the partial derivatives of U and V in D (which means continuity of the second partial derivatives of φ in D), condition (V.2.1) follows easily.

V.2.2. Remark

The validity of condition (V.2.1) is necessary for the field $\vec{f}$ to be potential in a domain D. However, it is not sufficient. This means that if we do not know in advance whether $\vec{f} = (U; V)$ is a potential field in D, then the validity of (V.2.1) alone does not allow us to conclude it. This can be illustrated by the example of the vector field

$$\vec{f}(x, y) = (U; V) = \left(-\frac{y}{x^2 + y^2}; \frac{x}{x^2 + y^2}\right)$$

from Section III.4.12 in the domain $D = \mathbb{E}_2 \setminus \{[0; 0]\}$. ($D$ is the entire plane $\mathbb{E}_2$ except for the origin.) By an easy computation you can verify that the functions U and V satisfy condition (V.2.1) in D. Nevertheless, the field $\vec{f}$ is not a potential field in D. In Example III.4.12 we found that the circulation of $\vec{f}$ along the circle $C:\ x^2 + y^2 = r^2$ ($r > 0$), oriented clockwise, equals -2π. If the field $\vec{f}$ were potential in D, its circulation along every closed curve in D would have to be zero (Theorem V.1.7).

On the other hand, if some vector field does not satisfy condition (V.2.1), we can immediately declare that it is not potential.

V.2.3. Example

Decide whether the vector field $\vec{f}(x, y) = (x + y; x^2 + y^2)$ is a potential field in $\mathbb{E}_2$.

S o l u t i o n:
Here $U(x, y) = x + y$ and $V(x, y) = x^2 + y^2$. Since

$$\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} = 2x - 1$$

and $2x - 1$ is not equal to zero everywhere in $\mathbb{E}_2$, the vector field $\vec{f}$ is not a potential field in $\mathbb{E}_2$.

Our next goal is to formulate a sufficient condition whose validity already guarantees that a given two-dimensional vector field $\vec{f}$ is potential in a domain $D \subset \mathbb{E}_2$. For this we need the notion of a so-called "simply connected domain" in $\mathbb{E}_2$.

V.2.4. Simply connected domain in $\mathbb{E}_2$

A domain $D \subset \mathbb{E}_2$ is called *simply connected* if every closed curve C in D can be continuously deformed (contracted) to a point in D without ever leaving the domain D.

A simply connected domain in $\mathbb{E}_2$ can also be defined as a domain D in which the interior of every closed curve in D is a subset of D.

Roughly speaking, domains that have bounded "holes" (and look like Swiss cheese) are not simply connected. On the other hand, domains without such holes are simply connected. Examples of simply connected domains in $\mathbb{E}_2$ are the whole $\mathbb{E}_2$, a half-plane, $\mathbb{E}_2$ minus a half-line, the interior of a rectangle, square, or any other closed curve, etc. Examples of domains in $\mathbb{E}_2$ that are not simply connected are $\mathbb{E}_2$ minus one point, $\mathbb{E}_2$ minus a bounded closed set, an annulus, etc.

V.2.5.* Remark

In Definition V.2.4 we used the phrase "continuously contract to a point A in D without leaving D", whose meaning we did not explain precisely. Nevertheless, we believe that the reader will correctly understand what we mean. However, for completeness of the exposition let us add that this phrase can be expressed in precise mathematical language as follows: "There exists a continuous mapping $F : C \times \langle 0; 1\rangle \longrightarrow D$ such that for all $X \in C$ we have $F(X, 0) = X$ and $F(X, 1) = A$."

V.2.6. Theorem. (Potential fields in $\mathbb{E}_2$ – sufficient condition.) *Let*

a) D be a simply connected domain in $\mathbb{E}_2$, and

b) $\vec{f} = (U; V)$ be a vector field in D whose coordinate functions U and V have continuous partial derivatives in D and satisfy condition (V.2.1):

$$\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} = 0 \quad \text{in } D.$$

Then $\vec{f}$ is a potential field in D.

P r o o f:
We will show that the circulation of the vector field $\vec{f}$ along every closed curve in D is zero. From this it then follows that $\vec{f}$ is a potential field in D. (See Theorem V.1.7.) Let C be an arbitrary closed curve in D. Since D is simply connected, $Int\, C \subset D$. Using Green's theorem we obtain:

$$\oint_C \vec{f} \cdot d\vec{s} = \pm \iint_{Int\ C} \left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} \right) dx\, dy = 0.$$

(The sign "+" holds if the curve C is positively oriented and "−" holds if C is negatively oriented. However, since the integral on the right-hand side equals zero, the sign plays no role.)

In the proof you can see why we need the domain D to be simply connected: without that we do not know whether $Int\, C \subset D$ and whether we may use Green's theorem.

V.2.7. Example

The vector field $\vec{f} = (U; V)$ from Example III.4.12 and from Remark V.2.2 is not a potential field in the domain $D = \{[x; y] \in \mathbb{E}_2 : x^2 + y^2 > 0\}$, even though the functions U and V

satisfy condition (V.2.1) in D. (We stated the reason in Remark V.2.2.) However, by Theorem V.2.6 the field $\vec{f}$ is a potential field in every domain $D' \subset D$ that is simply connected. For example, $\vec{f}$ is a potential field in the first quadrant.

Although we now know that $\vec{f}$ is a potential field in the first quadrant, we do not know its potential. Methods that make it possible to find the potential will be discussed in Sections V.2.10 and V.2.11, where among other things we will return to this example.

V.2.8. How do we recognize whether a given vector field is potential?
In this section we briefly summarize how we can determine whether a given vector field $\vec{f} = (U; V)$ is a potential field in a domain $D \subset \mathbb{E}_2$. For simplicity, assume that the coordinate functions U and V have continuous partial derivatives in D. Then we can proceed according to the following scheme:

Does the equality

$$\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} = 0$$

hold in the domain D?

- *YES* — *Is D a simply connected domain?*
  - *YES* — *Then $\vec{f}$ is potential in D.* (Theorem V.2.6.)
  - *NO* — *By Theorem V.2.6 one cannot decide whether $\vec{f}$ is potential in D.*
- *NO* — *Then $\vec{f}$ is not potential in D.* (Theorem V.2.1.)

V.2.9. Example
$\vec{f} = \left(y^2 + y\,\cos\,x + 6x;\ 2xy + \sin\,x + 5\right)$ Decide whether $\vec{f}$ is a potential field in $\mathbb{E}_2$.

S o l u t i o n:
The coordinate functions $U(x, y) = y^2 + y\,\cos\,x + 6x$ and $V(x, y) = 2xy + \sin\,x + 5$ have continuous partial derivatives in $\mathbb{E}_2$. You can easily verify that they satisfy condition (V.2.1):

$$\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} = 2y + \cos\,x - 2y - \cos\,x = 0$$

in $\mathbb{E}_2$. The whole plane $\mathbb{E}_2$ is a simply connected domain. Therefore, by Theorem V.2.6, $\vec{f}$ is a potential field in $\mathbb{E}_2$.

V.2.10. How to find a potential
In this section we present two methods by which we can find the potential φ of a vector field $\vec{f}$ which we know to be potential. We explain both methods using the vector field $\vec{f}$ from the preceding section.

Method 1
From the definition of the potential of a vector field it follows that $\vec{f} = \operatorname{grad}\varphi$ in $\mathbb{E}_2$, i.e.

(V.2.2) $$\frac{\partial\varphi}{\partial x} = y^2 + y\,\cos\,x + 6x \quad \text{and} \quad \frac{\partial\varphi}{\partial y} = 2xy + \sin\,x + 5.$$

If we integrate the first equation in (V.2.2) with respect to x, we obtain:

(V.2.3) $$\varphi(x,y) = xy^2 + y\,\sin\,x + 3x^2 + c_1(y).$$

c_1 is the constant of integration that arose when integrating with respect to x. It is therefore a constant with respect to x, but it may depend on y. By integrating the second equation in (V.2.2) with respect to y we obtain:

(V.2.4) $$\varphi(x,y) = xy^2 + y\,\sin\,x + 5y + c_2(x).$$

c_2 is the constant of integration that arose when integrating with respect to y. It is therefore a constant with respect to y, but in general it may depend on x. Now it remains to determine $c_1(y)$ and $c_2(x)$. For this purpose we compare the right-hand sides of (V.2.3) and (V.2.4). We obtain:

$$\begin{aligned} xy^2 + y\,\sin\,x + 3x^2 + c_1(y) &= xy^2 + y\,\sin\,x + 5y + c_2(x), \\ 3x^2 + c_1(y) &= 5y + c_2(x). \end{aligned}$$

This is satisfied, for example, if we set $c_1(y) = 5y$ and $c_2(x) = 3x^2$. We can now substitute into (V.2.3) or into (V.2.4) and express φ. If we also realize that the potential is determined uniquely up to an additive constant, we obtain:

(V.2.5) $$\varphi(x,y) = xy^2 + y\,\sin\,x + 3x^2 + 5y + const.$$

Method 2

This method is based on the proof of Theorem V.1.6, part b), and the potential is constructed using formula (V.1.4). Choose a point $A \in \mathbb{E}_2$, for example $A = O = [0;0]$. Denote the coordinates of the point X, at which we compute the value of the potential, by x_0 and y_0. During the computation we regard A and X as fixed points, i.e. we treat x_0 and y_0 as constants. Set $\varphi(X) = \varphi(x_0,y_0) = \int_C \vec{f}\cdot d\vec{s}$, where C is a curve in $\mathbb{E}_2$ with initial point A and terminal point X. Choose the curve C so that the computation of the line integral along it is as simple as possible. For example, take: $C = \overline{AB} \cup \overline{BX}$, where $\overline{AB}$ is the segment from $A = O = [0;0]$ to $B = [x_0,0]$ and $\overline{BX}$ is the segment from $B = [x_0,0]$ to $X = [x_0;y_0]$. For $\varphi(x_0,y_0)$ we have:

$$\varphi(x_0,y_0) = \int_{\overline{AB}} \vec{f}\cdot d\vec{s} + \int_{\overline{BX}} \vec{f}\cdot d\vec{s}.$$

We can parametrize the segment $\overline{AB}$ by the mapping $x = t$, $y = 0$; $t \in \langle 0; x_0\rangle$. Therefore on this segment $dx = dt$, $dy = 0$ and

$$\begin{aligned} \int_{\overline{AB}} \vec{f}\cdot d\vec{s} &= \int_{\overline{AB}} (y^2 + y\,\cos\,x + 6x)\,dx + (2xy + \sin\,x + 5)\,dy = \\ &= \int_0^{x_0} 6t\,dt = 3x_0^2. \end{aligned}$$

The segment $\overline{BX}$ can be parametrized by the mapping $x = x_0$, $y = t$; $t \in \langle 0; y_0\rangle$. Hence on this segment $dx = 0$, $dy = dt$ and

$$\int_{\overline{BX}} \vec{f}\cdot d\vec{s} = \int_{\overline{BX}} (y^2 + y\,\cos\,x + 6x)\,dx + (2xy + \sin\,x + 5)\,dy =$$

$$= \int_0^{y_0} (2x_0 t + \sin\, x_0 + 5)\, dt \;=\; x_0 y_0^2 + y_0\, \sin\, x_0 + 5y_0.$$

Therefore the value of φ at the point X is

$$\varphi(x_0, y_0) \;=\; 3x_0^2 + x_0 y_0^2 + y_0\, \sin\, x_0 + 5y_0.$$

Writing x instead of x_0, y instead of y_0, and taking into account that the potential is determined uniquely up to an additive constant, we obtain the same expression for the potential as in (V.2.5).

V.2.11. Example

Let us return to the vector field

$$\vec{f} \;=\; (U; V) \;=\; \left(-\frac{y}{x^2+y^2}; \frac{x}{x^2+y^2} \right),$$

which from Example V.2.7 we know is a potential field in the first quadrant $D' = \big\{[x; y] \in \mathbb{E}_2 : x > 0,\ y > 0\big\}$. What is the potential φ of the field $\vec{f}$ in the region D'?

Solution:
To find φ let us use, for instance, the first method from Section V.2.10. We start from the equations

$$\frac{\partial \varphi}{\partial x} = -\frac{y}{x^2+y^2}\,, \qquad \frac{\partial \varphi}{\partial y} = \frac{x}{x^2+y^2}\,. \tag{V.2.6}$$

By integrating the first equation with respect to x we obtain:

$$\varphi(x,y) \;=\; -\int \frac{y}{x^2+y^2}\, dx \;=\; -\frac{1}{y}\int \frac{1}{(x/y)^2+1}\, dx \;=^{1)}\; -\int \frac{dt}{t^2+1} =$$

$$= \operatorname{arccotan} t + c_1(y) \;=^{2)}\; \arctan\, \frac{1}{t} + c_1(y) \;=\; \arctan\, \frac{y}{x} + c_1(y).$$

[1]) We used the substitution $t = x/y$.

[2]) We used the formula $\operatorname{arccotan} t = \arctan(1/t)$, which holds for $t > 0$.

By integrating the second equation with respect to y we get:

$$\varphi(x,y) \;=\; \int \frac{x}{x^2+y^2}\, dy \;=\; \frac{1}{x}\int \frac{1}{1+(y/x)^2}\, dy \;=^{3)}\; \int \frac{dr}{r^2+1} =$$

$$= \arctan\, r + c_2(x) \;=\; \arctan\, \frac{y}{x} + c_2(x).$$

[3]) We used the substitution $r = y/x$.

We determine the integration constants $c_1(y)$ and $c_2(x)$ by comparing both expressions for $\varphi(x, y)$. It is clear that we may choose $c_1(y) = c_2(x) = c$, where c is a constant independent of both x and y. Substituting into either expression for $\varphi(x, y)$ we get: $\varphi(x, y) = \arctan\,(y/x) + c$. The function $\arctan\,(y/x)$ has a clear geometric meaning: it is the angle between the positive part of the x-axis and the ray starting at the origin and passing through the point $[x; y]$.

V.3 Potential fields in $\mathbb{E}_3$

V.3.1. Theorem. (Potential field in $\mathbb{E}_3$ – necessary condition.)
Suppose that $\vec{f}$ is a potential field in a region $D \subset \mathbb{E}_3$ and has continuous partial derivatives in D. Then

(V.3.1) $$\operatorname{curl} \vec{f} = \vec{0} \quad \text{in } D.$$

Proof:
If φ is a potential of $\vec{f}$ in D, then $\vec{f} = \operatorname{grad} \varphi$ in D and condition (V.3.1) follows immediately from formula (IV.5.1).

V.3.2. Remark

In Section V.1.7 we explained why a potential field is called irrotational. Condition (V.3.1) provides another reason for this terminology.

V.3.3. Remark

Condition (V.3.1) is a necessary condition for the field $\vec{f}$ to be potential in a region D. However, it is not sufficient. That is, if we do not know in advance whether $\vec{f}$ is a potential field in D, we cannot conclude this merely from the validity of condition (V.2.1). This can be shown by the example of the vector field

$$\vec{f}(x,y,z) = \left(-\frac{y}{x^2+y^2}; \frac{x}{x^2+y^2}; 0 \right)$$

in the region $D = \{[x;y;z] \in \mathbb{E}_3 : x^2+y^2 > 0\}$. ($D$ is all of $\mathbb{E}_3$ except the z-axis.) By an easy computation you can verify that the vector function $\vec{f}$ satisfies condition (V.3.1) in D. Nevertheless, the field $\vec{f}$ is not a potential field in D. Similarly to Example III.4.12 one can compute that the circulation of $\vec{f}$ along the circle $C : x^2+y^2 = r^2$ ($r > 0$), $z = 0$ (i.e. the circle centered at the origin with radius r in the xy-plane), oriented clockwise when viewed from the positive part of the z-axis, equals -2π. Hence $\vec{f}$ is not potential in D. (Otherwise its circulation along every closed curve in D would have to be zero – see Theorem V.1.7.)

However, if a vector field does not satisfy condition (V.3.1), it means that it is not potential.

V.3.4. Example

Decide whether the vector field $\vec{f}(x,y,z) = (y+z;\ x^2+z^2;\ x^3+y^3)$ is a potential field in $\mathbb{E}_3$.

Solution:
Computing the curl of $\vec{f}$ we obtain:

$$\operatorname{curl} \vec{f}(x,y,z) = \begin{vmatrix} \vec{\imath}, & \vec{\jmath}, & \vec{k} \\ \dfrac{\partial}{\partial x}, & \dfrac{\partial}{\partial y}, & \dfrac{\partial}{\partial z} \\ y+z, & x^2+z^2, & x^3+y^3 \end{vmatrix} = \big(3y^2-2z;\ 1-3x^3;\ 2x-1\big).$$

Since the result is not the zero vector field in $\mathbb{E}_3$, the vector field $\vec{f}$ is not a potential field in $\mathbb{E}_3$.

In the following sections we will formulate a sufficient condition whose validity guarantees that a given three-dimensional vector field $\vec{f}$ is potential in a region $D \subset \mathbb{E}_3$. For that we will

need the notion of a "simply connected region" in $\mathbb{E}_3$.

V.3.5. Simply connected region in $\mathbb{E}_3$
A region $D \subset \mathbb{E}_3$ is called *simply connected* if every closed curve C in D can be continuously deformed (contracted) to a point in D without ever leaving D.

Examples of simply connected regions in $\mathbb{E}_3$ are the whole space $\mathbb{E}_3$, a half-space, $\mathbb{E}_3$ minus a half-line, the interior of a ball, a rectangular box, a cube, or the interior of any other closed surface, etc. Examples of regions in $\mathbb{E}_3$ that are not simply connected are $\mathbb{E}_3$ minus an entire line, the exterior of the cylindrical surface $x^2 + y^2 = r^2$ $(r > 0)$, the interior or the exterior of a so-called anuloid, etc.

We have mentioned several times that in $\mathbb{E}_3$ there occur sets of substantially more diverse shapes and properties than in $\mathbb{E}_2$. This is also the reason why in $\mathbb{E}_3$ one can define yet another and different type of simply connected regions. We discuss this in Section V.4.8.

V.3.6.* Remark

In Definition V.3.5, similarly to Definition V.2.4, we again rely on the reader's intuition and use the statement "continuously contract to a point in the region D without leaving D", whose meaning was not defined precisely. More demanding readers can find a precise definition in Section V.2.5.

V.3.7. Theorem. (Potential field in $\mathbb{E}_3$ – sufficient condition.) *Let*

a) D be a simply connected region in $\mathbb{E}_3$, and

b) $\vec{f}$ be a vector field in D with continuous partial derivatives in D that satisfies condition (V.3.1): $\operatorname{curl} \vec{f} = \vec{0}$ in D.

Then $\vec{f}$ is a potential field in D.

We do not present the proof of this theorem. Its idea is, as in the proof of Theorem V.2.6, to demonstrate that the circulation of the vector field $\vec{f}$ along every closed curve in D is zero. The procedure is, however, more complicated than in the case of Theorem V.2.6 and, instead of Green's theorem, it is based on the use of Stokes' theorem.

V.3.8. How to recognize whether a given vector field is potential.
Assume that the field $\vec{f}$ has continuous partial derivatives in a region $D \subset \mathbb{E}_3$. Then we can proceed according to the following scheme:

Is the equality $\operatorname{curl} \vec{f} = \vec{0}$ *satisfied in* D*?*

- *YES* — *Is* D *a simply connected region?*
  - *YES* — *Then* $\vec{f}$ *is potential in* D. (Theorem V.3.7.)
  - *NO* — *By Theorem V.3.7 it is not possible to decide whether* $\vec{f}$ *is potential in* D.
- *NO* — *Then* $\vec{f}$ *is not potential in* D. (Theorem V.3.1.)

V.3.9. How to find a potential

Up to small and obvious modifications we can use both methods explained in Section V.2.10. The use of the first method is shown in Example V.3.10.

V.3.10. Example – gravitational field

The intensity of the gravitational field created by a material point of mass m, placed in $\mathbb{E}_3$ at the origin of the coordinate system, is

$$\vec{g}(x,y,z) = -\frac{\kappa m\,(x\widehat{\imath} + y\widehat{\jmath} + z\widehat{k})}{(x^2+y^2+z^2)^{3/2}} \quad (\text{for } x^2+y^2+z^2 > 0),$$

where κ is Newton's gravitational constant. The region $D = \{[x;y;z] \in \mathbb{E}_3 : x^2+y^2+z^2 > 0\}$ ($\mathbb{E}_3$ minus the origin) is simply connected. By computing the curl of $\vec{g}$ we find that $\operatorname{curl} \vec{g} = \vec{0}$ in D. By Theorem V.3.7, the field $\vec{g}$ is a potential field in D.

Let us compute the potential φ of the field $\vec{g}$. From the equality $\vec{g} = \operatorname{grad} \varphi$ three equations follow:

$$\frac{\partial\varphi}{\partial x} = -\frac{\kappa m\, x}{(x^2+y^2+z^2)^{3/2}}, \quad \frac{\partial\varphi}{\partial y} = -\frac{\kappa m\, y}{(x^2+y^2+z^2)^{3/2}}, \quad \frac{\partial\varphi}{\partial z} = -\frac{\kappa m\, z}{(x^2+y^2+z^2)^{3/2}}.$$

By integrating the first equation with respect to x we obtain:

$$\varphi(x,y,z) = -\int \frac{\kappa m\, x\, dx}{(x^2+y^2+z^2)^{3/2}} \overset{1)}{=} -\int \frac{\kappa m}{2t^{3/2}}\, dt = \frac{\kappa m}{t^{1/2}} + c_1(y,z) =$$
$$= \frac{\kappa m}{\sqrt{x^2+y^2+z^2}} + c_1(y,z).$$

[1]) We used the substitution $t = x^2+y^2+z^2$.

Similarly, by integrating the second equation with respect to y and the third equation with respect to z, we obtain:

$$\varphi(x,y,z) = \frac{\kappa m}{\sqrt{x^2+y^2+z^2}} + c_2(x,z) \quad \text{and} \quad \varphi(x,y,z) = \frac{\kappa m}{\sqrt{x^2+y^2+z^2}} + c_3(x;y).$$

By comparing all three expressions for $\varphi(x,y,z)$ we find that the integration constants can be chosen so that $c_1(y,z) = c_2(x,z) = c_3(x;y) = c$, where c is a constant independent of x, y,

and z. Thus a potential of the gravitational field $\vec{g}$ in the region D is the function

$$\varphi(x,y,z) \;=\; \frac{\kappa m}{\sqrt{x^2+y^2+z^2}} + c.$$

In physics the vector function $x\widehat{\imath} + y\widehat{\jmath} + z\widehat{k} \equiv (x;y;z)$ is often called the "radius vector" and is denoted by $\vec{r}$. Using this function we can write $\vec{g}$ and φ as:

$$\vec{g}(\vec{r}) \;=\; -\frac{\kappa m\,\vec{r}}{\|\vec{r}\|^3}, \qquad \varphi(\vec{r}) \;=\; \frac{\kappa m}{\|\vec{r}\|} + c.$$

An electrostatic field created by a point charge placed at the origin has an analogous form to $\vec{g}$. Such an electrostatic field is likewise a potential field in D, and its potential again has a similar form to φ.

V.4 * Solenoidal fields

The Greek word "solenoid" in physics denotes a coil. In mathematics, a "solenoidal field" began to mean a vortical, rotational vector field. Later the term "solenoidal field" became established as a designation for a field that is, in a certain sense, the opposite of a potential field: whereas a potential field may contain sources (springs), but not vortices (see Sections V.1.7 and V.3.2), a solenoidal field may contain vortices, but not sources. (You will see this in Section V.4.3.)

In this text we restrict ourselves to the study of solenoidal fields in $\mathbb{E}_3$. Two-dimensional solenoidal fields are discussed in Remark V.4.13.

V.4.1. Solenoidal vector field in $\mathbb{E}_3$
We call a vector field $\vec{f}$ in a domain $D \subset \mathbb{E}_3$ a *solenoidal field* in D if there exists a vector field (= a vector function) $\vec{w}$ in D such that

(V.4.1) $$\vec{f} \;=\; \operatorname{curl} \vec{w}$$

in D. The vector function $\vec{w}$ is then called a *vector potential* of the field $\vec{f}$ in D.

You have surely noticed a certain similarity between the definition of a potential field and a solenoidal field in a domain $D \subset \mathbb{E}_3$: the field $\vec{f}$ is potential if it can be written in the form $\vec{f} = \operatorname{grad}\varphi$, and solenoidal if it can be written in the form $\vec{f} = \operatorname{curl}\vec{w}$. (The functions φ and $\vec{w}$ are then called the potential and the vector potential, respectively, of the field $\vec{f}$.) We shall meet another similarity in the following theorem: whereas $\vec{f}$ is potential in a domain D precisely when the circulation of $\vec{f}$ along every closed curve in D is zero (see Theorem V.1.7), the field $\vec{f}$ is solenoidal in a domain D precisely when its flux through every closed surface in D is zero.

V.4.2. Theorem
Suppose that $\vec{f}$ is a continuous vector field in a domain $D \subset \mathbb{E}_3$. Then the following two statements are equivalent:

a) $\vec{f}$ is a solenoidal field in D.

b) The flux of the field $\vec{f}$ through every closed surface in D is zero.

P r o o f :
The implication a) $\Longrightarrow$ b) is an immediate consequence of Remark IV.5.11: assume that the

field $\vec{f}$ is solenoidal in D and that $\vec{w}$ is its vector potential. The flux of $\vec{f}$ through a closed surface σ in D equals

$$\iint_\sigma \vec{f}\cdot d\vec{p} = \iint_\sigma \operatorname{curl}\vec{w}\cdot d\vec{p} = 0.$$

The proof of the converse implication, i.e. b) $\Longrightarrow$ a), is more demanding and we do not present it in this text.

V.4.3. Remark

Solenoidal fields are also often referred to by the term **source-free fields**. Theorem V.4.2 shows the reason for this name. Indeed, if there were some sources inside the field $\vec{f}$, then the flux of $\vec{f}$ through a closed surface enclosing these sources could not be zero, and hence the field would not be solenoidal.

V.4.4. Theorem. (Solenoidal field – necessary condition.)
Suppose that $\vec{f}$ is a solenoidal field in a domain $D \subset \mathbb{E}_3$ whose partial derivatives are continuous in D. Then

(V.4.2) $$\operatorname{div}\vec{f} = 0 \quad \text{in } D.$$

P r o o f :
If $\vec{w}$ is a vector potential of $\vec{f}$ in D, then $\vec{f} = \operatorname{curl}\vec{w}$ in D and condition (V.4.2) is an immediate consequence of formula (IV.5.2).

V.4.5. Remark

Condition (V.4.2) is a necessary condition for the field $\vec{f}$ to be solenoidal in a domain D. However, it is not sufficient. This means that from the validity of condition (V.4.2) alone one cannot decide whether $\vec{f}$ is solenoidal in D or not. This can be illustrated by the example of the gravitational field

$$\vec{g}(x,y,z) = -\frac{\kappa m\,(x\hat{\imath}+y\hat{\jmath}+z\hat{k})}{(x^2+y^2+z^2)^{3/2}} \quad (\text{for } x^2+y^2+z^2>0)$$

in the domain $D = \{[x;y;z]\in\mathbb{E}_3 :\ x^2+y^2+z^2>0\}$ from Section V.3.10. By computing the divergence of $\vec{g}$ you can find that

$$\operatorname{div}\vec{g} = -\frac{\partial}{\partial x}\frac{\kappa m\,x}{(x^2+y^2+z^2)^{3/2}} - \frac{\partial}{\partial y}\frac{\kappa m\,y}{(x^2+y^2+z^2)^{3/2}} - \frac{\partial}{\partial z}\frac{\kappa m\,z}{(x^2+y^2+z^2)^{3/2}} =$$

$$= -\frac{3\kappa m}{(x^2+y^2+z^2)^{3/2}} + \frac{3\kappa m\,(x^2+y^2+z^2)}{(x^2+y^2+z^2)^{5/2}} = 0$$

in D. Now let us compute the flux of the field $\vec{g}$ through the sphere $\sigma :\ x^2+y^2+z^2 = R^2$ ($R>0$), oriented outward. We cannot use the Gauss–Ostrogradsky theorem, because the vector field does not have continuous partial derivatives — in fact it is not even defined — at the point $[0;0;0]$, which lies in $Int\,\sigma$. Nevertheless, we can easily compute the flux of $\vec{g}$ through σ if we observe that on σ the normal vector is

$$\hat{n}(x,y,z) = \frac{x\hat{\imath}+y\hat{\jmath}+z\hat{k}}{\sqrt{x^2+y^2+z^2}} = \frac{x\hat{\imath}+y\hat{\jmath}+z\hat{k}}{R},$$

the vector function $\vec{g}$ has on σ the form

$$\vec{g}(x,y,z) \;=\; -\frac{\kappa m\,(x\widehat{\imath}+y\widehat{\jmath}+z\widehat{k})}{(x^2+y^2+z^2)^{3/2}} \;=\; -\frac{\kappa m\,(x\widehat{\imath}+y\widehat{\jmath}+z\widehat{k})}{R^3}$$

and therefore the scalar product $\vec{g}\cdot\widehat{n}$ is

$$\vec{g}(x,y,z)\cdot\widehat{n}(x,y,z) \;=\; -\frac{\kappa m\,(x\widehat{\imath}+y\widehat{\jmath}+z\widehat{k})}{R^3}\cdot\frac{(x\widehat{\imath}+y\widehat{\jmath}+z\widehat{k})}{R} \;=\; -\frac{\kappa m}{R^2}.$$

Hence we obtain

$$\iint_\sigma \vec{g}\cdot d\vec{p} \;=\; \iint_\sigma \vec{g}\cdot\widehat{n}\;dp \;=\; -\frac{\kappa m}{R^2}\iint_\sigma dp \;=\; -\frac{\kappa m}{R^2}\,4\pi R^2 \;=\; -4\kappa m\pi.$$

(We used the well-known formula for the area of a sphere.) Since the flux of $\vec{g}$ through the closed surface σ, lying in the domain D, is nonzero, the field $\vec{g}$ cannot be solenoidal in D. (That would contradict Theorem V.4.2.)

However, condition (V.4.2) can be used as follows: if a vector field in a domain D does not satisfy condition (V.4.2), then it is not solenoidal in D.

V.4.6. Example

Decide whether the vector field $\vec{f}(x,y,z) = (x+y;\; y^2+z^2;\; z^3+x^3)$ is a solenoidal field in $\mathbb{E}_3$.

S o l u t i o n :
By computing the divergence of $\vec{f}$ we obtain

$$\operatorname{div}\vec{f}(x,y,z) \;=\; \frac{\partial}{\partial x}(x+y)+\frac{\partial}{\partial y}(y^2+z^2)+\frac{\partial}{\partial z}(z^3+x^3) \;=\; 1+2y+3z^2.$$

Since the result is not the zero function in $\mathbb{E}_3$, the vector field $\vec{f}$ is not solenoidal in $\mathbb{E}_3$.

In the following sections we shall formulate a sufficient condition whose validity guarantees that a given vector field $\vec{f}$ is solenoidal in a domain $D\subset\mathbb{E}_3$. For this we will need the notion of a "surface-simply connected domain" in $\mathbb{E}_3$.

V.4.7. Surface-simply connected domain in $\mathbb{E}_3$

A domain $D\subset\mathbb{E}_3$ is called *surface-simply connected* if every closed surface σ in D can be continuously deformed (contracted) to a point in D without ever leaving the domain D.

A surface-simply connected domain in $\mathbb{E}_3$ could also be defined as a domain D with the property that the interior of every closed surface in D is a subset of D.

Examples of surface-simply connected domains in $\mathbb{E}_3$ are the whole space $\mathbb{E}_3$, a half-space, $\mathbb{E}_3$ minus a ray, $\mathbb{E}_3$ minus an entire line, the interior of a torus, a ball, a rectangular box, a cube, or any other closed surface, etc. Examples of domains in $\mathbb{E}_3$ that are not surface-simply connected are $\mathbb{E}_3$ minus a point, $\mathbb{E}_3$ minus a bounded closed set, the exterior of any closed surface, etc.

V.4.8. Remark

We warn the reader that the term "surface-simply connected domain" is not standardized. Some authors omit the word "surface". Then, however, confusion may arise with a "simply connected domain", defined in Section V.3.6.

In Definition V.4.7, as in Definitions V.2.4 and V.3.6, we again rely on the reader's intuition and believe that they will correctly understand what we mean by the statement "to contract continuously to a point in D without leaving D".

V.4.9. Theorem. (Solenoidal field – sufficient condition.) *Let*

a) D *is a surface-simply connected domain in* $\mathbb{E}_3$, *and*

b) $\vec{f}$ *is a vector field in* D *whose partial derivatives are continuous in* D *and which satisfies condition (V.4.2):* $\operatorname{div} \vec{f} = 0$ *in* D.

Then $\vec{f}$ *is a solenoidal field in* D.

P r o o f :
We show that the flux of the vector field $\vec{f}$ through every closed surface in D is zero. This implies that $\vec{f}$ is a solenoidal field in D. (See Theorem V.4.2.) Let σ be an arbitrary closed surface in D. Since D is surface-simply connected, $Int\,\sigma \subset D$. By the Gauss–Ostrogradsky theorem we obtain

$$\iint_{\sigma} \vec{f} \cdot d\vec{p} = \pm \iiint_{Int\,\sigma} \operatorname{div} \vec{f}\, dx\, dy\, dz = 0.$$

(The sign "+" applies if σ is oriented outward and "−" applies if σ is oriented inward. Since the integral on the right-hand side equals zero, the sign is not important.)

From the proof it is clear why we assume that the domain D is surface-simply connected: without this we do not know whether $Int\,\sigma \subset D$ and whether we may apply the Gauss–Ostrogradsky theorem.

V.4.10. How to determine whether a given vector field is solenoidal

Suppose that the vector function $\vec{f}$ has continuous partial derivatives in a domain $D \subset \mathbb{E}_3$. If we want to decide whether $\vec{f}$ is a solenoidal field in D, we may proceed according to the following scheme:

Is the equality $\operatorname{div} \vec{f} = 0$ *satisfied in* D*?*

- *YES* — *Is* D *surface-simply connected domain?*
  - *YES* — *Then* $\vec{f}$ *is solenoidal in* D. (Theorem V.4.9.)
  - *NO* — *By Theorem V.4.9 it is impossible to decide whether* $\vec{f}$ *is solenoidal in* D.
- *NO* — *Then* $\vec{f}$ *is not solenoidal in* D. (Theorem V.4.4.)

V.4.11. Example

$\vec{f}(x,y,z) = \big(2x-y; z-y; x-z\big)$ Investigate whether $\vec{f}$ is a solenoidal field in $\mathbb{E}_3$.

S o l u t i o n :
Condition (V.4.2) is satisfied, since

$$\operatorname{div}\vec{f}(x,y,z) \;=\; \frac{\partial}{\partial x}(2x-y) \;+\; \frac{\partial}{\partial y}(z-y) \;+\; \frac{\partial}{\partial z}(x-z) \;=\; 2-1-1 \;=\; 0$$

in $\mathbb{E}_3$. The whole space $\mathbb{E}_3$ is a surface-simply connected domain. Hence, by Theorem V.4.9, the field $\vec{f}$ is solenoidal in $\mathbb{E}_3$.

V.4.12. Example – electrostatic and magnetic field

Imagine that a domain $D \subset \mathbb{E}_3$ is filled with a homogeneous medium with permittivity ϵ and permeability μ. An electric current with density $\mathbf{J}$ flows in D, and electric charges are distributed with density ρ. Let $\mathbf{E}$ denote the electric field intensity and $\mathbf{B}$ the magnetic field intensity in D. Assume that neither $\mathbf{E}$ nor $\mathbf{B}$ depends on time and thus they are vector functions only of the spatial variables x, y, and z. In such a case, instead of "electric field" we more often use the term "electrostatic field". In electromagnetic theory one proves that the vector fields $\mathbf{E}$ and $\mathbf{B}$ satisfy in the domain D the so-called *Maxwell equations*:

$$\begin{aligned} \operatorname{curl}\mathbf{E} &= \vec{0}, \qquad & \operatorname{div}\mathbf{E} &= \rho/\epsilon, \\ \operatorname{curl}\mathbf{B} &= \mu\,\mathbf{J}, \qquad & \operatorname{div}\mathbf{B} &= 0. \end{aligned}$$

Let us think about what the Maxwell equations say about the nature of the fields $\mathbf{E}$ and $\mathbf{B}$. For simplicity, assume that the domain D is both simply connected and surface-simply connected.

From the equation $\operatorname{curl}\mathbf{E} = \vec{0}$ it follows that $\mathbf{E}$ is a potential field in D. *(Theorem V.3.7.)* The equation $\operatorname{div}\mathbf{E} = \rho/\epsilon$ shows that $\mathbf{E}$ in general need not be a solenoidal field in D. *(Theorem V.4.3.)* The field $\mathbf{E}$ is solenoidal in D if and only if $\rho = 0$ in D. *(Theorem V.4.9.)*

The equation $\operatorname{curl}\mathbf{B} = \mu\,\mathbf{J}$ shows that the field $\mathbf{B}$ is in general not a potential field in D. *(Theorem V.3.1.)* The field $\mathbf{B}$ is a potential field in D only if the current density $\mathbf{J}$ is zero in D. *(Theorem V.3.7.)* From the equation $\operatorname{div}\mathbf{B} = 0$ it follows that $\mathbf{B}$ is a solenoidal field in D. *(Theorem V.4.9.)*

Thus, the field $\mathbf{E}$ is irrotational in D, and the field $\mathbf{B}$ has no sources (= source points) in D. If $\rho = 0$ in D, then the field $\mathbf{E}$ also has no sources in D. If the electric current density in D is zero, then the field $\mathbf{B}$ is also irrotational in D.

V.4.13. Two-dimensional solenoidal field

Let D be a domain in $\mathbb{E}_2$ and let a vector potential $\vec{w}$ in the domain $\mathcal{D} = D \times \mathbb{E}_1 = \{[x;y;z] \in \mathbb{E}_3 \,:\, [x;y] \in D,\ z \in \mathbb{E}_1\}$ have the form $\vec{w} = \big(0;\,0;\,\psi(x,y)\big)$. Then the corresponding solenoidal vector field in $\mathcal{D}$ is

$$\vec{f} \;=\; \operatorname{curl}\vec{w} \;=\; \left(\frac{\partial\psi}{\partial y}(x,y);\; -\frac{\partial\psi}{\partial x}(x,y);\; 0\right).$$

We see that $\vec{f}$ has its third component equal to zero and the first two components depend only on the variables x and y. Such a vector field can be regarded as a two-dimensional field, i.e. as a vector field in the domain D in $\mathbb{E}_2$. If we denote by U and V the first two component functions of $\vec{f}$, then

(V.4.3) $$U(x,y) = \frac{\partial\psi}{\partial y}(x,y), \quad V(x,y) = -\frac{\partial\psi}{\partial x}(x,y)$$

for $[x;y] \in D$. A vector field $(U;V)$ whose component functions U and V can be expressed in this way (i.e. there exists a scalar function ψ in D such that U and V satisfy equations (V.4.3)) is called a *solenoidal field* in the domain D in $\mathbb{E}_2$.

From equations (V.4.3) it is immediately clear that the field $(U;V)$ is solenoidal in a domain $D \subset \mathbb{E}_2$ if and only if the corresponding perpendicular field $(-V;U)$ is potential in the same domain D. (The potential of the field $(-V;U)$ is evidently the function ψ.) The questions of how to recognize whether a given field $(U;V)$ is solenoidal in a domain D, what is the necessary condition for $(U;V)$ to be solenoidal in D, what is a sufficient condition, etc., can now be answered easily. We simply replace the field $(U;V)$ by $(-V;U)$, replace the term "solenoidal" by "potential", and use the results of Chapter V.2 on two-dimensional potential fields. If, for simplicity, we restrict ourselves to a vector field $(U;V)$ whose component functions U and V have continuous partial derivatives in D, then in particular we obtain:

a) A necessary condition for the field $(U;V)$ to be solenoidal in D is the validity of the equation

(V.4.4) $$\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} = 0 \quad \text{in } D.$$

b) A sufficient condition guaranteeing that the field $(U;V)$ is solenoidal in D is the simultaneous validity of equation (V.4.4) and the assumption that the domain D is simply connected.

Finally, let us mention another property that characterizes a two-dimensional solenoidal field $(U;V)$. From Theorem V.1.7 we know that the field $(-V;U)$ is potential in D precisely when its circulation along every closed curve C in D is zero, that is,

(V.4.5) $$0 = \oint_C (-V;U)\cdot\widehat{\tau}\,ds = \oint_C (U;V)\cdot\widehat{\nu}\,ds,$$

where $\widehat{\tau} = (\tau_1;\tau_2)$ is the unit tangent vector to the curve C and $\widehat{\nu} = (\tau_2;-\tau_1)$ is the unit normal (perpendicular) vector to the curve C. The last integral on the right-hand side of (V.4.5) can, in analogy with the name of the surface integral on the right-hand side of (IV.4.1), be called the *flux of the vector field $(U;V)$ through the curve C*. We have thus derived the statement: *The vector field $(U;V)$ is solenoidal in a domain $D \subset \mathbb{E}_2$ precisely when its flux through every closed curve C in D is zero.* (In other words, and not quite precisely: whatever flows into the interior of the curve C also flows out.)

V.5 Exercises

1. Find the maximal domains in $\mathbb{E}_2$ in which the vector field $\vec{f}$ is defined. In each domain found, verify whether $\vec{f}$ is a potential field there. If it is, find the potential φ and compute the integral $\int_A^B \vec{f}\cdot d\vec{s}$. *(See Sections V.2.8, V.2.10 and V.1.5.)*

a) $\vec{f}(x,y) = (y^2; 2xy),\ A=[1;3],\ B=[3;2]$,

b) $\vec{f}(x,y) = \dfrac{-\widehat{\imath}+\widehat{\jmath}}{(x-y)^2},\ A=[1;2],\ B=[4;1]$,

c) $\vec{f}(x,y) = (x^2; y^2),\ A=[0;0],\ B=[3;5]$,

d) $\vec{f}(x,y) = \left(y^2 - \dfrac{x}{\sqrt{y-x^2}} - 1;\ 2xy + \dfrac{1}{2\sqrt{y-x^2}}\right),\ A=[0;1],\ B=[1;2]$,

e) $\vec{f}(x,y) = \dfrac{y^2}{\sqrt{x}}\widehat{\imath} + 4y\sqrt{x}\,\widehat{\jmath},\ A=[1;2],\ B=[4;-2]$,

f) $\vec{f}(x,y) = \left(1 - y^2 + \dfrac{1}{2\sqrt{y^2+x}};\ \dfrac{y}{\sqrt{y^2+x}} - 2xy\right),\ A=[-3;2],\ B=[3;1]$,

g) $\vec{f}(x,y) = (xy; x+y),\ A=[0;0],\ B=[1;1]$,

h) $\vec{f}(x,y) = \dfrac{-y\widehat{\imath}+x\widehat{\jmath}}{(x-y)^2},\ A=[2;1],\ B=[6;2]$,

i) $\vec{f}(x,y) = \left(x^3y^2 + x;\ y^2 + yx^4\right),\ A=[3;-1],\ B=[1;5]$,

j) $\vec{f}(x,y) = \left(1 + y^2\,\sin\,2x;\ -2y\,\cos^2 x\right),\ A=[\pi;1],\ B=[\pi/2;2]$,

k) $\vec{f}(x,y) = \left(\ln\,y - \dfrac{\mathrm{e}^y}{x^2};\ \dfrac{\mathrm{e}^y}{x} + \dfrac{x}{y}\right),\ A=[1;1],\ B=[1;2]$,

l) $\vec{f}(x,y) = \left(\dfrac{x-2y}{(y-x)^2} + x;\ \dfrac{y}{(y-x)^2} - y^2\right),\ A=[0;1],\ B=[1;4]$,

m) $\vec{f}(x,y) = \dfrac{4x\widehat{\imath}+y\widehat{\jmath}}{4x^2+y^2-4},\ A=[0;0],\ B=[0;2]$,

n) $\vec{f}(x,y) = (y\,\sin\,x;\ y-\cos\,x),\ A=[0;1],\ B=[5;2]$,

o) $\vec{f}(x,y) = (\cos(2y)+y+x;\ y-2x\,\sin(2y)+x),\ A=[0;0],\ B=[-2;2]$.

2. In what maximal domain $D\subset\mathbb{E}_3$ is the function φ a potential of the vector field $\vec{f}$? Determine the field $\vec{f}$? Compute the work done by the force $\vec{f}$ along a curve leading from point A to point B. *(See Sections V.1.3 and V.1.5.)*

a) $\varphi(x,y,z) = xy+xz+yz,\ A=[-1;2;-1],\ B=[3;4;1]$,

b) $\varphi(x,y,z) = \ln\,|x^2+y^2+z^2-1|,\ A=[-1;1;2],\ B=[-3;4;-1]$.

3. Find the maximal domains in $\mathbb{E}_3$ in which the vector field $\vec{f}$ is defined. In each domain verify whether the field $\vec{f}$ is potential. If it is, determine a potential.
(See Sections V.3.8 and V.2.10.)

a) $\vec{f}(x,y,z) = \left(\dfrac{y^2}{z};\ \dfrac{2xy}{z};\ -\dfrac{xy^2}{z^2}\right)$, b) $\vec{f}(x,y,z) = \left(\dfrac{x}{x^2+y^2};\ \dfrac{y}{x^2+y^2};\ 2z\right)$,

c) $\vec{f}(x,y,z) = \dfrac{-x\,\widehat{\imath} - y\,\widehat{\jmath} + \widehat{k}/2}{\sqrt{z - x^2 - y^2}}$, d) $\vec{f}(x,y,z) = \left(\dfrac{2x-y}{x^2+y^2};\ \dfrac{x+2y}{x^2+y^2};\ \ln z\right)$,

e) $\vec{f}(x,y,z) = \left(\dfrac{1}{\sqrt{x}}\sin z - y^2;\ -2xy;\ 2\sqrt{x}\cos z\right)$,

f) $\vec{f}(x,y,z) = y^2\widehat{\imath} + z^2\widehat{\jmath} + x^2\widehat{k}$,

g) $\vec{f}(x,y,z) = \left(\dfrac{z}{x-y};\ \dfrac{z}{y-x};\ \ln(x-y) + \dfrac{1}{\sqrt{z}}\right)$,

h) $\vec{f}(x,y,z) = \dfrac{x\,\widehat{\imath} + y\,\widehat{\jmath} + z\,\widehat{k}}{\sqrt{x^2+y^2+z^2}}$,

i) $\vec{f}(x,y,z) = \left(\mathrm{e}^x(y+z^2);\ \mathrm{e}^x;\ z\mathrm{e}^x\right)$,

j) $\vec{f}(x,y,z) = \left(x-y;\ y^2;\ x+z\right)$,

k) $\vec{f}(x,y,z) = \left(\dfrac{-y}{x^2+y^2};\ \dfrac{x}{x^2+y^2};\ 2z\right)$,

l) $\vec{f}(x,y,z) = \left(\dfrac{x}{\sqrt{x^2+y^2}};\ \dfrac{y}{\sqrt{x^2+y^2}};\ \dfrac{1}{z}\right)$.

4.* In what maximal domains in $\mathbb{E}_3$ is the vector field $\vec{f}$ defined? In each domain found, verify whether the field $\vec{f}$ is solenoidal. *(See V.4.11.)*

a) $\vec{f}(x,y,z) = \left(y^2; z^2; x^2\right)$, b) $\vec{f}(x,y,z) = \left(z-y;\ x-z;\ y-z\right)$,

c) $\vec{f}(x,y,z) = \left(x;\ y;\ -2z\right)$, d) $\vec{f}(x,y,z) = \left(\dfrac{x}{y};\ \sqrt{z-x^2};\ -\dfrac{z}{y}\right)$,

e) $\vec{f}(x,y,z) = \left(y;\ z;\ x^2\right)$, f) $\vec{f}(x,y,z) = \left(xy;\ 1-y^2;\ yz\right)$,

g) $\vec{f}(x,y,z) = \dfrac{(yz;\ xz;\ -xy)}{y^2+z^2-1}$, h) $\vec{f}(x,y,z) = \left(5x^2;\ y-z;\ \ln z\right)$.

5.* $\vec{f}(x,y,z) = \left(\dfrac{x-y+z}{(x^2+y^2+z^2)^{3/2}};\ \dfrac{x+y-z}{(x^2+y^2+z^2)^{3/2}};\ \dfrac{-x+y+z}{(x^2+y^2+z^2)^{3/2}}\right)$

a) Show that $\operatorname{div}\vec{f} = 0$ in $\mathbb{E}_3\backslash\{[0;0;0]\}$. *(See Section IV.5.1.)*

b) Compute the flux of the field $\vec{f}$ through the spherical surface centered at the origin with radius $r = 1$, oriented outward. *(See Section V.4.5.)*

c) Decide whether $\vec{f}$ is a solenoidal field in $\mathbb{E}_3\backslash\{[0;0;0]\}$. Justify your answer. *(See Section V.4.5.)*

d) Determine a domain in $\mathbb{E}_3$ in which the field $\vec{f}$ is solenoidal. *(See Section V.4.11.)*

Recommended literature

1. E. Brožíková, M. Kittlerová: *Collection of Solved Problems in Mathematics II. (Sbírka řešených příkladů z Matematiky II.)* Publishing House of CTU, Prague 2001. (A collection of solved as well as unsolved problems, intended for classwork and for independent study.)
2. B. Budínský, J. Charvát: *Matematika II.* Publishing House of CTU, Prague 2002. (A detailed textbook for faculties of civil engineering.)
3. J. Neustupa: *Mathematics I (Matematika I)* Publishing House of CTU, Prague 2014. (Recommended lecture notes for the study of Mathematics I – in Czech.)
4. J. Neustupa, S. Kračmar S.: *Mathematics II, CTU Publishing House, Prague, 1998* (Original lecture notes for the study of Mathematics II.)

Supplementary literature

5. J. Neustupa: *Mathematics I*, CTU Publishing House, Prague, 1996 (Recommended lecture notes for the study of Mathematics I.)
6. Deborah Hughes-Hallett, Andrew M. Gleason, William G. McCallum *Calculus: Single and Multivariable, 8th Edition* (One of many textbooks on the fundamentals of differential and integral calculus, used especially at American universities.)