Prefaces

Preface to the English Edition

An entire generation of mathematicians has grown up during the time between the appearance of the first edition of this textbook and the publication of the fourth edition, a translation of which is before you. The book is familiar to many people, who either attended the lectures on which it is based or studied out of it, and who now teach others in universities all over the world. I am glad that it has become accessible to English-speaking readers.

This textbook consists of two parts. It is aimed primarily at university students and teachers specializing in mathematics and natural sciences, and at all those who wish to see both the rigorous mathematical theory and examples of its effective use in the solution of real problems of natural science.

The textbook exposes classical analysis as it is today, as an integral part of Mathematics in its interrelations with other modern mathematical courses such as algebra, differential geometry, differential equations, complex and functional analysis.

The two chapters with which this second book begins, summarize and explain in a general form essentially all most important results of the first volume concerning continuous and differentiable functions, as well as differential calculus. The presence of these two chapters makes the second book formally independent of the first one. This assumes, however, that the reader is sufficiently well prepared to get by without introductory considerations of the first part, which preceded the resulting formalism discussed here. This second book, containing both the differential calculus in its generalized form and integral calculus of functions of several variables, developed up to the general formula of Newton–Leibniz–Stokes, thus acquires a certain unity and becomes more self-contained.

More complete information on the textbook and some recommendations for its use in teaching can be found in the translations of the prefaces to the first and second Russian editions.

Moscow, 2003 *V. Zorich*

Preface to the Fourth Russian Edition

In the fourth edition all misprints that the author is aware of have been corrected.

Moscow, 2002 *V. Zorich*

Preface to the Third Russian Edition

The third edition differs from the second only in local corrections (although in one case it also involves the correction of a proof) and in the addition of some problems that seem to me to be useful.

Moscow, 2001 *V. Zorich*

Preface to the Second Russian Edition

In addition to the correction of all the misprints in the first edition of which the author is aware, the differences between the second edition and the first edition of this book are mainly the following. Certain sections on individual topics – for example, Fourier series and the Fourier transform – have been recast (for the better, I hope). We have included several new examples of applications and new substantive problems relating to various parts of the theory and sometimes significantly extending it. Test questions are given, as well as questions and problems from the midterm examinations. The list of further readings has been expanded.

Further information on the material and some characteristics of this second part of the course are given below in the preface to the first edition.

Moscow, 1998 *V. Zorich*

Preface to the First Russian Edition

The preface to the first part contained a rather detailed characterization of the course as a whole, and hence I confine myself here to some remarks on the content of the second part only.

The basic material of the present volume consists on the one hand of multiple integrals and line and surface integrals, leading to the generalized Stokes' formula and some examples of its application, and on the other hand the machinery of series and integrals depending on a parameter, including

Fourier series, the Fourier transform, and the presentation of asymptotic expansions.

Thus, this Part 2 basically conforms to the curriculum of the second year of study in the mathematics departments of universities.

So as not to impose rigid restrictions on the order of presentation of these two major topics during the two semesters, I have discussed them practically independently of each other.

Chapters 9 and 10, with which this book begins, reproduce in compressed and generalized form, essentially all of the most important results that were obtained in the first part concerning continuous and differentiable functions. These chapters are starred and written as an appendix to Part 1. This appendix contains, however, many concepts that play a role in any exposition of analysis to mathematicians. The presence of these two chapters makes the second book formally independent of the first, provided the reader is sufficiently well prepared to get by without the numerous examples and introductory considerations that, in the first part, preceded the formalism discussed here.

The main new material in the book, which is devoted to the integral calculus of several variables, begins in Chapter 11. One who has completed the first part may begin the second part of the course at this point without any loss of continuity in the ideas.

The language of differential forms is explained and used in the discussion of the theory of line and surface integrals. All the basic geometric concepts and analytic constructions that later form a scale of abstract definitions leading to the generalized Stokes' formula are first introduced by using elementary material.

Chapter 15 is devoted to a similar summary exposition of the integration of differential forms on manifolds. I regard this chapter as a very desirable and systematizing supplement to what was expounded and explained using specific objects in the mandatory Chapters 11–14.

The section on series and integrals depending on a parameter gives, along with the traditional material, some elementary information on asymptotic series and asymptotics of integrals (Chap. 19), since, due to its effectiveness, the latter is an unquestionably useful piece of analytic machinery.

For convenience in orientation, ancillary material or sections that may be omitted on a first reading, are starred.

The numbering of the chapters and figures in this book continues the numbering of the first part.

Biographical information is given here only for those scholars not mentioned in the first part.

As before, for the convenience of the reader, and to shorten the text, the end of a proof is denoted by $\square$. Where convenient, definitions are introduced by the special symbols $:=$ or $=:$ (equality by definition), in which the colon stands on the side of the object being defined.

Continuing the tradition of Part 1, a great deal of attention has been paid to both the lucidity and logical clarity of the mathematical constructions themselves and the demonstration of substantive applications in natural science for the theory developed.

Moscow, 1982 *V. Zorich*

10 *Differential Calculus from a more General Point of View

10.1 Normed Vector Spaces

Differentiation is the process of finding the best local linear approximation of a function. For that reason any reasonably general theory of differentiation must be based on elementary ideas connected with linear functions. From the course in algebra the reader is well acquainted with the concept of a *vector space*, as well as linear dependence and independence of systems of vectors, bases and dimension of a vector space, vector subspaces, and so forth. In the present section we shall present vector spaces with a norm, or as they are described, *normed vector spaces*, which are widely used in analysis. We begin, however, with some examples of vector spaces.

10.1.1 Some Examples of Vector Spaces in Analysis

Example 1. The real vector space $\mathbb{R}^n$ and the complex vector space $\mathbb{C}^n$ are classical examples of vector spaces of dimension n over the fields of real and complex numbers respectively.

Example 2. In analysis, besides the spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ exhibited in Example 1, we encounter the space closest to them, which is the space ℓ of sequences $x = (x^1, \dots, x^n, \dots)$ of real or complex numbers. The vector-space operations in ℓ, as in $\mathbb{R}^n$ and $\mathbb{C}^n$, are carried out coordinatewise. One peculiarity of this space, when compared with $\mathbb{R}^n$ or $\mathbb{C}^n$ is that any finite subsystem of the countable system of vectors $\{x_i = (0, \dots, 0, x^i = 1, 0, \dots),\ i \in \mathbb{N}\}$ is linearly independent, that is, ℓ is an infinite-dimensional vector space (of countable dimension in the present case).

The set of finite sequences (all of whose terms are zero from some point on) is a vector subspace $\underset{0}{\ell}$ of the space ℓ, also infinite-dimensional.

Example 3. Let $F[a, b]$ be the set of numerical-valued (real- or complex-valued) functions defined on the closed interval $[a, b]$. This set is a vector space over the corresponding number field with respect to the operations of addition of functions and multiplication of a function by a number.

The set of functions of the form

$$e_\tau(x) = \begin{cases} 0\,, & \text{if} \quad x \in [a,b] \quad \text{and} \quad x \neq \tau\,, \\ 1\,, & \text{if} \quad x \in [a,b] \quad \text{and} \quad x = \tau \end{cases}$$

is a continuously indexed system of linearly independent vectors in $F[a,b]$.

The set $C[a,b]$ of continuous functions is obviously a subspace of the space $F[a,b]$ just constructed.

Example 4. If X_1 and X_2 are two vector spaces over the same field, there is a natural way of introducing a vector-space structure into their direct product $X_1 \times X_2$, namely by carrying out the vector-space operations on elements $x = (x_1, x_2) \in X_1 \times X_2$ coordinatewise.

Similarly one can introduce a vector-space structure into the direct product $X_1 \times \cdots \times X_n$ of any finite set of vector spaces. This is completely analogous to the cases of $\mathbb{R}^n$ and $\mathbb{C}^n$.

10.1.2 Norms in Vector Spaces

We begin with the basic definition.

Definition 1. Let X be a vector space over the field of real or complex numbers.

A function $\| \ \| : X \to \mathbb{R}$ assigning to each vector $x \in X$ a real number $\|x\|$ is called a *norm* in the vector space X if it satisfies the following three conditions:

a) $\|x\| = 0 \Leftrightarrow x = 0$ (nondegeneracy);
b) $\|\lambda x\| = |\lambda|\, \|x\|$ (homogeneity);
c) $\|x_1 + x_2\| \leq \|x_1\| + \|x_2\|$ (the triangle inequality).

Definition 2. A vector space with a norm defined on it is called a *normed vector space.*

Definition 3. The value of the norm at a vector is called the *norm of that vector.*

The norm of a vector is always nonnegative and, as can be seen by a), equals zero only for the zero vector.

Proof. Indeed, by c), taking account of a) and b), we obtain for every $x \in X$,

$$0 = \|0\| = \|x + (-x)\| \leq \|x\| + \|-x\| = \|x\| + |-1|\, \|x\| = 2\|x\|\,. \ \square$$

By induction, condition c) implies the following general inequality.

$$\|x_1 + \cdots + x_n\| \leq \|x_1\| + \cdots + \|x_n\|\,, \tag{10.1}$$

and taking account of b), one can easily deduce from c) the following useful inequality.

$$\big| \|x_1\| - \|x_2\| \big| \le \|x_1 - x_2\| \ . \tag{10.2}$$

Every normed vector space has a natural metric

$$d(x_1, x_2) = \|x_1 - x_2\| \ . \tag{10.3}$$

The fact that the function $d(x_1, x_2)$ just defined satisfies the axioms for a metric follows immediately from the properties of the norm. Because of the vector-space structure in X the metric d in X has two additional special properties:

$$d(x_1 + x, x_2 + x) = \|(x_1 + x) - (x_2 + x)\| = \|x_1 - x_2\| = d(x_1, x_2) \ ,$$

that is, the metric is translation-invariant, and

$$d(\lambda x_1, \lambda x_2) = \|\lambda x_1 - \lambda x_2\| = \|\lambda(x_1 - x_2)\| = |\lambda|\, \|x_1 - x_2\| = |\lambda|\, d(x_1, x_2) \ ,$$

that is, it is homogeneous.

Definition 4. If a normed vector space is complete as a metric space with the natural metric (10.3), it is called a *complete normed vector space* or *Banach space.*

Example 5. If for $p \ge 1$ we set

$$\|x\|_p := \left(\sum_{i=1}^{n} |x^i|^p \right)^{\frac{1}{p}} \tag{10.4}$$

for $x = (x^1, \dots, x^n) \in \mathbb{R}^n$, it follows from Minkowski's inequality that we obtain a norm on $\mathbb{R}^n$. The space $\mathbb{R}^n$ endowed with this norm will be denoted $\mathbb{R}^n_p$.

One can verify that

$$\|x\|_{p_2} \le \|x\|_{p_1} \ , \ \text{ if } 1 \le p_1 \le p_2 \ , \tag{10.5}$$

and that

$$\|x\|_p \to \max\left\{|x^1|, \dots, |x^n|\right\} \tag{10.6}$$

as $p \to +\infty$. Thus, it is natural to set

$$\|x\|_\infty := \max\left\{|x^1|, \dots, |x^n|\right\} \ . \tag{10.7}$$

It then follows from (10.4) and (10.5) that

$$\|x\|_\infty \le \|x\|_p \le \|x\|_1 \le n\|x\|_\infty \ \text{ for } p \ge 1 \ . \tag{10.8}$$

It is clear from this inequality, as in fact it is from the very definition of the norm $\|x\|_p$ in Eq. (10.4), that $\mathbb{R}^n_p$ is a complete normed vector space.

Example 6. The preceding example can be usefully generalized as follows. If $X = X_1 \times \cdots \times X_n$ is the direct product of normed vector spaces, one can introduce the norm of a vector $x = (x_1, \ldots, x_n)$ in the direct product by setting

$$\|x\|_p := \left(\sum_{i=1}^{n} \|x_i\|^p \right)^{\frac{1}{p}}, \quad p \geq 1, \tag{10.9}$$

where $\|x_i\|$ is the norm of the vector $x_i \in X_i$.

Naturally, inequalities (10.8) remain valid in this case as well.

From now on, when the direct product of normed spaces is considered, unless the contrary is explicitly stated, it is assumed that the norm is defined in accordance with formula (10.9) (including the case $p = +\infty$).

Example 7. Let $p \geq 1$. We denote by ℓ_p the set of sequences $x = (x^1, \ldots, x^n, \ldots)$ of real or complex numbers such that the series $\sum\limits_{n=1}^{\infty} |x^n|^p$ converges, and for $x \in \ell_p$ we set

$$\|x\|_p := \left(\sum_{n=1}^{\infty} |x^n|^p \right)^{\frac{1}{p}}. \tag{10.10}$$

Using Minkowski's inequality, one can easily see that ℓ_p is a normed vector space with respect to the standard vector-space operations and the norm (10.10). This is an infinite-dimensional space with respect to which $\mathbb{R}^n_p$ is a vector subspace of finite dimension.

All the inequalities (10.8) except the last are valid for the norm (10.10). It is not difficult to verify that ℓ_p is a Banach space.

Example 8. In the vector space $C[a,b]$ of numerical-valued functions that are continuous on the closed interval $[a,b]$, one usually considers the following norm:

$$\|f\| := \max_{x \in [a,b]} |f(x)|. \tag{10.11}$$

We leave the verification of the norm axioms to the reader. We remark that this norm generates a metric on $C[a,b]$ that is already familiar to us (see Sect. 9.5), and we know that the metric space that thereby arises is complete. Thus the vector space $C[a,b]$ with the norm (10.11) is a Banach space.

Example 9. One can also introduce another norm in $C[a,b]$

$$\|f\|_p := \left(\int_a^b |f|^p(x)\,dx \right)^{\frac{1}{p}}, \quad p \geq 1, \tag{10.12}$$

which becomes (10.11) as $p \to +\infty$.

It is easy to see (for example, Sect. 9.5) that the space $C[a,b]$ with the norm (10.12) is not complete for $1 \leq p < +\infty$.

10.1.3 Inner Products in Vector Spaces

An important class of normed spaces is formed by the spaces with an inner product. They are a direct generalization of Euclidean spaces.

We recall their definition.

Definition 5. We say that a *Hermitian form* is defined in a vector space X (over the field of complex numbers) if there exists a mapping $\langle\, ,\, \rangle : X \times X \to \mathbb{C}$ having the following properties:

a) $\langle x_1, x_2 \rangle = \overline{\langle x_2, x_1 \rangle}$,
b) $\langle \lambda x_1, x_2 \rangle = \lambda \langle x_1, x_2 \rangle$,
c) $\langle x_1 + x_2, x_3 \rangle = \langle x_1, x_3 \rangle + \langle x_2, x_3 \rangle$,
where x_1, x_2, x_3 are vectors in X and $\lambda \in \mathbb{C}$.

It follows from a), b), and c), for example, that

$$\langle x_1, \lambda x_2 \rangle = \overline{\langle \lambda x_2, x_1 \rangle} = \overline{\lambda \langle x_2, x_1 \rangle} = \overline{\lambda}\, \overline{\langle x_2, x_1 \rangle} = \overline{\lambda} \langle x_1, x_2 \rangle \ ;$$
$$\langle x_1, x_2 + x_3 \rangle = \overline{\langle x_2 + x_3, x_1 \rangle} = \overline{\langle x_2, x_1 \rangle} + \overline{\langle x_3, x_1 \rangle} = \langle x_1, x_2 \rangle + \langle x_1, x_3 \rangle \ ;$$
$$\langle x, x \rangle = \overline{\langle x, x \rangle}\ , \ \text{ that is, } \langle x, x \rangle \ \text{ is a real number.}$$

A Hermitian form is called *nonnegative* if
d) $\langle x, x \rangle \geq 0$
and *nondegenerate* if
e) $\langle x, x \rangle = 0 \Leftrightarrow x = 0$.

If X is a vector space over the field of real numbers, one must of course consider a real-valued form $\langle x_1, x_2 \rangle$. In this case a) can be replaced by $\langle x_1, x_2 \rangle = \langle x_2, x_1 \rangle$, which means that the form is symmetric with respect to its vector arguments x_1 and x_2.

An example of such a form is the dot product familiar from analytic geometry for vectors in three-dimensional Euclidean space. In connection with this analogy we make the following definition.

Definition 6. A nondegenerate nonnegative Hermitian form in a vector space is called an *inner product* in the space.

Example 10. An inner product of vectors $x = (x^1, \ldots, x^n)$ and $y = (y^1, \ldots, y^n)$ in $\mathbb{R}^n$ can be defined by setting

$$\langle x, y \rangle := \sum_{i=1}^{n} x^i y^i \ , \tag{10.13}$$

and in $\mathbb{C}^n$ by setting

$$\langle x, y \rangle := \sum_{i=1}^{n} x^i \overline{y^i} \ . \tag{10.14}$$

Example 11. In ℓ_2 the inner product of the vectors x and y can be defined as

$$\langle x, y \rangle := \sum_{i=1}^{\infty} x^i \overline{y^i} \ ,$$

The series in this expression converges absolutely since

$$2 \sum_{i=1}^{\infty} |x^i \overline{y^i}| \le \sum_{i=1}^{\infty} |x^i|^2 + \sum_{i=1}^{\infty} |y^i|^2 \ .$$

Example 12. An inner product can be defined in $C[a, b]$ by the formula

$$\langle f, g \rangle := \int_a^b (f \cdot \bar{g})(x)\, dx \ . \tag{10.15}$$

It follows easily from properties of the integral that all the requirements for an inner product are satisfied in this case.

The following important inequality, known as the *Cauchy–Bunyakovskii inequality*, holds for the inner product:

$$|\langle x, y \rangle|^2 \le \langle x, x \rangle \cdot \langle y, y \rangle \ , \tag{10.16}$$

where equality holds if and only if the vectors x and y are collinear.

Proof. Indeed, let $a = \langle x, x \rangle$, $b = \langle x, y \rangle$, and $c = \langle y, y \rangle$. By hypothesis $a \ge 0$ and $c \ge 0$. If $c > 0$, the inequalities

$$0 \le \langle x + \lambda y, x + \lambda y \rangle = a + \bar{b}\lambda + b\bar{\lambda} + c\lambda\bar{\lambda}$$

with $\lambda = -\frac{b}{c}$ imply

$$0 \le a - \frac{\bar{b}b}{c} - \frac{b\bar{b}}{c} + \frac{b\bar{b}}{c}$$

or

$$0 \le ac - b\bar{b} = ac - |b|^2 \ , \tag{10.17}$$

which is the same as (10.16).

The case $a > 0$ can be handled similarly.

If $a = c = 0$, then, setting $\lambda = -b$ in (10.17), we find $0 \le -\bar{b}b - b\bar{b} = -2|b|^2$, that is, $b = 0$, and (10.16) is again true.

If x and y are not collinear, then $0 < \langle x + \lambda y, x + \lambda y \rangle$ and consequently inequality (10.16) is a strict inequality in this case. But if x and y are collinear, it becomes equality as one can easily verify. □

A vector space with an inner product has a natural norm:

$$\|x\| := \sqrt{\langle x, x\rangle} \tag{10.18}$$

and metric

$$d(x, y) := \|x - y\| \ .$$

Using the Cauchy–Bunyakovskii inequality, we verify that if $\langle x, y\rangle$ is a nondegenerate nonnegative Hermitian form, then formula (10.18) does indeed define a norm.

Proof. In fact,

$$\|x\| = \sqrt{\langle x, x\rangle} = 0 \Leftrightarrow x = 0 \ ,$$

since the form $\langle x, y\rangle$ is nondegenerate.

Next,

$$\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\lambda\bar{\lambda}\langle x, x\rangle} = |\lambda|\sqrt{\langle x, x\rangle} = |\lambda|\,\|x\| \ .$$

We verify finally that the triangle inequality holds:

$$\|x + y\| \le \|x\| + \|y\| \ .$$

Thus, we need to show that

$$\sqrt{\langle x + y, x + y\rangle} \le \sqrt{\langle x, x\rangle} + \sqrt{\langle y, y\rangle} \ ,$$

or, after we square and cancel, that

$$\langle x, y\rangle + \langle y, x\rangle \le 2\sqrt{\langle x, x\rangle \cdot \langle y, y\rangle} \ .$$

But

$$\langle x, y\rangle + \langle y, x\rangle = \langle x, y\rangle + \overline{\langle x, y\rangle} = 2\mathrm{Re}\,\langle x, y\rangle \le 2|\langle x, y\rangle| \ ,$$

and the inequality to be proved now follows immediately from the Cauchy–Bunyakovskii inequality (10.16). □

In conclusion we note that finite-dimensional vector spaces with an inner product are usually called *Euclidean* or *Hermitian* (*unitary*) spaces according as the field of scalars is $\mathbb{R}$ or $\mathbb{C}$ respectively. If a normed vector space is infinite-dimensional, it is called a *Hilbert space* if it is complete in the metric induced by the natural norm and a *pre-Hilbert space* otherwise.

10.1.4 Problems and Exercises

1. a) Show that if a translation-invariant homogeneous metric $d(x_1, x_2)$ is defined in a vector space X, then X can be normed by setting $\|x\| = d(0, x)$.

b) Verify that the norm in a vector space X is a continuous function with respect to the topology induced by the natural metric (10.3).

c) Prove that if X is a finite-dimensional vector space and $\|x\|$ and $\|x\|'$ are two norms on X, then one can find positive numbers M, N such that

$$M\|x\| \le \|x\|' \le N\|x\| \tag{10.19}$$

for any vector $x \in X$.

d) Using the example of the norms $\|x\|_1$ and $\|x\|_\infty$ in the space ℓ, verify that the preceding inequality generally does not hold in infinite-dimensional spaces.

2. a) Prove inequality (10.5).

b) Verify relation (10.6).

c) Show that as $p \to +\infty$ the quantity $\|f\|_p$ defined by formula (10.12) tends to the quantity $\|f\|$ given by formula (10.11).

3. a) Verify that the normed space ℓ_p considered in Example 7 is complete.

b) Show that the subspace of ℓ_p consisting of finite sequences (ending in zeros) is not a Banach space.

4. a) Verify that relations (10.11) and (10.12) define a norm in the space $C[a, b]$ and convince yourself that a complete normed space is obtained in one of these cases but not in the other.

b) Does formula (10.12) define a norm in the space $\mathcal{R}[a, b]$ of Riemann-integrable functions?

c) What factorization (identification) must one make in $\mathcal{R}[a, b]$ so that the quantity defined by (10.12) will be a norm in the resulting vector space?

5. a) Verify that formulas (10.13)–(10.15) do indeed define an inner product in the corresponding vector spaces.

b) Is the form defined by formula (10.15) an inner product in the space $\mathcal{R}[a, b]$ of Riemann-integrable functions?

c) Which functions in $\mathcal{R}[a, b]$ must be identified so that the answer to part b) will be positive in the quotient space of equivalence classes?

6. Using the Cauchy–Bunyakovskii inequality, find the greatest lower bound of the values of the product $\left(\int\limits_a^b f(x)\,dx\right)\left(\int\limits_a^b (1/f)(x)\,dx\right)$ on the set of continuous real-valued functions that do not vanish on the closed interval $[a, b]$.

10.2 Linear and Multilinear Transformations

10.2.1 Definitions and Examples

We begin by recalling the basic definition.

Definition 1. If X and Y are vector spaces over the same field (in our case, either $\mathbb{R}$ or $\mathbb{C}$), a mapping $A : X \to Y$ is *linear* if the equalities

$$\begin{aligned} A(x_1 + x_2) &= A(x_1) + A(x_2)\ , \\ A(\lambda x) &= \lambda A(x) \end{aligned}$$

hold for any vectors x, x_1, x_2 in X and any number λ in the field of scalars.

For a linear transformation $A : X \to Y$ we often write Ax instead of $A(x)$.

Definition 2. A mapping $A : X_1 \times \cdots \times X_n \to Y$ of the direct product of the vector spaces $X_1, \dots, X_n$ into the vector space Y is *multilinear* (n*-linear*) if the mapping $y = A(x_1, \dots, x_n)$ is linear with respect to each variable for all fixed values of the other variables.

The set of n-linear mappings $A : X_1 \times \cdots \times X_n \to Y$ will be denoted $\mathcal{L}(X_1, \dots, X_n; Y)$.

In particular for $n = 1$ we obtain the set $\mathcal{L}(X; Y)$ of linear mappings from $X_1 = X$ into Y.

For $n = 2$ a multilinear mapping is called *bilinear*, for $n = 3$, *trilinear*, and so forth.

One should not confuse an n-linear mapping $A \in \mathcal{L}(X_1, \dots, X_n; Y)$ with a linear mapping $A \in \mathcal{L}(X; Y)$ of the vector space $X = X_1 \times \cdots \times X_n$ (in this connection see Examples 9–11 below).

If $Y = \mathbb{R}$ or $Y = \mathbb{C}$, linear and multilinear mappings are usually called linear or multilinear *functionals*. When Y is an arbitrary vector space, a linear mapping $A : X \to Y$ is usually called a *linear transformation* from X into Y, and a *linear operator* in the special case when $X = Y$.

Let us consider some examples of linear mappings.

Example 1. Let $\underset{0}{\ell}$ be the vector space of finite numerical sequences. We define a transformation $A : \underset{0}{\ell} \to \underset{0}{\ell}$ as follows:

$$A\big((x_1, x_2, \dots, x_n, 0, \dots)\big) := (1x_1, 2x_2, \dots, nx_n, 0, \dots)\ .$$

Example 2. We define the functional $A : C[a, b] \to \mathbb{R}$ by the relation

$$A(f) := f(x_0)\ ,$$

where $f \in C\big([a, b], \mathbb{R}\big)$ and x_0 is a fixed point of the closed interval $[a, b]$.

Example 3. We define the functional $A : C([a, b], \mathbb{R}) \to \mathbb{R}$ by the relation

$$A(f) := \int_a^b f(x)\,dx\ .$$

Example 4. We define the transformation $A : C([a, b], \mathbb{R}) \to C([a, b], \mathbb{R})$ by the formula

$$A(f) := \int_a^x f(t)\,dt\ ,$$

where x is a point ranging over the closed interval $[a, b]$.

All of these transformations are obviously linear.

Let us now consider some familiar examples of multilinear mappings.

Example 5. The usual product $(x_1, \dots, x_n) \mapsto x_1 \cdot \ldots \cdot x_n$ of n real numbers is a typical example of an n-linear functional $A \in \mathcal{L}(\underbrace{\mathbb{R}, \dots, \mathbb{R}}_{n}; \mathbb{R})$.

Example 6. The inner product $(x_1, x_2) \overset{A}{\longmapsto} \langle x_1, x_2 \rangle$ in a Euclidean vector space over the field $\mathbb{R}$ is a bilinear function.

Example 7. The cross product $(x_1, x_2) \overset{A}{\longmapsto} [x_1, x_2]$ of vectors in three-dimensional Euclidean space E^3 is a bilinear transformation, that is, $A \in \mathcal{L}(E^3, E^3; E^3)$.

Example 8. If X is a finite-dimensional vector space over the field $\mathbb{R}$, $\{e_1, \dots, e_n\}$ is a basis in X, and $x = x^i e_i$ is the coordinate representation of the vector $x \in X$, then, setting

$$A(x_1, \dots, x_n) = \det \begin{pmatrix} x_1^1 & \cdots & x_1^n \\ \multicolumn{3}{c}{\dots\dots\dots\dots} \\ x_n^1 & \cdots & x_n^n \end{pmatrix}\ ,$$

we obtain an n-linear function $A : X^n \to \mathbb{R}$.

As a useful supplement to the examples just given, we investigate in addition the structure of the linear mappings of a product of vector spaces into a product of vector spaces.

Example 9. Let $X = X_1 \times \cdots \times X_m$ be the vector space that is the direct product of the spaces $X_1, \dots, X_m$, and let $A : X \to Y$ be a linear mapping of X into a vector space Y. Representing every vector $x = (x_1, \dots, x_m) \in X$ in the form

$$\begin{aligned} x &= (x_1, \dots, x_m) = \\ &= (x_1, 0, \dots, 0) + (0, x_2, 0, \dots, 0) + \dots + (0, \dots, 0, x_m) \end{aligned} \quad (10.20)$$

and setting

$$A_i(x_i) := A\big((0, \dots, 0, x_i, 0, \dots, 0)\big) \quad (10.21)$$

for $x_i \in X_i$, $i = \{1, \dots, m\}$, we observe that the mappings $A_i : X_i \to Y$ are linear and that

$$A(x) = A_1(x_1) + \dots + A_m(x_m) . \quad (10.22)$$

Since the mapping $A : X = X_1 \times \dots \times X_m \to Y$ is obviously linear for any linear mappings $A_i : X_i \to Y$, we have shown that formula (10.22) gives the general form of any linear mapping $A \in \mathcal{L}(X = X_1 \times \dots \times X_m; Y)$.

Example 10. Starting from the definition of the direct product $Y = Y_1 \times \dots \times Y_n$ of the vector spaces $Y_1, \dots, Y_n$ and the definition of a linear mapping $A : X \to Y$, one can easily see that any linear mapping

$$A : X \to Y = Y_1 \times \dots \times Y_n$$

has the form $x \mapsto Ax = (A_1 x, \dots, A_n x) = (y_1, \dots, y_n) = y \in Y$, where $A_i : X \to Y_i$ are linear mappings.

Example 11. Combining Examples 9 and 10, we conclude that any linear mapping

$$A : X_1 \times \dots \times X_m = X \to Y = Y_1 \times \dots \times Y_n$$

of the direct product $X = X_1 \times \dots \times X_m$ of vector spaces into another direct product $Y = Y_1 \times \dots \times Y_n$ has the form

$$y = \begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \multicolumn{3}{c}{\dots\dots\dots\dots} \\ A_{n1} & \cdots & A_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \cdots \\ x_m \end{pmatrix} = Ax , \quad (10.23)$$

where $A_{ij} : X_j \to Y_i$ are linear mappings.

In particular, if $X_1 = X_2 = \dots = X_m = \mathbb{R}$ and $Y_1 = Y_2 = \dots = Y_n = \mathbb{R}$, then $A_{ij} : X_j \to Y_i$ are the linear mappings $\mathbb{R} \ni x \mapsto a_{ij} x \in \mathbb{R}$, each of which is given by a single number a_{ij}. Thus in this case relation (10.23) becomes the familiar numerical notation for a linear mapping $A : \mathbb{R}^m \to \mathbb{R}^n$.

10.2.2 The Norm of a Transformation

Definition 3. Let $A : X_1 \times \dots \times X_n \to Y$ be a multilinear transformation mapping the direct product of the normed vector spaces $X_1, \dots, X_n$ into a normed space Y.

The quantity

$$\|A\| := \sup_{\substack{x_1,\dots,x_n \\ x_i \neq 0}} \frac{|A(x_1,\dots,x_n)|_Y}{|x_1|_{X_1} \times \cdots \times |x_n|_{X_n}} , \tag{10.24}$$

where the supremum is taken over all sets $x_1, \dots, x_n$ of nonzero vectors in the spaces $X_1, \dots, X_n$, is called the *norm* of the multilinear transformation A.

On the right-hand side of Eq. (10.24) we have denoted the norm of a vector x by the symbol $|\cdot|$ subscripted by the symbol for the normed vector space to which the vector belongs, rather than the usual symbol $\|\cdot\|$ for the norm of a vector. From now on we shall adhere to this notation for the norm of a vector; and, where no confusion can arise, we shall omit the symbol for the vector space, taking for granted that the norm (absolute value) of a vector is always computed in the space to which it belongs. In this way we hope to introduce for the time being some distinction in the notation for the norm of a vector and the norm of a linear or multilinear transformation acting on a normed vector space.

Using the properties of the norm of a vector and the properties of a multilinear transformation, one can rewrite formula (10.24) as follows:

$$\|A\| = \sup_{\substack{x_1,\dots,x_n \\ x_i \neq 0}} \left| A\Big(\frac{x_1}{|x_1|}, \dots, \frac{x_n}{|x_n|}\Big) \right| = \sup_{e_1,\dots,e_n} |A(e_1,\dots,e_n)| , \tag{10.25}$$

where the last supremum extends over all sets $e_1, \dots, e_n$ of unit vectors in the spaces $X_1, \dots, X_n$ respectively (that is, $|e_i| = 1$, $i = 1, \dots, n$).

In particular, for a linear transformation $A : X \to Y$, from (10.24) and (10.25) we obtain

$$\|A\| = \sup_{x \neq 0} \frac{|Ax|}{|x|} = \sup_{|e|=1} |Ae| . \tag{10.26}$$

It follows from Definition 3 for the norm of a multilinear transformation A that if $\|A\| < \infty$, then the inequality

$$|A(x_1,\dots,x_n)| \leq \|A\| \, |x_1| \times \cdots \times |x_n| \tag{10.27}$$

holds for any vectors $x_i \in X_i$, $i = 1, \dots, n$.

In particular, for a linear transformation we obtain

$$|Ax| \leq \|A\| \, |x| . \tag{10.28}$$

In addition, it follows from Definition 3 that if the norm of a multilinear transformation is finite, it is the greatest lower bound of all numbers M for which the inequality

$$|A(x_1,\dots,x_n)| \leq M |x_1| \times \cdots \times |x_n| \tag{10.29}$$

holds for all values of $x_i \in X_i$, $i = 1, \dots, n$.

Definition 4. A multilinear transformation $A : X_1 \times \cdots \times X_n \to Y$ is *bounded* if there exists $M \in \mathbb{R}$ such that inequality (10.29) holds for all values of $x_1, \ldots, x_n$ in the spaces $X_1, \ldots, X_n$ respectively.

Thus the bounded transformations are precisely those that have a finite norm.

On the basis of relation (10.26) one can easily understand the geometric meaning of the norm of a linear transformation in the familiar case $A : \mathbb{R}^m \to \mathbb{R}^n$. In this case the unit sphere in $\mathbb{R}^m$ maps under the transformation A into some ellipsoid in $\mathbb{R}^n$ whose center is at the origin. Hence the norm of A in this case is simply the largest of the semiaxes of the ellipsoid.

On the other hand, one can also interpret the norm of a linear transformation as the least upper bound of the coefficients of dilation of vectors under the mapping, as can be seen from the first equality in (10.26).

It is not difficult to prove that for mappings of finite-dimensional spaces the norm of a multilinear transformation is always finite, and hence in particular the norm of a linear transformation is always finite. This is no longer true in the case of infinite-dimensional spaces, as can be seen from the first of the following examples.

Let us compute the norms of the transformations considered in Examples 1–8.

Example 1$'$. If we regard $\underset{0}{\ell}$ as a subspace of the normed space ℓ_p, in which the vector $e_n = (\underbrace{0, \ldots, 0}_{n-1}, 1, 0, \ldots)$ has unit norm, then, since $Ae_n = ne_n$, it is clear that $\|A\| = \infty$.

Example 2$'$. If $|f| = \max\limits_{a \le x \le b} |f(x)| \le 1$, then $|Af| = |f(x_0)| \le 1$, and $|Af| = 1$ if $f(x_0) = 1$, so that $\|A\| = 1$.

We remark that if we introduce, for example, the integral norm

$$|f| = \int\limits_a^b |f|(x)\,dx$$

on the same vector space $C\big([a,b],\mathbb{R}\big)$, the result of computing $\|A\|$ may change considerably. Indeed, set $[a,b] = [0,1]$ and $x_0 = 1$. The integral norm of the function $f_n = x^n$ on $[0,1]$ is obviously $\frac{1}{n+1}$, while $Af_n = Ax^n = x^n\big|_{x=1} = 1$. It follows that $\|A\| = \infty$ in this case.

Throughout what follows, unless the contrary is explicitly stated, the space $C\big([a,b],\mathbb{R}\big)$ is assumed to have the norm defined by the maximum of the absolute value of the function on the closed interval $[a,b]$.

Example 3′. If $|f| = \max_{a \le x \le b} |f(x)| \le 1$, then

$$|Af| = \left| \int_a^b f(x)\,dx \right| \le \int_a^b |f|(x)\,dx \le \int_a^b 1\,dx = b - a\ .$$

But for $f(x) \equiv 1$, we obtain $|A1| = b - a$, and therefore $\|A\| = b - a$.

Example 4′. If $|f| = \max_{a \le x \le b} |f(x)| \le 1$, then

$$\max_{a \le x \le b} \left| \int_a^x f(t)\,dt \right| \le \max_{a \le x \le b} \int_a^x |f|(t)\,dt \le \max_{a \le x \le b} (x - a) = b - a\ .$$

But for $|f(t) \equiv 1$, we obtain

$$\max_{a \le x \le b} \int_a^x 1\,dt = b - a\ ,$$

and therefore in this example $\|A\| = b - a$.

Example 5′. We obtain immediately from Definition 3 that $\|A\| = 1$ in this case.

Example 6′. By the Cauchy–Bunyakovskii inequality

$$|\langle x_1, x_2 \rangle| \le |x_1| \cdot |x_2|\ ,$$

and if $x_1 = x_2$, this inequality becomes equality. Hence $\|A\| = 1$.

Example 7′. We know that

$$\big|[x_1, x_2]\big| = |x_1|\,|x_2| \sin\varphi\ ,$$

where φ is the angle between the vectors x_1 and x_2, and therefore $\|A\| \le 1$. At the same time, if the vectors x_1 and x_2 are orthogonal, then $\sin\varphi = 1$. Thus $\|A\| = 1$.

Example 8′. If we assume that the vectors lie in a Euclidean space of dimension n, we note that $A(x_1, \dots, x_n) = \det(x_1, \dots, x_n)$ is the volume of the parallelepiped spanned by the vectors $x_1, \dots, x_n$, and this volume is maximal if the vectors $x_1, \dots, x_n$ are made pairwise orthogonal while keeping their lengths constant.

Thus,

$$|\det(x_1, \dots, x_n)| \le |x_1| \cdot \ldots \cdot |x_n|\ ,$$

equality holding for orthogonal vectors. Hence in this case $\|A\| = 1$.

Let us now estimate the norms of the operators studied in Examples 9–11. We shall assume that in the direct product $X = X_1 \times \cdots \times X_m$ of the normed spaces $X_1, \ldots, X_m$ the norm of the vector $x = (x_1, \ldots, x_m)$ is introduced in accordance with the convention in Sect. 10.1 (Example 6).

Example 9$'$. Defining a linear transformation

$$A : X_1 \times \cdots \times X_m = X \to Y ,$$

as has been shown, is equivalent to defining the m linear transformations $A_i : X_i \to Y$ given by the relations $A_i x_i = A\bigl((0, \ldots, 0, x_i, 0, \ldots, 0)\bigr)$, $i = 1, \ldots, m$. When this is done, formula (10.22) holds, by virtue of which

$$|Ax|_Y \leq \sum_{i=1}^{m} |A_i x_i|_Y \leq \sum_{i=1}^{m} \|A_i\| \, |x_i|_{X_i} \leq \Bigl(\sum_{i=1}^{m} \|A_i\| \Bigr) |x|_X .$$

Thus we have shown that

$$\|A\| \leq \sum_{i=1}^{m} \|A_i\| .$$

On the other hand, since

$$|A_i x_i| = \bigl| A\bigl((0, \ldots, 0, x_i, 0, \ldots, 0)\bigr) \bigr| \leq$$
$$\leq \|A\| \, \bigl|(0, \ldots, 0, x_i, 0, \ldots, 0)\bigr|_X = \|A\| \, |x_i|_{X_i} ,$$

we can conclude that the estimate

$$\|A_i\| \leq \|A\|$$

also holds for all $i = 1, \ldots, m$.

Example 10$'$. Taking account of the norm introduce in $Y = Y_1 \times \cdots \times Y_n$, in this case we immediately obtain the two-sided estimates

$$\|A_i\| \leq \|A\| \leq \sum_{i=1}^{n} \|A_i\| .$$

Example 11$'$. Taking account of the results of Examples 9 and 10, one can conclude that

$$\|A_{ij}\| \leq \|A\| \leq \sum_{i=1}^{m} \sum_{j=1}^{n} \|A_{ij}\| .$$

10.2.3 The Space of Continuous Transformations

From now on we shall not be interested in all linear or multilinear transformations, only continuous ones. In this connection it is useful to keep in mind the following proposition.

Proposition 1. *For a multilinear transformation $A : X_1 \times \cdots \times X_n \to Y$ mapping a product of normed spaces $X_1, \dots, X_n$ into a normed space Y the following conditions are equivalent:*

a) A has a finite norm,
b) A is a bounded transformation,
c) A is a continuous transformation,
d) A is continuous at the point $(0, \dots, 0) \in X_1 \times \cdots \times X_n$.

Proof. We prove a closed chain of implications $a) \Rightarrow b) \Rightarrow c) \Rightarrow d) \Rightarrow a)$.

It is obvious from relation (10.27) that $a) \Rightarrow b)$.

Let us verify that $b) \Rightarrow c)$, that is, that (10.29) implies that the operator A is continuous. Indeed, taking account of the multilinearity of A, we can write that

$$\begin{aligned} &A(x_1 + h_1, x_2 + h_2, \dots, x_n + h_n) - A(x_1, x_2, \dots, x_n) = \\ &\qquad = A(h_1, x_2, \dots, x_n) + \cdots + A(x_1, x_2, \dots, x_{n-1}, h_n) = \\ &\qquad + A(h_1, h_2, x_3, \dots, x_n) + \cdots + A(x_1, \dots, x_{n-2}, h_{n-1}, h_n) + \\ &\qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\qquad\qquad\qquad\qquad + A(h_1, \dots, h_n)\ . \end{aligned}$$

From (10.29) we now obtain the estimate

$$\begin{aligned} &|A(x_1 + h_1, x_2 + h_2, \dots, x_n + h_n) - A(x_1, x_2, \dots, x_n) \le \\ &\quad \le M\big(|h_1| \cdot |x_2| \cdot \ldots \cdot |x_n| + \cdots + |x_1| \cdot |x_2| \cdot \ldots \cdot |x_{n-1}| \cdot |h_n| + \\ &\qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\qquad\qquad\qquad\qquad + |h_1| \cdot \ldots \cdot |h_n|\big)\ , \end{aligned}$$

from which it follows that A is continuous at each point $(x_1, \dots, x_n) \in X_1 \times \cdots \times X_n$.

In particular, if $(x_1, \dots, x_n) = (0, \dots, 0)$ we obtain d) from c).

It remains to be shown that $d) \Rightarrow a)$.

Given $\varepsilon > 0$ we find $\delta = \delta(\varepsilon) > 0$ such that $|A(x_1, \dots, x_n)| < \varepsilon$ when $\max\{|x_1|, \dots, |x_n|\} < \delta$. Then for any set $e_1, \dots, e_n$ of unit vectors we obtain

$$|A(e_1, \dots, e_n)| = \frac{1}{\delta^n} |A(\delta e_1, \dots, \delta e_n)| < \frac{\varepsilon}{\delta^n}\ ,$$

that is, $\|A\| < \frac{\varepsilon}{\delta^n} < \infty$. $\square$

We have seen above (Example 1) that not every linear transformation has a finite norm, that is, a linear transformation is not always continuous. We have also pointed out that continuity can fail for a linear transformation only when the transformation is defined on an infinite-dimensional space.

From here on $\mathcal{L}(X_1, \dots, X_n; Y)$ will denote the set of c o n t i n u o u s multilinear transformations mapping the direct product of the normed vector spaces $X_1, \dots, X_n$ into the normed vector space Y.

In particular, $\mathcal{L}(X; Y)$ is the set of continuous linear transformations from X into Y.

In the set $\mathcal{L}(X_1, \dots, X_n; Y)$ we introduce a natural vector-space structure:

$$(A + B)(x_1, \dots, x_n) := A(x_1, \dots, x_n) + B(x_1, \dots, x_n)$$

and

$$(\lambda A)(x_1, \dots, x_n) := \lambda A(x_1, \dots, x_n) .$$

It is obvious that if $A, B \in \mathcal{L}(X_1, \dots, X_n; Y)$, then $(A + B) \in \mathcal{L}(X_1, \dots, X_n; Y)$ and $(\lambda A) \in \mathcal{L}(X_1, \dots, X_n; Y)$.

Thus $\mathcal{L}(X_1, \dots, X_n; Y)$ can be regarded as a vector space.

Proposition 2. *The norm of a multilinear transformation is a norm in the vector space $\mathcal{L}(X_1, \dots, X_n; Y)$ of continuous multilinear transformations.*

Proof. We observe first of all that by Proposition 1 the nonnegative number $\|A\| < \infty$ is defined for every transformation $A \in \mathcal{L}(X_1, \dots, X_n; Y)$.

Inequality (10.27) shows that

$$\|A\| = 0 \Leftrightarrow A = 0 .$$

Next, by definition of the norm of a multilinear transformation

$$\|\lambda A\| = \sup_{\substack{x_1, \dots, x_n \\ x_i \neq 0}} \frac{(\lambda A)(x_1, \dots, x_n)|}{|x_1| \cdot \ldots \cdot |x_n|} =$$

$$= \sup_{\substack{x_1, \dots, x_n \\ x_i \neq 0}} \frac{|\lambda| \, |A(x_1, \dots, x_n)|}{|x_1| \cdot \ldots \cdot |x_n|} = |\lambda| \, \|A\| .$$

Finally, if A and B are elements of the space $\mathcal{L}(X_1, \dots, X_n; Y)$, then

$$\|A + B\| = \sup_{\substack{x_1, \dots, x_n \\ x_i \neq 0}} \frac{|(A + B)(x_1, \dots, x_n)|}{|x_1| \cdot \ldots \cdot |x_n|} =$$

$$= \sup_{\substack{x_1, \dots, x_n \\ x_i \neq 0}} \frac{|A(x_1, \dots, x_n) + B(x_1, \dots, x_n)|}{|x_1| \cdot \ldots \cdot |x_n|} \leq$$

$$\leq \sup_{\substack{x_1, \dots, x_n \\ x_i \neq 0}} \frac{|A(x_1, \dots, x_n)|}{|x_1| \cdot \ldots \cdot |x_n|} + \sup_{\substack{x_1, \dots, x_n \\ x_i \neq 0}} \frac{|B(x_1, \dots, x_n)|}{|x_1| \cdot \ldots \cdot |x_n|} = \|A\| + \|B\| . \quad \square$$

From now on when we use the symbol $\mathcal{L}(X_1, \ldots, X_n; Y)$ we shall have in mind the vector space of *continuous n-linear transformations* normed by this *transformation norm.* In particular $\mathcal{L}(X, Y)$ is the normed space of continuous linear transformations from X into Y.

We now prove the following useful supplement to Proposition 2.

Supplement. *If X, Y, and Z are normed spaces and $A \in \mathcal{L}(X;Y)$ and $B \in \mathcal{L}(Y;Z)$, then*

$$\|B \circ A\| \leq \|B\| \cdot \|A\| \ .$$

Proof. Indeed,

$$\|B \circ A\| = \sup_{x \neq 0} \frac{|(B \circ A)x|}{|x|} \leq \sup_{x \neq 0} \frac{\|B\|\, |Ax|}{|x|} = \\ = \|B\| \sup_{x \neq 0} \frac{|Ax|}{|x|} = \|B\| \cdot \|A\| \ . \ \square$$

Proposition 3. *If Y is a complete normed space, then $\mathcal{L}(X_1, \ldots, X_n; Y)$ is also a complete normed space.*

Proof. We shall carry out the proof for the space $\mathcal{L}(X;Y)$ of continuous linear transformations. The general case, as will be clear from the reasoning below, differs only in requiring a more cumbersome notation.

Let $A_1, A_2, \ldots, A_n \ldots$ be a Cauchy sequence in $\mathcal{L}(X;Y)$. Since for any $x \in X$ we have

$$|A_m x - A_n x| = \bigl|(A_m - A_n)x\bigr| \leq \|A_m - A_n\|\, |x| \ ,$$

it is clear that for any $x \in X$ the sequence $A_1 x$, $A_2 x$, $\ldots$, $A_n x$, $\ldots$ is a Cauchy sequence in Y. Since Y is complete, it has a limit in Y, which we denote by Ax.

Thus,

$$Ax := \lim_{n \to \infty} A_n x \ .$$

We shall show that $A : X \to Y$ is a continuous linear transformation.

The linearity of A follows from the relations

$$\lim_{n \to \infty} A_n(\lambda_1 x_1 + \lambda_2 x_2) = \lim_{n \to \infty} (\lambda_1 A_n x_1 + \lambda_2 A_n x_2) = \\ \lambda_1 \lim_{n \to \infty} A_n x_1 + \lambda_2 \lim_{n \to \infty} A_n x_2 \ .$$

Next, for any fixed $\varepsilon > 0$ and sufficiently large values of $m, n \in \mathbb{N}$ we have $\|A_m - A_n\| < \varepsilon$, and therefore

$$|A_m x - A_n x| \leq \varepsilon |x|$$

at each vector $x \in X$. Letting m tend to infinity in this last relation and using the continuity of the norm of a vector, we obtain

$$|Ax - A_n x| \leq \varepsilon |x| \ .$$

Thus $\|A - A_n\| \le \varepsilon$, and since $A = A_n + (A - A_n)$, we conclude that

$$\|A\| \le \|A_n\| + \varepsilon \ .$$

Consequently, we have shown that $A \in \mathcal{L}(X;Y)$ and $\|A - A_n\| \to 0$ as $n \to \infty$, that is, $A = \lim\limits_{n\to\infty} A_n$ in the sense of the norm of the space $\mathcal{L}(X;Y)$. □

In conclusion, we make one special remark relating to the space of multilinear transformations, which we shall need when studying higher-order differentials.

Proposition 4. *For each $m \in \{1, \dots, n\}$ there is a bijection between the spaces*

$$\mathcal{L}(X_1, \dots, X_m; \mathcal{L}(X_{m+1}, \dots, X_n; Y)) \ \text{ and } \ \mathcal{L}(X_1, \dots, X_n; Y)$$

that preserves the vector-space structure and the norm.

Proof. We shall exhibit this isomorphism.

Let $\mathfrak{B} \in \mathcal{L}(X_1, \dots, X_m; \mathcal{L}(X_{m+1}, \dots, X_n; Y))$, that is, $\mathfrak{B}(x_1, \dots, x_m) \in \mathcal{L}(X_{m+1}, \dots, X_n; Y)$.

We set

$$A(x_1, \dots, x_n) := \mathfrak{B}(x_1, \dots, x_m)(x_{m+1}, \dots, x_n) \ . \tag{10.30}$$

Then

$$\|\mathfrak{B}\| = \sup_{\substack{x_1, \dots, x_m \\ x_i \ne 0}} \frac{\|\mathfrak{B}(x_1, \dots, x_m)\|}{|x_1| \cdot \ldots \cdot |x_m|} =$$

$$= \sup_{\substack{x_1, \dots, x_m \\ x_i \ne 0}} \frac{\sup\limits_{\substack{x_{m+1}, \dots, x_n \\ x_j \ne 0}} \dfrac{|\mathfrak{B}(x_1, \dots, x_m)(x_{m+1}, \dots, x_n)|}{|x_{m+1}| \cdot \ldots \cdot |x_n|}}{|x_1| \cdot \ldots \cdot |x_m|} =$$

$$= \sup_{\substack{x_1, \dots, x_n \\ x_k \ne 0}} \frac{|A(x_1, \dots, x_n)|}{|x_1| \cdot \ldots \cdot |x_n|} = \|A\| \ .$$

We leave to the reader the verification that relation (10.30) defines an isomorphism of these vector spaces. □

Applying Proposition 4 n times, we find that the space

$$\mathcal{L}(X_1; \mathcal{L}(X_2; \dots; \mathcal{L}(X_n; Y)) \cdots)$$

is isomorphic to the space $\mathcal{L}(X_1, \dots, X_n; Y)$ of n-linear transformations.

10.2.4 Problems and Exercises

1. a) Prove that if $A : X \to Y$ is a linear transformation from the normed space X into the normed space Y and X is finite-dimensional, then A is a continuous operator.

b) Prove the proposition analogous to that stated in a) for a multilinear operator.

2. Two normed vector spaces are *isomorphic* if there exists an isomorphism between them (as vector spaces) that is continuous together with its inverse transformation.

a) Show that normed vector spaces of the same finite dimension are isomorphic.

b) Show that for the infinite-dimensional case assertion a) is generally no longer true.

c) Introduce two norms in the space $C\Big([a,b],\mathbb{R}\Big)$ in such a way that the identity mapping of $C\Big([a,b],\mathbb{R}\Big)$ is not a continuous mapping of the two resulting normed spaces.

3. Show that if a multilinear transformation of n-dimensional Euclidean space is continuous at some point, then it is continuous everywhere.

4. Let $A : E^n \to E^n$ be a linear transformation of n-dimensional Euclidean space and $A^* : E^n \to E^n$ the adjoint to this transformation.

Show the following.

a) All the eigenvalues of the operator $A \cdot A^* : E^n \to E^n$ are nonnegative.

b) If $\lambda_1 \le \cdots \le \lambda_n$ are the eigenvalues of the operator $A \cdot A^*$, then $\|A\| = \sqrt{\lambda_n}$.

c) If the operator A has an inverse $A^{-1} : E^n \to E^n$, then $\|A^{-1}\| = \frac{1}{\sqrt{\lambda_1}}$.

d) If (a^i_j) is the matrix of the operator $A : E^n \to E^n$ in some basis, then the estimates

$$\max_{1\le i\le n} \sqrt{\sum_{j=1}^{n} (a^i_j)^2} \le \|A\| \le \sqrt{\sum_{i,j=1}^{n} (a^i_j)^2} \le \sqrt{n}\|A\|$$

hold.

5. Let $\mathbb{P}[x]$ be the vector space of polynomials in the variable x with real coefficients. We define the norm of the vector $P \in \mathbb{P}[x]$ by the formula

$$|P| = \sqrt{\int_0^1 P^2(x)\,dx}.$$

a) Is the operator $D : \mathbb{P}[x] \to \mathbb{P}[x]$ given by differentiation $(D\Big(P(x)\Big) := P'(x))$ continuous in the resulting space?

b) Find the norm of the operator $F : \mathbb{P}[x] \to \mathbb{P}[x]$ of multiplication by x, which acts according to the rule $F\Big(P(x)\Big) = x \cdot P(x)$.

6. Using the example of projection operators in $\mathbb{R}^2$, show that the inequality $\|B \circ A\| \le \|B\| \cdot \|A\|$ may be a strict inequality.

10.3 The Differential of a Mapping

10.3.1 Mappings Differentiable at a Point

Definition 1. Let X and Y be normed spaces. A mapping $f : E \to Y$ of a set $E \subset X$ into Y is *differentiable at an interior point* $x \in E$ if there exists a continuous linear transformation $L(x) : X \to Y$ such that

$$f(x+h) - f(x) = L(x)h + \alpha(x;h) \ , \tag{10.31}$$

where $\alpha(x;h) = o(h)$ as $h \to 0$, $x + h \in E$.[1]

Definition 2. The function $L(x) \in \mathcal{L}(X;Y)$ that is linear with respect to h and satisfies relation (10.31) is called the *differential*, the *tangent mapping*, or the *derivative of the mapping* $f : E \to Y$ *at the point* x.

As before, we shall denote $L(x)$ by $\mathrm{d}f(x)$, $Df(x)$, or $f'(x)$.

We thus see that the general definition of differentiability of a mapping at a point is a nearly verbatim repetition of the one already familiar to us from Sect. 8.2, where it was considered in the case $X = \mathbb{R}^m$, $Y = \mathbb{R}^n$. For that reason, from now on we shall allow ourselves to use such concepts introduced there as *increment of a function*, *increment of the argument*, and *tangent space at a point* without repeating the explanations, preserving the corresponding notation.

We shall, however, verify the following proposition in general form.

Proposition 1. *If a mapping $f : E \to Y$ is differentiable at an interior point x of a set $E \subset X$, its differential $L(x)$ at that point is uniquely determined.*

Proof. Thus we are verifying the uniqueness of the differential.

Let $L_1(x)$ and $L_2(x)$ be linear mappings satisfying relation (10.31), that is

$$\begin{aligned} f(x+h) - f(x) - L_1(x)h &= \alpha_1(x;h) \ , \\ f(x+h) - f(x) - L_2(x)h &= \alpha_2(x;h) \ , \end{aligned} \tag{10.32}$$

where $\alpha_i(x;h) = o(h)$ as $h \to 0$, $x + h \in E$, $i = 1, 2$.

Then, setting $L(x) = L_2(x) - L_1(x)$ and $\alpha(x;h) = \alpha_2(x;h) - \alpha_1(x;h)$ and subtracting the second equality in (10.32) from the first, we obtain

$$L(x)h = \alpha(x;h) \ .$$

Here $L(x)$ is a mapping that is linear with respect to h, and $\alpha(x;h) = o(h)$ as $h \to 0$, $x + h \in E$. Taking an auxiliary numerical parameter λ, we can now

[1] The notation "$\alpha(x;h) = o(h)$ as $h \to 0$, $x + h \in E$", of course, means that
$$\lim_{h \to 0,\, x+h \in E} |\alpha(x;h)|_Y \cdot |h|_X^{-1} = 0.$$

write

$$|L(x)h| = \frac{|L(x)(\lambda h)|}{|\lambda|} = \frac{|\alpha(x;\lambda h)|}{|\lambda h|}|h| \to 0 \text{ as } \lambda \to 0 .$$

Thus $L(x)h = 0$ for any $h \neq 0$ (we recall that x is an interior point of E). Since $L(x)0 = 0$, we have shown that $L_1(x)h = L_2(x)h$ for every value of h. □

If E is an open subset of X and $f : E \to Y$ is a mapping that is differentiable at each point $x \in E$, that is, *differentiable on* E, by the uniqueness of the differential of a mapping at a point, which was just proved, a function $E \ni x \mapsto f'(x) \in \mathcal{L}(X;Y)$ arises on the set E, which we denote $f' : E \to \mathcal{L}(X;Y)$. This mapping is called the *derivative of* f, or the *derivative mapping* relative to the original mapping $f : E \to Y$. The value $f'(x)$ of this function at an individual point $x \in E$ is the continuous linear transformation $f'(x) \in \mathcal{L}(X;Y)$ that is the differential or derivative of the function f at the particular point $x \in E$.

We note that by the requirement of continuity of the linear mapping $L(x)$ Eq. (10.31) implies that a mapping that is differentiable at a point is necessarily continuous at that point.

The converse is of course not true, as we have seen in the case of numerical functions.

We now make one more important remark.

Remark. If the condition for differentiability of the mapping f at some point a is written as

$$f(x) - f(a) = L(x)(x-a) + \alpha(a;x) ,$$

where $\alpha(a;x) = o(x-a)$ as $x \to a$, it becomes clear that Definition 1 actually applies to a mapping $f : A \to B$ of any affine spaces (A,X) and (B,Y) whose vector spaces X and Y are normed. Such affine spaces, called *normed affine spaces*, are frequently encountered, so that it is useful to keep this remark in mind when using the differential calculus.

Everything that follows, unless specifically stated otherwise, applies equally to both normed vector spaces and normed affine spaces, and we use the notation for vector spaces only for the sake of simplicity.

10.3.2 The General Rules for Differentiation

The following general properties of the operation of differentiation follow from Definition 1. In the statements below X, Y, and Z are normed spaces and U and V open sets in X and Y respectively.

a. Linearity of Differentiation *If the mappings* $f_i : U \to Y$, $i = 1, 2$, *are differentiable at a point* $x \in U$, *a linear combination of them* $(\lambda_1 f_1 + \lambda_2 f_2) : U \to Y$ *is also differentiable at* x, *and*

$$(\lambda_1 f_1 + \lambda_2 f_2)'(x) = \lambda_1 f_1'(x) + \lambda_2 f_2'(x) .$$

Thus the differential of a linear combination of mappings is the corresponding linear combination of their differentials.

b. Differentiation of a Composition of Mappings (Chain Rule) *If the mapping* $f : U \to V$ *is differentiable at a point* $x \in U \subset X$, *and the mapping* $g : V \to Z$ *is differentiable at* $f(x) = y \in V \subset Y$, *then the composition* $g \circ f$ *of these mappings is differentiable at* x, *and*

$$(g \circ f)'(x) = g'\big(f(x)\big) \circ f'(x) .$$

Thus, the differential of a composition is the composition of the differentials.

c. Differentiation of the Inverse of a Mapping *Let* $f : U \to Y$ *be a mapping that is continuous at* $x \in U \subset X$ *and has an inverse* $f^{-1} : V \to X$ *that is defined in a neighborhood of* $y = f(x)$ *and continuous at that point.*

If the mapping f *is differentiable at* x *and its tangent mapping* $f'(x) \in \mathcal{L}(X; Y)$ *has a continuous inverse* $\big[f'(x)\big]^{-1} \in \mathcal{L}(Y; X)$, *then the mapping* f^{-1} *is differentiable at* $y = f(x)$ *and*

$$\big[f^{-1}\big]'\big(f(x)\big) = \big[f'(x)\big]^{-1} .$$

Thus, the differential of an inverse mapping is the linear mapping inverse to the differential of the original mapping at the corresponding point.

We omit the proofs of a, b, and c, since they are analogous to the proofs given in Sect. 8.3 for the case $X = \mathbb{R}^m$, $Y = \mathbb{R}^n$.

10.3.3 Some Examples

Example 1. If $f : U \to Y$ is a constant mapping of a neighborhood $U = U(x) \subset X$ of the point x, that is, $f(U) = y_0 \in Y$, then $f'(x) = 0 \in \mathcal{L}(X; Y)$.

Proof. Indeed, in this case it is obvious that

$$f(x + h) - f(x) - 0h = y_0 - y_0 - 0 = 0 = o(h) . \ \square$$

Example 2. If the mapping $f : X \to Y$ is a continuous linear mapping of a normed vector space X into a normed vector space Y, then $f'(x) = f \in \mathcal{L}(X; Y)$ at any point $x \in A$.

Proof. Indeed,

$$f(x+h) - f(x) - fh = fx + fh - fx - fh = 0 . \ \square$$

We remark that strictly speaking $f'(x) \in \mathcal{L}(TX_x; TY_{f(x)})$ here and h is a vector of the tangent space TX_x. But parallel translation of a vector to any point $x \in X$ is defined in a vector space, and this allows us to identify the tangent space TX_x with the vector space X itself. (Similarly, in the case of an affine space (A, X) the space TA_a of vectors "attached" to the point $a \in A$ can be identified with the vector space X of the given affine space.) Consequently, after choosing a basis in X, we can extend it to all the tangent spaces TX_x. This means that if, for example, $X = \mathbb{R}^m$, $Y = \mathbb{R}^n$, and the mapping $f \in \mathcal{L}(\mathbb{R}^m; \mathbb{R}^n)$ is given by the matrix (a_i^j), then at every point $x \in \mathbb{R}^m$ the tangent mapping $f'(x) : T\mathbb{R}^m_x \to T\mathbb{R}^n_{f(x)}$ will be given by the same matrix.

In particular, for a linear mapping $x \overset{f}{\longmapsto} ax = y$ from $\mathbb{R}$ to $\mathbb{R}$ with $x \in \mathbb{R}$ and $h \in T\mathbb{R}_x \sim \mathbb{R}$, we obtain the corresponding mapping $T\mathbb{R}_x \ni h \overset{f'}{\longmapsto} ah \in T\mathbb{R}_{f(x)}$.

Taking account of these conventions, we can provisionally state the result of Example 2 as follows: The mapping $f' : X \to Y$ that is the derivative of a linear mapping $f : X \to Y$ of normed spaces is constant, and $f'(x) = f$ at each point $x \in X$.

Example 3. From the chain rule for differentiating a composition of mappings and the result of Example 2 one can conclude that if $f : U \to Y$ is a mapping of a neighborhood $U = U(x) \subset X$ of the point $x \in X$ and is differentiable at x, while $A \in \mathcal{L}(Y; Z)$, then

$$(A \circ f)'(x) = A \circ f'(x) .$$

For numerical functions, when $Y = Z = \mathbb{R}$, this is simply the familiar possibility of moving a constant factor outside the differentiation sign.

Example 4. Suppose once again that $U = U(x)$ is a neighborhood of the point x in a normed space X, and let

$$f : U \to Y = Y_1 \times \cdots \times Y_n$$

be a mapping of U into the direct product of the normed spaces $Y_1, \dots, Y_n$.

Defining such a mapping is equivalent to defining the n mappings $f_i : U \to Y_i$, $i = 1, \dots, n$, connected with f by the relation

$$x \mapsto f(x) = y = (y_1, \dots, y_n) = \bigl(f_1(x), \dots, f_n(x)\bigr) ,$$

which holds at every point of U.

If we now take account of the fact that in formula (10.31) we have

$$f(x+h)-f(x)=\big(f_1(x+h)-f_1(x),\dots,f_n(x+h)-f_n(x)\big)\ ,$$
$$L(x)h=\big(L_1(x)h,\dots,L_n(x)h\big)\ ,$$
$$\alpha(x;h)=\big(\alpha_1(x;h),\dots,\alpha_n(x;h)\big)\ ,$$

then, referring to the results of Example 6 of Sect. 10.1 and Example 10 of Sect. 10.2, we can conclude that the mapping f is differentiable at x if and only if all of its components $f_i : U \to Y_i$ are differentiable at x, $i = 1,\dots,n$; and when the mapping f is differentiable, we have the equality

$$f'(x)=\big(f_1'(x),\dots,f_n'(x)\big)\ .$$

Example 5. Now let $A \in \mathcal{L}(X_1,\dots,X_n;Y)$, that is, A is a continuous n-linear transformation from the product $X_1\times\cdots\times X_n$ of the normed vector spaces $X_1,\dots,X_n$ into the normed vector space Y.

We shall prove that the mapping

$$A : X_1\times\cdots\times X_n = X \to Y$$

is differentiable and find its differential.

Proof. Using the multilinearity of A, we find that

$$\begin{aligned}A(x+h)-A(x) &= A(x_1+h_1,\dots,x_n+h_n)-A(x_1,\dots,x_n)=\\ &= A(x_1,\dots,x_n)+A(h_1,x_2,\dots,x_n)+\cdots+A(x_1,\dots,x_{n-1},h_n)+\\ &\quad + A(h_1,h_2,x_3,\dots,x_n)+\cdots+A(x_1,\dots,x_{n-2},h_{n-1},h_n)+\\ &\quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ &\quad + A(h_1,\dots,h_n)-A(x_1,\dots,x_n)\ .\end{aligned}$$

Since the norm in $X = X_1\times\cdots\times X_n$ satisfies the inequalities

$$|x_i|_{X_i}\le |x|_X\le\sum_{i=1}^n |x_i|_{X_i}\ ,$$

and the norm $\|A\|$ of the transformation A is finite and satisfies

$$|A(\xi_1,\dots,\xi_n)|\le\|A\|\,|\xi_1|\times\cdots\times|\xi_n|\ ,$$

we can conclude that

$$\begin{aligned}A(x+h)-A(x) &= A(x_1+h_1,\dots,x_n+h_n)-A(x_1,\dots,x_n)=\\ &= A(h_1,x_2,\dots,x_n)+\cdots+A(x_1,\dots,x_{n-1},h_n)+\alpha(x;h)\ ,\end{aligned}$$

where $\alpha(x;h)=o(h)$ as $h\to 0$.

But the transformation

$$L(x)h = A(h_1, x_2, \ldots, x_n) + \cdots + A(x_1, \ldots, x_{n-1}, h_n)$$

is a continuous transformation (because A is continuous) that is linear in $h = (h_1, \ldots, h_n)$.

Thus we have established that

$$A'(x)h = A'(x_1, \ldots, x_n)(h_1, \ldots, h_n) = \\ = A(h_1, x_2, \ldots, x_n) + \cdots + A(x_1, \ldots, x_{n-1}, h_n) \ ,$$

or, more briefly,

$$\mathrm{d}A(x_1, \ldots, x_n) = A(\mathrm{d}x_1, x_2, \ldots, x_n) + \cdots + A(x_1, \ldots x_{n-1}, \mathrm{d}x_n) \ . \ \square$$

In particular, if:

a) $x_1 \cdot \ldots \cdot x_n$ is the product of n numerical variables, then

$$\mathrm{d}(x_1 \cdot \ldots \cdot x_n) = \mathrm{d}x_1 \cdot x_2 \cdot \ldots \cdot x_n + \cdots + x_1 \cdot \ldots \cdot x_{n-1} \cdot \mathrm{d}x_n \ ;$$

b) $\langle x_1, x_2 \rangle$ is the inner product in E^3, then

$$\mathrm{d}\langle x_1, x_2 \rangle = \langle \mathrm{d}x_1, x_2 \rangle + \langle x_1, \mathrm{d}x_2 \rangle \ ;$$

c) $[x_1, x_2]$ is the vector cross product in E^3, then

$$\mathrm{d}[x_1, x_2] = [\mathrm{d}x_1, x_2] + [x_1, \mathrm{d}x_2] \ ;$$

d) (x_1, x_2, x_3) is the scalar triple product in E^3, then

$$\mathrm{d}(x_1, x_2, x_3) = (\mathrm{d}x_1, x_2, x_3) + (x_2, \mathrm{d}x_2, x_3) + (x_2, x_2, \mathrm{d}x_3) \ ;$$

e) $\det(x_1, \ldots, x_n)$ is the determinant of the matrix formed from the coordinates of n vectors $x_1, \ldots, x_n$ in an n-dimensional vector space X with a fixed basis, then

$$\mathrm{d}\big(\det(x_1, \ldots, x_n)\big) = \det(\mathrm{d}x_1, x_2, \ldots, x_n) + \cdots + \det(x_1, \ldots, x_{n-1}, \mathrm{d}x_n) \ .$$

Example 6. Let U be the subset of $\mathcal{L}(X; Y)$ consisting of the continuous linear transformations $A : X \to Y$ having continuous inverse transformations $A^{-1} : Y \to X$ (belonging to $\mathcal{L}(Y; X)$). Consider the mapping

$$U \ni A \mapsto A^{-1} \in \mathcal{L}(Y; X) \ ,$$

which assigns to each transformation $A \in U$ its inverse $A^{-1} \in \mathcal{L}(Y; X)$.

Proposition 2 proved below makes it possible to determine whether this mapping is differentiable.

Proposition 2. *If X is a complete space and $A \in U$, then for any $h \in \mathcal{L}(X;Y)$ such that $\|h\| < \|A^{-1}\|^{-1}$, the transformation $A+h$ also belongs to U and the following relation holds:*

$$(A+h)^{-1} = A^{-1} - A^{-1}hA^{-1} + o(h) \text{ as } h \to 0 . \tag{10.33}$$

Proof. Since

$$(A+h)^{-1} = \bigl(A(E + A^{-1}h)\bigr)^{-1} = (E + A^{-1}h)^{-1}A^{-1} , \tag{10.34}$$

it suffices to find the operator $(E+A^{-1}h)^{-1}$ inverse to $(E+A^{-1}h) \in \mathcal{L}(X;X)$, where E is the identity mapping e_X of X into itself.

Let $\Delta := -A^{-1}h$. Taking account of the supplement to Proposition 2 of Sect. 10.2, we can observe that $\|\Delta\| \le \|A^{-1}\| \cdot \|h\|$, so that by the assumptions made with respect to the operator h we may assume that $\|\Delta\| \le q < 1$.

We now verify that

$$(E - \Delta)^{-1} = E + \Delta + \Delta^2 + \cdots + \Delta^n + \cdots , \tag{10.35}$$

where the series on the right-hand side is formed from the linear operators $\Delta^n = (\Delta \circ \cdots \circ \Delta) \in \mathcal{L}(X;X)$.

Since X is a complete normed vector space, it follows from Proposition 3 of Sect. 10.2 that the space $\mathcal{L}(X;X)$ is also complete. It then follows immediately from the relation $\|\Delta^n\| \le \|\Delta\|^n \le q^n$ and the convergence of the series $\sum\limits_{n=0}^{\infty} q^n$ for $|q| < 1$ that the series (10.35) formed from the vectors in that space converges.

The direct verification that

$$\begin{aligned}(E + \Delta + \Delta^2 + \cdots)(E - \Delta) &= \\ &= (E + \Delta + \Delta^2 + \cdots) - (\Delta + \Delta^2 + \Delta^3 + \cdots) = E\end{aligned}$$

and

$$\begin{aligned}(E - \Delta)(E + \Delta + \Delta^2 + \cdots) &= \\ &= (E + \Delta + \Delta^2 + \cdots) - (\Delta + \Delta^2 + \Delta^3 + \cdots) = E\end{aligned}$$

shows that we have indeed found $(E - \Delta)^{-1}$.

It is worth remarking that the freedom in carrying out arithmetic operations on series (rearranging the terms!) in this case is guaranteed by the absolute convergence (convergence in norm) of the series under consideration.

Comparing relations (10.34) and (10.35), we conclude that

$$\begin{aligned}(A+h)^{-1} = A^{-1} - A^{-1}hA^{-1} \;+\; &(A^{-1}h)^2A^{-1} - \cdots \\ &\cdots + (-1)^n(A^{-1}h)^nA^{-1} + \cdots\end{aligned} \tag{10.36}$$

for $\|h\| \le \|A^{-1}\|^{-1}$.

Since

$$\left\| \sum_{n=2}^{\infty} (-A^{-1}h)^n A^{-1} \right\| \le \sum_{n=2}^{\infty} \|A^{-1}h\|^n \|A^{-1}\| \le$$
$$\le \|A^{-1}\|^3 \|h\|^2 \sum_{m=0}^{\infty} q^m = \frac{\|A^{-1}\|^3}{1-q} \|h\|^2 ,$$

Eq. (10.33) follows in particular from (10.36). □

Returning now to Example 6, we can say that when the space X is complete the mapping $A \overset{f}{\longmapsto} A^{-1}$ under consideration is necessarily differentiable, and

$$\mathrm{d}f(A)h = \mathrm{d}(A^{-1})h = -A^{-1}hA^{-1} .$$

In particular, this means that if A is a nonsingular square matrix and A^{-1} is its inverse, then under a perturbation of the matrix A by a matrix h whose elements are close to zero, we can write the inverse matrix $(A+h)^{-1}$ in first approximation in the following form:

$$(A+h)^{-1} \approx A^{-1} - A^{-1}hA^{-1} .$$

More precise formulas can obviously be obtained starting from Eq. (10.36).

Example 7. Let X be a complete normed vector space. The important mapping

$$\exp : \mathcal{L}(X;X) \to \mathcal{L}(X;X)$$

is defined as follows:

$$\exp A := E + \frac{1}{1!}A + \frac{1}{2!}A^2 + \cdots + \frac{1}{n!}A^n + \cdots , \qquad (10.37)$$

if $A \in \mathcal{L}(X;X)$.

The series in (10.37) converges, since $\mathcal{L}(X;X)$ is a complete space and $\|\frac{1}{n!}A^n\| \le \frac{\|A\|^n}{n!}$, while the numerical series $\sum\limits_{n=0}^{\infty} \frac{\|A\|^n}{n!}$ converges.

It is not difficult to verify that

$$\exp(A+h) = \exp A + L(A)h + o(h) \ \text{as} \ h \to \infty , \qquad (10.38)$$

where

$$L(A)h = h + \frac{1}{2!}(Ah + hA) + \frac{1}{3!}(A^2h + AhA + hA^2) + \cdots$$
$$\cdots + \frac{1}{n!}(A^{n-1}h + A^{n-2}hA + \cdots + AhA^{n-2} + hA^{n-1}) + \cdots$$

and $\|L(A)\| \le \exp\|A\| = \mathrm{e}^{\|A\|}$, that is, $L(A) \in \mathcal{L}\big(\mathcal{L}(X;X), \mathcal{L}(X;X)\big)$.

Thus, the mapping $\mathcal{L}(X;X) \ni A \mapsto \exp A \in \mathcal{L}(X;X)$ is differentiable at every value of A.

We remark that if the operators A and h commute, that is, $Ah = hA$, then, as one can see from the expression for $L(A)h$, in this case we have $L(A)h = (\exp A)h$. In particular, for $X = \mathbb{R}$ or $X = \mathbb{C}$, instead of (10.38) we again obtain

$$\exp(A+h) = \exp A + (\exp A)h + o(h) \text{ as } h \to 0 . \tag{10.39}$$

Example 8. We shall attempt to give a mathematical description of the instantaneous angular velocity of a rigid body with a fixed point o (a top). Consider an orthonormal frame $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ at the point o rigidly attached to the body. It is clear that the position of the body is completely characterized by the position of this orthoframe, and the triple $\{\dot{\mathbf{e}}_1, \dot{\mathbf{e}}_2, \dot{\mathbf{e}}_3\}$ of instantaneous velocities of the vectors of the frame obviously give a complete characterization of the instantaneous angular velocity of the body. The position of the frame itself $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ at time t can be given by an orthogonal matrix (α_i^j) $i, j = 1, 2, 3$ composed of the coordinates of the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ with respect to some fixed orthonormal frame in space. Thus, the motion of the top corresponds to a mapping $t \mapsto O(t)$ from $\mathbb{R}$ (the time axis) into the group $SO(3)$ of special orthogonal 3×3 matrices. Consequently, the angular velocity of the body, which we have agreed to describe by the triple $\{\dot{\mathbf{e}}_1, \dot{\mathbf{e}}_2, \dot{\mathbf{e}}_3\}$, is the matrix $\dot{O}(t) =: (\omega_i^j)(t) = (\dot{\alpha}_i^j)(t)$, which is the derivative of the matrix $O(t) = (\alpha_i^j)(t)$ with respect to time.

Since $O(t)$ is an orthogonal matrix, the relation

$$O(t)O^*(t) = E \tag{10.40}$$

holds at any time t, where $O^*(t)$ is the transpose of $O(t)$ and E is the identity matrix.

We remark that the product $A \cdot B$ of matrices is a bilinear function of A and B, and the derivative of the transposed matrix is obviously the transpose of the derivative of the original matrix. Differentiating (10.40) and taking account of these things, we find that

$$\dot{O}(t)O^*(t) + O(t)\dot{O}^*(t) = 0$$

or

$$\dot{O}(t) = -O(t)\dot{O}^*(t)O(t) , \tag{10.41}$$

since $O^*(t)O(t) = E$.

In particular, if we assume that the frame $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ coincides with the spatial frame of reference at time t, then $O(t) = E$, and it follows from (10.41) that

$$\dot{O}(t) = -\dot{O}^*(t) , \tag{10.42}$$

that is, the matrix $\dot{O}(t) =: \Omega(t) = (\omega_i^j)$ of coordinates of the vectors $\{\dot{\mathbf{e}}_1, \dot{\mathbf{e}}_2, \dot{\mathbf{e}}_3\}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ turns out to be skew-symmetric:

$$\Omega(t) = \begin{pmatrix} \omega_1^1 & \omega_1^2 & \omega_1^3 \\ \omega_2^1 & \omega_2^2 & \omega_2^3 \\ \omega_3^1 & \omega_3^2 & \omega_3^3 \end{pmatrix} = \begin{pmatrix} 0 & -\omega^3 & \omega^2 \\ \omega^3 & 0 & -\omega^1 \\ -\omega^2 & \omega^1 & 0 \end{pmatrix} .$$

Thus the instantaneous angular velocity of a top is actually characterized by three independent parameters, as follows in our line of reasoning from relation (10.40) and is natural from the physical point of view, since the position of the frame $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, and hence the position of the body itself, can be described by three independent parameters (in mechanics these parameters may be, for example, the Euler angles).

If we associate with each vector $\boldsymbol{\omega} = \omega^1\mathbf{e}_1 + \omega^2\mathbf{e}_2 + \omega^3\mathbf{e}_3$ in the tangent space at the point o a right-handed rotation of space with angular velocity $|\boldsymbol{\omega}|$ about the axis defined by this vector, it is not difficult to conclude from these results that at each instant of time t the body has an instantaneous angular velocity and that the velocity at that time can be adequately described by the instantaneous angular velocity vector $\boldsymbol{\omega}(t)$ (see Problem 5 below).

10.3.4 The Partial Derivatives of a Mapping

Let $U = U(a)$ be a neighborhood of the point $a \in X = X_1 \times \cdots \times X_m$ in the direct product of the normed spaces $X_1, \ldots, X_m$, and let $f : U \to Y$ be a mapping of U into the normed space V. In this case

$$y = f(x) = f(x_1, \ldots, x_m) , \tag{10.43}$$

and hence, if we fix all the variables but x_i in (10.43) by setting $x_k = a_k$ for $k \in \{1, \ldots, m\} \setminus i$, we obtain a function

$$f(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_m) =: \varphi_i(x_i) , \tag{10.44}$$

defined in some neighborhood U_i of a_i in X.

Definition 3. Relative to the original mapping (10.43) the mapping $\varphi_i : U_i \to Y$ is called the *partial mapping with respect to the variable x_i at $a \in X$.*

Definition 4. If the mapping (10.44) is differentiable at $x_i = a_i$, its derivative at that point is called the *partial derivative* or *partial differential of f at a with respect to the variable x_i.*

We usually denote this partial derivative by one of the symbols

$$\partial_i f(a) , \quad D_i f(a) , \quad \frac{\partial f}{\partial x_i}(a) , \quad f'_{x_i}(a) .$$

In accordance with these definitions $D_i f(a) \in \mathcal{L}(X_i; Y)$. More precisely, $D_i f(a) \in \mathcal{L}\big(TX_i(a_i); TY\big(f(a)\big)\big)$.

The differential $\mathrm{d}f(a)$ of the mapping (10.43) at the point a (if f is differentiable at that point) is often called the *total differential* in this situation in order to distinguish it from the partial differentials with respect to the individual variables.

We have already encountered all these concepts in the case of real-valued functions of m real variables, so that we shall not give a detailed discussion of them. We remark only that by repeating our earlier reasoning, taking account of Example 9 in Sect. 9.2, one can prove easily that the following proposition holds in general.

Proposition 3. *If the mapping* (10.43) *is differentiable at the point* $a = (a_1, \dots, a_m) \in X_1 \times \cdots \times X_m = X$*, it has partial derivatives with respect to each variable at that point, and the total differential and the partial differentials are related by the equation*

$$\mathrm{d}f(a)h = \partial_1 f(a)h_1 + \cdots + \partial_m f(a)h_m \,, \tag{10.45}$$

where $h = (h_1, \dots, h_m) \in TX_1(a_1) \times \cdots \times TX_m(a_m) = TX(a)$*.*

We have already shown by the example of numerical functions that the existence of partial derivatives does not in general guarantee the differentiability of the function (10.43).

10.3.5 Problems and Exercises

1. a) Let $A \in \mathcal{L}(X; X)$ be a *nilpotent operator*, that is, there exists $k \in \mathbb{N}$ such that $A^k = 0$. Show that the operator $(E - A)$ has an inverse in this case and that $(E - A)^{-1} = E + A + \cdots + A^{k-1}$.

b) Let $D : \mathbb{P}[x] \to \mathbb{P}[x]$ be the operator of differentiation on the vector space $\mathbb{P}[x]$ of polynomials. Remarking that D is a nilpotent operator, write the operator $\exp(aD)$, where $a \in \mathbb{R}$, and show that $\exp(aD)\Big(P(x)\Big) = P(x + a) =: T_a\Big(P(x)\Big)$.

c) Write the matrices of the operators $D : \mathbb{P}_n[x] \to \mathbb{P}_n[x]$ and $T_a : \mathbb{P}_n[x] \to \mathbb{P}_n[x]$ from part b) in the basis $e_i = \frac{x^{n-i}}{(n-i)!}$, $1 \le i \le n$, in the space $\mathbb{P}_n[x]$ of real polynomials of degree n in one variable.

2. a) If $A, B \in \mathcal{L}(X; X)$ and $\exists B^{-1} \in \mathcal{L}(X; X)$, then $\exp(B^{-1}AB) = B^{-1}(\exp A)B$.

b) If $AB = BA$, then $\exp(A + B) = \exp A \cdot \exp B$.

c) Verify that $\exp 0 = E$ and that $\exp A$ always has an inverse, namely $(\exp A)^{-1} = \exp(-A)$.

3. Let $A \in \mathcal{L}(X; X)$. Consider the mapping $\varphi_A : \mathbb{R} \to \mathcal{L}(X; X)$ defined by the correspondence $\mathbb{R} \ni t \mapsto \exp(tA) \in \mathcal{L}(X; X)$. Show the following.

a) The mapping φ_A is continuous.

b) φ_A is a homomorphism of $\mathbb{R}$ as an additive group into the multiplicative group of invertible operators in $\mathcal{L}(X; X)$.

4. Verify the following.

a) If $\lambda_1, \dots, \lambda_n$ are the eigenvalues of the operator $A \in \mathcal{L}(\mathbb{C}^n; \mathbb{C}^n)$, then $\exp \lambda_1, \dots, \exp \lambda_n$ are the eigenvalues of $\exp A$.

b) $\det(\exp A) = \exp(\operatorname{tr} A)$, where $\operatorname{tr} A$ is the trace of the operator $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^n)$.

c) If $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$, then $\det(\exp A) > 0$.

d) If A^* is the transpose of the matrix $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^n)$ and $\bar{A}$ is the matrix whose elements are the complex conjugates of those of A, then $(\exp A)^* = \exp A^*$ and $\overline{\exp A} = \exp \bar{A}$.

e) The matrix $\begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix}$ is not of the form $\exp A$ for any 2×2 matrix A.

5. We recall that a set endowed with both a group structure and a topology is called a *topological group* or *continuous group* if the group operation is continuous. If there is a sense in which the group operation is even analytic, the topological group is called a *Lie group*.[2]

A *Lie algebra* is a vector space X with an anticommutative bilinear operation $[\ ,\ ] : X \times X \to X$ satisfying the *Jacobi identity*: $[[a, b], c] + [[b, c], a] + [[c, a], b] = 0$ for any vectors $a, b, c \in X$. Lie groups and algebras are closely connected with each other, and the mapping exp plays an important role in establishing this connection (see Problem 1 above).

An example of a Lie algebra is the oriented Euclidean space E^3 with the operation of the vector cross product. For the time being we shall denote this Lie algebra by LA_1.

a) Show that the real 3×3 skew-symmetric matrices form a Lie algebra (which we denote LA_2) if the product of the matrices A and B is defined as $[A, B] = AB - BA$.

b) Show that the correspondence

$$\Omega = \begin{pmatrix} 0 & -\omega^3 & \omega^2 \\ \omega^3 & 0 & -\omega^1 \\ -\omega^2 & \omega^1 & 0 \end{pmatrix} \leftrightarrow (\omega_1, \omega_2, \omega_3) = \boldsymbol{\omega}$$

is an isomorphism of the Lie algebras LA_2 and LA_1.

c) Verify that if the skew-symmetric matrix Ω and the vector $\boldsymbol{\omega}$ correspond to each other as shown in b), then the equality $\Omega \mathbf{r} = [\boldsymbol{\omega}, \mathbf{r}]$ holds for any vector $\mathbf{r} \in E^3$, and the relation $P \Omega P^{-1} \leftrightarrow P\boldsymbol{\omega}$ holds for any matrix $P \in SO(3)$.

d) Verify that if $\mathbb{R} \ni t \mapsto O(t) \in SO(3)$ is a smooth mapping, then the matrix $\Omega(t) = O^{-1}(t)\dot{O}(t)$ is skew-symmetric.

e) Show that if $\mathbf{r}(t)$ is the radius vector of a point of a rotating top and $\Omega(t)$ is the matrix $(O^{-1}\dot{O})(t)$ found in d), then $\dot{\mathbf{r}}(t) = (\Omega \mathbf{r})(t)$.

f) Let $\mathbf{r}$ and $\boldsymbol{\omega}$ be two vectors attached at the origin of E^3. Suppose a right-handed frame has been chosen in E^3, and that the space undergoes a right-handed rotation with angular velocity $|\boldsymbol{\omega}|$ about the axis defined by $\boldsymbol{\omega}$. Show that $\dot{\mathbf{r}}(t) = [\boldsymbol{\omega}, \mathbf{r}(t)]$ in this case.

---

[2] For the precise definition of a Lie group and the corresponding reference see Problem 8 in Sect. 15.2.

g) Summarize the results of d), e), and f) and exhibit the instantaneous angular velocity of the rotating top discussed in Example 8.

h) Using the result of c), verify that the velocity vector $\boldsymbol{\omega}$ is independent of the choice of the fixed orthoframe in E^3, that is, it is independent of the coordinate system.

6. Let $\mathbf{r} = \mathbf{r}(s) = \big(x^1(s), x^2(s), x^3(s)\big)$ be the parametric equations of a smooth curve in E^3, the parameter being arc length along the curve (the *natural parametrization of the curve*).

a) Show that the vector $\mathbf{e}_1(s) = \frac{d\mathbf{r}}{ds}(s)$ tangent to the curve has unit length.

b) The vector $\frac{d\mathbf{e}_1}{ds}(s) = \frac{d^2\mathbf{r}}{ds^2}(s)$ is orthogonal to $\mathbf{e}_1$. Let $\mathbf{e}_2(s)$ be the unit vector formed from $\frac{d\mathbf{e}_1}{ds}(s)$. The coefficient $k(s)$ in the equality $\frac{d\mathbf{e}_1}{ds}(s) = k(s)\mathbf{e}_2(s)$ is called the *curvature* of the curve at the corresponding point.

c) By constructing the vector $\mathbf{e}_3(s) = [\mathbf{e}_1(s), \mathbf{e}_2(s)]$ we obtain a frame $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ at each point, called the *Frenet* frame[3] or *companion trihedral* of the curve. Verify the following Frenet formulas:

$$\begin{array}{rcl}
\frac{d\mathbf{e}_1}{ds}(s) &=& \qquad\qquad k(s)\mathbf{e}_2(s)\ , \\
\frac{d\mathbf{e}_2}{ds}(s) &=& -k(s)\mathbf{e}_1(s) \qquad\qquad\qquad \varkappa(s)\mathbf{e}_3(s)\ , \\
\frac{d\mathbf{e}_3}{ds}(s) &=& \qquad\qquad -\varkappa(s)\mathbf{e}_2(s)\ .
\end{array}$$

Explain the geometric meaning of the coefficient $\varkappa(s)$ called the *torsion* of the curve at the corresponding point.

10.4 The Finite-increment Theorem and some Examples of its Use

10.4.1 The Finite-increment Theorem

In our study of numerical functions of several variables in Subsect. 5.3.1 we proved the finite-increment theorem for them and discussed in detail various aspects of this important theorem of analysis. In the present section the finite-increment theorem will be proved in its general form. So that its meaning will be fully obvious, we advise the reader to recall the discussion in that subsection and also to pay attention to the geometric meaning of the norm of a linear operator (see Subsect. 10.2.2).

Theorem 1. (The finite-increment theorem). *Let $f : U \to Y$ be a continuous mapping of an open set U of a normed space X into a normed space Y.*

If the closed interval $[x, x+h] = \{\xi \in X \,|\, \xi = x + \theta h,\ 0 \le \theta \le 1\}$ is contained in U and the mapping f is differentiable at all points of the open interval $]x, x+h[= \{\xi \in X \,|\, \xi = x + \theta h,\ 0 < \theta < 1\}$, then the following

[3] J. F. Frenet (1816–1900) – French mathematician.

estimate holds:

$$|f(x+h) - f(x)|_Y \le \sup_{\xi \in]x,x+h[} \|f'(\xi)\|_{\mathcal{L}(X;Y)} |h|_X \ . \tag{10.46}$$

Proof. We remark first of all that if we could prove the inequality

$$|f(x'') - f(x')| \le \sup_{\xi \in [x',x'']} \|f'(\xi)\| \, |x'' - x'| \tag{10.47}$$

in which the supremum extends over the whole interval $[x', x'']$, for every closed interval $[x', x''] \subset]x, x+h[$, then, using the continuity of f and the norm together with the fact that

$$\sup_{\xi \in [x',x'']} \|f'(\xi)\| \le \sup_{\xi \in]x,x+h[} \|f'(\xi)\| \ ,$$

we would obtain inequality (10.46) in the limit as $x' \to x$ and $x'' \to x + h$.

Thus, it suffices to prove that

$$|f(x+h) - f(x)| \le M|h| \ , \tag{10.48}$$

where $M = \sup\limits_{0 \le \theta \le 1} \|f'(x + \theta h)\|$ and the function f is assumed differentiable on the entire closed interval $[x, x+h]$.

The very simple computation

$$\begin{aligned}|f(x_3) - f(x_1)| \le |f(x_3) - f(x_2)| + |f(x_2) - f(x_1)| \le \\ \le M|x_3 - x_2| + M|x_2 - x_1| = M\big(|x_3 - x_2| + |x_2 - x_1|\big) = \\ = M|x_3 - x_1| \ ,\end{aligned}$$

which uses only the triangle inequality and the properties of a closed interval, shows that if an inequality of the form (10.48) holds on the portions $[x_1, x_2]$ and $[x_2, x_3]$ of the closed interval $[x_1, x_3]$, then it also holds on $[x_1, x_3]$.

Hence, if estimate (10.48) fails for the closed interval $[x, x+h]$, then by successive bisections, one can obtain a sequence of closed intervals $[a_k, b_k] \subset]x, x+h[$ contracting to some point $x_0 \in [x, x+h]$ such that (10.48) fails on each interval $[a_k, b_k]$. Since $x_0 \in [a_k, b_k]$, consideration of the closed intervals $[a_k, x_0]$ and $[x_0, b_k]$ enables us to assume that we have found a sequence of closed intervals of the form $[x_0, x_0 + h_k] \subset [x, x+h]$, where $h_k \to 0$ as $k \to \infty$ on which

$$|f(x_0 + h_k) - f(x_0)| > M|h_k| \ . \tag{10.49}$$

If we prove (10.48) with M replaced by $M + \varepsilon$, where ε is any positive number, we will still obtain (10.48) as $\varepsilon \to 0$, and hence we can also replace (10.49) by

$$|f(x_0 + h_k) - f(x_0)| > (M + \varepsilon)|h_k| \tag{10.49'}$$

and we can now show that this is incompatible with the assumption that f is differentiable at x_0.

Indeed, by the assumption that f is differentiable,

$$|f(x_0 + h_k) - f(x_0)| = |f'(x_0)h_k + o(h_k)| \leq \\ \leq \|f'(x_0)\| \, |h_k| + o(|h_k|) \leq (M + \varepsilon)|h_k|$$

as $h_k \to 0$. $\square$

The finite-increment theorem has the following useful, purely technical corollary.

Corollary. *If $A \in \mathcal{L}(X;Y)$, that is, A is a continuous linear mapping of the normed space X into the normed space Y and $f : U \to Y$ is a mapping satisfying the hypotheses of the finite-increment theorem, then*

$$|f(x+h) - f(x) - Ah| \leq \sup_{\xi \in]x, x+h[} \|f'(\xi) - A\| \, |h| \; .$$

Proof. For the proof it suffices to apply the finite-increment theorem to the mapping

$$t \mapsto F(t) = f(x + th) - Ath$$

of the unit interval $[0,1] \subset \mathbb{R}$ into Y, since

$$\begin{gathered} F(1) - F(0) = f(x+h) - f(x) - Ah \; , \\ F'(\theta) = f'(x + \theta h)h - Ah \;\text{ for }\; 0 < \theta < 1 \; , \\ \|F'(\theta)\| \leq \|f'(x+\theta h) - A\| \, |h| \; , \\ \sup_{0<\theta<1} \|F'(\theta)\| \leq \sup_{\xi \in]x, x+h[} \|f'(\xi) - A\| \, |h| \; . \; \square \end{gathered}$$

Remark. As can be seen from the proof of Theorem 1, in its hypotheses there is no need to require that f be differentiable as a mapping $f : U \to Y$; it suffices that its restriction to the closed interval $[x, x+h]$ be a continuous mapping of that interval and differentiable at the points of the open interval $]x, x+h[$.

This remark applies equally to the corollary of the finite-increment theorem just proved.

10.4.2 Some Applications of the Finite-increment Theorem

a. Continuously Differentiable Mappings Let

$$f : U \to Y \tag{10.50}$$

be a mapping of an open subset U of a normed vector space X into a normed space Y. If f is differentiable at each point $x \in U$, then, assigning to the point x the mapping $f'(x) \in \mathcal{L}(X;Y)$ tangent to f at that point, we obtain the derivative mapping

$$f' : U \to \mathcal{L}(X;Y) . \tag{10.51}$$

Since the space $\mathcal{L}(X;Y)$ of continuous linear transformations from X into Y is, as we know, a normed space (with the transformation norm), it makes sense to speak of the continuity of the mapping (10.51).

Definition. When the derivative mapping (10.51) is continuous in U, the mapping (10.50), in complete agreement with our earlier terminology, will be said to be *continuously differentiable.*

As before, the set of continuously differentiable mappings of type (10.50) will be denoted by the symbol $C^{(1)}(U,Y)$, or more briefly, $C^{(1)}(U)$, if it is clear from the context what the range of the mapping is.

Thus, by definition

$$f \in C^{(1)}(U,Y) \Leftrightarrow f' \in C\big(U, \mathcal{L}(X;Y)\big) .$$

Let us see what continuous differentiability of a mapping means in different particular cases.

Example 1. Consider the familiar situation when $X = Y = \mathbb{R}$, and hence $f : U \to \mathbb{R}$ is a real-valued function of a real argument. Since any linear mapping $A \in \mathcal{L}(\mathbb{R};\mathbb{R})$ reduces to multiplication by some number $a \in \mathbb{R}$, that is, $Ah = ah$ and obviously $\|A\| = |a|$, we find that $f'(x)h = a(x)h$, where $a(x)$ is the numerical derivative of the function f at the point x.

Next, since

$$\begin{aligned}\big(f'(x+\delta) \,-\, f'(x)\big)h &= f'(x+\delta)h - f'(x)h = \\ &= a(x+\delta)h - a(x)h = \big(a(x+\delta) - a(x)\big)h ,\end{aligned} \tag{10.52}$$

it follows that

$$\|f'(x+\delta) - f'(x)\| = |a(x+\delta) - a(x)|$$

and hence in this case continuous differentiability of the mapping f is equivalent to the concept of a continuously differentiable numerical function (of class $C^{(1)}(U,\mathbb{R})$) studied earlier.

Example 2. This time suppose that X is the direct product $X_1 \times \cdots \times X_n$ of normed spaces. In this case the mapping (10.50) is a function $f(x) = f(x_1, \dots, x_m)$ of m variables $x_i \in X_i$, $i = 1, \dots, m$, with values in Y.

If the mapping f is differentiable at $x \in U$, its differential $\mathrm{d}f(x)$ at that point is an element of the space $\mathcal{L}(X_1 \times \cdots \times X_m = X; Y)$.

The action of $\mathrm{d}f(x)$ on a vector $h = (h_1, \dots, h_m)$, by formula (10.45), can be represented as

$$\mathrm{d}f(x)h = \partial_1 f(x)h_1 + \cdots + \partial_m f(x)h_m ,$$

where $\partial_i f(x) : X_i \to Y$, $i = 1, \dots, m$, are the partial derivatives of the mapping f at the point x under consideration.

Next,

$$\big(\mathrm{d}f(x+\delta) - \mathrm{d}f(x)\big)h = \sum_{i=1}^{m} \big(\partial_i f(x+\delta) - \partial_i f(x)\big)h_i \ . \tag{10.53}$$

But by the properties of the standard norm in the direct product of normed spaces (see Example 6 in Subsect. 10.1.2) and the definition of the norm of a transformation, we find that

$$\begin{aligned}\|\partial_i f(x+\delta) - \partial_i f(x)\|_{\mathcal{L}(X_i;Y)} &\leq \|\mathrm{d}f(x+\delta) - \mathrm{d}f(x)\|_{\mathcal{L}(X;Y)} \leq \\ &\leq \sum_{i=1}^{m} \|\partial_i f(x+\delta) - \partial_i f(x)\|_{\mathcal{L}(X_i;Y)} \ . \end{aligned} \tag{10.54}$$

Thus in this case the differentiable mapping (10.50) is continuously differentiable in U if and only if all its partial derivatives are continuous in U.

In particular, if $X = \mathbb{R}^m$ and $Y = \mathbb{R}$, we again obtain the familiar concept of a continuously differentiable numerical function of m real variables (a function of class $C^{(1)}(U, \mathbb{R})$, where $U \subset \mathbb{R}^m$).

Remark. It is worth noting that in writing (10.52) and (10.53) we have made essential use of the canonical identification $TX_x \sim X$, which makes it possible to compare or identify vectors lying in different tangent spaces.

We shall now show that continuously differentiable mappings satisfy a Lipschitz condition.

Proposition 1. *If K is a convex compact set in a normed space X and $f \in C^{(1)}(K, Y)$, where Y is also a normed space, then the mapping $f : K \to Y$ satisfies a Lipschitz condition on K, that is, there exists a constant $M > 0$ such that the inequality*

$$|f(x_2) - f(x_1)| \leq M|x_2 - x_1| \tag{10.55}$$

holds for any points $x_1, x_2 \in K$.

Proof. By hypothesis $f' : K \to \mathcal{L}(X; Y)$ is a continuous mapping of the compact set K into the metric space $\mathcal{L}(X; Y)$. Since the norm is a continuous function on a normed space with its natural metric, the mapping $x \mapsto \|f'(x)\|$, being the composition of continuous functions, is itself a continuous mapping of the compact set K into $\mathbb{R}$. But such a mapping is necessarily bounded. Let M be a constant such that $\|f'(x)\| \leq M$ at any point $x \in K$. Since K is convex, for any two points $x_1 \in K$ and $x_2 \in K$ the entire interval $[x_1, x_2]$ is contained in K. Applying the finite-increment theorem to that interval, we immediately obtain relation (10.55). $\square$

Proposition 2. *Under the hypotheses of Proposition* 1 *there exists a non-negative function* $\omega(\delta)$ *tending to* 0 *as* $\delta \to +0$ *such that*

$$|f(x+h) - f(x) - f'(x)h| \le \omega(\delta)|h| \tag{10.56}$$

at any point $x \in K$ *for* $|h| < \delta$ *if* $x + h \in K$.

Proof. By the corollary to the finite-increment theorem we can write

$$|f(x+h) - f(x) - f'(x)h| \le \sup_{0<\theta<1} \|f'(x+\theta h) - f'(x)\| \, |h|$$

and, setting

$$\omega(\delta) = \sup_{\substack{x_1, x_2 \in K \\ |x_1 - x_2| < \delta}} \|f'(x_2) - f'(x_1)\| \, ,$$

we obtain (10.56) in view of the uniform continuity of the function $x \mapsto f'(x)$, which is continuous on the compact set K. $\square$

b. A Sufficient Condition for Differentiability We shall now show that by using the general finite-increment theorem, we can obtain a general sufficient condition for differentiability of a mapping in terms of its partial derivatives.

Theorem 2. *Let* U *be a neighborhood of the point* x *in a normed space* $X = X_1 \times \cdots \times X_m$, *which is the direct product of the normed spaces* $X_1 \times \cdots \times X_m$, *and let* $f : U \to Y$ *be a mapping of* U *into a normed space* Y. *If the mapping* f *has partial derivatives with respect to all its variables in* U, *then it is differentiable at the point* x *if the partial derivatives are all continuous at that point.*

Proof. To simplify the writing we carry out the proof for the case $m = 2$. We verify immediately that the mapping

$$Lh = \partial_1 f(x)h_1 + \partial_2 f(x)h_2 \, ,$$

which is linear in $h = (h_1, h_2)$, is the total differential of f at x.

Making the elementary transformations

$$\begin{aligned}
&f(x+h) - f(x) - Lh = \\
&\quad = f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) - \partial_1 f(x)h_1 - \partial_2 f(x)h_2 = \\
&\quad = f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2 + h_2) - \partial_1 f(x_1, x_2)h_1 + \\
&\qquad\qquad + f(x_1, x_2 + h_2) - f(x_1, x_2) - \partial_2 f(x_1, x_2)h_2 \, ,
\end{aligned}$$

by the corollary to Theorem 1 we obtain

$$\begin{aligned}
&|f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) - \partial_1 f(x_1, x_2)h_1 - \partial_2 f(x_1, x_2)h_2| \le \\
&\quad \le \sup_{0<\theta_1<1} \|\partial_1 f(x_1 + \theta_1 h_1, x_2 + h_2) - \partial_1 f(x_1, x_2)\| \, |h_1| + \\
&\qquad + \sup_{0<\theta_2<1} \|\partial_2 f(x_1, x_2 + \theta_2 h_2) - \partial_2 f(x_1, x_2)\| \, |h_2| \, .
\end{aligned} \tag{10.57}$$

Since $\max\{|h_1|, |h_2\} \le |h|$, it follows obviously from the continuity of the partial derivatives $\partial_1 f$ and $\partial_2 f$ at the point $x = (x_1, x_2)$ that the right-hand side of inequality (10.57) is $o(h)$ as $h = (h_1, h_2) \to 0$. $\square$

Corollary. *A mapping $f : U \to Y$ of an open subset U of the normed space $X = X_1 \times \cdots \times X_m$ into a normed space Y is continuously differentiable if and only if all the partial derivatives of the mapping f are continuous.*

Proof. We have shown in Example 2 that when the mapping $f : U \to Y$ is differentiable, it is continuously differentiable if and only if its partial derivatives are continuous.

We now see that if the partial derivatives are continuous, then the mapping f is automatically differentiable, and hence (by Example 2) also continuously differentiable. $\square$

10.4.3 Problems and Exercises

1. Let $f : I \to Y$ be a continuous mapping of the closed interval $I = [0, 1] \subset \mathbb{R}$ into a normed space Y and $g : I \to \mathbb{R}$ a continuous real-valued function on I. Show that if f and g are differentiable in the open interval $]0, 1[$ and the relation $\|f'(t)\| \le g'(t)$ holds at points of this interval, then the inequality $|f(1) - f(0)| \le g(1) - g(0)$ also holds.

2. a) Let $f : I \to Y$ be a continuously differentiable mapping of the closed interval $I = [0, 1] \subset \mathbb{R}$ into a normed space Y. It defines a smooth path in Y. Define the length of that path.

b) Recall the geometric meaning of the norm of the tangent mapping and give an upper bound for the length of the path considered in a).

c) Give a geometric interpretation of the finite-increment theorem.

3. Let $f : U \to Y$ be a continuous mapping of a neighborhood U of the point a in a normed space X into a normed space Y. Show that if f is differentiable in $U \setminus a$ and $f'(x)$ has a limit $L \in \mathcal{L}(X; Y)$ as $x \to a$, then the mapping f is differentiable at a and $f'(a) = L$.

4. a) Let U be an open convex subset of a normed space X and $f : U \to Y$ a mapping of U into a normed space Y. Show that if $f'(x) \equiv 0$ on U, then the mapping f is constant.

b) Generalize the assertion of a) to the case of an arbitrary domain U (that is, when U is an open connected subset of X).

c) The partial derivative $\frac{\partial f}{\partial y}$ of a smooth function $f : D \to \mathbb{R}$ defined in a domain $D \subset \mathbb{R}^2$ of the xy-plane is identically zero. Is it true that f is then independent of y in this domain? For which domains D is this true?

10.5 Higher-order Derivatives

10.5.1 Definition of the nth Differential

Let U be an open set in a normed space X and

$$f : U \to Y \tag{10.58}$$

a mapping of U into a normed space Y.

If the mapping (10.58) is differentiable in U, then the derivative of f, given by

$$f' : U \to \mathcal{L}(X;Y)\ , \tag{10.59}$$

is defined in U.

The space $\mathcal{L}(X;Y) =: Y_1$ is a normed space relative to which the mapping (10.59) has the form (10.58), that is, $f' : U \to Y_1$, and it makes sense to speak of differentiability for it.

If the mapping (10.59) is differentiable, its derivative

$$(f')' : U \to \mathcal{L}(X;Y_1) = \mathcal{L}(X;\mathcal{L}(X;Y))$$

is called the *second derivative* or *second differential* of f and denoted f'' or $f^{(2)}$. In general, we adopt the following inductive definition.

Definition 1. The *derivative of order* $n \in \mathbb{N}$or *nth differential* of the mapping (10.58) at the point $x \in U$ is the mapping tangent to the derivative of f of order $n-1$ at that point.

If the derivative of order $k \in \mathbb{N}$ at the point $x \in U$ is denoted $f^{(k)}(x)$, Definition 1 means that

$$f^{(n)}(x) := \left(f^{(n-1)}\right)'(x)\ . \tag{10.60}$$

Thus, if $f^{(n)}(x)$ is defined, then

$$f^{(n)}(x) \in \mathcal{L}(X;Y_n) = \mathcal{L}(X;\mathcal{L}(X;Y_{n-1})) = \cdots$$
$$\cdots = \mathcal{L}(X;\mathcal{L}(X;\ldots;\mathcal{L}(X;Y))\ldots)\ .$$

Consequently, by Proposition 4 of Sect. 10.2, $f^{(n)}(x)$, the differential of order n of the mapping (10.58) at the point x can be interpreted as an element of the space $\mathcal{L}(\underbrace{X,\ldots,X}_{n\ \text{factors}};Y)$ of continuous n-linear transformations.

We note once again that the tangent mapping $f'(x) : TX_x \to TY_{f(x)}$ is a mapping of tangent spaces, each of which, because of the affine or vector-space structure of the spaces being mapped, we have identified with the corresponding vector space and said on that basis that $f'(x) \in \mathcal{L}(X;Y)$. It is this device of regarding elements $f'(x_1) \in \mathcal{L}(TX_{x_1};TY_{f(x_1)})$ and

$f'(x_2) \in \mathcal{L}(TX_{x_2}, TY_{f(x_2)})$, which lie in different spaces, as vectors in the same space $\mathcal{L}(X;Y)$ that provides the basis for defining higher-order differentials of mappings of normed vector spaces. In the case of an affine or vector space there is a natural connection between vectors in the different tangent spaces corresponding to different points of the original space. In the final analysis, it is this connection that makes it possible to speak of the continuous differentiability of both the mapping (10.58) and its higher-order differentials.

10.5.2 Derivative with Respect to a Vector and Computation of the Values of the nth Differential

When we are making the abstract Definition 1 specific, the concept of the derivative with respect to a vector may be used to advantage. This concept is introduced for the general mapping (10.58) just as was done earlier in the case $X = \mathbb{R}^m$, $Y = \mathbb{R}$.

Definition 2. If X and Y are normed vector spaces over the field $\mathbb{R}$, the *derivative of the mapping* (10.58) *with respect to the vector* $h \in TX_x \sim X$ *at the point* $x \in U$ is defined as the limit

$$D_h f(x) := \lim_{\mathbb{R}\ni t\to 0} \frac{f(x+th) - f(x)}{t} ,$$

provided this limit exists.

It can be verified immediately that

$$D_{\lambda h} f(x) = \lambda D_h f(x) \tag{10.61}$$

and that if the mapping f is differentiable at the point $x \in U$, it has a derivative at that point with respect to every vector; moreover

$$D_h f(x) = f'(x)h , \tag{10.62}$$

and, by the linearity of the tangent mapping,

$$D_{\lambda_1 h_1 + \lambda_2 h_2} f(x) = \lambda_1 D_{h_1} f(x) + \lambda_2 D_{h_2} f(x) . \tag{10.63}$$

It can also be seen from Definition 2 that the value $D_h f(x)$ of the derivative of the mapping $f : U \to Y$ with respect to a vector is an element of the vector space $TY_{f(x)} \sim Y$, and that if L is a continuous linear transformation from Y to a normed space Z, then

$$D_h(L \circ f)(x) = L \circ D_h f(x) . \tag{10.64}$$

We shall now try to give an interpretation to the value $f^{(n)}(h_1, \dots, h_n)$ of the nth differential of the mapping f at the point x on the set $(h_1, \dots, h_n)$ of vectors $h_i \in TX_x \sim X$, $i = 1, \dots, n$.

We begin with $n = 1$. In this case, by formula (10.62)

$$f'(x)(h) = f'(x)h = D_h f(x) .$$

We now consider the case $n = 2$. Since $f^{(2)}(x) \in \mathcal{L}(X; \mathcal{L}(X; Y))$, fixing a vector $h_1 \in X$, we assign a linear transformation $\left(f^{(2)}(x)h_1\right) \in \mathcal{L}(X; Y)$ to it by the rule

$$h_1 \mapsto f^{(2)}(x)h_1 .$$

Then, after computing the value of this operator at the vector $h_2 \in X$, we obtain an element of Y:

$$f^{(2)}(x)(h_1, h_2) := \left(f^{(2)}(x)h_1\right)h_2 \in Y . \tag{10.65}$$

But

$$f^{(2)}(x)h = (f')'(x)h = D_h f'(x) ,$$

and therefore

$$f^{(2)}(x)(h_1, h_2) = \left(D_{h_1} f'(x)\right)h_2 . \tag{10.66}$$

If $A \in \mathcal{L}(X; Y)$ and $h \in X$, this pairing with Ah can be regarded not only as a mapping $h \mapsto Ah$ from X into Y, but as a mapping $A \mapsto Ah$ from $\mathcal{L}(X; Y)$ into Y, the latter mapping being linear, just like the former.

Comparing relations (10.62), (10.64), and (10.66), we can write

$$\left(D_{h_1} f'(x)\right)h_2 = D_{h_1}\left(f'(x)h_2\right) = D_{h_1} D_{h_2} f(x) .$$

Thus we finally obtain

$$f^{(2)}(x)(h_1, h_2) = D_{h_1} D_{h_2} f(x) .$$

Similarly, one can show that the relation

$$f^{(n)}(x)(h_1, \dots, h_n) := \left(\dots (f^{(n)}(x)h_1) \dots h_n\right) = D_{h_1} D_{h_2} \cdots D_{h_n} f(x) \tag{10.67}$$

holds for any $n \in \mathbb{N}$, the differentiation with respect to the vectors being carried out sequentially, starting with differentiation with respect to h_n and ending with differentiation with respect to h_1.

10.5.3 Symmetry of the Higher-order Differentials

In connection with formula (10.67), which is perfectly adequate for computation as it now stands, the question naturally arises: To what extent does the result of the computation depend on the order of differentiation?

Proposition. *If the form $f^{(n)}(x)$ is defined at the point x for the mapping* (10.58), *it is symmetric with respect to any pair of its arguments.*

Proof. The main element in the proof is to verify that the proposition holds in the case $n = 2$.

Let h_1 and h_2 be two arbitrary fixed vectors in the space $TX_x \sim X$. Since U is open in X, the following auxiliary function of t is defined for all values of $t \in \mathbb{R}$ sufficiently close to zero:

$$F_t(h_1, h_2) = f(x + t(h_1 + h_2)) - f(x + th_1) - f(x + th_2) + f(x) .$$

We consider also the following auxiliary function:

$$g(v) = f(x + t(h_1 + v)) - f(x + tv) ,$$

which is certainly defined for vectors v that are collinear with the vector h_2 and such that $|v| \leq |h_2|$.

We observe that

$$F_t(h_1, h_2) = g(h_2) - g(0) .$$

We further observe that, since the function $f : U \to Y$ has a second differential $f''(x)$ at the point $x \in U$, it must be differentiable at least in some neighborhood of x. We shall assume that the parameter t is sufficiently small that the arguments on the right-hand side of the equality that defines $F_t(h_1, h_2)$ lie in that neighborhood.

We now make use of these observations and the corollary of the mean-value theorem in the following computations:

$$\begin{aligned}
&|F_t(h_1, h_2) - t^2 f''(x)(h_1, h_2)| = \\
&\qquad = |g(h_2) - g(0) - t^2 f''(x)(h_1, h_2)| \leq \\
&\qquad \leq \sup_{0<\theta_2<1} \|g'(\theta_2 h_2) - t^2 f''(x) h_1\| \, |h_2| = \\
&= \sup_{0<\theta_2<1} \|(f'(x + t(h_1 + \theta_2 h_2)) - f'(x + t\theta_2 h_2))t - t^2 f''(x) h_1\| \, |h_2| .
\end{aligned}$$

By definition of the derivative mapping we can write that

$$f'\big(x + t(h_1 + \theta_2 h_2)\big) = f'(x) + f''(x)\big(t(h_1 + \theta_2 h_2)\big) + o(t)$$

and

$$f'(x + t\theta_2 h_2) = f'(x) + f''(x)(t\theta_2 h_2) + o(t)$$

as $t \to 0$. Taking this relation into account, one can continue the preceding computation, finding after cancellation that

$$|F_t(h_1, h_2) - t^2 f''(x)(h_1, h_2)| = o(t^2)$$

as $t \to 0$. But this equality means that

$$f''(x)(h_2, h_2) = \lim_{t \to 0} \frac{F_t(h_1, h_2)}{t^2} .$$

Since it is obvious that $F_t(h_1, h_2) = F_t(h_2, h_1)$, it follows from this relation that $f''(x)(h_1, h_2) = f''(x)(h_2, h_1)$.

One can now complete the proof of the proposition by induction, repeating verbatim what was said in the proof that the values of the mixed partial derivatives are independent of the order of differentiation. □

Thus we have shown that the nth differential of the mapping (10.58) at the point $x \in U$ is a symmetric n-linear transformation

$$f^{(n)}(x) \in \mathcal{L}(TX_x, \dots, TX_x; TY_{f(x)}) \sim \mathcal{L}(X, \dots, X; Y)$$

whose value on the set $(h_1, \dots, h_n)$ of vectors $h_i \in TX_x = X$, $i = 1, \dots, n$, can be computed by formula (10.67).

If X is a finite-dimensional space having a basis $\{e_1, \dots, e_k\}$ and $h_j = h_j^i e_i$ is the expansion of the vector h_j, $j = 1, \dots, n$, with respect to that basis, then by the multilinearity of $f^{(n)}(x)$ we can write

$$\begin{aligned} f^{(n)}(x)(h_1, \dots, h_n) = f^{(n)}(x)(h_1^{i_1} e_{i_1}, \dots, h_n^{i_n} e_{i_n}) = \\ = f^{(n)}(x)(e_{i_1}, \dots, e_{i_n}) h_1^{i_1} \cdot \ldots \cdot h_n^{i_n} \ . \end{aligned}$$

Using our earlier notation $\partial_{i_1 \cdots i_n} f(x)$ for $D_{e_1} \cdots D_{e_n} f(x)$, we find finally that

$$f^{(n)}(x)(h_1, \dots, h_n) = \partial_{i_1 \cdots i_n} f(x) h_1^{i_1} \cdots h_n^{i_n} \ ,$$

where as usual summation extends over the repeated indices on the right-hand side within their range of variation, that is, from 1 to k.

Let us agree to use the following abbreviation:

$$f^{(n)}(x)(h, \dots, h) =: f^{(n)}(x) h^n \ . \tag{10.68}$$

In particular, if we are discussing a finite-dimensional space X and $h = h^i e_i$, then

$$f^{(n)}(x) h^n = \partial_{i_1 \cdots i_n} f(x) h^{i_1} \cdot \ldots \cdot h^{i_n} \ ,$$

which is already very familiar to us from the theory of numerical functions of several variables.

10.5.4 Some Remarks

In connection with the notation (10.68) consider the following example, which is quite useful and will be used in the next section.

Example. Let $A \in \mathcal{L}(X_1, \dots, X_n; Y)$, that is, $y = A(x_1, \dots, x_n)$ is a continuous n-linear transformation from the product of the normed vector spaces $X_1, \dots, X_n$ into the normed vector space Y.

It was shown in Example 5 of Sect. 10.4 that A is a differentiable mapping $A : X_1 \times \cdots \times X_n \to Y$ and

$$A'(x_1,\dots,x_n)(h_1,\dots,h_n) = \\ = A(h_1,x_2,\dots,x_n) + \dots + A(x_1,\dots,x_{n-1},h_n)\ .$$

Thus, if $X_1 = \dots = X_n = X$ and A is symmetric, then

$$A'(x,\dots,x)(h,\dots,h) = nA(\underbrace{x,\dots,x}_{n-1},h) =: (nAx^{n-1})h\ .$$

Hence, if we consider the function $F : X \to Y$ defined by the condition

$$X \ni x \mapsto F(x) = A(x,\dots,x) =: Ax^n\ ,$$

it turns out to be differentiable and

$$F'(x)h = (nAx^{n-1})h\ ,$$

that is, in this case

$$F'(x) = nAx^{n-1}\ ,$$

where $Ax^{n-1} := A(\underbrace{x,\dots,x}_{n-1},\cdot)$.

In particular, if the mapping (10.58) has a differential $f^{(n)}(x)$ at a point $x \in U$, then the function $F(h) = f^{(n)}(x)h^n$ is differentiable, and

$$F'(h) = nf^{(n)}(x)h^{n-1}\ . \tag{10.69}$$

To conclude our discussion of the concept of an nth-order derivative, it is useful to add the remark that if the original function (10.58) is defined on a set U in a space X that is the direct product of normed spaces $X_1,\dots,X_m$, one can speak of the first-order partial derivatives $\partial_1 f(x),\dots,\partial_m f(x)$ of f with respect to the variables $x_i \in X_i$, $i = 1,\dots,m$, and the higher-order partial derivatives $\partial_{i_1\cdots i_n} f(x)$.

On the basis of Theorem 2 of Sect. 10.4, we obtain by induction in this case that *if all the partial derivatives $\partial_{i_1\cdots i_n} f(x)$ of a mapping $f : U \to Y$ are continuous at a point $x \in X = X_1 \times \dots \times X_m$, then the mapping f has an n-th order differential $f^{(n)}(x)$ at that point.*

If we also take account of the result of Example 2 from the same section, we can conclude that *the mapping $U \ni x \mapsto f^{(n)}(x) \in \mathcal{L}(\underbrace{X,\dots,X}_{n\ \text{factors}};Y)$ is continuous if and only if all the nth-order partial derivatives $U \ni x \mapsto \partial_{i_1\cdots i_n} f(x) \in \mathcal{L}(X_{i_1},\dots,X_{i_n};Y)$ of the original mapping $f : U \to Y$ are continuous* (or, what is the same, the partial derivatives of all orders up to n inclusive are continuous).

The class of mappings (10.58) having continuous derivatives up to order n inclusive in U is denoted $C^{(n)}(U,Y)$, or, where no confusion can arise, by the briefer symbol $C^{(n)}(U)$ or even $C^{(n)}$.

In particular, if $X = X_1 \times \cdots \times X_n$, the conclusion reached above can be written in abbreviated form as

$$(f \in C^{(n)}) \Longleftrightarrow (\partial_{i_1 \cdots i_n} f \in C,\ i_1, \ldots, i_n = 1, \ldots, m)\ ,$$

where C, as always, denotes the corresponding set of continuous functions.

10.5.5 Problems and Exercises

1. Carry out the proof of Eq. (10.64) in full.

2. Give the details at the end of the proof that $f^{(n)}(x)$ is symmetric.

3. a) Show that if the functions $D_{h_1} D_{h_2} f$ and $D_{h_2} D_{h_1} f$ are defined and continuous at a point $x \in U$ for a pair of vectors h_1, h_2 and the mapping (10.58) in the domain U, then the equality $D_{h_1} D_{h_2} f(x) = D_{h_2} D_{h_1} f(x)$ holds.

b) Show using the example of a numerical function $f(x, y)$ that, although the continuity of the mixed partial derivatives $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ implies by a) that they are equal at this point, it does not in general imply that the second differential of the function exists at the point.

c) Show that, although the existence of $f^{(2)}(x, y)$, guarantees that the mixed partial derivatives $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ exist and are equal, it does not in general guarantee that they are continuous at that point.

4. Let $A \in \mathcal{L}(X, \ldots, X; Y)$ where A is a symmetric n-linear transformation. Find the successive derivatives of the function $x \mapsto Ax^n := A(x, \ldots, x)$ up to order $n+1$ inclusive.

10.6 Taylor's Formula and the Study of Extrema

10.6.1 Taylor's Formula for Mappings

Theorem 1. *If a mapping $f : U \to Y$ from a neighborhood $U = U(x)$ of a point x in a normed space X into a normed space Y has derivatives up to order $n - 1$ inclusive in U and has an n-th order derivative $f^{(n)}(x)$ at the point x, then*

$$f(x+h) = f(x) + f'(x)h + \cdots + \frac{1}{n!} f^{(n)}(x) h^n + o(|h|^n) \tag{10.70}$$

as $h \to 0$.

Equality (10.70) is one of the varieties of Taylor's formula, written here for rather general classes of mappings.

Proof. We prove Taylor's formula by induction.

For $n = 1$ it is true by definition of $f'(x)$.

Assume formula (10.70) is true for some $(n-1) \in \mathbb{N}$.

Then by the mean-value theorem, formula (10.69) of Sect. 10.5, and the induction hypothesis, we obtain

$$\Big| f(x+h) - \big(f(x) + f'(x)h + \cdots + \frac{1}{n!} f^{(n)}(x)h^n\big) \Big| \le$$
$$\le \sup_{0<\theta<1} \Big\| f'(x+\theta h) - \Big(f'(x) + f''(x)(\theta h) + \cdots$$
$$\cdots + \frac{1}{(n-1)!} f^{(n)}(x)(\theta h)^{n-1}\Big) \Big\| \, |h| = o\big(|\theta h|^{n-1}\big)|h| = o\big(|h|^n\big)$$

as $h \to 0$. $\square$

We shall not take the time here to discuss other versions of Taylor's formula, which are sometimes quite useful. They were discussed earlier in detail for numerical functions. At this point we leave it to the reader to derive them (see, for example, Problem 1 below).

10.6.2 Methods of Studying Interior Extrema

Using Taylor's formula, we shall exhibit necessary conditions and also sufficient conditions for an interior local extremum of real-valued functions defined on an open subset of a normed space. As we shall see, these conditions are analogous to the differential conditions already known to us for an extremum of a real-valued function of a real variable.

Theorem 2. *Let $f : U \to \mathbb{R}$ be a real-valued function defined on an open set U in a normed space X and having continuous derivatives up to order $k-1 \ge 1$ inclusive in a neighborhood of a point $x \in U$ and a derivative $f^{(k)}(x)$ of order k at the point x itself.*

If $f'(x) = 0, \dots, f^{(k-1)}(x) = 0$ and $f^{(k)}(x) \neq 0$, then for x to be an extremum of the function f it is:

n e c e s s a r y that k be even and that the form $f^{(k)}(x)h^k$ be semidefinite,[4] *and*

s u f f i c i e n t that the values of the form $f^{(k)}(x)h^k$ on the unit sphere $|h| = 1$ be bounded away from zero; moreover, x is a local minimum if the inequalities

$$f^{(k)}(x)h^k \ge \delta > 0$$

hold on that sphere, and a local maximum if

$$f^{(k)}(x)h^k \le \delta < 0 \,.$$

[4] This means that the form $f^{(k)}(x)h^k$ cannot take on values of opposite signs, although it may vanish for some values $h \neq 0$. The equality $f^{(i)}(x) = 0$, as usual, is understood to mean that $f^{(i)}(x)h = 0$ for every vector h.

Proof. For the proof we consider the Taylor expansion (10.70) of f in a neighborhood of x. The assumptions enable us to write

$$f(x+h)-f(x)=\frac{1}{k!}f^{(k)}(x)h^k+\alpha(h)|h|^k\ ,$$

where $\alpha(h)$ is a real-valued function, and $\alpha(h)\to 0$ as $h\to 0$.

We first prove the necessary conditions.

Since $f^{(k)}(x)\neq 0$, there exists a vector $h_0\neq 0$ on which $f^{(k)}(x)h_0^k\neq 0$. Then for values of the real parameter t sufficiently close to zero,

$$f(x+th_0)-f(x)=\frac{1}{k!}f^{(k)}(x)(th_0)^k+\alpha(th_0)|th_0|^k=$$
$$=\Big(\frac{1}{k!}f^{(k)}(x)h_0^k\pm\alpha(th_0)|h_0|^k\Big)t^k$$

and the expression in the outer parentheses has the same sign as $f^{(k)}(x)h_0^k$.

For x to be an extremum it is necessary for the left-hand side (and hence also the right-hand side) of this last equality to be of constant sign when t changes sign. But this is possible only if k is even.

This reasoning shows that if x is an extremum, then the sign of the difference $f(x+th_0)-f(x)$ is the same as that of $f^{(k)}(x)h_0^k$ for sufficiently small t; hence in that case there cannot be two vectors h_0, h_1 at which the form $f^{(k)}(x)$ assumes values with opposite signs.

We now turn to the proof of the sufficiency conditions. For definiteness we consider the case when $f^{(k)}(x)h^k\geq\delta>0$ for $|h|=1$. Then

$$f(x+h)-f(x)=\frac{1}{k!}f^{(k)}(x)h^k+\alpha(h)|h|^k=$$
$$=\Big(\frac{1}{k!}f^{(k)}(x)\Big(\frac{h}{|h|}\Big)^k+\alpha(h)\Big)|h|^k\geq\Big(\frac{1}{k!}\delta+\alpha(h)\Big)|h|^k\ ,$$

and, since $\alpha(h)\to 0$ as $h\to 0$, the last term in this inequality is positive for all vectors $h\neq 0$ sufficiently close to zero. Thus, for all such vectors h,

$$f(x+h)-f(x)>0\ ,$$

that is, x is a strict local minimum.

The sufficient condition for a strict local maximum is verified similiarly. □

Remark 1. If the space X is finite-dimensional, the unit sphere $S(x,1)$ with center at $x\in X$, being a closed bounded subset of X, is compact. Then the continuous function $f^{(k)}(x)h^k=\partial_{i_1\dots i_k}f(x)h^{i_1}\cdot\ldots\cdot h^{i_k}$ (a k-form) has both a maximal and a minimal value on $S(x,1)$. If these values are of opposite sign, then f does not have an extremum at x. If they are both of the same sign, then, as was shown in Theorem 2, there is an extremum. In the latter case, a sufficient condition for an extremum can obviously be stated as the equivalent requirement that the form $f^{(k)}(x)h^k$ be either positive- or negative-definite.

It was this form of the condition that we encountered in studying real-valued functions on $\mathbb{R}^n$.

Remark 2. As we have seen in the example of functions $f : \mathbb{R}^n \to \mathbb{R}$, the semi-definiteness of the form $f^{(k)}h^k$ exhibited in the necessary conditions for an extremum is not a sufficient criterion for an extremum.

Remark 3. In practice, when studying extrema of differentiable functions one normally uses only the first or second differentials. If the uniqueness and type of extremum are obvious from the meaning of the problem being studied, one can restrict attention to the first differential when seeking an extremum, simply finding the point x where $f'(x) = 0$.

10.6.3 Some Examples

Example 1. Let $L \in C^{(1)}(\mathbb{R}^3, \mathbb{R})$ and $f \in C^{(1)}\big([a,b], \mathbb{R}\big)$. In other words, $(u^1, u^2, u^3) \mapsto L(u^1, u^2, u^3)$ is a continuously differentiable real-valued function defined in $\mathbb{R}^3$ and $x \mapsto f(x)$ a smooth real-valued function defined on the closed interval $[a,b] \subset \mathbb{R}$.

Consider the function

$$F : C^{(1)}\big([a,b], \mathbb{R}\big) \to \mathbb{R} \tag{10.71}$$

defined by the relation

$$C^{(1)}\big([a,b], \mathbb{R}\big) \ni f \mapsto F(f) = \int_a^b L\big(x, f(x), f'(x)\big)\, \mathrm{d}x \in \mathbb{R}\,. \tag{10.72}$$

Thus, (10.71) is a real-valued functional defined on the set of functions $f \in C(1)\big([a,b], \mathbb{R}\big)$.

The basic variational principles connected with motion are known in physics and mechanics. According to these principles, the actual motions are distinguished among all the conceivable motions in that they proceed along trajectories along which certain functionals have an extremum. Questions connected with the extrema of functionals are central in optimal control theory. Thus, finding and studying the extrema of functionals is a problem of intrinsic importance, and the theory associated with it is the subject of a large area of analysis – the calculus of variations. We have already done a few things to make the transition from the analysis of the extrema of numerical functions to the problem of finding and studying extrema of functionals seem natural to the reader. However, we shall not go deeply into the special problems of variational calculus, but rather use the example of the functional

(10.72) to illustrate only the general ideas of differentiation and study of local extrema considered above.

We shall show that the functional (10.72) is a differentiable mapping and find its differential.

We remark that the function (10.72) can be regarded as the composition of the mappings

$$F_1 : C^{(1)}\big([a,b],\mathbb{R}\big) \to C\big([a,b],\mathbb{R}\big) \tag{10.73}$$

defined by the formula

$$F_1(f)(x) = L\big(x, f(x) f'(x)\big) \tag{10.74}$$

followed by the mapping

$$C\big([a,b],\mathbb{R}\big) \ni g \mapsto F_2(g) = \int\limits_a^b g(x)\,\mathrm{d}x \in \mathbb{R}\ . \tag{10.75}$$

By properties of the integral, the mapping F_2 is obviously linear and continuous, so that its differentiability is clear.

We shall show that the mapping F_1 is also differentiable, and that

$$F_1'(f)h(x) = \partial_2 L\big(x, f(x), f'(x)\big)h(x) + \partial_3 L\big(x, f(x)f'(x)\big)h'(x) \tag{10.76}$$

for $h \in C^{(1)}\big([a,b],\mathbb{R}\big)$.

Indeed, by the corollary to the mean-value theorem, we can write in the present case

$$\begin{aligned}\Big|L(u^1 \ + \ &\varDelta^1, u^2+\varDelta^2, u^3+\varDelta^3) - L(u^1,u^2,u^3) - \\ &- \sum_{i=1}^{3} \partial_i L(u^1,u^2,u^3)\varDelta^i\Big| \le \sup_{0<\theta<1} \|(\partial_1 L(u+\theta\varDelta) - \partial_1 L(u)\ , \\ &\partial_2\ L(u+\theta\varDelta) - \partial_2 L(u),\ \partial_3 L(u+\theta\varDelta) - \partial_3 L(u))\| \cdot |\varDelta| \le \\ &\qquad \le 3 \max_{\substack{0\le\theta\le1 \\ i=1,2,3}} |\partial_i L(u+\theta u) - \partial_i L(u)| \cdot \max_{i=1,2,3} |\varDelta^i|\ ,\end{aligned} \tag{10.77}$$

where $u = (u^1,u^2,u^3)$ and $\varDelta = (\varDelta^1,\varDelta^2,\varDelta^3)$.

If we now recall that the norm $|f|_{C^{(1)}}$ of the function f in $C^{(1)}\big([a,b],\mathbb{R}\big)$ is $\max\big\{|f|_C, |f'|_C\big\}$ (where $|f|_C$ is the maximum absolute value of the function on the closed interval $[a,b]$), then, setting $u^1 = x$, $u^2 = f(x)$, $u^3 = f'(x)$, $\varDelta^1 = 0$, $\varDelta^2 = h(x)$, and $\varDelta^3 = h'(x)$, we obtain from inequality (10.77), taking account of the uniform continuity of the functions $\partial_i L(u^1,u^2,u^3)$, $i = 1,2,3$, on bounded subsets of $\mathbb{R}^3$, that

$$\begin{aligned}\max_{a\le x\le b} &|L(x, f(x)+h(x), f'(x)+h'(x)) - L(x,f(x),f'(x)) - \\ &- \partial_2 L(x,f(x),f'(x))h(x) - \partial_3 L(X, f(x)f'(x))h'(x)| = \\ &\qquad\qquad = o\big(|h|_{C^{(1)}}\big) \ \text{ as } \ |h|_{C^{(1)}} \to 0\ .\end{aligned}$$

But this means that Eq. (10.76) holds.

By the chain rule for differentiating a composite function, we now conclude that the functional (10.72) is indeed differentiable, and

$$F'(f)h = \int_a^b \Big(\big(\partial_2 L(x, f(x) f'(x)\big)h(x) + \partial_3 L\big(x, f(x), f'(x)\big)\big)h'(x)\Big)\,\mathrm{d}x\ . \tag{10.78}$$

We often consider the restriction of the functional (10.72) to the affine space consisting of the functions $f \in C^{(1)}\big([a,b],\mathbb{R}\big)$ that assume fixed values $f(a) = A$, $f(b) = B$ at the endpoints of the closed interval $[a,b]$. In this case, the functions h in the tangent space $TC^{(1)}_f$ must have the value zero at the endpoints of the closed interval $[a,b]$. Taking this fact into account, we may integrate by parts in (10.78) and bring it into the form

$$F'(f)h = \int_a^b \Big(\partial_2 L\big(x, f(x), f'(x)\big) - \frac{\mathrm{d}}{\mathrm{d}x}\partial_3 L\big(x, f(x) f'(x)\big)\Big)h(x)\,\mathrm{d}x\ , \tag{10.79}$$

of course under the assumption that L and f belong to the corresponding class $C^{(2)}$.

In particular, if f is an extremum (extremal) of such a functional, then by Theorem 2 we have $F'(f)h = 0$ for every function $h \in C^{(1)}\big([a,b],\mathbb{R}\big)$ such that $h(a) = h(b) = 0$. From this and relation (10.79) one can easily conclude (see Problem 3 below) that the function f must satisfy the equation

$$\partial_2 L(x, f(x), f'(x)) - \frac{\mathrm{d}}{\mathrm{d}x}\partial_3 L(x, f(x), f'(x)) = 0\ . \tag{10.80}$$

This is a frequently-encountered form of the equation known in the calculus of variations as the *Euler–Lagrange equation.*

Let us now consider some specific examples.

Example 2. The shortest-path problem.

Among all the curves in a plane joining two fixed points, find the curve that has minimal length.

The answer in this case is obvious, and it rather serves as a check on the formal computations we will be doing later.

We shall assume that a fixed Cartesian coordinate system has been chosen in the plane, in which the two points are, for example, $(0,0)$ and $(1,0)$. We confine ourselves to just the curves that are the graphs of functions $f \in C^{(1)}\big([0,1],\mathbb{R}\big)$ assuming the value zero at both ends of the closed interval $[0,1]$. The length of such a curve

$$F(f) = \int_0^1 \sqrt{1 + (f')^2(x)}\,\mathrm{d}x \tag{10.81}$$

depends on the function f and is a functional of the type considered in Example 1. In this case the function L has the form

$$L(u^1, u^2, u^3) = \sqrt{1 + (u^3)^2} \;,$$

and therefore the necessary condition (10.80) for an extremal here reduces to the equation

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f'(x)}{\sqrt{1 + (f')^2(x)}}\right) = 0 \;,$$

from which it follows that

$$\frac{f'(x)}{\sqrt{1 + (f')^2(x)}} \equiv \text{const} \tag{10.82}$$

on the closed interval $[0, 1]$.

Since the function $\frac{u}{\sqrt{1+u^2}}$ is not constant on any interval, Eq. (10.82) is possible only if $f'(x) \equiv$ const on $[a, b]$. Thus a smooth extremal of this problem must be a linear function whose graph passes through the points $(0, 0)$ and $(1, 0)$. It follows that $f(x) \equiv 0$, and we arrive at the closed interval of the line joining the two given points.

Example 3. The brachistochrone problem.

The classical brachistochrone problem, posed by Johann Bernoulli I in 1696, was to find the shape of a track along which a point mass would pass from a prescribed point P_0 to another fixed point P_1 at a lower level under the action of gravity in the shortest time.

We neglect friction, of course. In addition, we shall assume that the trivial case in which both points lie on the same vertical line is excluded.

In the vertical plane passing through the points P_0 and P_1 we introduce a rectangular coordinate system such that P_0 is at the origin, the x-axis is directed vertically downward, and the point P_1 has positive coordinates (x_1, y_1). We shall find the shape of the track among the graphs of smooth functions defined on the closed interval $[0, x_1]$ and satisfying the condition $f(0) = 0$, $f(x_1) = y_1$. At the moment we shall not take time to discuss this by no means uncontroversial assumption (see Problem 4 below).

If the particle began its descent from the point P_0 with zero velocity, the law of variation of its velocity in these coordinates can be written as

$$v = \sqrt{2gx} \tag{10.83}$$

Recalling that the differential of the arc length is computed by the formula

$$\mathrm{d}s = \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2} = \sqrt{1 + (f')^2(x)}\,\mathrm{d}x \;, \tag{10.84}$$

we find the time of descent

$$F(f) = \frac{1}{\sqrt{2g}} \int_0^{x_1} \sqrt{\frac{1+(f')^2(x)}{x}}\, \mathrm{d}x \tag{10.85}$$

along the trajectory defined by the graph of the function $y = f(x)$ on the closed interval $[0, x_1]$.

For the functional (10.85)

$$L(u^1, u^2, u^3) = \sqrt{\frac{1+(u^3)^2}{u^1}}\ ,$$

and therefore the condition (10.80) for an extremum reduces in this case to the equation

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f'(x)}{\sqrt{x(1+(f')^2(x))}}\right) = 0\ ,$$

from which it follows that

$$\frac{f'(x)}{\sqrt{1+(f')^2(x)}} = c\sqrt{x}\ , \tag{10.86}$$

where c is a nonzero constant, since the points are not both on the same vertical line.

Taking account of (10.84), we can rewrite (10.86) in the form

$$\frac{\mathrm{d}y}{\mathrm{d}s} = c\sqrt{x}\ . \tag{10.87}$$

However, from the geometric point of view

$$\frac{\mathrm{d}x}{\mathrm{d}s} = \cos\varphi\ , \quad \frac{\mathrm{d}y}{\mathrm{d}s} = \sin\varphi\ , \tag{10.88}$$

where φ is the angle between the tangent to the trajectory and the positive x-axis.

By comparing Eq. (10.87) with the second equation in (10.88), we find

$$x = \frac{1}{c^2}\sin^2\varphi\ . \tag{10.89}$$

But it follows from (10.88) and (10.89) that

$$\frac{\mathrm{d}y}{\mathrm{d}\varphi} = \frac{\mathrm{d}y}{\mathrm{d}x}\cdot\frac{\mathrm{d}x}{\mathrm{d}\varphi} = \tan\varphi\frac{\mathrm{d}x}{\mathrm{d}\varphi} = \tan\varphi\frac{\mathrm{d}}{\mathrm{d}\varphi}\left(\frac{\sin^2\varphi}{c^2}\right) = 2\frac{\sin^2\varphi}{c^2}\ ,$$

from which we find

$$y = \frac{2}{c^2}(2\varphi - \sin 2\varphi) + b\ . \tag{10.90}$$

Setting $2/c^2 =: a$ and $2\varphi =: t$, we write relations (10.89) and (10.90) as

$$\begin{aligned} x &= a(1-\cos t)\,, \\ y &= a(t-\sin t)+b\,. \end{aligned} \tag{10.91}$$

Since $a \neq 0$, it follows that $x = 0$ only for $t = 2k\pi$, $k \in \mathbb{Z}$. It follows from the form of the function (10.91) that we may assume without loss of generality that the parameter value $t = 0$ corresponds to the point $P_0 = (0,0)$. In this case Eq. (10.90) implies $b = 0$, and we arrive at the simpler form

$$\begin{aligned} x &= a(1-\cos t)\,, \\ y &= a(t-\sin t) \end{aligned} \tag{10.92}$$

for the parametric definition of this curve.

Thus the brachistochrone is a cycloid having a cusp at the initial point P_0 where the tangent is vertical. The constant a, which is a scaling coefficient, must be chosen so that the curve (10.92) also passes through the point P_1. Such a choice, as one can see by sketching the curve (10.92), is by no means always unique, and this shows that the necessary condition (10.80) for an extremum is in general not sufficient. However, from physical considerations it is clear which of the possible values of the parameter a should be preferred (and this, of course, can be confirmed by direct computation).

10.6.4 Problems and Exercises

1. Let $f : U \to Y$ be a mapping of class $C^{(n)}(U;Y)$ from an open set U in a normed space X into a normed space Y. Suppose the closed interval $[x, x+h]$ is entirely contained in U, that f has a differential of order $(n+1)$ at the points of the open interval $]x, x+h[$, and that $\|f^{(n+1)}(\xi)\| \le M$ at every point $\xi \in]x, x+h[$.

a) Show that the function

$$g(t) = f(x+th) - \Big(f(x) + f'(x)(th) + \cdots + \frac{1}{n!}f^{(n)}(x)(th)^n\Big)$$

is defined on the closed interval $[0,1] \subset \mathbb{R}$ and differentiable on the open interval $]0,1[$, and that the estimate

$$\|g'(t)\| \le \frac{1}{n!}M|th|^n|h|$$

holds for every $t \in]0,1[$.

b) Show that $|g(1) - g(0)| \le \frac{1}{(n+1)!}M|h|^{n+1}$.

c) Prove the following version of Taylor's formula:

$$\left| f(x+h) - \Big(f(x) + f'(x)h + \cdots + \frac{1}{n!}f^{(n)}(x)h^n\Big)\right| \le \frac{M}{(n+1)!}|h|^{n+1}\,.$$

d) What can be said about the mapping $f : U \to Y$ if it is known that $f^{(n+1)}(x) \equiv 0$ in U?

2. a) If a symmetric n-linear operator A is such that $Ax^n = 0$ for every vector $x \in X$, then $A(x_1, \dots, x_n) \equiv 0$, that is, A equals zero on every set $x_1, \dots, x_n$ of vectors in X.

b) If a mapping $f : U \to Y$ has an nth-order differential $f^{(n)}(x)$ at a point $x \in U$ and satisfies the condition

$$f(x+h) = L_0 + L_1 h + \cdots + \frac{1}{n!} L_n h^n + \alpha(h)|h|^n ,$$

where L_i, $i = 0, 1, \dots, n$, are i-linear operators, and $\alpha(h) \to 0$ as $h \to 0$, then $L_i = f^{(i)}(x)$, $i = 0, 1, \dots, n$.

c) Show that the existence of the expansion for f given in the preceding problem does not in general imply the existence of the n-th order differential $f^{(n)}(x)$ (for $n > 1$) for the function at the point x.

d) Prove that the mapping $\mathcal{L}(X;Y) \ni A \mapsto A^{-1} \in \mathcal{L}(X;Y)$ is infinitely differentiable in its domain of definition, and that $(A^{-1})^{(n)}(A)(h_1, \dots, h_n) = (-1)^n A^{-1} h_1 A^{-1} h_2 \cdot \ldots \cdot A^{-1} h_n A^{-1}$.

3. a) Let $\varphi \in C\Big([a,b], \mathbb{R}\Big)$. Show that if the condition

$$\int\limits_a^b \varphi(x) h(x)\,\mathrm{d}x = 0$$

holds for every function $h \in C^{(2)}\Big([a,b], \mathbb{R}\Big)$ such that $h(a) = h(b) = 0$, then $\varphi(x) \equiv 0$ on $[a,b]$.

b) Derive the Euler–Lagrange equation (10.80) as a necessary condition for an extremum of the functional (10.72) restricted to the set of functions $f \in C^{(2)}\Big([a,b], \mathbb{R}\Big)$ assuming prescribed values at the endpoints of the closed interval $[a,b]$.

4. Find the shape $y = f(x)$, $a \le x \le b$, of a meridian of the surface of revolution (about the x-axis) having minimal area among all surfaces of revolution having circles of prescribed radius r_a and r_b as their sections by the planes $x = a$ and $x = b$ respectively.

5. a) The function L in the brachistochrone problem does not satisfy the conditions of Example 1, so that we cannot justify a direct application of the results of Example 1 in this case. Show by repeating the derivation of formula (10.79) with necessary modifications that this equation and Eq. (10.80) remain valid in this case.

b) Does the equation of the brachistochrone change if the particle starts from the point P_0 with a nonzero initial velocity (the motion is frictionless in a closed pipe)?

c) Show that if P is an arbitrary point of the brachistochrone corresponding to the pair of points P_0, P_1, the arc of that brachistochrone from P_0 to P is the brachistochrone of the pair P_0, P.

d) The assumption that the brachistochrone corresponding to a pair of points P_0, P_1 can be written as $y = f(x)$, is not always justified, as was revealed by the final formulas (10.92). Show by using the result of c) that the derivation of (10.92) can be carried out without any such assumption as to the global structure of the brachistochrone.

e) Locate a point P_1 such that the brachistochrone corresponding to the pair of points P_0, P_1 in the coordinate system introduced in Example 3 cannot be written in the form $y = f(x)$.

f) Locate a point P_1 such that the brachistochrone corresponding to the pair of points P_0, P_1 in the coordinate system introduced in Example 3) has the form $y = f(x)$, and $f \notin C^{(1)}\Big([a,b],\mathbb{R}\Big)$. Thus it turns out that in this case the functional (10.85) we are interested in has a greatest lower bound on the set $C^{(1)}\Big([a,b],\mathbb{R}\Big)$, but not a minimum.

g) Show that the brachistochrone of a pair of points P_0, P_1 of space is a smooth curve.

6. Let us measure the distance $d(P_0, P_1)$ of the point P_0 of space from the point P_1 in a homogeneous gravitational field by the time required for a point mass to move from one point to the other along the brachistochrone corresponding to the points.

a) Find the distance from the point P_0 to a fixed vertical line, measured in this sense.

b) Find the asymptotic behavior of the function $d(P_0, P_1)$ as the point P_1 is raised along a vertical line, approaching the height of the point P_0.

c) Determine whether the function $d(P_0, P_1)$ is a metric.

10.7 The General Implicit Function Theorem

In this concluding section of the chapter we shall illustrate practically all of the machinery we have developed by studying an implicitly defined function. The reader already has some idea of the content of the implicit theorem, its place in analysis, and its applications from Chap. 8. For that reason, we shall not go into detail here with preliminary explanations of the essence of the matter preceding the formalism. We note only that this time the implicitly defined function will be constructed by an entirely different method, one that relies on the contraction mapping principle. This method is often used in analysis and is quite useful because of its computational efficiency.

Theorem. *Let X, Y, and Z be normed spaces (for example, $\mathbb{R}^m$, $\mathbb{R}^n$, and $\mathbb{R}^k$), Y being a complete space. Let $W = \{(x,y) \in X \times Y \mid |x - x_0| < \alpha \wedge |y - y_0| < \beta\}$ be a neighborhood of the point (x_0, y_0) in the product $X \times Y$ of the spaces X and Y.*

Suppose that the mapping $F : W \to Z$ satisfies the following conditions:

1. $F(x_0, y_0) = 0$;
2. *$F(x, y)$ is continuous at (x_0, y_0);*

3. $F'(x,y)$ *is defined in* W *and continuous at* (x_0, y_0);
4. $F'_y(x_0, y_0)$ *is an invertible*[5] *transformation.*

Then there exists a neighborhood $U = U(x_0)$ *of* $x_0 \in X$, *a neighborhood* $V = V(y_0)$ *of* $y_0 \in Y$, *and a mapping* $f : U \to V$ *such that:*

1′. $U \times V \subset W$;
2′. $\big(F(x,y) = 0 \ \textit{in} \ U \times V\big) \Leftrightarrow \big(y = f(x), \ \textit{where}\ x \in U \ \textit{and}\ f(x) \in V\big)$;
3′. $y_0 = f(x_0)$;
4′. f *is continuous at* x_0.

In essence, this theorem asserts that if the linear mapping F'_y is invertible at a point (hypothesis 4), then in a neighborhood of this point the relation $F(x,y) = 0$ is equivalent to the functional dependence $y = f(x)$ (conclusion 2′).

Proof. 1^0 To simplify the notation and obviously with no loss of generality, we may assume that $x_0 = 0$, $y_0 = 0$, and consequently

$$W = \{(x,y) \in X \times Y | \, |x| < \alpha \wedge |y| < \beta\} \, .$$

2^0 The main role in the proof is played by the auxiliary family of functions

$$g_x(y) := y - \big(F'_y(0,0)\big)^{-1} \cdot F(x,y) \, , \tag{10.93}$$

which depend on the parameter $x \in S$, $|x| < \alpha$, and are defined on the set $\{y \in Y| \, |y| < \beta\}$.

Let us discuss formula (10.93). We first determine whether the mappings g_x are unambiguously defined and where their values lie.

The mapping F is defined for $(x,y) \in W$, and its value $F(x,y)$ at the pair (x,y) lies in Z. The partial derivative $F'_y(x,y)$ at any point $(x,y) \in W$, as we know, is a continuous linear mapping from Y into Z.

By hypothesis 4 the mapping $F'_y(0,0) : Y \to Z$ has a continuous inverse $\big(F'_y(0,0)\big)^{-1} : Z \to Y$. Hence the composition $\big(F'_y(0,0)\big)^{-1} \cdot F(x,y)$ really is defined, and its values lie in Y.

Thus, for any x in the α-neighborhood $B_X(0,\alpha) := \{x \in X| \, |x| < \alpha\}$ of the point $0 \in X$, the function g_x is a mapping $g_x : B_Y(0,\beta) \to Y$ from the β-neighborhood $B_Y(0,\beta) := \{y \in Y| \, |y| < \beta\}$ of the point $0 \in Y$ into Y.

The connection of the mappings (10.93) with the problem of solving the equation $F(x,y) = 0$ for y obviously consists of the following: the point y_x is a fixed point of g_x if and only if $F(x, y_x) = 0$.

Let us state this important observation firmly:

$$g_x(y_x) = y_x \Longleftrightarrow F(x, y_x) = 0 \, . \tag{10.94}$$

Thus, finding and studying the implicitly defined function $y = y_x = f(x)$ reduces to finding the fixed points of the mappings (10.93) and studying the way in which they depend on the parameter x.

[5] That is, $\exists \, [F'_y(x_0, y_0)]^{-1} \in \mathcal{L}(Z; Y)$.

3^0 We shall show that there exists a positive number $\gamma < \min\{\alpha, \beta\}$ such that for each $x \in X$ satisfying the condition $|x| < \gamma < \alpha$, the mapping $g_x : B_Y(0,\gamma) \to Y$ of the ball $B_Y(0,\gamma) := \{y \in Y | \, |y| < \gamma < \beta\}$ into Y is a contraction with a coefficient of contraction that does not exceed, say 1/2. Indeed, for each fixed $x \in B_X(0,\alpha)$ the mapping $g_x : B_Y(0,\beta) \to Y$ is differentiable, as follows from hypothesis 3 and the theorem on differentiation of a composite mapping. Moreover,

$$g'_x(y) = e_Y - \big(F'_y(0,0)\big)^{-1} \cdot \big(F'_y(x,y)\big) = \\ = \big(F'_y(0,0)\big)^{-1}\big(F'_y(0,0) - F'_y(x,y)\big) . \quad (10.95)$$

By the continuity of $F'_y(x,y)$ at the point $(0,0)$ (hypothesis 3), there exists a neighborhood $\{(x,y) \in X \times Y | \, |x| < \gamma < \alpha \wedge |y| < \gamma < \beta\}$ of $(0,0) \in X \times Y$ in which

$$\|g'_x(y)\| \le \|(F'_y(0,0))^{-1}\| \cdot \|F'_y(0,0) - F'_y(x,y)\| < \frac{1}{2} . \quad (10.96)$$

Here we are using the relation

$$\big(F'_y(0,0)\big)^{-1} \in \mathcal{L}(Z;Y) , \text{ that is, } \|(F'_y(0,0))^{-1}\| < \infty .$$

Throughout the following we shall assume that $|x| < \gamma$ and $|y| < \gamma$, so that estimate (10.96) holds.

Thus, at any $x \in B_X(0,\gamma)$ and for any $y_1, y_2 \in B_Y(0,\gamma)$, by the mean-value theorem, we indeed now find that

$$|g_x(y_1) - g_x(y_2)| \le \sup_{\xi \in]y_1,y_2[} \|g'(\xi)\| \, |y_1 - y_2| < \frac{1}{2}|y_1 - y_2| . \quad (10.97)$$

4^0. In order to assert the existence of a fixed point y_x for the mapping g_x, we need a complete metric space that maps into (but not necessarily onto) itself under this mapping.

We shall verify that for any ε satisfying $0 < \varepsilon < \gamma$ there exists $\delta = \delta(\varepsilon)$ in the open interval $]0,\gamma[$ such that for any $x \in B_X(0,\delta)$ the mapping g_x maps the closed ball $\overline{B}_y(0,\varepsilon)$ into itself, that is, $g_x\big(\overline{B}_Y(0,\varepsilon)\big) \subset \overline{B}_Y(0,\varepsilon)$.

Indeed, we first choose a number $\delta \in]0,\gamma[$ depending on ε such that

$$|g_x(0)| = |(F'_y(0,0))^{-1} \cdot F(x,0)\| \le \|(F'_y(0,0))^{-1}\| \, |F(x,0)| < \frac{1}{2}\varepsilon \quad (10.98)$$

for $|x| < \delta$.

This can be done by hypotheses 1 and 2, which guarantee that $F(0,0) = 0$ and $F(x,y)$ is continuous at $(0,0)$.

Now if $|x| < \delta(\varepsilon) < \gamma$ and $|y| \le \varepsilon < \gamma$, we find by (10.97) and (10.98) that

$$|g_x(y)| \le |g_x(y) - g_x(0)| + |g_x(0)| < \frac{1}{2}|y| + \frac{1}{2}\varepsilon < \varepsilon ,$$

and hence for $|x| < \delta(\varepsilon)$

$$g_x\big(\overline{B}_Y(0,\varepsilon)\big) \subset B_Y(0,\varepsilon) . \quad (10.99)$$

Being a closed subset of the complete metric space Y, the closed ball $\overline{B}_Y(0,\varepsilon)$ is itself a complete metric space.

5^0 Comparing relations (10.97) and (10.99), we can now assert by the fixed-point principle (Sect. 9.7) that for each $x \in B_X(0,\delta(\varepsilon)) =: U$ there exists a unique point $y = y_x =: f(x) \in B_Y(0,\varepsilon) =: V$ that is a fixed point of the mapping $g_x : \overline{B}_Y(0,\varepsilon) \to \overline{B}_Y(0,\varepsilon)$.

By the basic relation (10.94), it follows from this that the function $f : U \to V$ so constructed has property $2'$ and hence also property $3'$, since $F(0,0) = 0$ by hypothesis 1.

Property $1'$ of the neighborhoods U and V follows from the fact that, by construction, $U \times V \subset B_X(0,\alpha) \times B_Y(0,\beta) = W$.

Finally, the continuity of the function $y = f(x)$ at $x = 0$, that is, property $4'$, follows from $2'$ and the fact that, as was shown in part 4^0 of the proof, for every $\varepsilon > 0$ $(\varepsilon < \gamma)$ there exists $\delta(\varepsilon) > 0$ $(\delta(\varepsilon) < \gamma)$ such that $g_x\big(\overline{B}_Y(0,\varepsilon)\big) \subset B_Y(0,\varepsilon)$ for any $x \in B_X(0,\delta(\varepsilon))$, that is, the unique fixed point $y_x = f(x)$ of the mapping $g_x : \overline{B}_Y(0,\varepsilon) \to \overline{B}_Y(0,\varepsilon)$ satisfies the condition $|f(x)| < \varepsilon$ for $|x| < \delta(\varepsilon)$. □

We have now proved the existence of the implicit function. We now prove a number of extensions of these properties of the function, generated by properties of the original function F.

Extension 1. (Continuity of the implicit function.) *If in addition to hypotheses 2 and 3 of the theorem it is known that the mappings $F : W \to Z$ and F'_y are continuous not only at the point (x_0, y_0) but in some neighborhood of this point, then the function $f : U \to V$ will be continuous not only at $x_0 \in U$ but in some neighborhood of this point.*

Proof. By properties of the mapping $\mathcal{L}(Y;Z) \ni A \mapsto A^{-1} \in \mathcal{L}(Z;Y)$ it follows from hypotheses 3 and 4 of the theorem (see Example 6 of Sect. 10.3) that at each point (x,y) in some neighborhood of (x_0,y_0) the transformation $f'_y(x,y) \in \mathcal{L}(Y;Z)$ is invertible. Thus under the additional hypothesis that F is continuous all points $(\tilde{x},\tilde{y})$ of the form $(x, f(x))$ in some neighborhood of (x_0,y_0) satisfy hypotheses 1–4, previously satisfied only by the point (x_0,y_0).

Repeating the construction of the implicit function in a neighborhood of these points $(\tilde{x},\tilde{y})$, we would obtain a function $y = \tilde{f}(x)$ that is continuous at $\tilde{x}$ and by $2'$ would coincide with the function $y = f(x)$ in some neighborhood of x. But that means that f itself is continuous at $\tilde{x}$. □

Extension 2. (Differentiability of the implicit function.) *If in addition to the hypotheses of the theorem it is known that a partial derivative $F'_x(x,y)$ exists in some neighborhood W of (x_0,y_0) and is continuous at (x_0,y_0), then the function $y = f(x)$ is differentiable at x_0, and*

$$f'(x_0) = -\big(F'_y(x_0,y_0)\big)^{-1} \cdot \big(F'_x(x_0,y_0)\big)\ . \tag{10.100}$$

Proof. We verify immediately that the linear transformation $L \in \mathcal{L}(X;Y)$ on the right-hand side of formula (10.100) is indeed the differential of the function $y = f(x)$ at x_0.

As before, to simplify the notation, we shall assume that $x_0 = 0$ and $y_0 = 0$, so that $f(0) = 0$.

We begin with a preliminary computation.

$$
\begin{aligned}
|f(x) - f(0) - Lx| = |f(x) - Lx| = \\
= \left|f(x) + \left(F'_y(0,0)\right)^{-1} \cdot \left(F'_x(0,0)\right)x\right| = \\
= \left|\left(F'_y(0,0)\right)^{-1}\left(F'_x(0,0)x + F'_y(0,0)f(x)\right)\right| = \\
= \left|\left(F'_y(0,0)\right)^{-1}\left(F(x,f(x)) - F(0,0) - F'_x(0,0)x - F'_y(0,0)f(x)\right)\right| \le \\
\le \left\|\left(F'_y(0,0)\right)^{-1}\right\|\left|\left(F(x,f(x)) - F(0,0) - F'_x(0,0)x - F'_y(0,0)f(x)\right)\right| \le \\
\le \left\|\left(F'_y(0,0)\right)^{-1}\right\| \cdot \alpha\big(x,f(x)\big)\big(|x| + |f(x)|\big)\ ,
\end{aligned}
$$

where $\alpha(x,y) \to 0$ as $(x,y) \to (0,0)$.

These relations have been written taking account of the relation $F\big(x,f(x)\big) \equiv 0$ and the fact that the continuity of the partial derivatives F'_x and F'_y at $(0,0)$ guarantees the differentiability of the function $F(x,y)$ at that point.

For convenience in writing we set $a := \|L\|$ and $b := \left\|\left(F'_y(0,0)\right)^{-1}\right\|$.

Taking account of the relations

$$
|f(x)| = |f(x) - Lx + Lx| \le |f(x) - Lx| + |Lx| \le |f(x) - Lx| + a|x|\ ,
$$

we can extend the preliminary computation just done and obtain the relation

$$
|f(x) - Lx| \le b\alpha(x,f(x))\big((a+1)|x| + |f(x) - Lx|\big)\ ,
$$

or

$$
|f(x) - Lx| \le \frac{(a+1)b}{1 - b\alpha(x,f(x))}\alpha(x,f(x))|x|\ .
$$

Since f is continuous at $x = 0$ and $f(0) = 0$, we also have $f(x) \to 0$ as $x \to 0$, and therefore $\alpha(x,f(x)) \to 0$ as $x \to 0$.

It therefore follows from the last inequality that

$$
|f(x) - f(0) - Lx| = |f(x) - Lx| = o\big(|x|\big) \text{ as } x \to 0.\ \square
$$

Extension 3. (Continuous differentiability of the implicit function.) *If in addition to the hypotheses of the theorem it is known that the mapping F has continuous partial derivatives F'_x and F'_y in some neighborhood W of (x_0, y_0), then the function $y = f(x)$ is continuously differentiable in some neighborhood of x_0, and its derivative is given by the formula*

$$
f'(x) = -\big(F'_y(x,f(x))\big)^{-1} \cdot \big(F'_x(x,f(x))\big)\ . \tag{10.101}
$$

Proof. We already know from formula (10.100) that the derivative $f'(x)$ exists and can be expressed in the form (10.101) at an individual point x at which the transformation $F'_y(x, f(x))$ is invertible.

It remains to be verified that under the present hypotheses the function $f'(x)$ is continuous in some neighborhood of $x = x_0$.

The bilinear mapping $(A, B) \mapsto A \cdot B$ – the product of linear transformations A and B – is a continuous function.

The transformation $B = -F'_x(x, f(x))$ is a continuous function of x, being the composition of the continuous functions $x \mapsto (x, f(x)) \mapsto -F'_x(x, f(x))$.

The same can be said about the linear transformation $A^{-1} = F'_y(x, f(x))$.

It remains only to recall (see Example 6 of Sect. 10.3) that the mapping $A^{-1} \mapsto A$ is also continuous in its domain of definition.

Thus the function $f'(x)$ defined by formula (10.101) is continuous in some neighborhood of $x = x_0$, being the composition of continuous functions. □

We can now summarize and state the following general proposition.

Proposition. *If in addition to the hypotheses of the implicit function theorem it is known that the function F belongs to the class $C^{(k)}(W, Z)$, then the function $y = f(x)$ defined by the equation $F(x, y) = 0$ belongs to $C^{(k)}(U, Y)$ in some neighborhood U of x_0.*

Proof. The proposition has already been proved for $k = 0$ and $k = 1$. The general case can now be obtained by induction from formula (10.101) if we observe that the mapping $\mathcal{L}(Y; Z) \ni A \mapsto A^{-1} \in \mathcal{L}(Z; Y)$ is (infinitely) differentiable and that when Eq. (10.101) is differentiated, the right-hand side always contains a derivative of f one order less than the left-hand side. Thus, successive differentiation of Eq. (10.101) can be carried out a number times equal to the order of smoothness of the function F. □

In particular, if

$$f'(x)h_1 = -\big(F'_y(x, f(x))\big)^{-1} \cdot \big(F'_x(x, f(x))\big)h_1 ,$$

then

$$\begin{aligned}
f''(x)(h_1, h_2) &= -\mathrm{d}\big(F'_y(x, f(x))\big)^{-1} h_2 F'_x(x, f(x))h_1 - \\
&\qquad - \big(F'_y(x, f(x))\big)^{-1} \mathrm{d}\big(F'_x(x, f(x))h_1\big)h_2 = \\
&= \big(F'_y(x, f(x))\big)^{-1} \mathrm{d}F'_y(x, f(x))h_2 \big(F'_y(x, f(x))\big)^{-1} F'_x(x, f(x))h_1 - \\
&\qquad - \big(F'_y(x, f(x))\big)^{-1} \big((F''_{xx}(x, f(x)) + F''_{yy}(x, f(x))f'(x))h_1\big)h_2 = \\
&= \big(F'_y(x, f(x))\big)^{-1} \big((F''_{yx}(x, f(x)) + F''_{yy}(x, f(x))f'(x))h_2\big) \times \\
&\qquad \times \big(F'_y(x, f(x))\big)^{-1} F'_x(x, f(x))h_1\big)\big(F'_y(x, f(x))\big)^{-1} \times \\
&\qquad \times \big((F''_{xx}(x, f(x)) + F''_{xy}(x, f(x))f'(x))h_1\big)h_2 .
\end{aligned}$$

In less detailed, but more readable notation, this means that

$$f''(x)(h_1,h_2) = (F'_y)^{-1}\big[\big((F''_{yx} + F''_{yy}f')h_2\big)(F'_y)^{-1}F'_x h_1 - \\ - \big((F''_{xx} + F''_{yy}f')h_1\big)h_2\big] . \quad (10.102)$$

In this way one could theoretically obtain an expression for the derivative of an implicit function to any order; however, as can be seen even from formula (10.102), these expressions are generally too cumbersome to be conveniently used. Let us now see how these results can be made specific in the important special case when $X = \mathbb{R}^m$, $Y = \mathbb{R}^n$, and $Z = \mathbb{R}^n$.

In this case the mapping $z = F(x,y)$ has the coordinate representation

$$\begin{aligned} z^1 &= F^1(x^1,\dots,x^m,y^1,\dots,y^n) , \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ z^n &= F^n(x^1,\dots,x^m,y^1,\dots,y^n) . \end{aligned} \quad (10.103)$$

The partial derivatives $F'_x \in \mathcal{L}(\mathbb{R}^m;\mathbb{R}^n)$ and $F'_y \in \mathcal{L}(\mathbb{R}^n;\mathbb{R}^n)$ of the mapping are defined by the matrices

$$F'_x = \begin{pmatrix} \frac{\partial F^1}{\partial x^1} & \cdots & \frac{\partial F^1}{\partial x^m} \\ \dots & \dots & \dots \\ \frac{\partial F^n}{\partial x^1} & \cdots & \frac{\partial F^n}{\partial x^m} \end{pmatrix} , \qquad F'_y = \begin{pmatrix} \frac{\partial F^1}{\partial y^1} & \cdots & \frac{\partial F^1}{\partial y^n} \\ \dots & \dots & \dots \\ \frac{\partial F^n}{\partial y^1} & \cdots & \frac{\partial F^n}{\partial y^n} \end{pmatrix} ,$$

computed at the corresponding point (x,y).

As we know, the condition that F'_x and F'_y be continuous is equivalent to the continuity of all the entries of these matrices.

The invertibility of the linear transformation $F'_y(x_0,y_0) \in \mathcal{L}(\mathbb{R}^n;\mathbb{R}^n)$ is equivalent to the nonsingularity of the matrix that defines this transformation.

Thus, in the present case the implicit function theorem asserts that if

1) $F^1(x_0^1,\dots,x_0^m,y_0^1,\dots,y_0^n) = 0$,
$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$
$F^n(x_0^1,\dots,x_0^m,y_0^1,\dots,y_0^n) = 0$;

2) $F^i(x^1,\dots,x^m,y^1,\dots,y^n)$, $i = 1,\dots,n$, are continuous functions at the point $(x_0^1,\dots,x_0^m,y_0^1,\dots,y_0^n) \in \mathbb{R}^m \times \mathbb{R}^n$;

3) all the partial derivatives $\frac{\partial F^i}{\partial y^j}(x^1,\dots,x^m,y^1,\dots,y^n)$, $i = 1,\dots,n$, $j = 1,\dots,n$, are defined in a neighborhood of $(x_0^1,\dots,x_0^m,y_0^1,\dots,y_0^n)$ and are continuous at this point;

4) the determinant

$$\begin{vmatrix} \frac{\partial F^1}{\partial y^1} & \cdots & \frac{\partial F^1}{\partial y^n} \\ \multicolumn{3}{c}{\dotfill} \\ \frac{\partial F^n}{\partial y^1} & \cdots & \frac{\partial F^n}{\partial y^n} \end{vmatrix}$$

of the matrix F'_y is nonzero at the point $(x_0^1, \dots, x_0^m, y_0^1, \dots, y_0^n)$;

then there exist a neighborhood U of $x_0 = (x_0^1, \dots, x_0^m) \in \mathbb{R}^m$, a neighborhood V of $y_0 = (y_0^1, \dots, y_0^n) \in \mathbb{R}^n$, and a mapping $f : U \to V$ having a coordinate representation

$$\begin{aligned} y^1 &= f^1(x^1, \dots, x^m) , \\ &\dots\dots\dots\dots\dots\dots \\ y^n &= f^n(x^1, \dots, x^m) , \end{aligned} \tag{10.104}$$

such that

1′) inside the neighborhood $U \times V$ of $(x_0^1, \dots, x_0^m, y_0^1, \dots, y_0^n) \in \mathbb{R}^m \times \mathbb{R}^n$ the system of equations

$$\begin{cases} F^1(x^1, \dots, x^m, y^1, \dots, y^n) = 0 , \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ F^n(x^1, \dots, x^m, y^1, \dots, y^n) = 0 \end{cases}$$

is equivalent to the functional relation $f : U \to V$ expressed by (10.104);

2′)
$$\begin{aligned} y_0^1 &= f^1(x_0^1, \dots, x_0^m) , \\ &\dots\dots\dots\dots\dots\dots \\ y_0^n &= f^n(x_0^1, \dots, x_0^m) ; \end{aligned}$$

3′) the mapping (10.104) is continuous at $(x_0^1, , \dots, x_0^m, y_0^1, \dots, y_0^n)$.

If in addition it is known that the mapping (10.103) belongs to the class $C^{(k)}$, then, as follows from the proposition above, the mapping (10.104) will also belong to $C^{(k)}$, of course within its own domain of definition.

In this case formula (10.101) can be made specific, becoming the matrix equality

$$\begin{pmatrix} \frac{\partial f^1}{\partial x^1} & \cdots & \frac{\partial f^1}{\partial x^m} \\ \multicolumn{3}{c}{\dotfill} \\ \frac{\partial f^n}{\partial x^1} & \cdots & \frac{\partial f^n}{\partial x^m} \end{pmatrix} = - \begin{pmatrix} \frac{\partial F^1}{\partial y^1} & \cdots & \frac{\partial F^1}{\partial y^n} \\ \multicolumn{3}{c}{\dotfill} \\ \frac{\partial F^n}{\partial y^1} & \cdots & \frac{\partial F^n}{\partial y^n} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial F^1}{\partial x^1} & \cdots & \frac{\partial F^1}{\partial x^m} \\ \multicolumn{3}{c}{\dotfill} \\ \frac{\partial F^n}{\partial x^1} & \cdots & \frac{\partial F^n}{\partial x^m} \end{pmatrix} ,$$

in which the left-hand side is computed at $(x^1, \dots, x^m)$ and the right-hand side at the corresponding point $(x^1, \dots, x^m, y^1, \dots, y^n)$, where $y^i = f^i(x^1, \dots, x^m)$, $i = 1, \dots, n$.

If $n = 1$, that is, when the equation

$$F(x^1, \dots, x^m, y) = 0$$

is being solved for y, the matrix F'_y consists of a single entry – the number $\frac{\partial F}{\partial y}(x^1, \dots, x^m, y)$. In this case $y = f(x^1, \dots, x^m)$, and

$$\left(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^m}\right) = -\left(\frac{\partial F}{\partial y}\right)^{-1} \left(\frac{\partial F}{\partial x^1}, \dots, \frac{\partial F}{\partial x^m}\right). \tag{10.105}$$

In this case formula (10.102) also simplifies slightly; more precisely, it can be written in the following more symmetric form:

$$f''(x)(h_1, h_2) = -\frac{(F''_{xx} + F''_{xy} f')h_1 F'_y h_2 - (F''_{yx} + F''_{yy} f')h_2 F'_x h_1}{(F'_y)^2}. \tag{10.106}$$

And if $n = 1$ and $m = 1$, then $y = f(x)$ is a real-valued function of one real argument, and formulas (10.105) and (10.106) simplify to the maximum extent, becoming the numerical equalities

$$\begin{aligned} f'(x) &= -\frac{F'_x}{F'_y}(x, y), \\ f''(x) &= -\frac{(F''_{xx} + F''_{xy} f')F'_y - (F''_{yx} + F''_{yy} f')F'_x}{(F'_y)^2}(x, y) \end{aligned}$$

for the first two derivatives of the implicit function defined by the equation $F(x, y) = 0$.

10.7.1 Problems and Exercises

1. a) Assume that, along with the function $f : U \to Y$ given by the implicit function theorem, we have a function $\tilde{f} : \widetilde{U} \to Y$ defined in some neighborhood $\widetilde{U}$ of x_0 and satisfying $y_0 = \tilde{f}(x_0)$ and $F(x, \tilde{f}(x)) \equiv 0$ in $\widetilde{U}$. Prove that if $\tilde{f}$ is continuous at x_0, then the functions f and $\tilde{f}$ are equal on some neighborhood of x_0.

b) Show that the assertion in a) is generally not true without the assumption that $\tilde{f}$ is continuous at x_0.

2. Analyze once again the proof of the implicit function theorem and the extensions to it, and show the following.

a) If $z = F(x, y)$ is a continuously differentiable complex-valued function of the complex variables x and y, then the implicit function $y = f(x)$ defined by the equation $F(x, y) = 0$ is differentiable with respect to the complex variable x.

b) Under the hypotheses of the theorem X is not required to be a normed space, and may be any topological space.

3. a) Determine whether the form $f''(x)(h_1, h_2)$ defined by relation (10.102) is symmetric.

b) Write the forms (10.101) and (10.102) for the case of numerical functions $F(x^1, x^2, y)$ and $F(x, y^1, y^2)$ in matrix form.

c) Show that if $\mathbb{R} \ni t \mapsto A(t) \in \mathcal{L}(\mathbb{R}^n; \mathbb{R}^n)$ is family of nonsingular matrices $A(t)$ depending on the parameter t in an infinitely smooth manner, then

$$\frac{\mathrm{d}^2 A^{-1}}{\mathrm{d}t^2} = 2A^{-1}\left(\frac{\mathrm{d}A}{\mathrm{d}t}A^{-1}\right)^2 - A^{-1}\frac{\mathrm{d}^2 A}{\mathrm{d}t^2}A^{-1}\ , \quad \text{where } A^{-1} = A^{-1}(t)$$

denotes the inverse of the matrix $A = A(t)$.

4. a) Show that Extension 1 to the theorem is an immediate corollary of the stability conditions for the fixed point of the family of contraction mappings studied in Sect. 9.7.

b) Let $\{A_t : X \to X\}$ be a family of contraction mappings of a complete normed space into itself depending on the parameter t, which ranges over a domain Ω in a normed space T. Show that if $A_t(x) = \varphi(t, x)$ is a function of class $C^{(n)}(\Omega \times X, X)$, then the fixed point $x(t)$ of the mapping A_t belongs to class $C^{(n)}(\Omega, X)$ as a function of t.

5. a) Using the implicit function theorem, prove the following *inverse function theorem.*

Let $g : G \to X$ be a mapping from a neighborhood G of a point y_0 in a complete normed space Y into a normed space X. If

1^0 the mapping $x = g(y)$ is differentiable in G,

2^0 $g'(y)$ is continuous at y_0,

3^0 $g'(y_0)$ is an invertible transformation,

then there exists a neighborhood $V \subset Y$ of y_0 and a neighborhood $U \subset X$ of x_0 such that $g : V \to U$ is bijective, and its inverse mapping $f : U \to V$ is continuous in U and differentiable at x_0; moreover,

$$f'(x_0) = \big(g'(y_0)\big)^{-1} .$$

b) Show that if it is known, in addition to the hypotheses given in a), that the mapping g belongs to the class $C^{(n)}(V, U)$, then the inverse mapping f belongs to $C^{(n)}(U, V)$.

c) Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a smooth mapping for which the matrix $f'(x)$ is nonsingular at every point $x \in \mathbb{R}^n$ and satisfies the inequality $\|(f')^{-1}(x)\| > C > 0$ with a constant C that is independent of x. Show that f is a bijective mapping.

d) Using your experience in solving c), try to give an estimate for the radius of a spherical neighborhood $U = B(x_0, r)$ centered at x_0 in which the mapping $f : U \to V$ studied in the inverse function theorem is necessarily defined.

6. a) Show that if the linear mappings $A \in \mathcal{L}(X; Y)$ and $B \in \mathcal{L}(X; \mathbb{R})$ are such that $\ker A \subset \ker B$ (here ker, as usual, denotes the kernel of a transformation), then there exists a linear mapping $\lambda \in \mathcal{L}(Y; \mathbb{R})$, such that $B = \lambda \cdot A$.

b) Let X and Y be normed spaces and $f : X \to \mathbb{R}$ and $g : X \to Y$ smooth functions on X with values in $\mathbb{R}$ and Y respectively. Let S be the smooth surface defined in X by the equation $g(x) = y_0$. Show that if $x_0 \in S$ is an extremum of the function $f\big|_S$, then any vector h tangent to S at X_0 simultaneously satisfies two conditions: $f'(x_0)h = 0$ and $g'(x_0)h = 0$.

c) Prove that if $x_0 \in S$ is an extremum of the function $f\big|_S$ then $f'(x_0) = \lambda \cdot g'(x_0)$, where $\lambda \in \mathcal{L}(Y;\mathbb{R})$.

d) Show how the classical *Lagrange necessary condition for an extremum with constraint* of a function on a smooth surface in $\mathbb{R}^n$ follows from the preceding result.

7. As is known, the equation $z^n + c_1 z^{n-1} + \cdots + c_n = 0$ with complex coefficients has in general n distinct complex roots. Show that the roots of the equation are smooth functions of the coefficients, at least where all the roots are distinct.