Finite-Dimensional Spaces

Algebra, Geometry, and Analysis
Volume I

By

Walter Noll

Department of Mathematics, Carnegie Mellon University, Pittsburgh, PA 15213 USA

This book was published originally by Martinus Nijhoff Publishers in 1987. This is a corrected reprint, posted in 2006 on my website math.cmu.edu/wn0g/noll.
Introduction

A. Audience. This treatise (consisting of the present Vol.I and of Vol.II, to be published) is primarily intended to be a textbook for a core course in mathematics at the advanced undergraduate or the beginning graduate level. The treatise should also be useful as a textbook for selected students in honors programs at the sophomore and junior level. Finally, it should be of use to theoretically inclined scientists and engineers who wish to gain a better understanding of those parts of mathematics that are most likely to help them gain insight into the conceptual foundations of the scientific discipline of their interest.

B. Prerequisites. Before studying this treatise, a student should be familiar with the material summarized in Chapters 0 and 1 of Vol.I. Three one-semester courses in serious mathematics should be sufficient to gain such familiarity. The first should be an introduction to contemporary mathematics and should cover sets, families, mappings, relations, number systems, and basic algebraic structures. The second should be an introduction to rigorous real analysis, dealing with real numbers and real sequences, and with limits, continuity, differentiation, and integration of real functions of one real variable. The third should be an introduction to linear algebra, with emphasis on concepts rather than on computational procedures.

C. Organization. There are ten chapters in Vol.I, numbered from 0 to 9. A chapter contains from 4 to 12 sections. The first digit of a section number indicates the chapter to which the section belongs; for example, Sect.611 is the 11th section of Chap.6. The appropriate section title and number are printed on the top of each odd-numbered page.

A descriptive name is used for each important theorem. Less important results are called Propositions and are enumerated in each section; for example, Prop.5 of Sect.83 refers to the 5th proposition of the 3rd section of Chap.8. Similar enumerations are used, if needed, for formal Definitions, Remarks, and Pitfalls. The term Pitfall is used for comments designed to prevent possible misconceptions.

At the end of most sections there are notes in small print. Their purpose is to relate the notations and terms used in the text to other notations and terms in the mathematical literature, and to comment on symbols, terms, and procedures that appear in print here for the first time (to the best of my knowledge).
A list of problems is presented at the end of each chapter. Some of the problems deal with examples that should help the students to better understand the concepts explained in the text. Some of the problems present additional easy results. A few problems present harder results and may require ingenuity to solve, although hints are often given.

Theorems, Propositions, and Definitions are printed in italic type. Important terms are printed in boldface type when they are defined, so that their definitions can be easily located. The end of a Proof, Remark, or Pitfall is indicated by a block: ⫷. Words enclosed in brackets indicate substitutes that can be used without invalidating the statement. For example, “The maximum [minimum] of the set $S$ is denoted by $\max S$[min $S$]” is a shorthand for two statements. The first is obtained by omitting the words in brackets and the second is “The minimum of the set $S$ is denoted by $\min S$”.

D. Style. I do not believe that it is a virtue to present merely the logical skeleton of an argument and to hide carefully all motivations from students by refusing to draw pictures and analogies. Unfortunately, it is much easier to write down formal definitions, theorems, and proofs than to describe motivations. I wish I had been able to do more of the latter. In this regard, a large contribution must be made by the instructor who uses this treatise as a textbook.

I have tried to be careful and honest with wording. For example, the phrases “it is evident” and “clearly” mean that a student who has understood the terms used in the statement should be able to agree immediately. The phrases “it is easily seen” and “an easy calculation shows” mean that it should take a good student no more than fifteen minutes with pencil and paper to verify the result. The phrase “it can be shown” is for information only and is not intended to imply anything about the difficulty of the proof.

A mathematical symbol should be used only in one of three ways. It should denote a specific object (e.g. $0, 1, N, \mathbb{R}$), or it should denote a definite but unspecified object (e.g. $n$ in “Let $n \in N$ be given”), or it should be a dummy (e.g. $n$ in “$2n$ is even for all $n \in N$”). I have tried to avoid any other uses. Thus, there should be no “dangling dummies” or expressions such as $\frac{dy}{dx}$, to which I cannot assign a reasonable, precise meaning, although they appear very often in the literature.

E. Genesis. About 25 years ago I started to write notes for a course for seniors and beginning graduate students at Carnegie Institute of Technology.
(renamed Carnegie-Mellon University in 1968). At first, the course was entitled “Tensor Analysis”. I soon realized that what usually passes for “Tensor Analysis” is really an undigested mishmash of linear and multilinear algebra, differential calculus in finite-dimensional spaces, manipulation of curvilinear coordinates, and differential geometry on manifolds, all treated with mindless formalisms and without real insight. As a result, I omitted the abstract differential geometry, which is too difficult to be treated properly at this level, and renamed the course “Multidimensional Algebra, Geometry, and Analysis”, and later “Finite-Dimensional Spaces”. The notes were rewritten several times. They were widely distributed and they served as the basis for appendices to the books Viscometric Flows of Non-Newtonian Fluids by B. D. Coleman, H. Markovitz, and W. Noll (Springer-Verlag 1966) and A First Course in Rational Continuum Mechanics by C. Truesdell (Academic Press 1977).

Since 1973 my notes have also been used by J. J. Schäffer and me in an undergraduate honors program entitled “Mathematical Studies”. One of the purposes of the program has been to present mathematics as an integrated whole and to avoid its traditional division into separate and seemingly unrelated courses. In this connection, Schäffer and I gradually developed a system of notation and terminology that we believe is useful for all branches of mathematics. My involvement in the Mathematical Studies Program has had a profound influence on my thinking; it has led to radical revisions of my notes and finally to this treatise.

Chapter 9 of Vol. I is an adaptation of notes entitled “On the Structure of Linear Transformations”, which were written for a course in “Modern Algebra”. (They were issued as Report 70–12 of the Department of Mathematics, Carnegie-Mellon University, in March 1970.)

F. Apologia. I wish to list certain features which make this treatise different from much, and in some cases most, of the existing literature.

Much of the substance of this treatise is covered in textbooks with titles such as “Linear Algebra”, “Analytic Geometry”, “Finite-Dimensional Vector Spaces”, “Modern Algebra”, “Vector and Tensor Analysis”, “Advanced Calculus”, “Functions of Several Variables”, or “Elementary Differential Geometry”. However, I believe this treatise to be the first that deals with finite-dimensional spaces in a unified way and that emphasizes the interplay between algebra, geometry, and analysis.
The approach of this treatise is conceptual, geometric, and uncompromisingly “coordinate-free”. In some of the literature, “tensors” are still defined in terms of coordinates and their transformations. To me, this is like looking at shadows dancing on the wall rather than at reality itself. Coordinates have no place in the definition of concepts. Of course, when it comes to dealing with specific problems, coordinates are sometimes useful. For this reason, I have included a chapter in which I show how to handle coordinates efficiently.

The space $\mathbb{R}^n$, with $n \in \mathbb{N}$, is very rarely mentioned in this treatise. It is misused nearly every time it appears in the literature, because it is only a special model for the structure that is appropriate in most situations, and as a special model $\mathbb{R}^n$ contains extraneous features that impede geometric insight. Thus any textbook on finite-dimensional calculus with a title like “Functions of Several Variables” must be defective. I consider it a travesty to call $\mathbb{R}^n$ “the Euclidean $n$-space”, as so many do. To quote N. D. Goodman: “ Obviously, this is not what Euclid meant” (in “Mathematics as an objective science”, Am. Math. Monthly, Vol. 86, p. 549, 1979).

In this treatise, I have tried to present every mathematical topic in a setting that fits the topic naturally and hence leads to a maximum of insight. For example, the structure of a flat (a.k.a. affine) space is the most natural setting for the differential and integral calculus. Most treatments use $\mathbb{R}^n$, a linear space, a normed linear space, or a Euclidean space as the setting. Each of them has extraneous structure which conceals the true nature of the calculus. On the other hand, the structure of a differentiable manifold is too impoverished to be a setting for many aspects of calculus.

In this treatise, a very careful distinction is made between a set and a family (see Sect.02). Almost all the literature is very sloppy on this point. I have found it liberating to resist the compulsion to think of finite sets always in enumerated form and thus to confuse them with lists. Also, I have found it very useful to be able to use a single symbol for a family and to use the same symbol, with an index, for the terms of the family. For example, I use $M_{ij}$ for the $(i,j)$-term of the matrix $M$. It seems nonsensical to me to change from an upper case letter $M$ to the lower case letter $m$ when changing from the matrix to its terms. A notation such as $(m_{ij})$ for a matrix, often seen in textbooks, is poison to me because it contains the dangling dummies $i$ and $j$ (dangling dummies are like cigarettes: both are poison, but I used
both when young.)

In this treatise, I have paid more careful attention than is usual to the specification of domains and codomains of mappings. In particular, the adjustment of domains and codomains is done explicitly (see Sect.03). Most authors are either ambiguous on this matter or talk around it clumsily.

The terminology and notation used in this treatise have been very carefully selected. They often do not conform with what a particular reader might consider “standard”, especially since what is “standard” for an expert in category theory may differ from what is “standard” for an expert in engineering mechanics. When the more common terminology or notation is obscure, clumsy, or leads to clashes, I have introduced new terminology or notation. Perhaps the most conspicuous example is the term “lineon” for the too cumbersome “linear transformation”. Some terms, such as the noun “tensor”, have been used with so many different meanings in the past that I found it wise to avoid them altogether.

I have avoided naming concepts and theorems after their purported inventors or discoverers. For one thing, there is often much doubt about who the originators were and frequently credit is given to the wrong people. Secondly, the use of descriptive names makes it easier to learn the material. I have introduced such descriptive names in almost all cases. The experienced mathematician will rarely have any difficulty in understanding my meaning. For example, I have yet to find a mathematician who could not tell immediately that my “Inner-Product Inequality” is what is commonly called the “Cauchy-Schwarz Inequality”. The notes at the end of each section list the names that have been used elsewhere for the concepts and theorems introduced.
Chapter 0

Basic Mathematics

In this chapter, we introduce the notation and terminology used throughout the book. Also, we give a brief explanation of the basic concepts of contemporary mathematics to the extent needed in this book. Finally, we give a summary of those topics of elementary algebra and analysis that are a prerequisite for the remainder of the book.

00 Notations

The equality sign $=$ is used to express the assertion that on either side of $=$ are symbolic names (possibly very complicated) for one and the same object. Thus, $a = b$ means that $a$ and $b$ are names for the same object; $a \neq b$ means that $a$ and $b$ are names for distinct objects. The symbol $:=$ is used to mean that the left side is defined by the right side, that the left side is an abbreviation of the right side, or that the right side is to be substituted for the left side. The symbol $=:$. has an analogous meaning.

The logical equivalence sign $\iff$ is used to indicate logical equivalence of statements. The symbol $:\iff$ is used to define a phrase or property; it may be read as “means by definition that” or “is equivalent by definition to”.

Given a set $S$ and a property $p$ that any given member of $S$ may or may not have, we use the shorthand

$$? \ x \in \ S, \ x \ has \ the \ property \ p$$

(00.1)

to describe the problem “Find all $x \in S$, if any, such that $x$ has the property $p$”. An element of $S$ having the property $p$ is called a solution of the problem. Often, the property $p$ involves an equality; then the problem is called an equation.
On occasion we use one and the same symbol for two different objects. Of course, this is permissible only if the two objects are related by some natural correspondence and if the context makes it clear which of the meanings of the symbol is appropriate in each instance. Suppose that \( S \) and \( T \) are two sets and that there is a natural one-to-one correspondence between them. We say that we identify \( S \) and \( T \) and we write \( S \cong T \) if we wish to use the same symbol for an element of \( S \) and the corresponding element of \( T \). Identification must be handled with great care to avoid notational clashes, i.e. instances in which the meaning of a symbol becomes truly ambiguous, even in context. On the other hand, total avoidance of identifications would lead to staggering notational complexity.

We consider 0 to be a real number that is both positive and negative. If we wish to exclude 0, we use the terms “strictly positive” and “strictly negative”. The following is a list of notations for specific number sets.

<table>
<thead>
<tr>
<th>Set</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbb{N} )</td>
<td>set of all natural numbers 0, 1, 2, \ldots, including 0.</td>
</tr>
<tr>
<td>( \mathbb{Z} )</td>
<td>set of all integers \ldots − 2, −1, 0, 1, 2 \ldots</td>
</tr>
<tr>
<td>( \mathbb{Q} )</td>
<td>set of all rational numbers.</td>
</tr>
<tr>
<td>( \mathbb{R} )</td>
<td>set of all real numbers.</td>
</tr>
<tr>
<td>( \mathbb{P} )</td>
<td>set of all positive real numbers, including 0.</td>
</tr>
<tr>
<td>( \mathbb{C} )</td>
<td>set of all complex numbers.</td>
</tr>
<tr>
<td>( n! )</td>
<td>given ( n \in \mathbb{N} ), the set of the first ( n ) natural numbers 0, 1, \ldots, ( n-1 ), starting with 0.</td>
</tr>
<tr>
<td>( n^\dag )</td>
<td>given ( n \in \mathbb{N} ), the set of the first ( n ) natural numbers 1, 2, \ldots, ( n ), starting with 1.</td>
</tr>
</tbody>
</table>

It is useful to read \( n^\dag \) as “\( n \) out” and \( n! \) as “\( n \) in”.

We use a superscript cross to indicate that 0 has been taken out of a set. For example, \( \mathbb{P}^{\times} \) denotes the set of all strictly positive numbers and \( \mathbb{N}^{\times} \) the set of all non-zero natural numbers. This cross notation may be used on any set that has a special “zero-element”, not necessarily the number zero.

Notes 00

(1) The “quantifier” \( \exists \) as in (00.1) was introduced by J. J. Schäffer in about 1973.

(2) Many people say “positive” when we say “strictly positive” and use the awkward “non-negative” when we say “positive”.

(3) The use of the special letters \( \mathbb{N}, \ldots, \mathbb{C} \) for the various number sets has become fairly standard in recent years. However, some people use \( \mathbb{N} \) for what we denote by \( \mathbb{N}^{\times} \),
the set of non-zero natural numbers; they do not consider zero to be a natural number. Sometimes \( \mathbb{P} \) is used for what we call \( \mathbb{P}_+ \), the set of strictly positive real numbers. The notation \( \mathbb{R}^+ \) for what we denote by \( \mathbb{P} \) is often used. Older textbooks often use boldface letters or script letters instead of the special letters now common.

(4) In most of the literature, \( S \cong T \) is used to indicate that \( S \) is isomorphic to \( T \). I prefer to use \( \cong \) only when the isomorphism is natural and used for identification.

(5) The notations \( n \uparrow \) and \( n \downarrow \) were invented by J. J. Schäffer in about 1973. I cannot understand any more how I ever got along without them. Some textbooks use the boldface \( n \) for what we call \( n \uparrow \). I do not consider the change from lightface to boldface a legitimate notation for a functorial process, quite apart from the fact that it is impossible to produce on a blackboard.

(6) The use of the superscript \( \times \) to indicate the removal of 0 was introduced by J. J. Schäffer and me in about 1973. It has turned out to be a very effective notation. It is consistent with the commonly found notation \( \mathbb{F}^\times \) for the multiplicative group of a field \( \mathbb{F} \) (see Sect.06).

01 Sets, Partitions

To specify a set \( S \), one must have a criterion for deciding whether any given object \( x \) belongs to \( S \). If it does, we write \( x \in S \) and say that \( x \) is a member or an element of \( S \), that \( S \) contains \( x \), or that \( x \) is in \( S \) or contained in \( S \). If \( x \) does not belong to \( S \) we write \( x \notin S \). We use abbreviations such as “\( x, y \in S \)” for “\( x \in S \) and \( y \in S \)”.

Let \( S \) and \( T \) be sets. If every member of \( S \) is also a member of \( T \) we write \( S \subseteq T \) or \( T \supseteq S \) and say that \( S \) is a subset of \( T \), that \( S \) is included in \( T \), or that \( T \) includes \( S \). We have

\[
S = T \iff (S \subseteq T \text{ and } T \subseteq S).
\]  

(01.1)

If \( S \subseteq T \) but \( S \neq T \) we write \( S \subsetneq T \) and say that \( S \) is a proper subset of \( T \) or that \( S \) is properly included in \( T \).

There is exactly one set having no members at all; it is denoted by \( \emptyset \) and called the empty set. The empty set is a subset of every set. A set having exactly one member is called a singleton; it is denoted by \( \{a\} \) if \( a \) denotes its only member. If the set \( S \) is known to be a singleton, we write \( a \in S \) to indicate that we wish to denote the only member of \( S \) by \( a \). A set having exactly two members is called a doubleton; it is denoted by \( \{a, b\} \) if \( a \) and \( b \) denote its two (distinct) members.

A set \( \mathcal{C} \) whose members are themselves sets is often called a collection of sets. The collection of all subsets of a given set \( S \) is denoted by \( \text{Sub} \, S \).
Hence, if $T$ is a set, then
\[
T \subset S \iff T \in \text{Sub} S. \tag{01.2}
\]

Many sets in mathematics are specified by naming an encompassing set $A$ and a property $p$ that any given member of $A$ may or may not have. The set $S$ of all members of $A$ that have this property $p$ is denoted by
\[
S := \{ x \in A \mid x \text{ has the property } p \}, \tag{01.3}
\]
which is read as “$S$ is the set of all $x$ in $A$ such that $x$ has the property $p$”. Occasionally, one has no encompassing set and $p$ is a property that any object may or may not have. In this case, (01.3) is replaced by
\[
S := \{ x \mid x \text{ has the property } p \}. \tag{01.4}
\]

**Remark:** Definitions of sets of the type (01.4) must be treated with caution. Indiscriminate use of (01.4) can lead to difficulties known as “paradoxes”.

Sometimes, a set with only few members is specified by an explicit listing of its members and by enclosing the list in braces $\{\}$. Thus, $\{a, b, c, d\}$ denotes the set whose members are $a, b, c$ and $d$.

Given any sets $S$ and $T$, one can form their **union** $S \cup T$, consisting of all objects that belong either to $S$ or to $T$ (or to both), and one can form their **intersection** $S \cap T$, consisting of all objects that belong to both $S$ and $T$. We say that $S$ and $T$ are **disjoint** if they have no elements in common; i.e. if $S \cap T = \emptyset$. The following rules (01.5)–(01.10) are valid for any sets $S$, $T$, $U$.

\[
S \cup S = S \cap S = S \cup \emptyset = S, \quad S \cap \emptyset = \emptyset, \tag{01.5}
\]
\[
T \subset S \iff T \cup S = S \iff T \cap S = T. \tag{01.6}
\]

The following rules remain valid if $\cap$ and $\cup$ are interchanged.

\[
S \cup T = T \cup S, \tag{01.7}
\]
\[
(S \cup T) \cup U = S \cup (T \cup U), \tag{01.8}
\]
\[
(S \cup T) \cap U = (S \cap U) \cup (T \cap U), \tag{01.9}
\]
\[
T \subset S \implies T \cup U \subset S \cup U. \tag{01.10}
\]

Given any sets $S$ and $T$, the set of all members of $S$ that do not belong to $T$ is called the **set-difference** of $S$ and $T$ and is denoted by $S \setminus T$, so that
\[
S \setminus T := \{ x \in S \mid x \notin T \}. \tag{01.11}
\]
01. SETS, PARTITIONS

We read \( S \setminus T \) as “\( S \) without \( T \)”. The following rules (01.12)–(01.16) are valid for any sets \( S, T, U \).

\[
\begin{align*}
S \setminus S &= \emptyset, \quad S \setminus \emptyset = S, \quad (01.12) \\
S \setminus (T \setminus U) &= (S \setminus T) \cup (S \cap U), \quad (01.13) \\
(S \setminus T) \setminus U &= (S \setminus T) \cap (S \setminus U). \quad (01.14)
\end{align*}
\]

The following rules remain valid if \( \cap \) and \( \cup \) are interchanged.

\[
\begin{align*}
(S \cup T) \setminus U &= (S \setminus U) \cup (T \setminus U), \quad (01.15) \\
S \setminus (T \cup U) &= (S \setminus T) \cap (S \setminus U). \quad (01.16)
\end{align*}
\]

If \( T \) is a subset of \( S \), then \( S \setminus T \) is also called the complement of \( T \) in \( S \). The complement \( S \setminus T \) is the largest (with respect to inclusion) among all the subsets of \( S \) that are disjoint from \( T \).

The union \( \bigcup \mathcal{C} \) of a collection \( \mathcal{C} \) of sets is defined to be the set of all objects that belong to at least one member-set of the collection \( \mathcal{C} \). For any sets \( S \) and \( T \) we have

\[
\bigcup \emptyset = \emptyset, \quad \bigcup \{S\} = S, \quad \bigcup \{S, T\} = S \cup T. \quad (01.17)
\]

If \( \mathcal{C} \) and \( \mathcal{D} \) are collections of sets then

\[
\bigcup (\mathcal{C} \cup \mathcal{D}) = (\bigcup \mathcal{C}) \cup (\bigcup \mathcal{D}). \quad (01.18)
\]

The intersection \( \bigcap \mathcal{C} \) of a non-empty collection \( \mathcal{C} \) of sets is defined to be the set of all objects that belong to each of the member-sets of the collection \( \mathcal{C} \). For any sets \( S \) and \( T \) we have

\[
\bigcap \{S\} = S, \quad \bigcap \{S, T\} = S \cap T. \quad (01.19)
\]

If \( \mathcal{C} \) and \( \mathcal{D} \) are non-empty collections, then

\[
\bigcap (\mathcal{C} \cup \mathcal{D}) = (\bigcap \mathcal{C}) \cap (\bigcap \mathcal{D}). \quad (01.20)
\]

We say that a collection \( \mathcal{C} \) of sets is disjoint if any two distinct member-sets of \( \mathcal{C} \) are disjoint. We say that \( \mathcal{C} \) covers a given set \( S \) if \( S \subset \bigcup \mathcal{C} \).

Let a set \( S \) be given. A disjoint collection \( \mathcal{P} \) of non-empty subsets of \( S \) that covers \( S \) is called a partition of \( S \). The member-sets of \( \mathcal{P} \) are called the pieces of the partition \( \mathcal{P} \). The empty collection is the only partition of the empty set. If \( S \) is any set then \( \{E \in \text{Sub } S \mid E \text{ is a singleton }\} \) is
CHAPTER 0. BASIC MATHEMATICS

a partition of $S$, called the **singleton-partition** of $S$. If $S \neq \emptyset$, then $\{S\}$ is also a partition of $S$, called the **trivial partition**. If $T$ is a non-empty proper subset of $S$, then $\{T, S \setminus T\}$ is a partition of $S$.

Let $S$ be a set and let $\sim$ be a relation on $S$, i.e. a fragment that becomes a statement $x \sim y$, true or not, when $x, y \in S$. We say that $\sim$ is an **equivalence relation** if for all $x, y, z \in S$ we have

(i) $x \sim x$ (reflexivity),

(ii) $x \sim y \implies y \sim x$ (symmetry), and

(iii) $(x \sim y$ and $y \sim z) \implies x \sim z$ (transitivity).

If $\sim$ is an equivalence relation on $S$, then

$$\mathcal{P} := \{P \in \text{Sub} S \mid P = \{x \in S \mid x \sim y\} \text{ for some } y \in S\}$$

is a partition of $S$; its pieces are called the **equivalence classes** of the relation $\sim$, and we have $x \sim y$ if and only if $x$ and $y$ belong to the same piece of $\mathcal{P}$. Conversely, if $\mathcal{P}$ is a partition of $S$, then

$$x \sim y : \iff \text{ (for some } P \in \mathcal{P}, x, y \in P)$$

defines an equivalence relation on $S$ whose equivalence classes are the pieces of $\mathcal{P}$.

Notes 01

(1) Some authors use $\subseteq$ when we use $\subset$, and $\subset$ when we use $\subseteq$. There is some confusion in the literature concerning the use of “contain” and “include”. We carefully observe the distinction.

(2) The term “null set” is often used for what we call the “empty set”. Also the phrase “$S$ is void” instead of “$S$ is empty” can often be found.

(3) The notation $\text{Sub} S$ is used here for the first time. The notations $\wp(S)$ and $2^S$ are common. The collection $\text{Sub} S$ is often called the “power set” of $S$.

(4) The notation $S - T$ instead of $S \setminus T$ is used by some people. It clashes with the member-wise difference notation (06.16). If $T$ is a subset of $S$, the notations $C_S T$ or $T^c$ are used by some people for the complement $S \setminus T$ of $T$ in $S$. 
02. Families, Lists, Matrices

A **family** $a$ is specified by a procedure by which one associates with each member $i$ of a given set $I$ an object $a_i$. The given set $I$ is called the **index set** of the family and the object $a_i$ is called the **term** of index $i$ or simply the $i$-**term** of the family $a$. If $a$ and $b$ are families with the same index set $I$ and if $a_i = b_i$ for all $i \in I$, then $a$ and $b$ are considered to be the same, i.e. $a = b$. The notation $(a_i \mid i \in I)$ is often used to denote a family, especially if no name is available a priori.

The set of all terms of a family $a$ is called the **range** of $a$ and is denoted by $\text{Rng} a$ or $\{a_i \mid i \in I\}$, so that

$$ \text{Rng} a = \text{Rng} (a_i \mid i \in I) = \{a_i \mid i \in I\}.$$  (02.1)

Many sets in mathematics are specified by naming a family $a$ and by letting the set be the range (02.1) of $a$. We say that a family $a = (a_i \mid i \in I)$ is **injective** if, for all $i, j \in I$,

$$a_i = a_j \implies i = j.$$  

Roughly, a family is injective if there is no repetition of terms.

The concept of a family may be viewed as a generalization of the concept of a set. With each set $S$ one can associate a family by letting the index set be $S$ itself and by letting the term corresponding to any given $x \in S$ be $x$ itself. Thus, the family corresponding to $S$ is $(x \mid x \in S)$. We identify this family with $S$ and refer to it as "$S$ self-indexed." In this manner, every assertion involving arbitrary families includes, as a special case, an assertion involving sets. The empty set $\emptyset$ is identified with the empty family, which is the only family whose index set is empty.

If all the terms of a given family $a$ belong to a given set $S$; i.e. if $\text{Rng} a \subset S$, we say that $a$ is a **family in** $S$. The set of all families in $S$ with a given index set $I$ is denoted by $S^I$ and called the **I-set-power** of $S$. If $T \subset S$, then $T^I \subset S^I$. We have $S^\emptyset = \{\emptyset\}$.

Let $n \in \mathbb{N}$ be given. A family whose index set is $n^I$ or $n^I$ is called a **list of length** $n$. If $n$ is small, a list $a$ indexed on $n^I$ can often be specified by a bookkeeping scheme of the form

$$(a_1, a_2, \ldots, a_n) := (a_i \mid i \in n^I).$$  (02.2)

where $a_1, a_2, \ldots, a_n$ should be replaced by specific names of objects, to be filled in an actual use. For each $i \in n^I$, we call $a_i$ the $i$'th term of $a$. The only list of length 0 is the empty family $\emptyset$. A list of length 1 is called a **singlet**, **
a list of length 2 is called a **pair**, a list of length 3 is called a **triple**, and a list of length 4 a **quadruple**.

If \( S \) is a set and \( n \in \mathbb{N} \), we use the abbreviation \( S^n := S \uparrow \) for the set of all lists of length \( n \) in \( S \). We call \( S^n \) the **\( n \)'th set-power** of \( S \).

Let \( S \) and \( T \) be sets. The set of all pairs whose first term is in \( S \) and whose second term is in \( T \) is called the **set-product** of \( S \) and \( T \) and is denoted by \( S \times T \); i.e.

\[
S \times T := \{(x, y) \mid x \in S, \ y \in T\}. \tag{02.3}
\]

We have \( S \times T = \emptyset \) if and only if \( S = \emptyset \) or \( T = \emptyset \). We have \( S \times S = S^{2|} = S^2 \); we call it the **set-square** of \( S \).

A family whose index set is the set-square \( I \times J \) of given sets \( I \) and \( J \) is called an **\( (I \times J) \)-matrix**. The \((i, j)\)-term of an \( (I \times J)\)-matrix \( M \) is usually written \( M_{i,j} \) instead of \( M((i,j)) \). For each \( i \in I \), the family \( (M_{i,j} \mid j \in J) \) is called the **\( i \)'-row** of \( M \), and for each \( j \in J \), the family \( (M_{i,j} \mid i \in I) \) is called the **\( j \)'-column** of \( M \). The \((J \times I)\)-matrix \( M^\top \) defined by \( (M^\top)_{j,i} := M_{i,j} \) for all \((j, i) \in J \times I \) is called the **transpose** of \( M \). If \( m, n \in \mathbb{N} \), an \((m| \times n|)\)-matrix \( M \) is called an **\( m \)-by-\( n \)-matrix**. Its rows are lists of length \( n \) and its columns are lists of length \( m \). If \( m \) and \( n \) are small, an \( m \)-by-\( n \)-matrix can often be specified by a bookkeeping scheme of the form

\[
\begin{bmatrix}
M_{1,1} & M_{1,2} & \cdots & M_{1,n} \\
M_{2,1} & M_{2,2} & \cdots & M_{2,n} \\
\vdots & \vdots & & \vdots \\
M_{m,1} & M_{m,2} & \cdots & M_{m,n}
\end{bmatrix} := M. \tag{02.4}
\]

A family \( M \) whose index set is the set-square \( I^2 \) of a given set \( I \) is called a **square matrix**. We then say that \( M \) is **symmetric** if \( M = M^\top \), i.e. if \( M_{i,j} = M_{j,i} \) for all \( i, j \in I \). The family \( (M_{i,i} \mid i \in I) \) is called the **diagonal** of \( M \). We say that a square matrix \( M \) in \( \mathbb{R} \) (or in any set containing a zero element \( 0 \)) is a **diagonal matrix** if \( M_{i,j} = 0 \) for all \( i, j \in I \) with \( i \neq j \).

Let a set \( S \) be given. For every \( U \in \text{Sub} S \) we define the **characteristic family** \( \text{ch}_{U \subset S} \) of \( U \) in \( S \) by

\[
(\text{ch}_{U \subset S})_x := \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{if } x \in S \setminus U \end{cases}. \tag{02.5}
\]

If the context makes it clear what \( S \) is, we abbreviate \( \text{ch}_{U \subset S} \) to \( \text{ch}_U \).

A family whose index set is \( \mathbb{N} \) or \( \mathbb{N}^\times \) is called a **sequence**.
03. MAPPINGS

Notes 02

(1) A family is sometimes called an “indexed set”. The trouble with this term is that an “indexed set” is not a set. The notation \( \{a_i\}_{i \in I} \) is used by some people for what we denote by \( \{a_i : i \in I\} \). Some people even use just \( \{a_i\} \), which is poison because of the dangling dummy \( i \) and because it also denotes a singleton with member \( a_i \).

(2) The terms “member” or “entry of a family” are often used instead of “term of a family”. The use of “member” can lead to confusion with “member of a set”.

(3) One often finds the barbarism “n-tuple” for what we call “list of length \( n \)”. The term “finite sequence” is sometimes used for what we call “list”. A list of numbers is very often called a “vector”. I prefer to use the term “vector” only when it has its original geometric meaning (see Def.1 of Sect.32).

(4) The terms “Cartesian product” and “direct product” are often used for what we call “set product”.

(5) In most of the literature, the use of the term “matrix” is confined to the case when the index set is of the form \( n \times m \) and when the terms are numbers of some kind. The generalization used here turns out to be very useful.

03 Mappings

In order to specify a mapping \( f \), one first has to prescribe two sets, say \( D \) and \( C \), and then some kind of prescription, called the assignment rule of \( f \), by which one can assign to every element \( x \in D \) an element \( f(x) \in C \). We call \( f(x) \) the value of \( f \) at \( x \). It is very important to distinguish very carefully between the mapping \( f \) and its values \( f(x) \), \( x \in D \). The set \( D \) of objects to which the prescription embodied in \( f \) can be applied is called the domain of the mapping \( f \) and is denoted by \( \text{Dom } f := D \). The set \( C \) to which the values of \( f \) must belong is called the codomain of \( f \) and is denoted by \( \text{Cod } f := C \). In order to put \( C \) and \( D \) into full view, we often write

\[
 f : D \to C \quad \text{or} \quad D \xrightarrow{f} C
\]

instead of just \( f \) and we say that \( f \) maps \( D \) to \( C \) or that \( f \) is a mapping from \( D \) to \( C \). The phrase “\( f \) is defined on \( D \)” expresses the assertion that \( D \) is the domain of \( f \). If \( f \) and \( g \) are mappings with \( \text{Dom } f = \text{Dom } g \), \( \text{Cod } f = \text{Cod } g \), and \( f(x) = g(x) \) for all \( x \in \text{Dom } f \), then \( f \) and \( g \) are considered to coincide, i.e. \( f = g \).

Terms such as “function”, “map”, “functional”, “transformation”, and “operator” are often used to mean the same thing as “mapping”. The term “function” is preferred when the codomain is the set of real or complex
numbers or a subset thereof. A still greater variety of names is used for mappings having special properties. Also, in some contexts, the value of \( f \) at \( x \) is not written \( f(x) \) but \( f x \), \( xf \), \( f x \) or \( xf \).

In order to specify a mapping \( f \) explicitly without introducing unnecessary symbols, it is often useful to employ the notation \((x \mapsto f(x)) : \text{Dom } f \rightarrow \text{Cod } f \) instead of just \( f \). (Note the use of \( \mapsto \) instead of \( \to \).) For example, \((x \mapsto \frac{1}{x}) : \mathbb{R}^\times \rightarrow \mathbb{R} \) denotes the function \( f \) with \( \text{Dom } f := \mathbb{R}^\times \), \( \text{Cod } f := \mathbb{R} \) and evaluation rule

\[
f(x) := \frac{1}{x} \quad \text{for all } x \in \mathbb{R}^\times .
\]

The graph of a mapping \( f : D \rightarrow C \) is the subset \( \text{Gr } (f) \) of the set-product \( D \times C \) defined by

\[
\text{Gr } (f) := \{ (x, y) \in D \times C \mid y = f(x) \} .
\] (03.1)

The mappings \( f \) and \( g \) coincide if and only if they have the same domain, codomain, and graph.

**Remark:** Very often, a mapping is specified by two sets \( D \) and \( C \) and a statement scheme \( F(x, y) \), which may be become valid or not, depending on what elements of \( D \) and \( C \) are substituted for \( x \) and \( y \), respectively. If, for every \( x \in D \), there is exactly one \( y \in C \) such that \( F(x, y) \) is valid, then \( F \) defines a mapping \( f : D \rightarrow C \), namely by the prescription that assigns to \( x \in D \) the unique \( y \in C \) that makes \( F(x, y) \) valid. Then

\[
\text{Gr}(f) = \{ (x, y) \mid F(x, y) \text{ is valid } \} .
\]

In some cases, given \( x \in D \), one can define or obtain \( f(x) \) by a formula, algorithm, or other procedure. Finding an efficient procedure to this end is often a difficult task.

With every mapping \( f \) we can associate the family \( \{ f(x) \mid x \in \text{Dom } f \} \) of its values. Roughly, the family is obtained from the mapping by forgetting the codomain. Conversely, with every family \( a := (a_i \mid i \in I) \) and every set \( C \) that includes \( \text{Rng } a \) we can associate the mapping \( (i \mapsto a_i) : I \rightarrow C \). Roughly, the mapping is obtained from the family by specifying the codomain \( C \). For example, if \( U \) is a subset of a given set \( S \), we can obtain from the characteristic family of \( U \) in \( S \), defined by (02.5), the **characteristic function** of \( U \) in \( S \), also denoted by \( \text{ch}_{U \subset S} \) or simply \( \text{ch}_U \), by specifying a codomain, usually \( \mathbb{R} \).
03. MAPPINGS

The range of a mapping \( f \) is defined to be the range of the family of its values; i.e., with the notation (02.1), we have

\[
\text{Rng } f := \{ f(x) \mid x \in \text{Dom } f \}. \tag{03.2}
\]

We say that \( f \) is \textbf{surjective} if its codomain and range coincide, i.e. if \( \text{Cod } f = \text{Rng } f \).

We say that a mapping \( f \) is \textbf{injective} if the family of its values is injective; i.e. if for all \( x, y \in D \)

\[
f(x) = f(y) \implies x = y.
\]

A mapping \( f : D \to C \) is said to be \textbf{invertible} if it is both injective and surjective. This is the case if and only if there is a mapping \( g : C \to D \) such that for all \( x \in D \) and \( y \in C \)

\[
y = f(x) \iff x = g(y).
\]

The mapping \( g \) is then uniquely determined by \( f \); it is called the \textbf{inverse} of \( f \) and is denoted by \( f^{-1} \). A given mapping \( f : D \to C \) is invertible if and only if, for each \( y \in C \), the problem

\[
? x \in D, \quad f(x) = y
\]

has exactly one solution. This solution is then given by \( f^{-1}(y) \). Invertible mappings are also called \textbf{bijections} and injective mappings \textbf{injections}.

Let \( U \) be a subset of a given set \( S \). The \textbf{inclusion mapping} of \( U \) into \( S \) is the mapping \( 1_{U \subseteq S} : U \to S \) defined by the rule

\[
1_{U \subseteq S}(x) := x \quad \text{for all } x \in U. \tag{03.3}
\]

The inclusion mapping of the set \( S \) into itself is called the \textbf{identity mapping} of \( S \) and is denoted by \( 1_S := 1_{S \subseteq S} \). Inclusion mappings are injective. An identity mapping is invertible and equal to its own inverse.

A \textbf{constant} is a mapping whose range is a singleton. We sometimes use the symbol \( c_{D \to C} \) to denote the constant with domain \( D \), codomain \( C \), and range \( \{ c \} \). In most cases, we can identify \( c_{D \to C} \) with \( c \) itself, thus using the same symbol for the constant and its only value.

If \( f \) and \( g \) are mappings such that \( \text{Dom } g = \text{Cod } f \), we define the \textbf{composite} \( g \circ f : \text{Dom } f \to \text{Cod } g \) of \( f \) and \( g \) by the evaluation rule

\[
(g \circ f)(x) := g(f(x)) \quad \text{for all } x \in \text{Dom } f. \tag{03.4}
\]
If \( f, g \) and \( h \) are mappings with \( \text{Dom} \, g = \text{Cod} \, f \) and \( \text{Cod} \, g = \text{Dom} \, h \), then
\[
(h \circ g) \circ f = h \circ (g \circ f).
\]  
(03.5)

Because of this rule, we may omit parentheses and write \( h \circ g \circ f \).

Let \( f \) be a mapping from a set \( D \) to itself. For every \( n \in \mathbb{N} \), the \( n \)’th iterate \( f^n : D \to D \) of \( f \) is defined recursively by \( f^0 = 1_D \), and \( f^{k+1} = f \circ f^k \) for all \( k \in \mathbb{N} \). We have \( f^2 = f \), \( f^3 = f \circ f \), etc. An element \( z \in D \) is called a fixed point of \( f \) if \( f(z) = z \). We say that given mappings \( f \) and \( g \), both from \( D \) to itself, commute if \( f \circ g = g \circ f \).

Let a mapping \( f : D \to C \) be given. A mapping \( g : C \to D \) is called a right-inverse of \( f \) if \( f \circ g = 1_C \), a left-inverse of \( f \) if \( g \circ f = 1_D \). If \( f \) has a right-inverse, it must be surjective; if \( f \) has a left-inverse, it must be injective. If \( g \) is a right-inverse of \( f \) and \( h \) a left-inverse of \( f \), then \( f \) is invertible and \( g = h = f^{-1} \). We always have
\[
f \circ 1_{\text{Dom} \, f} = 1_{\text{Cod} \, f} \circ f = f.\]  
(03.6)

Again, let a mapping \( f : D \to C \) be given. We define the image mapping \( f_\uparrow : \text{Sub} \, D \to \text{Sub} \, C \) of \( f \) by the evaluation rule
\[
f_\uparrow(U) := \{ f(x) \mid x \in U \} \quad \text{for all} \quad U \in \text{Sub} \, D
\]  
(03.7)

and the pre-image mapping \( f_\uparrow : \text{Sub} \, C \to \text{Sub} \, D \) of \( f \) by the rule
\[
f_\downarrow(V) := \{ x \in D \mid f(x) \in V \} \quad \text{for all} \quad V \in \text{Sub} \, C.
\]  
(03.8)

The mappings \( f_\uparrow \) and \( f_\downarrow \) satisfy the following rules for all subsets \( U \) and \( U' \) of \( D \) and all subsets \( V \) and \( V' \) of \( C \):
\[
U \subset U' \quad \implies \quad f_\uparrow(U) \subset f_\uparrow(U'), \quad (03.9)
\]
\[
V \subset V' \quad \implies \quad f_\downarrow(V) \subset f_\downarrow(V'), \quad (03.10)
\]
\[
U \subset f_\downarrow(f_\uparrow(U)), \quad f_\uparrow(f_\downarrow(V)) = V \cap \text{Rng} \, f, \quad (03.11)
\]
\[
f_\uparrow(U \cup U') = f_\downarrow(U) \cup f_\downarrow(U'), \quad f_\downarrow(U \cap U') \subset f_\downarrow(U) \cap f_\downarrow(U'), \quad (03.12)
\]
\[
f_\downarrow(V \cup V') = f_\downarrow(V) \cup f_\downarrow(V'), \quad f_\downarrow(V \cap V') = f_\downarrow(V) \cap f_\downarrow(V'), \quad (03.13)
\]
\[
f_\downarrow(C \setminus V) = D \setminus f_\downarrow(V). \quad (03.14)
\]

The inclusions \( \subset \) in (03.11) and (03.12) become equalities if \( f \) is injective. If \( f \) is injective, so is \( f_\uparrow \), and \( f_\downarrow \) is a left-inverse of \( f_\uparrow \). If \( f \) is surjective, so is \( f_\uparrow \), and \( f_\downarrow \) is a right-inverse of \( f_\uparrow \). If \( f \) is invertible, then \((f_\uparrow)^{-1} = f_\downarrow \).

If \( f \) and \( g \) are mappings such that \( \text{Dom} \, g = \text{Cod} \, f \), then
\[
(g \circ f)_\uparrow = g_\uparrow \circ f_\uparrow, \quad (g \circ f)_\downarrow = f_\downarrow \circ g_\downarrow.
\]  
(03.15)
03. MAPPINGS

\[ \text{Rng} \,(g \circ f) = g_{>}\text{(Rng } f). \]  \hfill (03.16)

If \( \text{ch}_{V \subset C} \) is the characteristic function of a subset \( V \) of a given set \( C \) and if \( f : D \to C \) is given, we have

\[ \text{ch}_{V \subset C} \circ f = \text{ch}_{f^{-1}(V) \subset D}. \]  \hfill (03.17)

Let a mapping \( f \) and sets \( A \) and \( B \) be given. We define the restriction \( f|_A \) of \( f \) to \( A \) by \( \text{Dom } f|_A := A \cap \text{Dom } f, \text{ Cod } f|_A := \text{Cod } f \), and the evaluation rule

\[ f|_A(x) := f(x) \quad \text{for all } \quad x \in A \cap \text{Dom } f. \]  \hfill (03.18)

We define the mapping

\[ f|_B^A : A \cap f_{<}(B \cap \text{Cod } f) \to B \]  \hfill (03.19)

by the rule

\[ f|_B^A(x) := f(x) \quad \text{for all } \quad x \in A \cap f_{<}(B \cap \text{Cod } f). \]  \hfill (03.20)

We say that \( f|_B^A \) is an adjustment of \( f \). We have \( f|_A = f|_{A^{\text{Cod } f}} \). We use the abbreviations

\[ f|_B := f|_{\text{Dom } f}, \quad f|_{\text{Rng } f} := f|_{\text{Dom } f}. \]  \hfill (03.21)

We have \( \text{Dom } (f|_B) = \text{Dom } f \) if and only if \( \text{Rng } f \subset B \). We note that

\[ f|_A|_B := (f|_A)|_B = (f|_B)|_A. \]  \hfill (03.22)

Let \( f \) be a mapping from a set \( D \) to itself. We say that a subset \( A \) of \( D \) is \( f \)-invariant if \( f_{>}(A) \subset A \). If this is the case, we define the \( A \)-adjustment \( f|_A : A \to A \) of \( f \) by

\[ f|_A := f|_A^A. \]  \hfill (03.23)

Let a set \( A \) and a collection \( \mathcal{C} \) of subsets of \( A \) be given such that \( A \in \mathcal{C} \). For every \( S \in \text{Sub } A \), the subcollection \( \{ U \in \mathcal{C} \mid S \subset U \} \) of \( \mathcal{C} \) then contains \( A \) and hence is not empty. We define \( \text{Sp} : \text{Sub } A \to \text{Sub } A \), the span-mapping corresponding to \( \mathcal{C} \), by the rule

\[ \text{Sp}(S) := \bigcap \{ U \in \mathcal{C} \mid S \subset U \} \quad \text{for all } \quad S \in \text{Sub } A. \]  \hfill (03.24)

The following rules hold for all \( S, T \in \text{Sub } A \):

\[ S \subset \text{Sp}(S), \]  \hfill (03.25)
\[ \text{Sp}(\text{Sp}(S)) = \text{Sp}(S), \quad (03.26) \]
\[ S \subset T \implies \text{Sp}(S) \subset \text{Sp}(T). \quad (03.27) \]

If the collection \( \mathcal{C} \) is intersection-stable i.e. if for every non-empty subcollection \( \mathcal{D} \) of \( \mathcal{C} \) we have \( \bigcap \mathcal{D} \in \mathcal{C} \), then, for all \( U \in \text{Sub} A \),
\[ \text{Sp}(U) = U \iff U \in \mathcal{C}, \quad (03.28) \]
and we have \( \text{Rng} \text{Sp} = \mathcal{C} \). Also, for every \( S \in \text{Sub} A \), \( \text{Sp}(S) \) is the smallest (with respect to inclusion) member of \( \mathcal{C} \) that includes \( S \).

Many of the definitions and rules of this section can be applied if one or more of the mappings involved are replaced by a family. For example, if \( a = (a_i \mid i \in I) \) is a family in a set \( S \) and if \( f \) is a mapping with \( \text{Dom} f = S \), then \( f \circ a \) denotes the family given by \( f \circ a := (f(a_i) \mid i \in I) \).

**Notes 03**

1. Some of the literature still confuses a mapping \( f \) with its value \( f(x) \), and one still finds phrases such as “Consider the function \( f(x) \)”, which contains the dangling dummy \( x \) and hence is poison.

2. There is some confusion in the literature concerning the term “image of a mapping \( f \)”. Some people use it for what we call “value of \( f \)” and others for what we call “range of \( f \)”. Occasionally, the term “range” is used for what we call “codomain”.

3. Many people do not consider the specification of a codomain as part of the specification of a mapping. If we did likewise, we would have no formal distinction between a family and a mapping. I believe it is very useful to have such a distinction.

4. The terms “onto” for “surjective” and “one-to-one” for “injective” are very often used. Also “one-to-one correspondence” is a common term for what we call “bijection” or “invertible mapping”. We sometimes use “one-to-one correspondence” informally.

5. The notation \( f^-\) for the inverse of the mapping was introduced in about 1973 by J. J. Schäffer and me. The notation \( f^{-1} \) is more common, but it clashes with notations for value-wise reciprocals. However, for linear mappings, it is useful to revert to the more traditional notation (see Sect.13).

6. The notation \( \text{id}_S \) is often used for the identity mapping \( 1_S \) of the set \( S \).

7. The notations \( 1_{U \in \mathcal{G}} \) for an inclusion mapping and \( c_{D \to C} \) for a constant were introduced by J. J. Schäffer and me in about 1973. Also, we started to write \( f^{\circ n} \) instead of the more common \( f^n \) for the \( n \)th iterate of \( f \) to avoid a clash with notations for value-wise powers.
(8) The notations \( f > \) and \( f < \) for the image mapping and pre-image mapping of a mapping \( f \) were introduced by J. J. Schäffer and me in about 1973. Many years before, I had introduced the notations \( f^* \) and \( f^* \), which were also introduced, independently, by S. MacLane and G. Birkhoff in their book “Algebra” (MacMillan, 1967). In most of the literature, the same symbol \( f \) is used for the image mapping as for the mapping \( f \) itself. The pre-image mapping of \( f \) is often denoted by \( f^{-1} \), the latter leading to confusion with the inverse (if there is one). A distinctive notation for image and pre-image mappings avoids a lot of confusion and leads to great economy when expressing relations (see, for example, (56.3)).

(9) The definition of the adjustment \( f|_A \) of a mapping \( f \), in the generality given here, is new. For the case when \( f>(A) \subset B \) it has been used by J. J. Schäffer and me since about 1973. I used the notation \( B|_A \) instead of \( f|_A \) in a paper published in 1971. Most people “talk around” such adjustments and do not use an explicit notation. The notation \( f|_A \) for the restriction is in fairly common use. The notation \( f|_A \) for \( f|_A \) was introduced recently by J. J. Schäffer.

04 Families of Sets; Families and Sets of Mappings

Let \((A_i \mid i \in I)\) be a family of sets, i.e. a family whose terms \( A_i \) are all sets. The range of this family is the collection \( \{A_i \mid i \in I\} \) of sets; it is non-empty if and only if \( I \) is non-empty. We define the union and intersection of the family to be, respectively, the union and intersection of the range. We use the notations

\[
\bigcup_{i \in I} A_i = \bigcup \{A_i \mid i \in I\},
\]

\[
\bigcap_{i \in I} A_i = \bigcap \{A_i \mid i \in I\} \quad \text{if} \quad I \neq \emptyset.
\]

Given any set \( S \), we have the following generalizations of (01.9) and (01.16):

\[
S \cap \bigcup_{i \in I} A_i = \bigcup_{i \in I} (S \cap A_i), \quad S \cup \bigcap_{i \in I} A_i = \bigcap_{i \in I} (S \cup A_i),
\]

\[
S \setminus \bigcup_{i \in I} A_i = \bigcap_{i \in I} (S \setminus A_i), \quad S \setminus \bigcap_{i \in I} A_i = \bigcup_{i \in I} (S \setminus A_i).
\]

Given any mapping \( f \) with \( \bigcup_{i \in I} A_i \subset \text{Dom} f \) we have the following generalizations of (03.12):

\[
f>(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f>(A_i), \quad f>(\bigcap_{i \in I} A_i) \subset \bigcap_{i \in I} f>(A_i),
\]
and the inclusion becomes an equality if \( f \) is injective. If \( f \) is a mapping
with \( \bigcup (A_i \mid i \in I) \subset \text{Cod} \ f \), we have the following generalizations of (03.13):

\[
f^<(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f^<(A_i), \quad f^<(\bigcap_{i \in I} A_i) = \bigcap_{i \in I} f^<(A_i). \tag{04.6}
\]

The **set-product** of the family \( (A_i \mid i \in I) \) of sets is the set of all families
\( a \) with index set \( I \) such that \( a_i \in A_i \) for all \( i \in I \). This set product is denoted by

\[
\prod_{i \in I} A_i = \prod (A_i \mid i \in I)
\]

It generalizes the set-product \( A_1 \times A_2 \) of sets \( A_1, A_2 \) as defined in Sect.02
because \( A_1 \times A_2 \) is the set-product of the pair \( (A_1, A_2) \) in the sense just
defined. If the terms of a family \( (A_i \mid i \in I) \) are all the same, i.e. if there is
a set \( S \) such that \( A_i = S \) for all \( i \in I \), then the set-product reduces to the
set-power defined in Sect.02, i.e. \( \prod (A_i \mid i \in I) = S^I \).

Let a family \( (A_i \mid i \in I) \) of sets and \( j \in I \) be given. Then the mapping

\[
(a \mapsto (a_j, a \mid_{i \neq j})) : \prod_{i \in I} A_i \to A_j \times \prod_{i \in I \setminus \{j\}} A_i \tag{04.7}
\]

is a natural bijection.

Let \( A_1 \) and \( A_2 \) be sets. We define the two **evaluations**

\[
ev_1 : A_1 \times A_2 \to A_1 \quad \text{and} \quad ev_2 : A_1 \times A_2 \to A_2
\]

by the rules

\[
\ev_1(a) := a_1, \quad \ev_2(a) := a_2 \quad \text{for all} \quad a \in A_1 \times A_2. \tag{04.8}
\]

We have

\[
a = (\ev_1(a), \ ev_2(a)) \quad \text{for all} \quad a \in A_1 \times A_2.
\]

More generally, given a family \( (A_i \mid i \in I) \) of sets, we can define, for each
\( j \in I \), the evaluation

\[
\ev_j : \prod_{i \in I} A_i \to A_j
\]

by the rule

\[
\ev_j(a) := a_j \quad \text{for all} \quad a \in \prod_{i \in I} A_i, \tag{04.9}
\]
and we have
\[ a = \left( ev_i(a) \mid i \in I \right) \quad \text{for all} \quad a \in \prod_{i \in I} A_i. \]

Given any sets \( S \) and \( T \), we denote the set of all mappings \( f \) with \( \text{Dom} f = S \) and \( \text{Cod} f = T \) by \( \text{Map}(S,T) \). The set of all injective mappings from \( S \) to \( T \) will be denoted by \( \text{Inj}(S,T) \). The set of all invertible mappings from a set \( S \) to itself is denoted by \( \text{Perm} S \) and its members are called permutations of \( S \).

Let \( S \) and \( T \) be sets. For each \( x \in S \), we define the evaluation at \( x \), \( ev_x : \text{Map}(S,T) \to T \), by the rule
\[ ev_x(f) := f(x) \quad \text{for all} \quad f \in \text{Map}(S,T). \] (04.10)

Assume now that a subset \( F \) of \( \text{Map}(S,T) \) is given. Then the restriction \( ev_x \, |_F \) is also called the evaluation at \( x \) and is simply denoted by \( ev_x \) if the context makes clear what \( F \) is. We define the mapping \( ev^F : S \to \text{Map}(F,T) \) by
\[ (ev^F(x))(f) := ev_x(f) = f(x) \quad \text{for all} \quad x \in S, f \in F. \] (04.11)

This mapping \( ev^F \), or a suitable adjustment of it, is called an evaluation mapping; it is simply denoted by \( ev \) if the context makes clear what \( F \) and the adjustments are.

There is a natural bijection from the set \( \text{Map}(S,T) \) of all mappings from \( S \) to \( T \) onto the set \( T^S \) of all families in \( T \) with index set \( S \), as described at the end of Sect.03. If \( U \) is a subset of \( T \) we have \( U^S \subset T^S \), but \( \text{Map}(S,U) \) is not a subset of \( \text{Map}(S,T) \). However, we have the natural injection
\[ (f \mapsto f^T) : \text{Map}(S,U) \to \text{Map}(S,T). \]

Let \( f \) and \( g \) be mappings having the same domain \( D \). The value-wise pair formation
\[ (x \mapsto (f(x),g(x))) : D \to \text{Cod} f \times \text{Cod} g \] (04.12)
will be identified with the pair \( (f,g) \), so that
\[ (f,g)(x) := (f(x),g(x)) \quad \text{for all} \quad x \in D. \] (04.13)

Thus, given any sets \( D, C_1, C_2 \), we obtain the identification
\[ \text{Map}(D,C_1) \times \text{Map}(D,C_2) \cong \text{Map}(D,C_1 \times C_2). \] (04.14)
More generally, given any family \((f_i \mid i \in I)\) of mappings, all having the same domain \(D\), we identify the **term-wise evaluation**

\[
(x \mapsto (f_i(x) \mid i \in I)) : D \to \prod_{i \in I} \text{Cod } f_i
\]  

with the family itself, so that

\[
(f_i \mid i \in I)(x) := (f_i(x) \mid i \in I) \quad \text{for all } x \in D.
\]

Thus, given any set \(D\) and any family \((C_i \mid i \in I)\) of sets, we obtain the identification

\[
\prod_{i \in I} \text{Map } (D, C_i) \cong \text{Map } (D, \prod_{i \in I} C_i).
\]

Let \(f\) and \(g\) be any mappings. We call the mapping

\[
f \times g : \text{Dom } f \times \text{Dom } g \to \text{Cod } f \times \text{Cod } g
\]

defined by

\[
(f \times g)(x, y) := (f(x), g(y)) \quad \text{for all } x \in \text{Dom } f, y \in \text{Dom } g
\]

the **cross-product** of \(f\) and \(g\).

More generally, given any family \((f_i \mid i \in I)\) of mappings, we call the mapping

\[
\prod_{i \in I} f_i = \prod_{i \in I} (f_i \mid i \in I) : \prod_{i \in I} \text{Dom } f_i \to \prod_{i \in I} \text{Cod } f_i
\]

defined by

\[
(\prod_{i \in I} f_i)(x) := (f_i(x) \mid i \in I) \quad \text{for all } x \in \prod_{i \in I} \text{Dom } f_i
\]

the **cross-product** of the family \((f_i \mid i \in I)\). If the terms of the family are all the same, i.e., if there is a mapping \(g\) such that \(f_i = g\) for all \(i \in I\), we write \(g^{\times I} := \prod_{i \in I} f_i \mid i \in I\) and call it the **I-cross-power** of \(g\). If no confusion can arise, we write simply \(g(c) := g^{\times I}(c) = (g(c_i) \mid i \in I)\) when \(c \in (\text{Dom } g)^I\).

Given any families \((D_i \mid i \in I)\) and \((C_i \mid i \in I)\) of sets, the mapping

\[
((f_i \mid i \in I) \mapsto \prod_{i \in I} f_i) : \prod_{i \in I} \text{Map } (D_i, C_i) \to \text{Map } (\prod_{i \in I} D_i, \prod_{i \in I} C_i)
\]
is a natural injection.

Let sets $S$ and $T$ be given. For every $a \in S$ and every $b \in T$ we define mappings $(a, \cdot) : T \to S \times T$ and $(\cdot, b) : S \to S \times T$ by the rules

$$
(a, \cdot)(y) := (a, y) \quad \text{and} \quad (\cdot, b)(x) := (x, b)
$$

for all $x \in S, y \in T$. A mapping $f$ whose domain $\text{Dom} \ f$ is a subset of a set product $S \times T$ is often called a “function of two variables”. Its value at $(x, y) \in S \times T$ is written simply $f(x, y)$ rather than $f((x, y))$. Given such a mapping and any $a \in S, b \in T$, we define

$$
(f(a, \cdot)) := f \circ (a, \cdot) |_{\text{Dom} \ f}, \quad f(\cdot, b) := f \circ (\cdot, b) |_{\text{Dom} \ f}.
$$

Hence we have

$$
f(\cdot, y)(x) = f(x, y) = f(x, \cdot)(y) \quad \text{for all } (x, y) \in \text{Dom} \ f.
$$

More generally, if $(A_i \mid i \in I)$ is a family of sets, we define, for each $j \in I$ and each $c \in \prod \ (A_i \mid i \in I \{j\})$, the mapping

$$(c, j) : A_j \to \prod_{i \in I} A_i$$

by the rule

$$
((c, j)(z))_i := \begin{cases} c_i & \text{if } i \in I \{j\} \\ z & \text{if } i = j \end{cases}.
$$

If $a \in \prod (A_i \mid i \in I)$, we abbreviate $(a, j) := (a \mid_{\{j\}}, j)$. If $I = \{1, 2\}$, we have $(a.1) = (a_2), (a.2) &= (a_1, \cdot)$. Given a mapping $f$ whose domain is a subset of $\prod (A_i \mid i \in I)$ and given $j \in I$ and $c \in \prod (A_i \mid i \in I \{j\})$ or $c \in \prod (A_i \mid i \in I)$, we define

$$
f(c, j) := f \circ (c, j) |_{\text{Dom} \ f}.
$$

We have

$$
f(x, j)(x_j) = f(x) \quad \text{for all } x \in \prod_{i \in I} A_i.
$$

Let $S, T$ and $C$ be sets. Given any mapping $f : S \times T \to C$, we identify the mapping

$$(x \mapsto f(x, \cdot)) : S \to \text{Map} (T, C)$$
with $f$ itself, so that
\[ (f(x, \cdot))(y) := f(x, y) \quad \text{for all } x \in S, y \in T. \quad (04.28) \]
Thus, we obtain the identification
\[ \text{Map}(S \times T, C) \cong \text{Map}(S, \text{Map}(T, C)). \quad (04.29) \]

Notes 04

1. The notation $\bigcup (A_i \mid i \in I)$ for the union of the family $(A_i \mid i \in I)$ is used when it occurs in the text rather than in a displayed formula. For typographical reasons, the notation on the left of (04.1) can be used only in displayed formulas. A similar remark applies to intersections, set-products, sums, products, and other operations on families (see (04.2) and (07.1)).

2. The term “projection” is often used for what we call “evaluation”, in particular when the domain is a set-product rather than a set of mappings.

3. The notation $\prod (A_i \mid i \in I)$ is sometimes used for the set-product $\bigotimes (A_i \mid i \in I)$, which is often called “Cartesian product” or “direct product” (see Note (4) to Sect.02).

4. The notations $(a, \cdot)$ and $(\cdot, b)$ as defined by (04.21) and the notation $(c, j)$ as defined by (04.24) were introduced by me a few years ago. The notations $f(\cdot, y)$ and $f(x, \cdot)$, using (04.23) as definitions rather than propositions, are fairly common in the literature.

5. Many people use the term “permutation” only if the domain is a finite set. I cannot see any advantage in such a limitation.

05 Finite Sets

The **cardinal** of a finite set $S$, i.e. the number of elements in $S$, will be denoted by $\# S \in \mathbb{N}$. Given $n \in \mathbb{N}$, the range $S := \{a_i \mid i \in n\}$ of an injective list $(a_i \mid i \in n)$ is a finite set with cardinal $\# S = n$. By an **enumeration** of a finite set $S$ we mean a list $(a_i \mid i \in n)$ of length $n := \# S$ whose range is $S$, or the corresponding mapping $(i \mapsto a_i) : n \to S$. There are $n!$ such enumerations. We have $\# (n!) = \# (n) = n$ for all $n \in \mathbb{N}$.

Let $S$ and $T$ be finite sets. Then $S \cap T$, $S \cup T$, $S \setminus T$, and $S \times T$ are again finite and their cardinals satisfy the rules
\[
\begin{align*}
\# (S \cup T) + \# (S \cap T) &= \# S + \# T, \quad (05.1) \\
\# (S \setminus T) &= \# S - \# (S \cap T) = \# (S \cup T) - \# T, \quad (05.2) \\
\# (S \times T) &= (\# S)(\# T). \quad (05.3)
\end{align*}
\]
We say that a family is finite if its index set is finite. If \((A_i \mid i \in I)\) is a finite family of finite sets, then \(\prod_{i \in I} (A_i)\) is finite and
\[
\# (\prod_{i \in I} (A_i)) = \prod_{i \in I} (\# A_i).
\] (05.4)

(See Sect.07 concerning the product \(\prod (n_i \mid i \in I)\) of a finite family \((n_i \mid i \in I)\) in \(\mathbb{N}\).)

Let \(\mathcal{P}\) be a partition of the given finite set \(S\). Then \(\mathcal{P}\) and the members of \(\mathcal{P}\) are all finite and
\[
\# S = \sum_{P \in \mathcal{P}} (\# P).
\] (05.5)

(See Sect.07 concerning the sum \(\sum (n_i \mid i \in I)\) of a finite family \((n_i \mid i \in I)\) in \(\mathbb{N}\).)

**Pigeonhole Principle:** Let \(f\) be a mapping with finite domain and codomain. If \(f\) is injective, then \(\# \text{Dom } f \leq \# \text{Cod } f\). If \(f\) is surjective, then \(\# \text{Dom } f \geq \# \text{Cod } f\). If \(\# \text{Dom } f = \# \text{Cod } f\) then the following are equivalent:

(i) \(f\) is injective;

(ii) \(f\) is surjective;

(iii) \(f\) is invertible.

If \(S\) and \(T\) are finite sets, then the set \(T^S\) of all families in \(T\) indexed on \(S\) and the set \(\text{Map}(S, T)\) of all mappings from \(S\) to \(T\) are finite and
\[
\# (T^S) = \# \text{Map}(S, T) = (\# T)^{\# S}.
\] (05.6)

If \(S\) is finite, so is \(\text{Sub } S\) and
\[
\# \text{Sub } S = 2^{(\# S)}.
\] (05.7)

Let \(S\) be any set. We denote the set of all finite subsets of \(S\) by \(\text{Fin } S\), and the set of all subsets of \(S\) having \(m\) elements by \(\text{Fin}_m S\), so that
\[
\text{Fin}_m S := \{U \in \text{Fin } S \mid \# U = m\}.
\] (05.8)

If \(S\) is finite, so is \(\text{Fin}_m S\). The notation
\[
\binom{n}{m} := \# \text{Fin}_m (n^\downarrow)
\]
(read “n choose m”) is customary, and we have

\[ \# \text{Fin}_m(S) = \binom{\# S}{m} \quad \text{for all } m \in \mathbb{N}. \] (05.9)

We have the following rules, valid for all \( n, m \in \mathbb{N} \):

\[ \binom{n}{m} = 0 \quad \text{if} \quad m > n, \quad \binom{n}{0} = \binom{n}{n} = 1, \] (05.10)

\[ \binom{n+m}{m} = \binom{n+m}{n}, \] (05.11)

\[ \binom{n+1}{m+1} = \binom{n}{m} + \binom{n}{m+1}, \] (05.12)

\[ \binom{m+n}{m} m! = \prod_{k \in [m]} (n + k) = \frac{(m+n)!}{n!}, \] (05.13)

\[ \sum_{k \in [n+1]} \binom{n}{k} = 2^n. \] (05.14)

If \( S \) is a finite set, then the set \( \text{Perm} S \) of all permutations of \( S \) is again finite and

\[ \# \text{Perm} S = (\# S)!. \] (05.15)

If \( S \) and \( T \) are finite sets, then the set \( \text{Inj} (S, T) \) of all injections from \( S \) to \( T \) is again finite and we have \( \text{Inj} (S, T) = \emptyset \) if \( \# S > \# T \) while

\[ \# \text{Inj} (S, T) = \frac{((\# T)!)}{(\# T - \# S)!} \quad \text{if} \quad \# S \leq \# T. \] (05.16)

Notes 05

(1) Many people use \( |S| \) for the cardinal \( \# S \) of a finite set.

(2) The notations \( \text{Fin} S \) and \( \text{Fin}_m S \) are used here for the first time. J. J. Schäffer and I have used \( \mathfrak{S}(S) \) and \( \mathfrak{S}_m(S) \), respectively, since about 1973.
A pre-monoid is a set $M$ endowed with structure by the prescription of a mapping $\text{cmb} : M \times M \to M$ called combination, which satisfies the associative law

$$\text{cmb}(\text{cmb}(a, b), c) = \text{cmb}(a, \text{cmb}(b, c)) \quad \text{for all} \quad a, b, c \in M. \quad (06.1)$$

A monoid $M$ is a pre-monoid, with combination $\text{cmb}$, endowed with additional structure by the prescription of a neutral $n \in M$ which satisfies the neutrality law

$$\text{cmb}(a, n) = \text{cmb}(n, a) = a \quad \text{for all} \quad a \in M. \quad (06.2)$$

A group $G$ is a monoid, with combination $\text{cmb}$ and neutral $n$, endowed with additional structure by the prescription of a mapping $\text{rev} : G \to G$, called reversion, which satisfies the reversion law

$$\text{cmb}(a, \text{rev}(a)) = \text{cmb}(\text{rev}(a), a) = n \quad \text{for all} \quad a \in G. \quad (06.3)$$

Let $H$ be a subset of a pre-monoid $M$. We say that $H$ is a sub-pre-monoid if it is stable under combination, i.e. if

$$\text{cmb}_{\lhd}(H \times H) \subset H. \quad (06.4)$$

If this is the case, then the adjustment $\text{cmb}_{\lhd}^{H}_{H \times H}$ endows $H$ with the natural structure of a pre-monoid.

Let $H$ be a subset of a monoid $M$. We say that $H$ is a submonoid of $M$ if it is a sub-pre-monoid of $M$ and if it contains the neutral $n$ of $M$. If this is the case, then the designation of $n$ as the neutral of $H$ endows $H$ with the natural structure of a monoid.

Let $H$ be a subset of a group $G$. We say that $H$ is a subgroup of $G$ if it is a submonoid of $G$ and if $H$ is reversion-invariant, i.e. if

$$\text{rev}_{\lhd}(H) \subset H.$$ If this is the case, then the adjustment $\text{rev}_{\lhd}^{H}$ (see (03.23)) endows $H$ with the natural structure of a group. The singleton $\{n\}$ is a subgroup of $G$, called the neutral-subgroup of $G$.

Let $M$ be a pre-monoid. Then $M$ contains at most one element $n$ which satisfies the neutrality law (06.2). If such an element exists, we say that $M$ is monoidable, because we can use $n$ to endow $M$ with the natural structure of a monoid.
Let \( M \) be a monoid. Then there exists at most one mapping \( \text{rev} : M \to M \) which satisfies the reversion law (06.3). If such a mapping exists, we say that \( M \) is \textit{groupable} because we can use \text{rev} to endow \( M \) with the natural structure of a group. We say that a pre-monoid \( M \) is groupable if it is monoidable and if the resulting monoid is groupable. If \( G \) is a group, then every groupable sub-pre-monoid of \( G \) is in fact a subgroup.

\textbf{Pitfall:} A monoidable sub-pre-monoid of a monoid need not be a submonoid. For example, the set of natural numbers \( \mathbb{N} \) with multiplication as combination and \( 1 \) as neutral is a monoid and the singleton \( \{0\} \) a sub-pre-monoid of \( \mathbb{N} \). Now, \( \{0\} \) is monoidable, in fact groupable, but it does not contain \( 1 \) and hence is not a submonoid of \( \mathbb{N} \).

Let \( E \) be a set. The set \( \text{Map}(E, E) \) of all mappings from \( E \) to itself becomes a monoid, if we define its combination by composition, i.e. by \( \text{cmb}(f, g) := f \circ g \) for all \( f, g \in \text{Map}(E, E) \), and if we designate the identity \( 1_E \) to be its neutral. We call this monoid the \textit{transformation-monoid} of \( E \). The set \( \text{Perm} E \) of all permutations of \( E \) is a groupable submonoid of \( \text{Map}(E, E) \). It becomes a group if we designate the reversion to be the inversion of mappings, i.e. if \( \text{rev}(f) := f^{-1} \) for all \( f \in \text{Perm} E \). We call this group \( \text{Perm} E \) the \textit{permutation-group} of \( E \).

Let \( G \) and \( G' \) be groups with combinations \( \text{cmb}, \text{cmb}' \), neutrals \( n, n' \) and reversions \( \text{rev}, \text{rev}' \), respectively. We say that \( \tau : G \to G' \) is a \textit{homomorphism} if it preserves combinations, i.e. if

\[
\text{cmb}'(\tau(a), \tau(b)) = \tau(\text{cmb}(a, b)) \quad \text{for all } a, b \in G.
\]  

(06.5)

It then also preserves neutrals and reversions, i.e.

\[
n' = \tau(n) \quad \text{and} \quad \text{rev}'(\tau(a)) = \tau(\text{rev}(a)) \quad \text{for all } a \in G.
\]

(06.6)

If \( \tau \) is invertible, then \( \tau^{-1} : G' \to G \) is again a homomorphism; we then say that \( \tau \) is a \textit{group-isomorphism}.

Let \( \tau : G \to G' \) be a homomorphism. Then the image \( \tau(G) \) under \( \tau \) of every subgroup \( H \) of \( G \) is a subgroup of \( G' \) and the pre-image \( \tau^{-1}(H') \) under \( \tau \) of every subgroup \( H' \) of \( G' \) is a subgroup of \( G \). The pre-image of the neutral-subgroup \( \{n'\} \) of \( G' \) is called the \textit{kernel} of \( \tau \) and is denoted by

\[
\text{Ker} \tau := \tau^{-1}(\{n'\}).
\]

(06.7)

The homomorphism \( \tau \) is injective if and only if its kernel is the neutral-subgroup of \( G \), i.e. \( \text{Ker} \tau = \{n\} \).
We say that a pre-monoid or monoid $M$ is cancellative if it satisfies the following cancellation laws for all $b, c \in M$:

\begin{align*}
(cmb(a, b) = cmb(a, c) & \text{ for some } a \in M) \implies b = c, \quad (06.8) \\
(cmb(b, a) = cmb(c, a) & \text{ for some } a \in M) \implies b = c. \quad (06.9)
\end{align*}

A group is always cancellative.

One often uses multiplicative terminology and notation when dealing with a pre-monoid, monoid, or group. This means that one uses the term "multiplication" for the combination, one writes $ab := cmb(a, b)$ and calls it the "product" of $a$ and $b$, one calls the neutral "unity" and denotes it by 1, and one writes $a^{-1} := rev(a)$ and calls it the "reciprocal" of $a$. The number sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{P}$, and $\mathbb{C}$ (see Sect.00) are all multiplicative monoids and $\mathbb{Q}^\times$, $\mathbb{R}^\times$, $\mathbb{P}^\times$, and $\mathbb{C}^\times$ are multiplicative groups. If $M$ is a multiplicative pre-monoid and if $S$ and $T$ are subsets of $M$, we write

$$ST := cmb_{>}(S \times T) = \{st \mid s \in S, t \in T\}. \quad (06.10)$$

and call it the \textbf{member-wise product} of $S$ and $T$.

If $t \in M$, we abbreviate $St := S\{t\}$ and $tS := \{t\}S$. If $G$ is a multiplicative group and $S$ a subset of $G$, we write

$$S^{-1} := rev_{>}(S) = \{s^{-1} \mid s \in S\}. \quad (06.11)$$

and call it the \textbf{member-wise reciprocal} of $S$.

A pre-monoid, monoid, or group $M$ is said to be \textbf{commutative} if

$$cmb(a, b) = cmb(b, a) \quad \text{for all } a, b \in M. \quad (06.12)$$

If this is the case, one often uses additive terminology and notation. This means that one uses the term "addition" for the combination, one writes $a + b := cmb(a, b)$ and calls it the "sum" of $a$ and $b$, one calls the neutral "zero" and denotes it by 0, and one writes $-a := rev(a)$ and calls it the "opposite" of $a$. The abbreviation $a - b := a + (-b)$ is customary. The number sets $\mathbb{N}$ and $\mathbb{P}$ are additive monoids while $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are additive groups. If $M$ is an additive pre-monoid and if $S$ and $T$ are subsets of $M$, we write

$$S + T := cmb_{>}(S \times T) = \{s + t \mid s \in S, t \in T\}. \quad (06.13)$$

and call it the \textbf{member-wise sum} of $S$ and $T$. If $t \in M$, we abbreviate

$$S + t := S + \{t\} = \{t\} + S =: t + S. \quad (06.14)$$
If $G$ is an additive group and if $S$ is a subset of $G$, we write

$$-S := \text{rev}_>(S) = \{ -s \mid s \in S \}. \quad (06.15)$$

and call it the \textbf{member-wise opposite} of $S$. If $S$ and $T$ are subsets of $M$, we write

$$T - S := \{ t - s \mid t \in T, s \in S \}. \quad (06.16)$$

and call it the \textbf{member-wise difference} of $T$ and $S$.

A ring is a set $R$ endowed with the structure of a commutative group, called the \textit{additive structure} of $R$ and described with additive terminology and notation, and the structure of a monoid, called the \textit{multiplicative structure} of $R$ and described with multiplicative terminology and notation, provided that the following \textit{distributive laws} hold

$$(a + b)c = ac + bc$$
$$c(a + b) = ca + cb$$

for all $a, b, c \in R$. \quad (06.17)

A ring then has a zero, denoted by 0, and a unity, denoted by 1. We have $0 \neq 1$ unless $R$ is a singleton. The following rules hold for all $a, b, c, \in R$:

$$(a - b)c = ac - bc, \quad c(a - b) = ca - cb,$$
$$(a - b)c = ac - bc, \quad c(a - b) = ca - cb,$$  
$$(a - b)c = ac - bc, \quad c(a - b) = ca - cb,$$
$$a0 = 0a = 0. \quad (06.20)$$

A ring $R$ is called a \textbf{commutative ring} if its multiplicative monoid-structure is commutative. A ring $R$ is called an \textbf{integral ring} if $R^\times$ is a submonoid of $R$ with respect to its multiplicative structure. This multiplicative monoid $R^\times$ is necessarily cancellative. A ring $F$ is called a \textbf{field} if it is commutative and integral and if the multiplicative submonoid $F^\times$ of $F$ is groupable. Endowed with its natural group structure, $F^\times$ is then called the \textbf{multiplicative group} of the field $F$.

A subset $S$ of a ring $R$ is called a \textbf{subring} if it is a subgroup of the additive group $R$ and a submonoid of the multiplicative monoid $R$. If this is the case, then $S$ has the natural structure of a ring. If a ring $R$ is commutative [integral], so is every subring of it. A subring of a field is called a subfield if its natural ring-structure is that of a field.

The set $\mathbb{C}$ of complex numbers is a field, $\mathbb{R}$ is a subfield of $\mathbb{C}$, and $\mathbb{Q}$ is a subfield of $\mathbb{R}$. The set $\mathbb{Z}$ is an (integral) subring of $\mathbb{Q}$. The set $\mathbb{P}$ is a submonoid of $\mathbb{R}$ both with respect to its additive structure and its multiplicative structure, but it is not a subring of $\mathbb{R}$. A similar statement applies to $\mathbb{N}$ and $\mathbb{Z}$ instead of $\mathbb{P}$ and $\mathbb{R}$. 
07. SUMMATIONS

Notes 06

(1) I am introducing the term “pre-monoid” here for the first time. In some textbooks, the term “semigroup” is used for it. However, “semigroup” is often used for what we call “monoid” and “monoid” is sometimes used for what we call “pre-monoid”.

(2) In much of the literature, no clear distinction is made between a group and a groupable pre-monoid or between a monoid and a monoidable pre-monoid. Some textbooks refer to an element satisfying the neutrality law (06.2) with the definite article before its uniqueness has been proved; such sloppiness should be avoided.

(3) In most textbooks, monoids and groups are introduced with multiplicative terminology and notation. I believe it is useful to at least start with an impartial terminology and notation. It is for this purpose that I introduce the terms “combination”, “neutral”, and “reversion”.

(4) Some people use the term “unit” or “identity” for what we call “unity”. This can lead to confusion because “unit” and “identity” also have other meanings.

(5) Some people use the term “inverse” for what we call “reciprocal” or the awkward “additive inverse” for what we call “opposite”. The use of “negative” for “opposite” and the reading of the minus-sign as “negative” are barbarisms that should be stamped out.

(6) The term “abelian” is often used instead of “commutative”. The latter term is more descriptive and hence preferable.

(7) Some people use the term “ring” for a slightly different structure: they assume that the multiplicative structure is not that of a monoid but only that of a pre-monoid. I call such a structure a pre-ring. Another author calls it “rng” (I assume as the result of a fit of humor).

(8) Most textbooks use the term “integral domain”, or even just “domain”, for what we call an “integral ring”. Often these terms are used only in the commutative case. I see good reason not to use the separate term “domain” for a ring, especially since “domain” has several other meanings.

07 Summations

Let $M$ be a commutative monoid, described with additive notation. Also, let a family $c = (c_i \mid i \in I)$ in $M$ be given. One can define, by recursion, for every finite subset $J$ of $I$, the sum

$$\sum_J c = \sum (c_i \mid i \in J) = \sum_{i \in J} c_i$$

(07.1)
CHAPTER 0. BASIC MATHEMATICS

of \( c \) over \( J \). This summation in \( M \) over finite index sets is characterized by the requirement that

\[
\sum_{\emptyset} c = 0 \quad \text{and} \quad \sum_J c = \sum_{J \setminus \{j\}} c. \tag{07.2}
\]

for all \( J \in \text{Fin}(I) \) and all \( j \in J \). If \( I \) itself is finite, we write \( \sum c := \sum_i c \). If the index set \( I \) is arbitrary, we then have \( \sum_J c = \sum c_J \) for all \( J \in \text{Fin}(I) \).

If \( K \) is a finite set and if \( \phi : K \to I \) is an injective mapping, we have

\[
\sum_K c \circ \phi = \sum_{k \in K} c_{\phi(k)} = \sum_{i \in \phi^{-1}(K)} c_i = \sum_{\phi^{-1}(K)} c. \tag{07.3}
\]

If \( J \) and \( K \) are disjoint finite subsets of \( I \), we have

\[
\sum_{J \cup K} c = \sum_J c + \sum_K c. \tag{07.4}
\]

More generally, if \( J \) is a finite subset of \( I \) and if \( \mathcal{P} \) is a partition of \( J \), then

\[
\sum_J c = \sum_{P \in \mathcal{P}} \left( \sum_P c \right). \tag{07.5}
\]

If \( J := \{j\} \) is a singleton or if \( J := \{j, k\} \) is a doubleton we have

\[
\sum_{\{j\}} c = c_j, \quad \sum_{\{j, k\}} c = c_j + c_k. \tag{07.6}
\]

If \( c, d \in M^I \), we have

\[
\sum_{i \in J} (c_i + d_i) = \left( \sum_{i \in J} c_i \right) + \left( \sum_{i \in J} d_i \right) \tag{07.7}
\]

for all \( J \in \text{Fin}\, I \). More generally, if \( m \in M^{I \times K} \) is a matrix in \( M \) we have

\[
\sum_{i \in J} \sum_{k \in L} m_{i,k} = \sum_{(i,k) \in J \times L} m_{i,k} = \sum_{k \in L} \sum_{i \in J} m_{i,k} \tag{07.8}
\]

for all \( J \in \text{Fin}\, I \) and all \( L \in \text{Fin}\, K \).

The support of a family \( c = (c_i \mid i \in I) \) in \( M \) is defined to be the set of all indices \( i \in I \) with \( c_i \neq 0 \) and is denoted by

\[
\text{Supt } c := \{ i \in I \mid c_i \neq 0 \}. \tag{07.9}
\]
We use the notation

\[ M^{(I)} := \{ c \in M^I \mid \text{Supp } c \text{ is finite} \} \]  

(07.10)

for the set of all families in \( M \) with index set \( I \) and finite support. For every \( c \in M^{(I)} \) we define

\[ \sum_J c := \sum_{(J \cap \text{Supp } c)} c \]  

(07.11)

for all \( J \in \text{Sub } I \). With this definition, the rules (07.1)–(07.5), (07.7), and (07.8) extend to the case when \( J \) and \( L \) are not necessarily finite, provided that \( c, d \) and \( m \) have finite support.

If \( n \in \mathbb{N} \) and \( a \in M \) then \((a \mid i \in n)\) is the constant list of length \( n \) and range \( \{a\} \) if \( n \neq 0 \), and the empty list if \( n = 0 \). We write

\[ na := \sum_{i \in n^1} (a \mid i \in n) = \sum_{i \in n^1} a. \]  

(07.12)

and call it the nth multiple of \( a \). If \((a \mid i \in I)\) is any constant finite family with range \( \{a\} \) and if \( I \neq \emptyset \), we have

\[ \sum_{i \in I} a = (\# I)a. \]  

(07.13)

If \((A_i \mid i \in I)\) is a finite family of subsets of \( M \) we write

\[ \sum_{i \in I} A_i := \{ \sum_J a \mid a \in \bigtimes_{i \in I} A_i \} \]  

(07.14)

and call it the member-wise sum of the family.

Let \( G \) be a commutative group and let \( c = (c_i \mid i \in I) \) be a family in \( G \). We then have

\[ \sum_{i \in J} (-c_i) = -\sum_{i \in J} c_i \]  

(07.15)

if \( J \in \text{Sub } I \) is finite or if \( c|_J \) has finite support.

If \( M \) is a commutative pre-monoid, described with additive notation, and if \( c = (c_i \mid i \in I) \) is a family in \( M \), one still can define the sum \( \sum_J c \) provided that \( J \) is a non-empty finite set. This summation is characterized by (07.6)\(_1 \) and (07.1)\(_2 \), restricted by the condition that \( J \setminus \{j\} \neq \emptyset \). All rules stated before remain valid for summations over non-empty finite index sets.

If \( M \) is a commutative monoid described with multiplicative rather than additive notation, the symbol \( \sum \) is replaced by \( \prod \) and

\[ \prod_J c = \prod_{i \in J} c_i = \prod_{c|_J} \]  

(07.16)
is called the **product** of the family $c = (c_i \mid i \in I)$ in $M$ over $J \in \text{Fin} I$. Of course, 0 must be replaced by 1 wherever it occurs. Also, if $n \in \mathbb{N}$ and $a \in M$, the nth multiple $na$ is replaced by the nth **power** $a^n$ of $a$.

Let $R$ be a commutative ring. If $c = (c_i \mid i \in I)$ is a family in $R$ and if $J \in \text{Fin} I$, it makes sense to form the product $\prod_J c$ as well as the sum $\sum_J c$ of $c$ over $J$. We list some generalized versions of the distributive law. If $c = (c_i \mid i \in I)$ and $d = (d_k \mid k \in K)$ are families in $R$ then

\[(\sum_J c)(\sum_L d) = \sum_{(i,k) \in J \times L} c_id_k \quad (07.17)\]

for all $J \in \text{Fin} I$ and $L \in \text{Fin} K$. If $m \in M^{I \times K}$ is a matrix in $R$ then

\[\prod_{i \in J}(\sum_{k \in L} m_{i,k}) = \sum_{s \in L^J}(\prod_{i \in J} m_{i,s_i}) \quad (07.18)\]

for all $J \in \text{Fin} I$ and $L \in \text{Fin} K$. If $I$ is finite and if $c,d \in R^I$, then

\[\prod_{i \in I}(c_i + d_i) = \sum_{J \in \text{Sub} I} (\prod_{j \in J} c_j)(\prod_{k \in I \setminus J} d_k). \quad (07.19)\]

If $a,b \in R$ and $n \in \mathbb{N}$, then

\[(a + b)^n = \sum_{k \in (n+1)!} \binom{n}{k} a^k b^{n-k}. \quad (07.20)\]

If $I$ is finite and $c \in R^I$, then

\[\prod_{I} c = (-1)^{|I|} \sum_{K \in \text{Sub} I} (-1)^{|K|} (\sum_{K} c)^{|I|}. \quad (07.21)\]

**Notes 07**

1. In much of the literature, summations are limited to those over sets of the form $\{i \in \mathbb{Z} \mid n \leq i \leq m\}$ with $n,m \in \mathbb{Z}$, $n \leq m$. The notation

\[\sum_{i=n}^{m} c_i := \sum_{i \in \mathbb{Z}} (c_i \mid i \in \mathbb{Z}, n \leq i \leq m)\]

is often used. Much flexibility is gained by allowing summations over arbitrary finite index sets.
08 Real Analysis

As mentioned in Sect.06, the set $\mathbb{R}$ of real numbers is a field and the set $\mathbb{P}$ of positive real numbers (which contains 0) is a submonoid of $\mathbb{R}$ both with respect to addition and multiplication. The set $\mathbb{P}^\times = \mathbb{P} \setminus \{0\}$ of strictly positive real numbers is a subgroup of $\mathbb{R}^\times = \mathbb{R} \setminus \{0\}$ with respect to multiplication and a sub-pre-monoid of $\mathbb{R}$ with respect to addition. The set of all negative [strictly negative] real numbers is $-\mathbb{P}$ [−$\mathbb{P}^\times$]. The collection \{−$\mathbb{P}^\times$, \{0\}, $\mathbb{P}^\times$\} is a partition of $\mathbb{R}$. The relation $<$ (read as “strictly less than”) on $\mathbb{R}$ is defined by

\[ x < y \iff y - x \in \mathbb{P}^\times \]  \hspace{3cm} (08.1)

for all $x, y \in \mathbb{R}$. For every $x, y \in \mathbb{R}$, exactly one of the three statements $x < y$, $y < x$, $x = y$ is valid. The relation $\leq$ (read as “less than”) on $\mathbb{R}$ is defined by

\[ x \leq y \iff y - x \in \mathbb{P}. \]  \hspace{3cm} (08.2)

We have $x \leq y \iff (x < y \text{ or } x = y)$ and

\[ (x \leq y \text{ and } y \leq x) \iff x = y. \]  \hspace{3cm} (08.3)

The following rules are valid for all $x, y, z \in \mathbb{R}$ and remain valid if $<$ is replaced by $\leq$:

\[ (x < y \text{ and } y < z) \implies x < z, \]  \hspace{3cm} (08.4)

\[ x < y \iff x + z < y + z, \]  \hspace{3cm} (08.5)

\[ x < y \iff -y < -x, \]  \hspace{3cm} (08.6)

\[ (x < y \iff ux < uy) \quad \text{for all } u \in \mathbb{P}^\times. \]  \hspace{3cm} (08.7)

The absolute value $|x| \in \mathbb{P}$ of a number $x \in \mathbb{R}$ is defined by

\[ |x| := \begin{cases} 
    x & \text{if } x \in \mathbb{P}^\times \\
    0 & \text{if } x = 0 \\
    -x & \text{if } x \in -\mathbb{P}^\times
\end{cases}. \]  \hspace{3cm} (08.8)

The following rules are valid for all $x, y \in \mathbb{R}$:

\[ |x| = 0 \iff x = 0, \]  \hspace{3cm} (08.9)

\[ |-x| = |x|, \]  \hspace{3cm} (08.10)

\[ ||x| - |y|| \leq |x + y| \leq |x| + |y|, \]  \hspace{3cm} (08.11)

\[ |xy| = |x||y|. \]  \hspace{3cm} (08.12)
The sign \( \text{sgn} \ x \in \{-1, 0, 1\} \) of a number \( x \in \mathbb{R} \) is defined by

\[
\text{sgn} \ x := \begin{cases} 
1 & \text{if } x \in \mathbb{P}^x \\
0 & \text{if } x = 0 \\
-1 & \text{if } x \in -\mathbb{P}^x 
\end{cases},
\]

(08.13)

We have, for all \( x, y \in \mathbb{R} \),

\[
\text{sgn} \ (xy) = (\text{sgn} \ x)(\text{sgn} \ y),
\]

(08.14)

\[
x = (\text{sgn} \ x)|x|.
\]

(08.15)

Let \( a, b \in \mathbb{R} \) be given. If \( a < b \), we use the following notations:

\[
]a, b[ := \{t \in \mathbb{R} \mid a < t < b\},
\]

(08.16)

\[
[a, b] := \{t \in \mathbb{R} \mid a \leq t \leq b\},
\]

(08.17)

\[
]a, b[ := \{t \in \mathbb{R} \mid a < t \leq b\},
\]

(08.18)

\[
[a, b[ := \{t \in \mathbb{R} \mid a \leq t < b\}.
\]

(08.19)

If \( a > b \), we sometimes write

\[
]a, b[ := b, a[, \ [a, b] := [b, a], \text{ etc.}
\]

(08.20)

A subset \( I \) of \( \mathbb{R} \) is called an interval if for all \( a, b \in I \) with \( a < b \) we have \([a, b] \subset I \). The empty set \( \emptyset \) and singleton subsets of \( \mathbb{R} \) are intervals. All other intervals are infinite sets and will be called genuine intervals. The whole of \( \mathbb{R} \) is a genuine interval. All other genuine intervals can be classified into eight types. The first four are described by (08.16)–(08.19). The term open interval is used if it is of the form (08.16), closed interval if it is of the form (08.17), and half-open interval if it is of the form (08.18) or (08.19). Intervals of the remaining four types have the form \( a + \mathbb{P} \), \( a + \mathbb{P}^x \), \( a - \mathbb{P} \), or \( a - \mathbb{P}^x \) for some \( a \in \mathbb{R} \); they are called half-lines.

Let \( S \) be a subset of \( \mathbb{R} \). A number \( a \in \mathbb{R} \) is called an upper bound of \( S \) if \( S \subset a - \mathbb{P} \) and a lower bound of \( S \) if \( S \subset a + \mathbb{P} \). There is at most one \( a \in S \) which is also an upper bound [lower bound] of \( S \). If it exists, it is called the maximum [minimum] of \( S \) and is denoted by \( \max S \) [\( \min S \)]. If \( S \) is non-empty and finite, then both \( \max S \) and \( \min S \) exist. We say that \( S \) is bounded above [bounded below] if the set of its upper bounds [lower bounds] is not empty. We say that \( S \) is bounded if it is both bounded above and below. This is the case if and only if there is a \( b \in \mathbb{P} \) such that \( |s| \leq b \) for all \( s \in S \). Every finite subset of \( \mathbb{R} \) is bounded. If \( S \) is bounded, then
every subset of \( S \) is bounded. The union of a finite collection of bounded subsets of \( \mathbb{R} \) is again bounded. A genuine interval is bounded if and only if it is of one of the types described by (08.16)–(08.19).

Let \( S \) be a non-empty subset of \( \mathbb{R} \). If \( S \) is bounded above [bounded below] then the set of all its upper bounds [lower bounds] has a minimum [maximum], which is called the \textbf{supremum} [\textbf{infimum}] of \( S \) and is denoted by \( \sup S \) [\( \inf S \)]. If \( S \) has a maximum [minimum] then \( \sup S = \max S \) [\( \inf S = \min S \)]. It is often useful to consider the extended-real-number set \( \overline{\mathbb{R}} := \mathbb{R} \cup \{ \infty, -\infty \} \) and then to express the assertion that \( S \) fails to be bounded above [bounded below] by writing \( \sup S = \infty \) [\( \inf S = -\infty \)]. In this sense \( \sup S \in \overline{\mathbb{R}} \) and \( \inf S \in \overline{\mathbb{R}} \) are well-defined if \( S \) is an arbitrary non-empty subset of \( \mathbb{R} \). We also use the notation \( \overline{P} := P \cup \{ \infty \} \). The relation \( < \) on \( \mathbb{R} \) is extended to \( \overline{\mathbb{R}} \) in such a way that \( -\infty < \infty \) and that \( t \in \overline{\mathbb{R}} \) satisfies \( -\infty < t < \infty \) if and only if \( t \in \mathbb{R} \).

We say that a family in \( \mathbb{R} \), or a function whose codomain is included in \( \mathbb{R} \), is \textbf{bounded} [\textbf{bounded above}, \textbf{bounded below}] if its range is bounded [bounded above, bounded below]. Let \( I \) be a subset of \( \mathbb{R} \). We say that the family \( a \in \mathbb{R}^I \) is \textbf{isotone} [\textbf{antitone}] if

\[
i \leq j \implies a_i \leq a_j \quad [a_j \leq a_i] \quad \text{for all} \quad i, j \in I. \tag{08.21}\]

We say that the family is \textbf{strictly isotone} [\textbf{strictly antitone}] if (08.21) still holds after \( \leq \) has been replaced by \( < \). This terminology applies, in particular, to lists and sequences (see Sect. 02). We say that a mapping \( f \) with \( \text{Dom} f, \text{Cod} f \in \text{Sub} \mathbb{R} \) is \textit{isotone}, \textit{antitone}, \textit{strictly isotone}, or \textit{strictly antitone} if the family \( (f(t) \mid t \in \text{Dom} f) \) of its values has the corresponding property.

Let \( a \) be a sequence in \( \mathbb{R} \), i.e. \( a \in \mathbb{R}^{\mathbb{N}} \) or \( a \in \mathbb{R}^{\mathbb{N}} \). We say that \( a \) \textbf{converges} to \( t \in \mathbb{R} \) if for every \( \varepsilon \in \mathbb{P} \) there is \( n \in \mathbb{N}^\times \) such that \( a_{n+1} \cup |t - \varepsilon, t + \varepsilon| \). A sequence in \( \mathbb{R} \) can converge to at most one \( t \in \mathbb{R} \). If it does, we call \( t \) the \textbf{limit} of \( a \) and write \( t = \lim a \) to express the assertion that \( a \) converges to \( t \). Every isotone [antitone] sequence that is bounded above [below] converges.

Let \( a \) be any sequence. We say that the sequence \( b \) is a \textbf{subsequence} of \( a \) if \( b = a \circ \sigma \) for some strictly isotone mapping \( \sigma \) whose domain, and whose codomain, is \( \mathbb{N} \) or \( \mathbb{N}^\times \), as appropriate. If \( a \) is a sequence in \( \mathbb{R} \) that converges to \( t \in \mathbb{R} \), then every subsequence of \( a \) also converges to \( t \).

Let \( a \) be a sequence in \( \mathbb{R} \). We say that \( t \in \mathbb{R} \) is a \textbf{cluster point} of \( a \), or that \( a \) \textbf{clusters} at \( t \), if for every \( \varepsilon \in \mathbb{P} \) and every \( n \in \mathbb{N}^\times \), we have \( a_{n+1} \cap |t - \varepsilon, t + \varepsilon| \neq \emptyset \). The sequence \( a \) clusters at \( t \) if and only if some subsequence of \( a \) converges to \( t \).
Cluster Point Theorem: Every bounded sequence in \( \mathbb{R} \) has at least one cluster point.

Let a sequence \( a \in \mathbb{R}^\mathbb{N} \) be given. The sum-sequence \( \text{ssq} \ a \in \mathbb{R}^\mathbb{N} \) of \( a \) is defined by

\[
(\text{ssq} \ a)_n := \sum_{k \in n^1} a_k \quad \text{for all} \quad n \in \mathbb{N}.
\]  

(08.22)

Let \( S \) be a subset of \( \mathbb{R} \). Given \( f, g \in \text{Map}(S, \mathbb{R}) \), we define the value-wise sum \( f + g \in \text{Map}(S, \mathbb{R}) \) and the value-wise product \( fg \in \text{Map}(S, \mathbb{R}) \) by

\[
(f + g)(t) := f(t) + g(t), \quad (fg)(t) := f(t)g(t) \quad \text{for all} \quad t \in S
\]  

(08.23)

The value-wise quotient \( \frac{f}{g} \in \text{Map}(g^<(\mathbb{R}^\times), \mathbb{R}) \) is defined by

\[
\left( \frac{f}{g} \right)(t) := \frac{f(t)}{g(t)} \quad \text{for all} \quad t \in g^<(\mathbb{R}^\times) = S \setminus g^<(\{0\}).
\]  

(08.24)

We write \( f \leq g \) \([f < g]\) to express the assertion that \( f(t) \leq g(t) \) \([f(t) < g(t)]\) for all \( t \in S \). Given \( f \in \text{Map}(S, \mathbb{R}) \) and \( n \in \mathbb{N} \), we define the value-wise opposite \(-f\), the value-wise absolute value \(|f|\) and the value-wise \( n \)'th power \( f^n \) of \( f \) by

\[
(-f)(t) := -f(t), \quad |f|(t) := |f(t)|, \quad f^n(t) = (f(t))^n \quad \text{for all} \quad t \in S.
\]  

(08.25)

If \( n \in -\mathbb{N}^\times \), we define \( f^n \in \text{Map}(f^<(\mathbb{R}^\times), \mathbb{R}) \) by \( f^n := \frac{1}{f^n} \).

The identity mapping of \( \mathbb{R} \) will be abbreviated by

\[
\iota := 1_{\mathbb{R}}.
\]  

(08.26)

The symbol \( \iota \) will also be used for various adjustments of \( 1_{\mathbb{R}} \) if the context makes it clear which adjustments are appropriate.

For every \( t \in \mathbb{R} \), the sum-sequence \( \text{ssq}(\frac{t^k}{k!} \mid k \in \mathbb{N}) \) converges; its limit is called the exponential of \( t \) and is denoted by \( e^t \). The exponential function \( \exp : \mathbb{R} \to \mathbb{R} \) can be defined by the rule

\[
\exp(t) := e^t \quad \text{for all} \quad t \in \mathbb{R}.
\]  

(08.27)

Let \( f \) be a function with \( \text{Dom} f, \text{Cod} f \in \text{Sub} \mathbb{R} \). Let \( t \in \mathbb{R} \) be an accumulation point of \( \text{Dom} f \), which means that

\[
(\{t\} - \sigma, \{t\} + \sigma \setminus \{t\}) \cap \text{Dom} f \neq \emptyset \quad \text{for all} \quad \sigma \in \mathbb{R}^\times.
\]  

(08.28)
We say that $\lambda \in \mathbb{R}$ is a **limit** of $f$ at $t$ if for every $\varepsilon \in \mathbb{P}^\times$, there is $\sigma \in \mathbb{P}^\times$ such that

$$|f|_{t-\sigma,t+\sigma\setminus\{t\}} - \lambda | < \varepsilon. \quad (08.29)$$

The function $f$ can have at most one limit at $t$. If it has, we write $\lim_{t} f = \lambda$ to express the assertion that $f$ has the limit $\lambda$ at $t$. We say that the function $f$ is **continuous at** $t$ if $t \in \text{Dom } f$ and $\lim_{t} f = f(t)$.

Let $I$ be a genuine interval. We say that $f : I \to \mathbb{R}$ is **continuous** if it is continuous at every $t \in I$. Let $t \in I$ be given. We say that $f$ is **differentiable at** $t$ if the limit

$$\partial_t f := \lim_{t} \left( \frac{f - f(t)}{t - t} \right) \quad (08.30)$$

exists. If it does, then $\partial_t f$ is called the **derivative** of $f$ at $t$. We say that $f$ is **differentiable** if it is differentiable at every $t \in I$. If this is the case, we define the derivative (-function) $\partial f : I \to \mathbb{R}$ of $f$ by

$$(\partial f)(t) := \partial_t f \quad \text{for all } t \in I. \quad (08.31)$$

If $f$ and $g$ are differentiable, so are $f + g$ and $fg$, and we have

$$\partial(f + g) = \partial f + \partial g, \quad \partial(fg) = (\partial f)g + f(\partial g). \quad (08.32)$$

Let $n \in \mathbb{N}$ be given. We say that $f : I \to \mathbb{R}$ is $n$ times **differentiable** if $\partial^n f : I \to \mathbb{R}$ can be defined by the recursion

$$\partial^0 f := f, \quad \partial^{k+1} f := \partial(\partial^k f) \quad \text{for all } k \in n^\downarrow. \quad (08.33)$$

We say that $f$ is of **class** $C^n$ if it is $n$ times differentiable and if $\partial^n f$ is continuous. We say that $f$ is of **class** $C^\infty$ if it is $n$ times differentiable for all $n \in \mathbb{N}$. We also use the notations

$$f^{(k)} := \partial^k f \quad \text{for all } k \in n^\downarrow \quad (08.34)$$

and

$$f^* := \partial f, \quad f^{**} := \partial^2 f, \quad f^{***} := \partial^3 f. \quad (08.35)$$

For each $n \in \mathbb{Z}$, the function $\nu^n$ is of class $C^\infty$ and we have

$$\partial^k \nu^n = \begin{cases} \prod_{j \in k^\downarrow} (n-j) \nu^{n-k} & \text{if } n \not\in k^\downarrow \\ 0 & \text{if } n \in k^\downarrow \end{cases} \quad \text{for all } k \in \mathbb{N}. \quad (08.36)$$
The exponential function defined by (08.27) is of class $C^\infty$ and we have $\partial^n \exp = \exp$ for all $n \in \mathbb{N}$.

**Difference-Quotient Theorem:** Let a genuine interval $I$ and $a, b \in I$ with $a < b$ and a function $f : I \to \mathbb{R}$ be given. If $f|_{[a,b]}$ is continuous and $f|_{[a,b]}$ differentiable, then

$$\frac{f(b) - f(a)}{b - a} \in \{\partial_t f \mid t \in [a,b]\}. \quad (08.37)$$

**Corollary:** If $I$ is a genuine interval and if $f : I \to \mathbb{R}$ is differentiable with $\partial f = 0$, then $f$ must be a constant.

Let $f$ be a function with $\text{Cod} f \subset \mathbb{R}$. We say that $f$ attains a maximum [minimum] (at $x \in \text{Dom} f$) if $\text{Rng} f$ has a maximum [minimum] (and if $f(x) = \max \text{Rng} f$ [min $\text{Rng} f$]). The term extremum is used for “maximum or minimum”.

**Extremum Theorem:** Let $I$ be an open interval. If $f : I \to \mathbb{R}$ attains an extremum at $t \in I$ and if $f$ is differentiable at $t$ then $\partial_t f = 0$.

**Theorem on Attainment of Extrema:** If $I$ is a closed and bounded interval and if $f : I \to \mathbb{R}$ is continuous, then $f$ attains a maximum and minimum.

Let $I$ be a genuine interval. Let $f : I \to \mathbb{R}$ be a continuous function. One can define, for every $a, b \in I$ the integral $\int_a^b f \in \mathbb{R}$. This integration is characterized by the requirement that

$$\int_a^b f = \int_a^c f + \int_c^b f \quad \text{for all} \quad a, b, c \in I \quad (08.38)$$

and

$$\min f>([a,b]) \leq \frac{1}{b - a} \int_a^b f \leq \max f>([a,b]) \quad \text{if} \quad a < b. \quad (08.39)$$

We have

$$\left| \int_a^b f \right| \leq \left| \int_a^b |f| \right| \quad \text{for all} \quad a, b \in I. \quad (08.40)$$

If $\lambda \in \mathbb{R}$, if $f, g \in \text{Map} (I, \mathbb{R})$ are both continuous, and if $a, b \in I$, then

$$\lambda \int_a^b f = \int_a^b (\lambda f), \quad \int_a^b (f + g) = \int_a^b f + \int_a^b g. \quad (08.41)$$

If $f, g \in \text{Map} (I, \mathbb{R})$ are both continuous and if $a \leq b$, we have

$$f \leq g \implies \int_a^b f \leq \int_a^b g. \quad (08.42)$$
Fundamental Theorem of Calculus: Let $I$ be a genuine interval, let $a \in I$, and let $f : I \to \mathbb{R}$ be a continuous function. Then the function $g : I \to \mathbb{R}$ defined by

$$g(t) := \int_a^t f \quad \text{for all} \quad t \in I$$

(08.43)

is differentiable and $\partial g = f$.

Let $f : I \to \mathbb{R}$ be a continuous function whose domain $I$ is a genuine interval and let $g : I \to \mathbb{R}$ be an antiderivative of $f$, i.e. a differentiable function such that $\partial g = f$. For every $a, b \in I$, we then have

$$\int_a^b f = g(b) - g(a).$$

(08.44)

If no dummy-free symbol for the function $f$ is available, it is customary to write

$$\int_a^b f(s)ds := \int_a^b f.$$  

(08.45)

Notes 08

(1) Many people use parentheses in notations such as (08.13) - (08.16) and write for example, $(a, b]$ for what we denote by $[a, b]$. This usage should be avoided because it leads to confusion between the pair $(a, b)$ and an interval.

(2) Many people say “less than” when we say “strictly less than” and use the awkward “less than or equal to” when we say “less than”.

(3) There are some disparities in the literature about the usage of “interval”. Often the term is restricted to genuine intervals or even to bounded genuine intervals.

(4) Some people use the term “limit point” for “cluster point” and sometimes also for “accumulation point”. This usage should be avoided because it can too easily be confused with “limit”. There is no agreement in the literature about the terms “cluster point” and “accumulation point”. Sometimes “cluster point” is used for what we call “accumulation point” and sometimes “accumulation point” for what we call “cluster point”.

(5) In the literature, the term “infinite series” is often used in connection with subsequences. There is no agreement on what precisely a “series” should be. Some textbooks contain a “definition” that makes no sense. We avoid the problem by not using the term at all.

(6) The Cluster Point Theorem, or a variation thereof, is very often called the “Bolzano-Weierstrass Theorem”. 
(7) Most of the literature on real analysis suffers from the lack of a one-symbol notation for $1_\mathbb{R}$, as has been noted by several authors in the past. I introduced the use of $\iota$ for $1_\mathbb{R}$ in 1971 (class notes). Some of my colleagues and I cannot understand any more how we could ever have lived without it (as without a microwave oven).

(8) I introduced the notation $\partial_t f$ for the derivative of $f$ at $t$ a few years ago. Most of the more traditional notations, such as $df(t)/dt$, turn out, upon careful examination, to make no sense. The notations $f'$ and $f$ for the derivative-function of $f$ are very common. The former clashes with the use of the prime as a mere distinction symbol (as in Def.1 of Sect.13), the latter makes it difficult to write derivatives of functions with compound symbols, such as $(\exp \circ f + \iota(\sin \circ h))^\prime$.

(9) The Difference-Quotient Theorem is most often called the "Mean-Value Theorem of Differential Calculus". I believe that "Difference-Quotient" requires no explanation while "Mean-Value" does.
Chapter 1

Linear Spaces

This chapter is a brief survey of basic linear algebra. It is assumed that the reader is already familiar with this subject, if not with the exact terminology and notation used here. Many elementary proofs are omitted, but the experienced reader will have no difficulty supplying these proofs for himself or herself.

In this chapter the letter $F$ denotes either the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers. (Actually, $F$ could be an arbitrary field. To preserve the validity of certain remarks, $F$ should be infinite.)

11 Basic Definitions

**Definition 1:** A linear space (over $F$) is a set $V$ endowed with structure by the prescription of

(i) an operation $\text{add}: V \times V \to V$, called the addition in $V$,

(ii) an operation $\text{sm}: F \times V \to V$, called the scalar multiplication in $V$,

(iii) an element $0 \in V$ called the zero of $V$,

(iv) a mapping $\text{opp}: V \to V$ called the opposition in $V$, provided that the following axioms are satisfied:

\begin{align*}
(A1) \quad & \text{add} (u, \text{add}(v, w)) = \text{add}(\text{add}(u, v), w) \text{ for all } u, v, w \in V, \\
(A2) \quad & \text{add}(u, v) = \text{add}(v, u) \text{ for all } u, v \in V, \\
(A3) \quad & \text{add}(u, 0) = u \text{ for all } u \in V, \\
(A4) \quad & \text{add}(u, \text{opp}(u)) = 0 \text{ for all } u \in V,
\end{align*}
(S1) \( \text{sm}(\xi, \text{sm}(\eta, \mathbf{u})) = \text{sm}(\xi \eta, \mathbf{u}) \) for all \( \xi, \eta \in \mathbb{F}, \mathbf{u} \in \mathcal{V} \),
(S2) \( \text{sm}(\xi + \eta, \mathbf{u}) = \text{add}(\text{sm}(\xi, \mathbf{u}), \text{sm}(\eta, \mathbf{u})) \) for all \( \xi, \eta \in \mathbb{F}, \mathbf{u} \in \mathcal{V} \),
(S3) \( \text{sm}(\xi, \text{add}(\mathbf{u}, \mathbf{v})) = \text{add}(\text{sm}(\xi, \mathbf{u}), \text{sm}(\xi, \mathbf{v})) \) for all \( \xi \in \mathbb{F}, \mathbf{u}, \mathbf{v} \in \mathcal{V} \),
(S4) \( \text{sm}(1 \mathbf{u}) = \mathbf{u} \) for all \( \mathbf{u} \in \mathcal{V} \).

The prescription of \( \text{add} \), \( 0 \), and \( \text{opp} \), subject to the axioms (A1)-(A4), endows \( \mathcal{V} \) with the structure of a commutative group (See Sect. 06). Thus, one can say that a linear space is a commutative group endowed with additional structure by the prescription of a scalar multiplication \( \text{sm}: \mathbb{F} \times \mathcal{V} \to \mathcal{V} \) subject to the conditions (S1)-(S4).

The zero \( 0 \) of \( \mathcal{V} \) and the opposition \( \text{opp} \) of \( \mathcal{V} \) are uniquely determined by the operation \( \text{add} \) (see the remark on groupable pre-monoids in Sect. 06).

In other words, if \( \text{add}, \text{sm}, 0 \), and \( \text{opp} \) endow \( \mathcal{V} \) with the structure of a linear space and if \( \text{add}, \text{sm}, 0' \), and \( \text{opp}' \) also endow \( \mathcal{V} \) with such a structure, then \( 0' = 0 \) and \( \text{opp}' = \text{opp} \) and hence the structures coincide. This fact enables one to say that the prescription of two operations, \( \text{add}: \mathcal{V} \times \mathcal{V} \to \mathcal{V} \) and \( \text{sm}: \mathbb{F} \times \mathcal{V} \to \mathcal{V} \), endow \( \mathcal{V} \) with the structure of a linear space if there exist \( 0 \in \mathcal{V} \) and \( \text{opp}: \mathcal{V} \to \mathcal{V} \) such that the conditions (A1)-(S4) are satisfied.

The following facts are easy consequences of the (A1)-(S4):

(I) For every \( \mathbf{u}, \mathbf{v} \in \mathcal{V} \) there is exactly one \( \mathbf{w} \in \mathcal{V} \) such that \( \text{add}(\mathbf{u}, \mathbf{w}) = \mathbf{v} \); in fact, \( \mathbf{w} \) is given by \( \mathbf{w} := \text{add}(\mathbf{v}, \text{opp}(\mathbf{u})) \).

(II) \( \text{sm}(\xi - \eta, \mathbf{u}) = \text{add}(\text{sm}(\xi, \mathbf{u}), \text{opp}(\text{sm}(\eta, \mathbf{u}))) \) for all \( \xi, \eta \in \mathbb{F}, \mathbf{u} \in \mathcal{V} \).

(III) \( \text{sm}(\xi, \text{add}(\mathbf{u}, \text{opp}(\mathbf{v}))) = \text{add}(\text{sm}(\xi, \mathbf{u}), \text{opp}(\text{sm}(\xi, \mathbf{v}))) \) for all \( \xi \in \mathbb{F}, \mathbf{u}, \mathbf{v} \in \mathcal{V} \).

(IV) \( \text{sm}(-\xi, \mathbf{u}) = \text{opp}(\text{sm}(\xi, \mathbf{u})) \) for all \( \xi \in \mathbb{F}, \mathbf{u} \in \mathcal{V} \).

(V) If \( \xi \in \mathbb{F} \) and \( \mathbf{u} \in \mathcal{V} \), then \( \text{sm}(\xi, \mathbf{u}) = 0 \) if and only if \( \xi = 0 \) or \( \mathbf{u} = 0 \).

The field \( \mathbb{F} \) acquires the structure of a linear space if we prescribe \( \text{add}, \text{sm}, 0 \) and \( \text{opp} \) for \( \mathbb{F} \) by putting \( \text{add}(\xi, \eta) := \xi + \eta, \text{sm}(\xi, \eta) := \xi \eta, 0 := 0, \text{opp}(\xi) := -\xi \), so that addition, zero, and opposition in the linear space \( \mathbb{F} \) have their ordinary meaning while scalar multiplication reduces to ordinary multiplication. The axioms (A1)-(S4) reduce to rules of elementary arithmetic in \( \mathbb{F} \).

It is customary to use the following simplified notations in an arbitrary linear space \( \mathcal{V} \):
11. BASIC DEFINITIONS

\[ u + v := \text{add}(u,v) \quad \text{when} \quad u, v \in V, \]
\[ \xi u := \text{sm}(\xi, u) \quad \text{when} \quad \xi \in F, u \in V, \]
\[ -u := \text{opp}(u) \quad \text{when} \quad u \in V, \]
\[ u - v := u + (-v) = \text{add}(u, \text{opp}(v)) \quad \text{when} \quad u, v \in V. \]

We will use these notations most of the time. If several linear spaces are considered at the same time, the context should make it clear which of the several addition, scalar multiplication, or opposition operations are meant when the symbols +, −, or juxtaposition are used. Also, when the symbols 0 or 0 are used, they denote the zero of whatever linear space the context requires. With the notations above, the axioms (A1)-(S4) and the consequences (I)-(V) translate into the following familiar rules, valid for all \( u, v, w \in V \) and all \( \xi, \eta \in F \):

\[ u + (v + w) = (u + v) + w, \quad (11.1) \]
\[ u + v = v + u, \quad (11.2) \]
\[ u + 0 = u, \quad (11.3) \]
\[ u - u = 0, \quad (11.4) \]
\[ \xi(\eta u) = (\xi \eta)u, \quad (11.5) \]
\[ (\xi + \eta)u = \xi u + \eta u, \quad (11.6) \]
\[ \xi(u + v) = \xi u + \xi v, \quad (11.7) \]
\[ 1u = u, \quad (11.8) \]
\[ u + w = v \iff w = v - u, \quad (11.9) \]
\[ (\xi - \eta)u = \xi u - \eta u, \quad (11.10) \]
\[ \xi(u - v) = \xi u - \xi v, \quad (11.11) \]
\[ \xi u = 0 \iff (\xi = 0 \text{ or } u = 0). \quad (11.12) \]

Since the linear space \( V \) is a commutative monoid described with additive notation, we can consider the addition of arbitrary families with finite support of elements of \( V \) as explained in Sect. 07.

Notes 11

(1) The term “vector space” is very often used for what we call a “linear space”. The trouble with “vector space” is that it leads one to assume that the elements are “vectors” in some sense, while in fact they very often are objects that could not be called “vectors” by any stretch of the imagination. I prefer to use “vector” only when it has its original geometric meaning (see Def. 1 or Sect. 32).

(2) Sometimes, one finds the term “origin” for what we call the “zero” of a linear space.
12 Subspaces

Definition 1: A non-empty subset \( U \) of a linear space \( V \) is called a sub-

space of \( V \) if it is stable under the addition \( \text{add} \) and scalar multiplication

\( \text{sm} \) in \( V \), i.e., if

\[
\text{add}_> (U \times U) \subset U \quad \text{and} \quad \text{sm}_>(\mathbb{F} \times U) \subset U.
\]

It is easily proved that a subspace \( U \) of \( V \) must contain the zero \( 0 \) of \( V \)

and must be invariant under opposition, so that

\[
0 \in U, \quad \text{opp}_> U \subset U.
\]

Moreover, \( U \) acquires the natural structure of a linear space if the addition, scalar multiplication, and opposition in \( U \) are taken to be the adjustments

\( \text{add}_U, \text{sm}_U, \text{opp}_U \), respectively, and if the zero of \( U \) is taken to be the zero \( 0 \) of \( V \).

Let a linear space \( V \) be given. Trivial subspaces of \( V \) are \( V \) itself and the

zero-space \( \{0\} \). The following facts are easily proved:

Proposition 1: The collection of all subspaces of \( V \) is intersection stable; i.e., the intersection of any collection of subspaces of \( 0 \) \( V \) is again a subspace

of \( V \).

We denote the span-mapping (see Sect. 03) associated with the collection

of all subspaces of \( V \) by \( \text{Lsp} \) and call its value \( \text{Lsp} S \) at a given \( S \in \text{Sub} V \) the linear span of \( S \). In view of Prop. 1, \( \text{Lsp} S \) is the smallest subspace

of \( V \) that includes \( S \) (see Sect. 03). If \( \text{Lsp} S = V \) we say that \( S \) spans \( V \)

or that \( V \) is spanned by \( S \). A subset of \( V \) is a subspace if and only if it

coincides with its own linear span. The linear span of the empty subset of \( V \)

is the zero-space \( \{0\} \) of \( V \), i.e., \( \text{Lsp} \emptyset = \{0\} \). The linear span of a singleton

\( \{v\}, v \in V \), is the set of all scalar multiples of \( v \), which we denote by \( \mathbb{F}v \):

\[
\text{Lsp}\{v\} = \mathbb{F}v := \{\xi v \mid \xi \in \mathbb{F}\}.
\] (12.1)

We note that the notations for member-wise sums of sets introduced in

Sects. 06 and 07 can be used, in particular, if the sets are subsets of a linear

space.

Proposition 2: If \( \mathcal{U}_1 \) and \( \mathcal{U}_2 \) are subspaces of \( V \), so is their sum and

\[
\text{Lsp}(\mathcal{U}_1 \cup \mathcal{U}_2) = \mathcal{U}_1 + \mathcal{U}_2.
\] (12.2)
More generally, if \( (U_i \mid i \in I) \) is a finite family of subspaces of \( V \), so is its sum and

\[
\text{Lsp} \left( \bigcup_{i \in I} U_i \right) = \sum_{i \in I} U_i. \tag{12.3}
\]

**Proposition 3:** If \( (S_i \mid i \in I) \) is a finite family of arbitrary subsets of \( V \), then

\[
\text{Lsp} \left( \bigcup_{i \in I} S_i \right) = \sum_{i \in I} \text{Lsp} S_i. \tag{12.4}
\]

**Definition 2:** We say that two subspaces \( U_1 \) and \( U_2 \) of \( V \) are **disjunct**, and that \( U_2 \) is **disjunct from** \( U_1 \), if \( U_1 \cap U_2 = \{0\} \).

We say that two subspaces \( U_1 \) and \( U_2 \) of \( V \) are **supplementary** in \( V \) and that \( U_2 \) is a **supplement of** \( U_1 \) in \( V \) if \( U_1 \cap U_2 = \{0\} \) and \( U_1 + U_2 = V \).

**Proposition 4:** Let \( U_1, U_2 \) be subspaces of \( V \). Then the following are equivalent:

(i) \( U_1 \) and \( U_2 \) are supplementary in \( V \).

(ii) To every \( v \in V \) corresponds exactly one pair \((u_1, u_2) \in U_1 \times U_2\) such that \( v = u_1 + u_2 \).

(iii) \( U_2 \) is maximal among the subspaces that are disjunct from \( U_1 \), i.e., \( U_2 \) is disjunct from \( U_1 \) and not properly included in any other subspace disjunct from \( U_1 \).

**Pitfall:** One should not confuse “disjunct” with “disjoint”, or “supplement” with “complement”. Two subspaces, even if disjunct, are never disjoint. The complement of a subspace is never a subspace, and there is no relation between this complement and any supplement. Moreover, supplements are not unique unless the given subspace is the whole space \( V \) or the zero-space \( \{0\} \).

To say that two subspaces \( U_1 \) and \( U_2 \) are disjunct is equivalent to saying they are supplementary in \( U_1 + U_2 \).
Notes 12

(1) The phrase “subspace generated by” is often used for what we call “linear span of”. The modifier “linear” is very often omitted and the linear span of a subset $S$ of $V$ is often simply denoted by $\text{Sp } S$. I think it is important to distinguish carefully between linear spans and other kinds of spans. (See the discussion of span-mappings in Sect. 03 and the definition of flat spans in Prop. 7 of Sect. 32).

(2) In many textbooks the term “disjoint” is used when we say “disjunct” and the term “complementary” when we say “supplementary”. Such usage clashes with the set-theoretical meanings of “disjoint” and “complementary” and greatly increases the danger of becoming a victim of the Pitfall above. It is for this reason that I introduced the terms “disjunct” and “supplementary” after consulting a thesaurus. Later, I realized that “supplementary” had already been used by others. In particular, Bourbaki has used “supplémentaire” since 1947, which I read in 1950 but had forgotten.

(3) Some people write $U_1 \oplus U_2$ instead of merely $U_1 + U_2$ when the two subspaces $U_1$ and $U_2$ are disjunct, and then call $U_1 \oplus U_2$ the “direct sum” rather than merely the “sum” of $U_1$ and $U_2$. This is really absurd; it indicates confusion between a property (namely disjunctness) of a pair $(U_1, U_2)$ of two subspaces and a property of their sum. The sum $U_1 + U_2$, as a subspace of $V$, has no special properties, because $U_1$ and $U_2$ cannot be recovered from $U_1 + U_2$ (except when they are all zero-spaces).

13 Linear Mappings

Definition 1: A mapping $L : V \to V'$ from a linear space $V$ to a linear space $V'$ is said to be linear if it preserves addition and scalar multiplication, i.e., if

$$\text{add}'(L(u), L(v)) = L(\text{add}(u, v)) \text{ for all } u, v \in V$$

and

$$\text{sm}'(\xi, L(u)) = L(\text{sm}(\xi, u)) \text{ for all } \xi \in \mathbb{F}, u \in V.$$  

where add, sm denote operations in $V$ and add', sm' operations in $V'$.

For linear mappings, it is customary to omit parentheses and write $Lu$ for the value of $L$ at $u$. In simplified notation, the rules that define linearity read

$$Lu + Lv = L(u + v), \quad (13.1)$$

$$L(\xi u) = \xi(Lu), \quad (13.2)$$
valid for all \( u, v \in V \) and all \( \xi \in F \). It is easily seen that linear mappings \( L : V \to V' \) also preserve zero and opposition, i.e., that

\[
L0 = 0'
\]  

(13.3)

when \( 0 \) and \( 0' \) denote the zeros of \( V \) and \( V' \) respectively, and that

\[
L(-u) = -(Lu)
\]  

(13.4)

for all \( u \in V \).

the constant mapping \( 0_{V' \to V} \), whose value is the zero \( 0' \) of \( V' \), is the only constant mapping that is linear. This mapping is called a zero-mapping and is denoted simply by \( 0 \). The identity mapping \( 1_V : V \to V \) of any linear space \( V \) is trivially linear.

The following facts are easily proved:

**Proposition 1:** The composite of two linear mappings is again linear. More precisely: If \( V, V' \) and \( V'' \) are linear spaces and if \( L : V \to V' \) and \( M : V' \to V'' \) are linear mappings, so is \( M \circ L : V \to V'' \).

**Proposition 2:** The inverse of an invertible linear mappings is again linear. More precisely: If \( L : V \to V' \) is linear and invertible, then \( L^{-1} : V' \to V \) is linear.

If \( L : V \to V' \) is linear, it is customary to write \( Lf := L \circ f \) when \( f \) is any mapping with codomain \( V \). In particular, we write \( ML \) instead of \( M \circ L \) when both \( L \) and \( M \) are linear.

If \( L : V \to V' \) is linear and invertible, it is customary to write \( L^{-1} := L^{-} \) for its inverse. By Prop. 2, both \( L \) and \( L^{-1} \) then preserve the linear-space structure. Invertible linear mappings are, therefore, linear-space isomorphisms, and we also call them linear isomorphisms. We say that two linear spaces \( V \) and \( V' \) are linearly isomorphic if there exists a linear isomorphism from \( V \) to \( V' \).

**Proposition 3:** If \( L : V \to V' \) is linear and if \( U \) and \( U' \) are subspaces of \( V \) and \( V' \), respectively, then the image \( L_>(U) \) of \( U \) is a subspace of \( V' \) and the pre-image \( L^<(U') \) of \( U' \) is a subspace of \( V' \).

In particular, \( \text{Rng } L = L_>(V) \) is a subspace of \( V' \) and \( L^<\{0\} \) is a subspace of \( V \).

**Definition:** The pre-image of the zero-subspace of the codomain of a linear mapping \( L \) is called the nullspace of \( L \) and is denoted by
Null \( L := L^< \{0\} = \{u \in \text{Dom } L \mid Lu = 0\} \). (13.5)

**Proposition 4:** A linear mapping \( L : \mathcal{V} \to \mathcal{V}' \) is injective if and only if \( \text{Null } L = \{0\} \).

If \( L : \mathcal{V} \to \mathcal{V}' \) is linear, if \( \mathcal{U} \) is a subspace of \( \mathcal{V} \), and if \( \mathcal{U}' \) is a linear space of which \( L|_{\mathcal{U}} \) is a subspace, then the adjustment \( L|_{\mathcal{U}} : \mathcal{U} \to \mathcal{U}' \) is evidently again linear. In particular, if \( \mathcal{U} \) is a subspace of \( \mathcal{V} \), the inclusion mapping \( 1_{\mathcal{U} \subset \mathcal{V}} \) is linear.

**Proposition 5:** If \( L : \mathcal{V} \to \mathcal{V}' \) is linear and \( \mathcal{U} \) is a supplement of Null \( L \) in \( \mathcal{V} \), then \( L|_{\mathcal{U}} : \mathcal{U} \to \mathcal{U}' \) is invertible. If \( v' \in \text{Rng } L \), then \( v \in \mathcal{V} \) is a solution of the linear equation

\[
? v \in \mathcal{V}, \quad Lv = v'
\]

if and only if \( v \in \left(L|_{\mathcal{U}}\right)^{-1} v' + \text{Null } L \).

Notes 13

1. The terms “linear transformation” or “linear operator” are sometimes used for what we call a “linear mapping”.

2. Some people use the term “kernel” for what we call “nullspace” and they use the notation \( \text{Ker } L \) for \( \text{Null } L \). Although the nullspace of \( L \) is a special kind of kernel (in the sense explained in Sect. 06), I believe it is useful to have a special term and a special notation to stress that one deals with linear mappings rather than homomorphisms in general. Notations such as \( N(N) \) and \( N(L) \) are often used for Null \( L \).

14 Spaces of Mappings, Product Spaces

Let \( \mathcal{V} \) be a linear space and let \( S \) be any set. The set Map\((S, \mathcal{V})\) of all mappings from \( S \) to \( \mathcal{V} \) can be endowed, in a natural manner, with the structure of a linear space by defining the operations in Map\((S, \mathcal{V})\) by value-wise application of the operations in \( \mathcal{V} \). Thus, if \( f, g \in \text{Map}(S, \mathcal{V}) \) and \( \xi \in \mathbb{F} \), then \( f + g \), \(-f\) and \( \xi f \) are defined by requiring

\[
(f + g)(s) := f(s) + g(s), \quad (14.1)
\]

\[
(-f)(s) := -(f(s)), \quad (14.2)
\]

\[
(\xi f)(s) := \xi(f(s)), \quad (14.3)
\]

to be valid for all \( s \in S \). The zero-element of Map\((S, \mathcal{V})\) is the constant mapping \( 0_{S \to \mathcal{V}} \) with value \( 0 \in \mathcal{V} \). We denote it simply by \( 0 \). It is immediate
that the axioms (A1)-(S4) of Sect. 01 for a linear space are, in fact, satisfied for the structure of Map(S, V) just described. Thus, we can talk about the (linear) space of mappings Map(S, V).

**Proposition 1:** Let S and S′ be sets, let h : S′ → S be a mapping, and let V be a linear space. Then

\[(f \mapsto f \circ h) : \text{Map}(S, V) \to \text{Map}(S′, V)\]

is a linear mapping, i.e.,

\[(f + g) \circ h = (f \circ h) + (g \circ h), \quad (\xi f) \circ h = \xi (f \circ h)\]  

hold for all f, g ∈ Map(S, V) and all ξ ∈ F.

Let V, V′ be linear spaces. We denote the set of all linear mappings from V into V′ by

\[\text{Lin}(V, V′) := \{L \in \text{Map}(V, V′) \mid L \text{ is linear}\}.\]

**Proposition 2:** Lin(V, V′) is a subspace of Map(V, V′).

**Proof:** Lin(V, V′) is not empty because the zero-mapping belongs to it. We must show that Lin(V, V′) is stable under addition and scalar multiplication. Let L, M ∈ Lin(V, V′) be given. Using first the definition (14.1) for L + M, then the axioms (A1) and (A2) in V′, then the linearity rule (13.1) for L and M, and finally the definition of L + M again, we obtain

\[
(L + M)(u) + (L + M)(v) = (Lu + Mu) + (Lv + Mv)
= (Lu + Lv) + (Mu + Mv)
= L(u + v) + M(u + v)
= (L + M)(u + v)
\]

for all u, v ∈ V. This shows that L + M satisfies the linearity rule (13.1). In a similar way, one proves that L + M satisfies the linearity rule (13.2) and hence that L + M ∈ Lin(V, V′). Since L, M ∈ Lin(calV, V′) were arbitrary, it follows that Lin(V, V′) is stable under addition.

The proof that Lin(V, V′) is stable under scalar multiplication is left to the reader. ■

We call Lin(V, V′) the space of linear mappings from V to V′. We denote the set of all invertible linear mappings from V to V′, i.e., the set of
all linear isomorphisms from $V$ to $V'$, by $\text{Lis}(V, V')$. This set $\text{Lis}(V, V')$ is a subset of $\text{Lin}(V, V')$ but not a subspace (except when both $V$ and $V'$ are zero-spaces).

**Proposition 3:** Let $S$ be a set, let $V$ and $V'$ be linear spaces, and let $L \in \text{Lin}(V, V')$ be given. Then

$$(f \mapsto Lf) : \text{Map}(S, V) \to \text{Map}(S, V')$$

is a linear mapping, i.e.,

$$L(f + g) = Lf + Lg, \quad (14.6)$$
$$L(\xi f) = \xi (Lf) \quad (14.7)$$

hold for all $f, g \in \text{Map}(S, V)$ and all $\xi \in F$.

Let $(V_1, V_2)$ be a pair of linear spaces. The set product $V_1 \times V_2$ (see Sect. 02) has the natural structure of a linear space whose operations are defined by term-wise application of the operations in $V_1$ and $V_2$, i.e., by

$$(u_1, u_2) + (v_1, v_2) := (u_1 + v_1, u_2 + v_2), \quad (14.8)$$
$$\xi (u_1, u_2) := (\xi u_1, \xi u_2), \quad (14.9)$$
$$-(u_1, u_2) := (-u_1, -u_2) \quad (14.10)$$

for all $u_1, v_1 \in V_1$, $u_2, v_2 \in V_2$, and $\xi \in F$. The zero of $V_1 \times V_2$ is the pair $(0_1, 0_2)$, where $0_1$ is the zero of $V_1$ and $0_2$ the zero of $V_2$. Thus, we may refer to $V_1 \times V_2$ as the (linear) **product-space** of $V_1$ and $V_2$.

The evaluations $\text{ev}_1 : V_1 \times V_2 \to V_1$ and $\text{ev}_2 : V_1 \times V_2 \to V_2$ associated with the product space $V_1 \times V_2$ (see Sect. 04) are obviously linear. So are the **insertion mappings**

$$\text{ins}_1 := (u_1 \mapsto (u_1, 0)) : V_1 \to V_1 \times V_2, \quad (14.11)$$
$$\text{ins}_2 := (u_2 \mapsto (0, u_2)) : V_2 \to V_1 \times V_2.$$  

If $U_1$ is a subspace of $V_1$ and $U_2$ a subspace of $V_2$, then $U_1 \times U_2$ is a subspace of $V_1 \times V_2$.

**Pitfall:** In general, the product-space $V_1 \times V_2$ has many subspaces that are **not** of the form $U_1 \times U_2$. ■

Let $W$ be a third linear space. Recall the identification
14. SPACES OF MAPPINGS, PRODUCT SPACES

\[ \text{Map}(W, V_1) \times \text{Map}(W, V_2) \cong \text{Map}(W, V_1 \times V_2) \]

(see Sect. 04). It turns out that this identification is such that the subspaces \( \text{Lin}(W, V_1), \text{Lin}(W, V_2) \), and \( \text{Lin}(W, V_1 \times V_2) \) are matched in the sense that the pairs \( (L_1, L_2) \in \text{Lin}(W, V_1) \times \text{Lin}(W, V_2) \) of linear mappings correspond to the linear mappings in \( \text{Lin}(W, V_1 \times V_2) \) defined by term-wise evaluation, i.e., by

\[ (L_1, L_2)w := (L_1w, L_2w) \text{ for all } w \in W. \]  

(14.12)

Thus, (14.12) describes the identification

\[ \text{Lin}(W, V_1) \times \text{Lin}(W, V_2) \cong \text{Lin}(W, V_1 \times V_2). \]

There is also a natural linear isomorphism

\[ ((L_1, L_2) \mapsto L_1 \oplus L_2) : \text{Lin}(V_1, W) \times \text{Lin}(V_2, W) \to \text{Lin}(V_1 \times V_2, W). \]

It is defined by

\[ (L_1 \oplus L_2)(v_1, v_2) := L_1v_1 + L_2v_2 \]  

(14.13)

for all \( v_1 \in V_1, v_2 \in V_2 \), which is equivalent to

\[ L_1 \oplus L_2 := L_1ev_1 + L_2ev_2, \]  

(14.14)

where \( ev_1 \) and \( ev_2 \) are the evaluation mappings associated with \( V_1 \times V_2 \) (see Sect. 04).

What we said about a pair of linear spaces easily generalizes to an arbitrary family \( (V_i \mid i \in I) \) of linear spaces. The set product \( \prod (V_i \mid i \in I) \) has the natural structure of a linear space whose operations are defined by term-wise application of the operations in the \( V_i, i \in I \). Hence, we may refer to \( \prod (V_i \mid i \in I) \) as the **product-space** of the family \( (V_i \mid i \in I) \). Given \( j \in I \), the evaluation

\[ ev_j : \prod_{i \in I} V_i \to V_j, \]

(see (04.9)) is linear. So is the insertion mapping

\[ \text{ins}_j : V_j \to \prod_{i \in I} V_i \]
CHAPTER 1. LINEAR SPACES

defined by

\[(\text{ins}_j \mathbf{v})_i := \begin{cases} 0 \in \mathcal{V}_i & \text{if } i \neq j \\ \mathbf{v} \in \mathcal{V}_j & \text{if } i = j \end{cases} \quad \text{for all } \mathbf{v} \in \mathcal{V}_j.\]  \hspace{1cm} (14.15)

It is evident that

\[ev_k \text{ ins}_j = \begin{cases} 0 \in \text{Lin}(\mathcal{V}_j, \mathcal{V}_k) & \text{if } j \neq k \\ 1_{\mathcal{V}_j} \in \text{Lin}(\mathcal{V}_j, \mathcal{V}_j) & \text{if } j = k \end{cases} \]  \hspace{1cm} (14.16)

for all \(j, k \in I\).

Let \(\mathcal{W}\) be an additional linear space. For families \((\mathbf{L}_i \mid i \in I)\) or linear mappings \(\mathbf{L}_i \in \text{Lin}(\mathcal{W}, \mathcal{V}_i)\) we use termwise evaluation

\[(\mathbf{L}_i \mid i \in I)\mathbf{w} := (\mathbf{L}_i \mathbf{w} \mid i \in I) \quad \text{for all } \mathbf{w} \in \mathcal{W},\]  \hspace{1cm} (14.17)

which describes the identification

\[\bigotimes_{i \in I} \text{Lin}(\mathcal{W}, \mathcal{V}_i) \cong \text{Lin}(\mathcal{W}, \bigotimes_{i \in I} \mathcal{V}_i).\]

If the index set \(I\) is finite, we also have a natural isomorphism

\[((\mathbf{L}_i \mid i \in I) \mapsto \bigoplus_{i \in I} \mathbf{L}_i) : \bigotimes_{i \in I} \text{Lin}(\mathcal{V}_i, \mathcal{W}) \to \text{Lin}(\bigotimes_{i \in I} \mathcal{V}_i, \mathcal{W})\]

defined by

\[\bigoplus_{i \in I} \mathbf{L}_i := \sum_{i \in I} \mathbf{L}_i ev_i.\]  \hspace{1cm} (14.18)

If the spaces in a family indexed on \(I\) all coincide with a given linear space \(\mathcal{V}\), then the product space reduces to the **power-space** \(\mathcal{V}^I\), which consist of all families in \(\mathcal{V}\) indexed on \(I\) (see Sect. 02). The set \(\mathcal{V}^I\) of all families contained in \(\mathcal{V}^I\) that have finite support (see Sect. 07) is easily seen to be a subspace of \(\mathcal{V}^I\). Of particular interest is the space \(\mathbb{F}^I\) of all families \(\lambda := (\lambda_i \mid i \in I)\) in \(\mathbb{F}\) with finite support. Also, if \(I\) and \(J\) are finite sets, it is useful to consider the linear space \(\mathbb{F}^{J \times I}\) of \(J \times I\)-matrices with terms in \(\mathbb{F}\) (see Sect. 02). Cross products of linear mappings, as defined in Sect. 04, are again linear mappings.
15. LINEAR COMBINATIONS, LINEAR INDEPENDENCE, BASES

Notes 14

(1) The notations $\mathcal{L}(\mathcal{V}, \mathcal{V}')$ and $L(\mathcal{V}, \mathcal{V}')$ for our $\operatorname{Lin}(\mathcal{V}, \mathcal{V}')$ are very common.

(2) The notation $\operatorname{Lin}(\mathcal{V}, \mathcal{V}')$ was apparently first introduced by S. Lang (Introduction to Differentiable Manifolds, Interscience 1966). In some previous work, I used $\operatorname{Invlin}(\mathcal{V}, \mathcal{V}')$.

(3) The product-space $\mathcal{V}_1 \times \mathcal{V}_2$ is sometimes called the "direct sum" of the linear spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ and it is then denoted by $\mathcal{V}_1 \oplus \mathcal{V}_2$. I believe such a notation is superfluous because the set-product $\mathcal{V}_1 \times \mathcal{V}_2$ carries the natural structure of a linear space and a special notation to emphasize this fact is redundant. A similar remark applies to product-spaces of families of linear spaces.

15 Linear Combinations, Linear Independence, Bases

Definition 1: Let $f := (f_i \mid i \in I)$ be a family of elements in a linear space $\mathcal{V}$. The mapping

$$\operatorname{lnc}_f^\mathcal{V} : \mathbb{F}(I) \to \mathcal{V}$$

defined by

$$\operatorname{lnc}_f^\mathcal{V} \lambda := \sum_{i \in I} \lambda_i f_i \quad (15.1)$$

for all $\lambda := (\lambda_i \mid i \in I) \in \mathbb{F}(I)$ is then called the linear-combination mapping for $f$. The value $\operatorname{lnc}_f^\mathcal{V} \lambda$ is called the linear combination of $f$ with coefficient family $\lambda$. (See (07.10) and (07.11) for the notation used here.)

It is evident from the rules that govern sums (see Sect. 07) that linear combination mappings are linear mappings.

In the special case when $\mathcal{V} := \mathbb{F}$ and when $f$ is the constant family whose terms are all $1 \in \mathbb{F}$, then $\operatorname{lnc}_f^\mathcal{V}$ reduces to the summation mapping

$$\operatorname{sum}_f : \mathbb{F}(I) \to \mathbb{F} \text{ given by }$$

$$\operatorname{sum}_f \lambda := \sum_{i \in I} \lambda_i \quad \text{for all } \lambda \in \mathbb{F}(I). \quad (15.2)$$

Let $\mathcal{U}$ be a subspace of the given linear space $\mathcal{V}$ and let $f$ be a family in $\mathcal{U}$. then $\operatorname{Rng} \operatorname{lnc}_f^\mathcal{V} \subset \mathcal{U}$ and $\operatorname{lnc}_f^\mathcal{U}$ is obtained from $\operatorname{lnc}_f^\mathcal{V}$ by adjustment of codomain

$$\operatorname{lnc}_f^\mathcal{U} = \operatorname{lnc}_f^\mathcal{V} \mid \mathcal{U} \quad \text{if } \operatorname{Rng} f \subset \mathcal{U}. \quad (15.3)$$
Definition 2: The family $f$ in a linear space $V$ is said to be linearly independent in $V$, spanning in $V$, or a basis of $V$ depending on whether the linear combination mapping $\text{lnc}_f$ is injective, surjective, or invertible, respectively. We say that $f$ is linearly dependent if it is not linearly independent.

If $f$ is a family in a given linear space $V$ and if there is no doubt what $V$ is, we simply write $\text{lnc}_f := \text{lnc}_V$. Also we then omit “in $V$” and “of $V$” when we say that $f$ is linear independent, spanning, or a basis. Actually, if $f$ is linearly independent in $V$, it is also linearly independent in any linear space that includes $\text{Rng } f$.

If $b := (b_i | i \in I)$ is a basis of $V$ and if $v \in V$ is given, then $\text{ln}_b^{-1} v \in F(I)$ is called the family components of $v$ relative to the basis $b$. Thus

$$v = \sum_{i \in I} \lambda_i b_i$$

holds if and only if $(\lambda_i | i \in I) \in F(I)$ is the family of components of $v$ relative to $b$.

An application of Prop. 4 of Sect. 13 gives:

Proposition 1: The family $f$ is linearly independent if and only if $\text{Null } (\text{lnc}_f) = \{0\}$.

Let $I'$ be a subset of the given index set $I$. We then define the insertion mapping

$$\text{ins}_{I' \subset I} : F(I') \rightarrow F(I)$$

by

$$(\text{ins}_{I' \subset I}(\lambda))_i = \begin{cases} \lambda_i & \text{if } i \in I' \\ 0 & \text{if } i \in I \setminus I' \end{cases}$$

(15.5)

It is clear that $\text{ins}_{I' \subset I}$ is an injective linear mapping. If $f$ is a family indexed on $I$ and if $f|_{I'}$ is its restriction to $I'$, we have

$$\text{lnc}_{f|_{I'}} = \text{ln}_f \text{ins}_{I' \subset I}.$$  

(15.6)

From this formula and the injectivity of $\text{ins}_{I' \subset I}$ we can read off the following:

Proposition 2: If the family $f$ is linearly independent, so are all its restrictions. If any restriction of $f$ is spanning, so is $f$. 


Let $S$ be a subset of $V$. In view of the identification of a set with the family obtained by self-indexing (see Sect. 02) we can consider the linear combination mapping $\text{lin}_S : F(S) \to V$, given by

$$\text{lin}_S \lambda := \sum_{u \in S} \lambda_u u$$

(15.7)

for all $\lambda := (\lambda_u | u \in S) \in F(S)$. Thus the definitions above apply not only to families in $V$ in general, but also to subsets of $V$ in particular.

The following facts are easily verified:

**Proposition 3:** A family in $V$ is linearly independent if and only if it is injective and its range is linearly independent. No term of a linearly independent family can be zero.

**Proposition 4:** A family in $V$ is spanning if and only if its range is spanning.

**Proposition 5:** A family in $V$ is a basis if and only if it is injective and its ranges is a basis-set.

The following result shows that the definition of a spanning set as a special case of a spanning family is not in conflict with the definition of a set that spans the space as given in Sect. 12.

**Proposition 6:** The set of all linear combinations of a family in $V$ is the linear span of the range of $f$, i.e., $\text{Rng lin}_f = \text{Lsp}(\text{Rng } f)$. In particular, if $S$ is a subset of $V$, then $\text{Rng lin}_S = \text{Lsp}S$.

The following is an immediate consequence of Prop. 6:

**Proposition 7:** A family $f$ is linearly independent if and only if it is a basis of $\text{LspRng } f$.

The next two results are not not hard to prove.

**Proposition 8:** A subset $b$ of $V$ is linearly dependent if and only if $\text{Lsp } b = \text{Lsp}(b \setminus \{v\})$ for some $v \in b$.

**Proposition 9:** If $b$ is a linearly independent subset of $V$ and $v \in V \setminus b$, then $b \cup \{v\}$ is linearly dependent if and only if $v \in \text{Lsp } b$.

By applying Props. 8 and 9, one easily proves the following important result:
Characterization of Bases: Let $b$ be subset of $V$. Then the following are equivalent:

(i) $b$ is a basis of $V$.
(ii) $b$ is both linearly independent and spanning.
(iii) $b$ is a maximal linearly independent set, i.e., $b$ is linearly independent and is not a proper subset of any other linearly independent set.
(iv) $b$ is a minimal spanning set, i.e., $b$ is spanning and has no proper subset that is also spanning.

We note that the zero-space $\{0\}$ has exactly one set basis, namely the empty set.

Pitfall: Every linear space other than a zero-space has infinitely many bases, if it has any at all. Unless the space has structure in addition to its structure as a linear space, all of these bases are of equal standing. The bases form a “democracy”. For example, every singleton $\{\xi\}$ with $\xi \in F^x$ is a basis set of the linear space $F$. The special role of the number 1 and hence the bases $\{1\}$ comes from the additional structure in $F$ given by the multiplication.

Notes 15

(1) In most textbooks, the definition of “linear combination” is rather muddled. Much of the confusion comes from a failure to distinguish a process (the linear-combination mapping) from its result (the linear-combination). Also, most authors fail to make a clear distinction between sets, lists, and families, a distinction that is crucial here. I believe that much precision, clarity, and insight are gained by the use of linear-combination mappings. Most textbooks only talk around these mappings without explicitly using them.

(2) The condition of Prop. 1 is very often used as the definition of linear linear independence.

(3) Many people say “coordinates” instead of “components” of $v$ relative to the basis of $b$. I prefer to use the term “coordinate” only when it has the meaning described in Chapter 7.

16 Matrices, Elimination of Unknowns

The following result states, roughly, that linear mappings preserve linear combinations.
Proposition 1: If $\mathcal{V}$ and $\mathcal{W}$ are linear spaces, $f := (f_i \mid i \in I)$ a family in $\mathcal{V}$, $L : \mathcal{V} \rightarrow \mathcal{W}$ a linear mapping, and $Lf := (Lf_i \mid i \in I)$, then

$$\ln c(Lf) = L \ln c f.$$ \hspace{1cm} (16.1)

Applying (16.1) to the case when $f$ is a basis, we obtain:

Proposition 2: Let $b := (b_i \mid i \in I)$ be a basis of the linear space $\mathcal{V}$ and let $g := (g_i \mid i \in I)$ be a family in the linear space $\mathcal{W}$, indexed on the same set $I$ as $b$. Then there is exactly one $L \in \text{Lin}(\mathcal{V}, \mathcal{W})$ such that $Lb = g$. This $L$ is injective, surjective, or invertible depending on whether $g$ is linearly independent, spanning, or a basis, respectively.

The first part of Prop. 2 states, roughly, that a linear mapping can be specified by prescribing its effect on a basis.

Let $I$ be any index set. We define a family $\delta^I := (\delta^I_i \mid i \in I)$ in $\mathbb{F}^I$ by

$$(\delta^I_i)_k = \delta_{i,k} := \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i \end{cases}$$ for all $i, k \in I$. \hspace{1cm} (16.2)

It is easily seen that $\ln c \delta^I = 1_{\mathbb{F}(I)}$, which is, of course, invertible. Hence $\delta^I$ is a basis of $\mathbb{F}(I)$. We call it the standard basis of $\mathbb{F}(I)$. If $f := (f_i \mid i \in I)$ is a family in a given linear space $\mathcal{V}$, then

$$f_j = \ln c f \delta^I_j$$ for all $j \in I$. \hspace{1cm} (16.3)

If we apply Prop. 2 to the case when $I$ is finite, when $b := \delta^I$ is the standard basis of $\mathcal{V} := \mathbb{F}^I$, and when $\mathcal{W} := \mathbb{F}^J$ for some finite index set $J$, we obtain:

Proposition 3: Let $I$ and $J$ be finite index sets. Then the mapping from $\text{Lin}(\mathbb{F}^I, \mathbb{F}^J)$ to $\mathbb{F}^{J \times I}$ which assigns to $M \in \text{Lin}(\mathbb{F}^I, \mathbb{F}^J)$ the matrix $((M\delta^I_j)_i \mid (j, i) \in J \times I)$ is a linear isomorphism.

We use the natural isomorphism described in Prop. 3 to identify $M \in \text{Lin}(\mathbb{F}^I, \mathbb{F}^J)$ with the corresponding matrix in $\mathbb{F}^{J \times I}$, so that

$$M_{j,i} = (M\delta^I_j)_i$$ for all $(j, i) \in J \times I$. \hspace{1cm} (16.4)

Thus, we obtain the identification

$$\text{Lin}(\mathbb{F}^I, \mathbb{F}^J) \cong \mathbb{F}^{J \times I}. \hspace{1cm} (16.5)$$

Proposition 4: Let $I, J$ be finite index sets. If $M \in \text{Lin}(\mathbb{F}^I, \mathbb{F}^J) \cong \mathbb{F}^{J \times I}$ and $\lambda \in \mathbb{F}^I$, then
(16.6)\[ (M\lambda)_j = \sum_{j\in J} N_{k,j} M_{j,i} \text{ for all } i \in I. \]

**Proposition 5:** Let $I, J, K$ be finite index sets. If $M \in \text{Lin}(F^I, F^J) \cong F^{J \times I}$ and $N \in \text{Lin}(F^J, F^K) \cong F^{K \times J}$, then $NM \in \text{Lin}(F^I, F^K) \cong F^{K \times I}$ is given by

\[ (NM)_{k,i} = \sum_{j \in J} N_{k,j} M_{j,i} \text{ for all } k, i \in K \times I. \] (16.7)

In the case when $I := n \mid J := m \mid K := p \mid$ and when the bookkeeping scheme (02.4) is used, then (16.6) and (16.7) can be represented in the forms

\[
\begin{bmatrix}
M_{11} & M_{12} & \cdots & M_{1n} \\
M_{21} & M_{22} & \cdots & M_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
M_{m1} & M_{m2} & \cdots & M_{mn}
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_n
\end{bmatrix}
= \begin{bmatrix}
\sum_{i \in I} M_{1i} \lambda_i \\
\sum_{i \in I} M_{2i} \lambda_i \\
\vdots \\
\sum_{i \in I} M_{ni} \lambda_i
\end{bmatrix}
\] (16.8)

and

\[
\begin{bmatrix}
N_{11} & \cdots & N_{1m} \\
\vdots & \ddots & \vdots \\
N_{p1} & \cdots & N_{pm}
\end{bmatrix}
\begin{bmatrix}
M_{11} & \cdots & M_{1n} \\
M_{m1} & \cdots & M_{mn}
\end{bmatrix}
= \begin{bmatrix}
\sum_{j \in m} N_{1j} M_{j1} & \cdots & \sum_{j \in m} N_{1j} M_{jn} \\
\vdots & \ddots & \vdots \\
\sum_{j \in m} N_{pj} M_{j1} & \cdots & \sum_{j \in m} N_{pj} M_{jn}
\end{bmatrix}
\] (16.9)

In particular, (16.9) states that the composition matrices when regarded as linear mappings corresponds to the familiar “row-by-column” multiplication of matrices.

Let $V$ and $W$ be linear spaces and let $b := (b_i \mid i \in I)$ and $c := (c_j \mid j \in J)$ be finite bases of $V$ and $W$, respectively. Then the mapping

\[ (L \mapsto (\text{Ln}_c)^{-1} \text{Ln}_b) : \text{Lin}(V, W) \to \text{Lin}(F^I, F^J) \] (16.10)

is a linear isomorphism. Given $L \in \text{Lin}(V, W)$ we say that $M := (\text{Ln}_c)^{-1} L \text{Ln}_b \in \text{Lin}(F^I, F^J) \cong F^{J \times I}$ is the **matrix of the linear mapping** $L$ relative to the bases $b$ and $c$. It is characterized by
16. MATRICES, ELIMINATION OF Unknowns

\[ \mathbf{Lb}_i = \sum_{j \in J} M_{j,i} c_j \text{ for all } i \in I. \quad (16.11) \]

Let \( I \) and \( J \) be finite index sets. If \( M \in \text{Lin}(\mathbb{F}^I, \mathbb{F}^J) \cong \mathbb{F}^{J \times I} \) and \( \mu \in \mathbb{F}^J \) are given and if we consider the problem \( \lambda \in \mathbb{F}^I, M\lambda = \mu \), i.e.,

\[ ?(\lambda_i \mid i \in I) \in \mathbb{F}^I, \sum_{i \in I} M_{j,i} \lambda_i = \mu_j \text{ for all } j \in J, \quad (16.12) \]

we say that we have the problem of solving a system of \( \#J \) linear equations with \( \#I \) unknowns. The following theorem describes a procedure, familiar from elementary algebra, called eliminations of unknowns, which enables one to reduce a system of linear equations to one having one equation less and one unknown less.

**Theorem on Elimination of Unknowns:** Let \( I \) and \( J \) be finite sets and let \( M \in \mathbb{F}^{J \times I} \) be such that \( M_{j_0,i_0} \neq 0 \) for given \( j_0 \in J, i_0 \in I \). Put \( I' := I \setminus \{i_0\}, J' := J \setminus \{j_0\} \) and define \( M' \in \mathbb{F}^{J' \times I'} \) by

\[ M'_{j,i} := M_{j,i} - \frac{M_{j_0,i} M_{j,i_0}}{M_{j_0,i_0}} \text{ for all } (j, i) \in J' \times I'. \quad (16.13) \]

Let \( \mu \in \mathbb{F}^J \) and \( \mu' \in \mathbb{F}^{J'} \) be related by

\[ \mu'_j = \mu_j - \frac{M_{j_0,i_0}}{M_{j_0,i_0}} j_0 \text{ for all } j \in J'. \quad (16.14) \]

Then \( \lambda \in \mathbb{F}^I \) is a solution of the equation

\[ ? \lambda \in \mathbb{F}^I, M\lambda = \mu \]

if and only if the restriction \( \lambda \mid \nu \in \mathbb{F}^{I'} \) of \( \lambda \) is a solution of the equation

\[ ? \lambda' \in \mathbb{F}^{I'}, M'\lambda' = \mu' \]

and \( \lambda_{i_0} \) is given by

\[ \lambda_{i_0} = \frac{1}{M_{j_0,i_0}} \left( \mu_{j_0} - \sum_{i \in I'} M_{j_0,i} \lambda_i \right). \quad (16.15) \]

**Corollary:** If \( \text{Lin}(\mathbb{F}^I, \mathbb{F}^J) \) contains an injective mapping, then \( \#J \geq \#I \).

**Proof:** We proceed by induction over \( \#I \). If \( \#I = 0 \) then the assertion is trivial. Assume, then, that \( \#I > 0 \) and that the assertion is valid if \( I \) is
CHAPTER 1. LINEAR SPACES

replaced by a subset \( I' \) of \( I \) having one element less. Assume that \( \text{Lin}(\mathcal{F}^I, \mathcal{F}^J) \)
contains an injective mapping \( M \), do that Null \( M = \{0\} \). Since \( \mathcal{F}^I \neq \{0\} \)
we must have \( M \neq 0 \). Hence we can apply the Theorem with the choice \( \mu := 0 \) to conclude that Null \( M' = \{0\} \), i.e., that \( M' \in \text{Lin}(\mathcal{F}^I', \mathcal{F}^J') \) must
be injective. By the induction hypothesis, we have \( \#J' \geq \#I' \), which means
\((\#J) - 1 \geq (\#I) - 1 \) and hence implies \( \#J \geq \#I \). □

Notes 16

(1) The notation \( \delta_{i,k} \) as defined by (16.2) is often attributed to Kronecker (“the Kro-
necker deltas”).

(2) The standard basis \( \delta^I \) is sometimes called the “natural basis” of \( \mathcal{F}^I \).

(3) Various algorithms, based on the Theorem on Elimination of Unknowns, for solving
systems of linear equations are often called “Gaussian elimination procedures”.

(4) A matrix can be interpreted as a non-invertible linear mapping by the identification
(16.5) is often called a “singular matrix”. In the same way, many people would use
the term “non-singular” when we speak of an “invertible” matrix.

17 Dimension

Definition: We say that a linear space \( \mathcal{F} \) is finite-dimensional if it is
spanned by some finite subset. The least among the cardinal numbers of
finite spanning subsets of \( \mathcal{V} \) is called the dimension of \( \mathcal{V} \) and is denoted by
\( \text{dim} \mathcal{V} \).

The following fundamental result gives a much stronger characterization
of the dimension than is given by the definition.

Characterization of Dimension: Let \( \mathcal{V} \) be a finite-dimensional linear
space and \( \mathcal{F} := (f_i \mid i \in I) \) a family of elements in \( \mathcal{V} \).

(a) If \( \mathcal{F} \) is linearly independent then \( I \) is finite and \( \#I \leq \text{dim} \mathcal{V} \), with
equality if and only if \( \mathcal{F} \) is a basis.

(b) If \( \mathcal{F} \) is spanning, then \( \#I \geq \text{dim} \mathcal{V} \), with equality if and only if \( \mathcal{F} \) is a
basis.

The proof of this theorem can easily be obtained from the Characteriza-
tion of Bases (Sect. 15) and the following:

Lemma: If \( (b_j \mid j \in J) \) is a finite basis of \( \mathcal{V} \) and if \( (f_i \mid i \in I) \) is any linearly
independent family in \( \mathcal{V} \), then \( I \) must be finite and \( \#I \leq \#J \).
Proof: Let \( I' \) be any finite subset of \( I \). By Prop. 2 of Sect. 15, the restriction \( f|_{I'} \) is still linearly independent, which means that 
\[ \text{rc}(f|_{I'}) : \mathbb{F}^{I'} \to \mathcal{V} \text{ is injective.} \]

Since \( b \) is a basis, \( \text{rc}(b) \) is invertible and hence 
\[ \text{rc}^{-1}(\text{rc}(f|_{I'})) : \mathbb{F}^{I'} \to \mathbb{F}^{J} \text{ is also injective.} \]
By the Corollary of the Theorem on Elimination of Unknowns, it follows that \( \#I' \leq \#J \). Since \( I' \) was an arbitrary finite subset of \( I \), it follows that \( I \) itself must be finite and \( \#I \leq \#J \).

The Theorem has the following immediate consequences:

**Corollary 1:** If \( \mathcal{V} \) is a finite-dimensional linear space, then \( \mathcal{V} \) has bases and every basis of \( \mathcal{V} \) has \( \dim \mathcal{V} \) terms.

**Corollary 2:** Two finite-dimensional spaces \( \mathcal{V} \) and \( \mathcal{V}' \) are linearly isomorphic if and only if they have the same dimension, i.e., \( \text{Lis}(\mathcal{V}, \mathcal{V}') \neq \emptyset \) if and only if \( \dim \mathcal{V} = \dim \mathcal{V}' \).

Now let \( \mathcal{V} \) be a finite-dimensional linear space. The following facts are not hard to prove with the help of the Characterization of Dimension.

**Proposition 1:** If \( \mathcal{S} \) is a linearly independent subset of \( \mathcal{V} \), then \( \mathcal{S} \) must be finite and 
\[ \dim(\text{Lsp}(\mathcal{S})) = \#\mathcal{S}. \quad (17.1) \]

**Proposition 2:** Every subspace \( \mathcal{U} \) of \( \mathcal{V} \) is finite-dimensional and satisfies 
\[ \dim \mathcal{U} \leq \dim \mathcal{V}, \] with equality only if \( \mathcal{U} = \mathcal{V} \).

**Proposition 3:** Every subspace of \( \mathcal{V} \) has a supplement in \( \mathcal{V} \). In fact, if \( \mathcal{U}_1 \) is a subspace of \( \mathcal{V} \), then \( \mathcal{U}_2 \) is a supplement of \( \mathcal{U}_1 \) in \( \mathcal{V} \) if and only if \( \mathcal{U}_2 \) is a space of greatest dimension among those that are disjunct from \( \mathcal{U}_1 \).

**Proposition 4:** Two subspaces \( \mathcal{U}_1 \) and \( \mathcal{U}_2 \) of \( \mathcal{V} \) are disjunct if and only if 
\[ \dim \mathcal{U}_1 + \dim \mathcal{U}_2 = \dim(\mathcal{U}_1 + \mathcal{U}_2). \quad (17.2) \]

**Proposition 5:** Two subspaces \( \mathcal{U}_1 \) and \( \mathcal{U}_2 \) of \( \mathcal{V} \) are supplementary in \( \mathcal{V} \) if and only if two of the following three conditions are satisfied:

(i) \( \mathcal{U}_1 \cap \mathcal{U}_2 = \{0\} \),

(ii) \( \mathcal{U}_1 + \mathcal{U}_2 = \mathcal{V} \),

(iii) \( \dim \mathcal{U}_1 + \dim \mathcal{U}_2 = \dim \mathcal{V} \).
Proposition 6: For any two subspaces \( U_1, U_2 \) of \( V \), we have

\[
\dim U_1 + \dim U_2 = \dim(U_1 + U_2) + \dim(U_1 \cap U_2). \tag{17.3}
\]

The following Theorem is perhaps the single most useful fact about linear mappings between finite-dimensional spaces:

**Theorem on Dimensions of Range and Nullspace:** If \( L \) is a linear mapping whose domain is finite-dimensional, then \( \text{Rng} \ L \) is finite dimensional and

\[
\dim(\text{Null} \ L) + \dim(\text{Rng} \ L) = \dim(\text{Dom} \ L). \tag{17.4}
\]

**Proof:** Put \( V := \text{Dom} \ L \). By Prop. 3 we may choose a supplement \( U \) of Null \( L \) in \( V \). By Prop. 5 (iii), we have \( \dim(\text{Null} \ L) + \dim U = \dim V \). By Prop. 5 of Sect. 13, \( L|_{\text{Rng} \ L} \) is invertible and hence a linear isomorphism. Since linear isomorphisms obviously preserve dimension, we infer that \( \dim U = \dim(\text{Rng} \ L) \) and hence that (17.4) holds.

The following result is an immediate consequence of the Theorem just stated and of Prop. 4 of Sect. 13. Its name comes from its analogy to the Pigeonhole Principle stated in Sect. 05.

**Pigeonhole Principle for Linear Mappings:** Let \( L \) be a linear mapping with finite-dimensional domain and codomain. If \( L \) is injective, then \( \dim(\text{Dom} \ L) \leq \dim(\text{Cod} \ L) \). If \( L \) is surjective, then \( \dim(\text{Cod} \ L) \leq \dim(\text{Dom} \ L) \). If \( \dim(\text{Dom} \ L) = \dim(\text{Cod} \ L) \), then the following are equivalent:

(i) \( L \) is invertible,
(ii) \( L \) is surjective,
(iii) \( L \) is injective,
(iv) \( \text{Null} \ L = \{0\} \).

Let \( I \) be a finite index set. Since the standard basis \( \delta^I \) of \( \mathbb{F}^I \) (see Sect. 16) has \( \#I \) terms, it follows from Cor. 1 to the Characterization of Dimension that

\[
\dim(\mathbb{F}^I) = \#I. \tag{17.5}
\]

If \( I \) is replaced by a set product \( J \times I \) of finite sets \( J \) and \( I \), (17.5) yields
\[ \dim(\text{Lin}(F^I, F^J)) = \dim(F^I \times F^J) = (#I)(#J). \]  
(17.6)

**Proposition 7:** If \( V \) and \( W \) are finite-dimensional linear spaces, so is \( \text{Lin}(V, W) \), and

\[ \dim(\text{Lin}(V, W)) = (\dim V)(\dim W). \]  
(17.7)

**Proof:** In view of Cor. 1, we may choose a basis \( (b_i \mid i \in I) \) of \( V \) and a basis \( (c_j \mid j \in J) \) of \( W \), so that \( \dim V = #I \) and \( \dim W = #J \). The desired result follows from (17.6) and the fact that \( \text{Lin}(F^I, F^J) \) is linearly isomorphic to \( \text{Lin}(V, W) \) by virtue of (16.10). \( \blacksquare \)

Notes 17

(1) The dimensions of the nullspace and of the range of a linear mapping \( L \) are often called the “nullity” and “rank” of \( L \), respectively. I believe that this terminology burdens the memory unnecessarily.

18 Lineons

Let \( V \) be a linear space. A linear mapping from \( V \) to itself will be called a **lineon** on \( V \) and the space of all lineons on \( V \) will be denoted by

\[ \text{Lin} V := \text{Lin}(V, V). \]

The composite of two lineons on \( V \) is again a lineon on \( V \) (see Prop. 1 of Sect. 13). Composition in \( \text{Lin} V \) plays a role analogous to multiplication in \( F \). For this reason, the composite of two lineons is also called their **product**, and composition in \( \text{Lin} V \) is also referred to as **multiplication**. The lineon \( 0 \) is the analogue of the number 0 in \( F \) and the identity-lineon \( 1_V \) is the analogue of the number 1 in \( F \). It follows from Props. 1 and 3 of Sect. 14 that the multiplication and the addition in \( \text{Lin} V \) are related by distributive laws and hence that \( \text{Lin} V \) has the structure of a ring (see Sect. 06). Moreover, the composition-multiplication and scalar multiplication in \( \text{Lin} V \) are related by the associative laws

\[ (\xi L) M = \xi (LM) = L(\xi M), \]  
(18.1)

valid for all \( L, M \in \text{Lin} V, \xi \in F \).

Since \( \text{Lin} V \) has, in addition to its structure as a linear space, also a structure given by the composition-multiplication, we refer to \( \text{Lin} V \) as the
algebra of lineons on \( \mathcal{V} \). Most of the rules of elementary algebra are also valid in Lin \( \mathcal{V} \). The most notable exception is the commutative law of multiplication: In general, if \( L, M \in \text{Lin} \mathcal{V} \), then \( LM \) is not the same as \( ML \). If it happens that \( LM = ML \), we say that \( L \) and \( M \) commute.

A subspace of Lin \( \mathcal{V} \) that contains \( 1_\mathcal{V} \) and is stable under multiplication is called a subalgebra of Lin \( \mathcal{V} \). For example

\[
\mathbb{F}1_\mathcal{V} := \{ \xi 1_\mathcal{V} \mid \xi \in \mathbb{F} \}
\]

is a subalgebra of Lin \( \mathcal{V} \). A subalgebra that contains a given \( L \in \text{Lin} \mathcal{V} \) is the set

\[
\text{Comm } L := \{ M \in \text{Lin} \mathcal{V} \mid ML = LM \}
\]

of all lineons that commute with \( L \). It is called the commutant-algebra of \( L \).

The set of all automorphisms of \( \mathcal{V} \), i.e., all linear isomorphisms from \( \mathcal{V} \) to itself, is denoted by

\[
\text{Lis } \mathcal{V} := \text{Lis}(\mathcal{V}, \mathcal{V}).
\]

Lis \( \mathcal{V} \) is a subgroup of the group Perm \( \mathcal{V} \) of all permutations of \( \mathcal{V} \). We call Lis \( \mathcal{V} \) the lineon-group of \( \mathcal{V} \).

**Pitfall**: If \( \mathcal{V} \) is not a zero-space, then Lis \( \mathcal{V} \) does not contain the zero-lineon and hence cannot be a subspace of Lin \( \mathcal{V} \). If \( \dim \mathcal{V} > 1 \), then Lis \( \mathcal{V} \neq (\text{Lin} \mathcal{V})^\times \), i.e., there are non-zero lineons that are not invertible.

**Definition 1**: Let \( L \) be a lineon on the given linear space \( \mathcal{V} \). We say that a subspace \( \mathcal{U} \) of \( \mathcal{V} \) is an \( L \)-subspace of \( \mathcal{V} \), or simply an \( L \)-space, if it is \( L \)-invariant, i.e., if \( L \mathcal{U} \subset \mathcal{U} \) (see Sect. 03).

The zero-space \( \{0\} \) and \( \mathcal{V} \) itself are \( L \)-spaces for every \( L \in \text{Lin} \mathcal{V} \). Also, Null \( L \) and Rng \( L \) are easily seen to be \( L \)-spaces for every \( L \in \text{Lin} \mathcal{V} \). Every subspace of \( \mathcal{V} \) is an \( (\lambda 1_\mathcal{V}) \)-space for every \( \lambda \in \mathbb{F} \). If \( \mathcal{U} \) is an \( L \)-space, it is also an \( L^m \)-space for every \( m \in \mathbb{N} \), where \( L^m \), the \( m \)-th lineonic power of \( L \), is defined by \( L^m := L \circ L \circ \ldots \circ L \), the \( m \)-th iterate of \( L \).

If \( \mathcal{U} \) is an \( L \)-subspace, then the adjustments \( L_{\mathcal{U}} := L \big|_{\mathcal{U}} \) is linear and hence a lineon on \( \mathcal{U} \). We have \( (L_{\mathcal{U}})^m = (L^m)_{\mathcal{U}} \) for all \( m \in \mathbb{N} \).

Let \( \mathcal{V} \), \( \mathcal{V}' \) be linear spaces and let \( A : \mathcal{V} \to \mathcal{V}' \) be a linear isomorphism. The mapping

\[
(L \mapsto ALA^{-1}) : \text{Lin} \mathcal{V} \to \text{Lin} \mathcal{V}'
\]
is then an algebra-isomorphism. Also, we have

\[ \text{Rng}(A L A^{-1}) = A > \text{Rng } L \]  

and

\[ \text{Null } (A L A^{-1}) = A > \text{Null } L \]  

for all \( L \in \text{Lin } V \).

We assume now that \( V \) is finite-dimensional. By Prop. 7 of Sect. 17 we then have

\[ \dim \text{Lin } V = (\dim V)^2. \]  

(18.5)

The Linear Pigeonhole Principle of Sect. 17 has the following immediate consequences:

**Proposition 1:** For a lineon \( L \) on \( V \), the following are equivalent.

(i) \( L \) is invertible,

(ii) \( L \) is surjective,

(iii) \( L \) is injective,

(iv) \( \text{Null } L = \{ 0 \} \).

**Proposition 2:** Let \( L \in \text{Lin } V \) be given. Assume that \( L \) is left-invertible or right-invertible, i.e., that \( M L = 1_V \) or \( L M = 1_V \) for some \( M \in \text{Lin } V \). Then \( L \) is invertible, and \( M = L^{-1} \).

Let \( b := (b_i | i \in I) \) be a basis of \( V \). If \( L \in \text{Lin } V \), we call

\[ [L]_b := \text{ln}_{b^{-1}} L \text{ ln}_b \]  

(18.6)

the **matrix of \( L \) relative to the basis \( b \).** (This is the matrix of \( L \), as defined in Sect. 16, when the bases \( b \) and \( c \) coincide.) This matrix is an element of \( \text{Lin } \mathbb{F}^I \cong \mathbb{F}^{I \times I} \) and \( (L \mapsto [L]_b) : \text{Lin } V \rightarrow \text{Lin } \mathbb{F}^I \) is an algebra-isomorphism, i.e., a linear isomorphism that also preserves multiplication and the identity. By (16.11), the matrix \([L]_b\) is characterized by

\[ L b_i = \sum_{j \in I} ([L]_b)_{j,i} b_j, \quad \text{for all } i \in I. \]  

(18.7)

The matrix of the identity-lineon is the **unit matrix** given by
\[ 1_{V} \mid _{U} = 1_{F^{I}} = (\delta_{i,j} \mid (i, j) \in I \times I), \] 

where \( \delta_{i,j} \) is 0 or 1 depending on whether \( i \neq j \) or \( i = j \), respectively.

If \( I \) is a finite set, then the identification \( \text{Lin} \ F^{I} \cong F^{I \times I} \) shows that a lineon on \( F^{I} \) may be regarded as an \( I \times I \)-matrix in \( F \). The algebra \( \text{Lin} \ F^{I} \) is therefore also called a matrix-algebra.

Notes 18

(1) The commonly accepted term for a linear mapping of a linear space to itself is “linear transformation”, but “linear operator”, “operator”, or “tensor” are also used in some contexts. I have felt for many years that there is a crying need for a short term with no other meanings. About three years ago my colleague Victor Mizel proposed to me the use of the contraction “lineon” for “linear transformation”, and his wife Phyllis pointed out that this lends itself to the formation of the adjective “lineonic”, which turned out to be extremely useful.

I conjecture that the lack of a short term such as “lineon” is one of the reasons why so many mathematicians talk so often about matrices when they really mean, or should mean, lineons.

(2) Let \( n \in \mathbb{N} \) be given. The group \( \text{Lis} \ F^{n} \) of lineons of \( F^{n} \), is often called the “group of invertible \( n \times n \) matrices of \( F \)” or the “general linear group \( \text{Gl}(n, F) \)”. Sometimes, the notations \( \text{Gl} \text{(V)} \) is used for the lineon-group \( \text{Lis} \text{V} \).

(3) What we call the “commutant-algebra” of \( L \) is often called the “centralizer” of \( L \).

19 Projections, Idempotents

Definition 1: Let \( V \) be a linear space. A linear mapping \( P : V \rightarrow U \) to a given subspace \( U \) of \( V \) is called a projection if \( P \mid _{U} = 1_{U} \). A lineon \( E \in \text{Lin} \ V \) is said to be idempotent (and is called an idempotent) if \( E^{2} = E \).

Proposition 1: A lineon \( E \in \text{Lin} \ V \) is idempotent if and only if \( E \mid _{\text{Rng}E} \) is a projection, i.e., if and only if \( E \mid _{\text{Rng}E} \mathbb{C}_{V} = 1_{\text{Rng}E} \mathbb{C}_{V} \).

Proof: Put \( U := \text{Rng} E \).

Assume that \( E \) is idempotent and let \( u \in U \) be given. We may choose \( v \in V \) such that \( u = Ev \). Then

\[ Eu = E(Ev) = E^{2}v = Ev = u. \]

Since \( u \in U \) was arbitrary, it follows that \( E \mid _{U} = 1_{U} \mathbb{C}_{V} \).
Now assume that $E|_U = 1_{U \subset V}$ and let $v \in V$ be given. Then $Ev \in \text{Rng } E = U$ and hence
\[ E^2v = E(El) = 1_{U \subset V}(Ev) = Ev. \]
Since $v \in V$ was arbitrary, it follows that $E^2 = E$. ■

**Proposition 2:** A linear mapping $P : V \to U$ from a linear space $V$ to a subspace $U$ of $V$ is a projection if and only if it is surjective and $P|_V \in \text{Lin } V$ is idempotent.

**Proof:** Apply Prop. 1 to the case when $E := P|_V$. ■

**Proposition 3:** Let $E \in \text{Lin } V$ be given. Then the following are equivalent:

(i) $E$ is idempotent.

(ii) $1_V - E$ is idempotent.

(iii) $\text{Rng}(1_V - E) = \text{Null } E$.

(iv) $\text{Rng } E = \text{Null } (1_V - E)$.

**Proof:** We observe that
\[ \text{Null } L \subset \text{Rng}(1_V - L) \quad (19.1) \]
is valid for all $L \in \text{Lin } V$.

(i) $\Leftrightarrow$ (iii): We have $E = E^2$ if and only if $(E - E^2)v = E(1_V - E)v = 0$ for all $v \in V$, which is the case if and only if $\text{Rng}(1_V - E) \subset \text{Null } E$. In view of (19.1), this is equivalent to (iii).

(ii) $\Leftrightarrow$ (iv): This follows by applying (i) $\Leftrightarrow$ (iii) with $E$ replaced by $1_V - E$.

(i) $\Leftrightarrow$ (ii): This follows from the identity $(1_V - E)^2 = (1_V - E) + (E^2 - E)$, valid for all $E \in \text{Lin } V$. ■

The following proposition shows how projections and idempotents are associated with pairs of supplementary subspaces.

**Proposition 4:** Let $U_1, U_2$ be subspaces of the linear space $V$. Then the following are equivalent:

(i) $U_1$ and $U_2$ are supplementary.
There are projections $P_1$ and $P_2$ onto $U_1$ and $U_2$, respectively, such that

$$v = P_1v + P_2v \quad \text{for all } v \in V.$$  \hfill (19.2)

There is a projection $P_1 : V \to U_1$ such that $U_2 = \text{Null } P_1$. \hfill (ii) \hfill (iii)

There are idempotents $E_1, E_2 \in \text{Lin } V$ such that $U_1 = \text{Rng } E_1$, $U_2 = \text{Rng } E_2$ and

$$E_1 + E_2 = 1_V.$$  \hfill (19.3)

There is an idempotent $E_1 \in \text{Lin } V$ such that $U_1 = \text{Rng } E_1$ and $U_2 = \text{Null } E_1$. \hfill (iv) \hfill (v)

The projections $P_1$ and $P_2$ and the idempotents $E_1$ and $E_2$ are uniquely determined by the subspaces $U_1, U_2$.

**Proof:** (i) $\iff$ (ii): If $U_1$ and $U_2$ are supplementary, it follows from Prop. 4, (ii) of Sect. 12 that every $v \in V$ uniquely determines $u_1 \in U_1$ and $u_2 \in U_2$ such that $v = u_1 + u_2$. This means that there are mappings $P_1 : V \to U_1$ and $P_2 : V \to U_2$ such that (19.2) holds. It is easily seen that $P_1$ and $P_2$ are projections.

(ii) $\Rightarrow$ (iii): It is clear from (19.2) that $v \in \text{Null } P_1$, i.e., $P_1 v = 0$, holds if and only if $v = P_2v$, which is the case if and only if $v \in U_2$.

(iii) $\Rightarrow$ (v): This is an immediate consequence of Prop. 2, with $E_1 := P_1 |_{V}$.\hfill (v) $\Rightarrow$ (iv): Put $E_2 := 1_V - E_1$. Then $E_2$ is idempotent and $\text{Rng } E_2 = \text{Null } E_1 = U_2$ by Prop. 2.

(iv) $\Rightarrow$ (i): We observe that $\text{Null } L \cap \text{Null } (1_V - L) = \{0\}$ and $V = \text{Rng } L + \text{Rng } (1_V - L)$ hold for all $L \in \text{Lin } V$. Using this observation when $L := E_1$ and hence $E_2 = 1_V - L$ we conclude that $V = U_1 + U_2$ and, from Prop. 3, that $\{0\} = U_1 \cap U_2$. By Def. 2 of Sect. 12 this means that (i) holds.

The proof of uniqueness of $P_1, P_2, E_1, E_2$ is left to the reader. \hfill □

The next result, which is easily proved, shows how linear mappings with domain $V$ are determined by their restrictions to each of two supplementary subspaces of $V$.

**Proposition 5:** Let $U_1$ and $U_2$ be supplementary subspaces of $V$. For every linear space $V'$ and every $L_1 \in \text{Lin } (U_1, V')$, $L_2 \in \text{Lin } (U_2, V')$, there is exactly one $L \in \text{Lin } (V, V')$ such that
It is given by

\[ \mathbf{L} := \mathbf{L}_1\mathbf{P}_1 + \mathbf{L}_2\mathbf{P}_2, \]

where \( \mathbf{P}_1 \) and \( \mathbf{P}_2 \) are the projections of Prop. 4, (ii).

Notes 19

(1) Some textbooks use the term “projection” in this context as a synonym for “idempotent”. Although the two differ only in the choice of codomain, I believe that the distinction is useful.

110 Problems for Chapter 1

1. Let \( \mathcal{V} \) and \( \mathcal{V}' \) be linear spaces. Show that a given mapping \( \mathbf{L} : \mathcal{V} \to \mathcal{V}' \) is linear if and only if its graph \( \text{Gr}(\mathbf{L}) \) (defined by (03.1)) is a subspace of \( \mathcal{V} \times \mathcal{V}' \).

2. Let \( \mathcal{V} \) be a linear space. For each \( \mathbf{L} \in \text{Lin}\mathcal{V} \), define the left-multiplication mapping

\[ \mathbf{L}_e\mathbf{L} : \text{Lin}\mathcal{V} \to \text{Lin}\mathcal{V} \]

by

\[ \mathbf{L}_e\mathbf{L}(\mathbf{M}) := \mathbf{L}\mathbf{M} \quad \text{for all } \mathbf{M} \in \text{Lin} \mathcal{V} \quad \text{(P1.1)} \]

(a) Show that \( \mathbf{L}_e\mathbf{L} \) is linear for all \( \mathbf{L} \in \text{Lin} \mathcal{V} \).

(b) Show that \( \mathbf{L}_e\mathbf{L}\mathbf{K} = \mathbf{L}_e\mathbf{K} \) for all \( \mathbf{L}, \mathbf{K} \in \text{Lin} \mathcal{V} \).

(c) Show that \( \mathbf{L}_e\mathbf{L} \) is invertible if and only if \( \mathbf{L} \) is invertible and that \( (\mathbf{L}_e\mathbf{L}^{-1}) = \mathbf{L}_e\mathbf{L}^{-1} \) if this is the case.

3. Consider

\[ \mathbf{L} := \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{2 \times 2} \cong \text{Lin}\mathbb{R}^2. \]

(a) Show that \( \mathbf{L} \) is invertible and find its inverse.
(b) Determine the matrix of \( L_{\mathbf{e}_L} \), as defined by (P1.1), relative to the list-basis

\[
B := \left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)
\]

of \( \text{Lin} \mathbb{R}^2 \).

(c) Determine the inverse of the matrix you found in (b).

4. Let \( C^\infty(\mathbb{R}) \) be the set of all functions \( f \in \text{Map}(\mathbb{R}, \mathbb{R}) \) such that \( f \) is \( k \)-times differentiable for all \( k \in \mathbb{N} \) (see Sect. 08). It is clear that \( C^\infty \mathbb{R} \) is a subspace of the linear space \( \text{Map} (\mathbb{R}, \mathbb{R}) \) (see Sect. 13). Define \( D \in \text{Lin} C^\infty(\mathbb{R}) \) by

\[
Df := \partial f \quad \text{for all} \quad f \in C^\infty(\mathbb{R})
\]

(see (08.31)) and, for each \( n \in \mathbb{N} \), the subspace \( \mathcal{P}_n \) of \( C^\infty(\mathbb{R}) \) by

\[
\mathcal{P}_n := \text{Null}D^n.
\]

(a) Show that the sequence

\[
p := (i^k \mid k \in \mathbb{N})
\]

is linearly independent (see (08.26) for notation). Is it a basis of \( C^\infty(\mathbb{R}) \)?

(b) Show that, for each \( n \in \mathbb{N} \),

\[
p \mid_{n!} := (i^k \mid k \in n!)
\]

is a basis of \( \mathcal{P}_n \) and hence that \( \dim \mathcal{P}_n = n \).

5. Let \( C^\infty(\mathbb{R}) \) and \( D \in \text{Lin} C^\infty(\mathbb{R}) \) by defined as in Problem 4. Define \( M \in \text{Lin} C^\infty(\mathbb{R}) \) by \( Mf := i f \) for all \( f \in C^\infty(\mathbb{R}) \) (see (08.26)). Prove that

\[
DM - MD = 1_{C^\infty(\mathbb{R})}.
\]

6. Let \( C^\infty(\mathbb{R}) \) be defined as in Problem 4 and define \( S \in \text{Lin} C^\infty(\mathbb{R}) \) by

\[
Sf := f \circ (i + 1) \quad \text{for all} \quad f \in C^\infty(\mathbb{R})
\]

(P1.7)

(a) Show that, for each \( n \in \mathbb{N} \), the subspace \( \mathcal{P}_n \) of \( C^\infty(\mathbb{R}) \) defined by (P1.3) is an \( S \)-space (see Sect. 18).

(b) Find the matrix of \( S \mid_{\mathcal{P}_n} \in \text{Lin} \mathcal{P}_n \) relative to the basis \( p \mid_{n!} \) of \( \mathcal{P}_n \) given by (P1.5).

7. Let \( \mathcal{V} \) be a finite-dimensional linear space and let \( \mathbf{L} \in \text{Lin} \mathcal{V} \) be given.
(i) \( \text{Rng } L \cap \text{Null } L = \{0\} \),
(ii) \( \text{Null } (L^2) \subset \text{Null } L \),
(iii) \( \text{Rng } L + \text{Null } L = \mathcal{V} \).

8. Consider the subspaces \( U_1 := \mathbb{R}(1, 1) \) and \( U_2 := \mathbb{R}(1, -1) \) of \( \mathbb{R}^2 \).

(a) Show that \( U_1 \) and \( U_2 \) are supplementary in \( \mathbb{R}^2 \).
(b) Find the idempotent matrices \( E_1, E_2 \in \mathbb{R}^{2 \times 2} \) which are determined by \( U_1, U_2 \) according to Prop. 4 of Sect. 19.
(c) Determine a basis \( b = (b_1, b_2) \) of \( \mathbb{R}^2 \) such that the matrices of \( E_1 \) and \( E_2 \) relative to \( b \) are
\[
\begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
0 & 0 \\
0 & 1
\end{bmatrix},
\] respectively.

9. Let a linear space \( \mathcal{V} \), a lineon \( L \) on \( \mathcal{V} \), and \( m \in \mathbb{N} \) be given.

(a) Show that \( \text{Rng}(L^m) \) is an \( L \)-subspace of \( \mathcal{V} \).
(b) Show that \( \text{Rng}(L^{m+1}) \subset \text{Rng}(L^m) \), with equality if and only if
the adjustment \( L|_{\text{Rng}(L^m)} \) is surjective.

10. Let \( \mathcal{V} \) be a finite-dimensional linear space and let \( L \) be a nilpotent lineon on \( \mathcal{V} \) with nilpotency \( m \), which means that \( L^m = 0 \) but \( L^k \neq 0 \) for all \( k \in m \) (see also Sect. 93). Prove:

(a) If \( m > 0 \), then Null \( L \neq \{0\} \).
(b) The only \( L \)-subspace \( \mathcal{U} \) for which the adjustment \( L|_{\mathcal{U}} \) is invertible
is the zero-space.
(c) The nilpotency cannot exceed \( \dim \mathcal{V} \). (Hint: Use Problem 9.)

11. Let a linear space \( \mathcal{V} \) and a lineon \( J \in \text{Lin } \mathcal{V} \) be given.

(a) Show that \( \text{Rng}(L^m) \) is an \( L \)-subspace of \( \mathcal{V} \).
(b) Show that \( \text{Rng}(J) \) is an \( L \)-subspace of \( \mathcal{V} \).
(c) The nilpotency cannot exceed \( \dim \mathcal{V} \). (Hint: Use Problem 9.)

12. Let a linear space \( \mathcal{V} \) and a lineon \( J \) on \( \mathcal{V} \) satisfying \( J^2 = -1_\mathcal{V} \) be given. Define \( E \in \text{Lin}(\text{Lin } \mathcal{V}) \) by
\[
E_L := \frac{1}{2}(L + JLJ)
\] for all \( L \in \text{Lin } \mathcal{V} \).
(a) Show that $E$ is idempotent.

(b) Show that $\text{Null } E = \text{Comm } J$, the commutant-algebra of $J$ (see (18.2)).

(c) Show that

$$\text{Rng } E = \{ L \in \text{Lin} V \mid LJ = -JL \}, \quad (P1.9)$$

the space of all lineons that “anticommute” with $J$.

(d) Prove: If $\dim V > 0$ and if $\{ \xi \in F \mid \xi^2 + 1 = 0 \} = \emptyset$, then the space ($P1.9$) is not the zero-space. (Hint: Use (b) of Problem 11 and Prop. 4 of Sect. 19.)
Chapter 2

Duality, Bilinearity

In this chapter, the phrase “let . . . be a linear space” will be used as a shorthand for “let . . . be a finite-dimensional linear space over \( \mathbb{R} \).” (Actually, many definitions remain meaningful and many results remain valid when the given spaces are infinite-dimensional or when \( \mathbb{R} \) is replaced by an arbitrary field. The interested reader will be able to decide for himself when this is the case.)

21 Dual Spaces, Transposition, Annihilators

Let \( V \) be a linear space. We write

\[ V^* := \text{Lin}(V, \mathbb{R}) \]

and call the linear space \( V^* \) the dual space of the space \( V \). The elements of \( V^* \) are often called linear forms or covectors, depending on context. It is evident from Prop.7 of Sect. 17 that

\[ \dim V^* = \dim V. \quad (21.1) \]

Let \( V \) and \( W \) be linear spaces and let \( L \in \text{Lin}(V, W) \) be given. It follows from Props.1 and 2 of Sect. 14 that the mapping \( \langle \mu \mapsto \mu L \rangle : W^* \rightarrow V^* \) is linear. We call this mapping the transpose of \( L \) and denote it by \( L^\top \), so that

\[ L^\top \in \text{Lin}(W^*, V^*) \quad \text{if} \quad L \in \text{Lin}(V, W) \quad (21.2) \]

and

\[ L^\top \mu = \mu L \quad \text{for all} \quad \mu \in W^*. \quad (21.3) \]
It is an immediate consequence of Prop. 3 of Sect. 14 that the mapping
\[(L \mapsto L^\top) : \text{Lin}(V, W) \to \text{Lin}(W^*, V^*)\] (21.4)
is linear. This mapping is called the **transposition** on $\text{Lin}(V, W)$. The following rules follow directly from (21.3):

**Proposition 1:** For every linear space $V$, we have
\[(1_V)^\top = 1_{V^*}.\] (21.5)

Let $V, V', V''$ be linear spaces. For all $L \in \text{Lin}(V, V')$ and all $M \in \text{Lin}(V', V'')$, we have
\[(LM)^\top = L^\top M^\top.\] (21.6)

If $L \in \text{Lin}(V, V')$ is invertible, so is $L^\top \in \text{Lin}(V', V^*)$, and
\[(L^\top)^{-1} = (L^{-1})^\top.\] (21.7)

**Definition:** Let $V$ be a linear space and let $S$ be a subset of $V$. We say that a linear form $\lambda \in V^*$ **annihilates** $S$ if $\lambda(S) \subset \{0\}$, or, equivalently, $\lambda|_S = 0$, or, equivalently, $S \subset \text{Null } \lambda$. The set of all linear forms that annihilate $S$ is called the **annihilator** of $S$ and is denoted by
\[S^\perp := \{\lambda \in V^* | \lambda(S) \subset \{0\}\}.\] (21.8)

The following facts are immediate consequences of the definition.

**Proposition 2:** $\emptyset^\perp = \{0\}^\perp = V^*$ and $V^\perp = \{0\}$. If $S_1$ and $S_2$ are subsets of $V$ then
\[S_1 \subset S_2 \implies S_2^\perp \subset S_1^\perp.\]

**Proposition 3:** For every subset $S$ of $V$, $S^\perp$ is a subspace of $V^*$ and $(\text{Lsp } S)^\perp = S^\perp$.

**Proposition 4:** If $(S_i | i \in I)$ is a family of subsets of $V$, then
\[
(\bigcup_{i \in I} S_i)^\perp = \bigcap_{i \in I} S_i^\perp.
\] (21.9)

Combining Prop.3 and Prop.4, and using Prop.2 of Sect. 12, we obtain the following:

**Proposition 5:** If $(U_i | i \in I)$ is a family of subspaces of $V$, then
\[
(\bigcap_{i \in I} U_i)^\perp = \bigcup_{i \in I} U_i^\perp.
\] (21.10)
In particular, if $U_1, U_2$ are subspaces of $V$, then

$$(U_1 + U_2)\perp = U_1\perp \cap U_2\perp.$$  \hfill (21.11)

The following result relates the annihilator of a subspace to the annihilator of the image of this subspace under a linear mapping.

**Theorem on Annihilators and Transposes:** Let $V, W$ be linear spaces and let $L \in \text{Lin}(V, W)$ be given. For every subspace $U$ of $V$, we then have

$$(L^\top(U))\perp = (L^\top)^{<}(U\perp).$$  \hfill (21.12)

In particular, we have

$$\text{Rng }L\perp = \text{Null } L^\top.$$  \hfill (21.13)

**Proof:** Let $\mu \in W^*$ be given. Then, by (21.8) and (21.3),

$$\mu \in (L^\top(U))\perp \iff \{0\} = \mu^\circ(L^\top(U)) = (\mu L)^\circ(U) = (L^\top \mu)^\circ(U) \iff L^\top \mu \in U\perp \iff \mu \in (L^\top)^{<}(U\perp).$$

Since $\mu \in W^*$ was arbitrary, (21.12) follows. Putting $U := V$ in (21.12) yields (21.13). \hfill $\blacksquare$

The following result states, among other things, that every linear form on a subspace of $V$ can be extended to a linear form on all of $V$.

**Proposition 6:** Let $U$ be a subspace of $V$. The mapping

$$(\lambda \mapsto \lambda|_U) : V^* \to U^*$$  \hfill (21.14)

is linear and surjective, and its nullspace is $U\perp$.

**Proof:** It is evident that the mapping (21.14) is linear and that its nullspace is $U\perp$. By Prop.3 of Sect. 17, we may choose a supplement $U'$ of $U$ in $V$. Now let $\mu \in U^*$ be given. By Prop.5 of Sect. 19 there is a $\lambda \in V^*$ such that $\lambda|_U = \mu$ (and $\lambda|_{U'} = 0$, say). Since $\mu \in U^*$ was arbitrary, it follows that the mapping (21.14) is surjective. \hfill $\blacksquare$

Using (21.1) and the Theorem on Dimensions of Range and Nullspace (Sect. 17), we see that Prop.6 has the following consequence:

**Formula for Dimension of Annihilators:** For every subspace $U$ of a given linear space $V$ we have

$$\dim U\perp = \dim V - \dim U.$$  \hfill (21.15)
(1) The notations $\mathcal{V}'$ or $\tilde{\mathcal{V}}$ are sometimes used for the dual $\mathcal{V}^*$ of a linear space $\mathcal{V}$.

(2) Some people use the term “linear functional” instead of “linear form”. I prefer to reserve “linear functional” for the case when the domain is infinite-dimensional.

(3) The terms “adjoint” or “dual” are often used in place of the “transpose” of a linear mapping $L$. Other notations for our $L^\top$ are $L^*, L^t, \text{t} L$, and $\tilde{L}$.

(4) The notation $S^0$ instead of $S^\perp$ is sometimes used for the annihilator of the set $S$.

22 The Second Dual Space

In view of (21.1) and Corollary 2 of the Characterization of Dimension (Sect. 17), it follows that there exist linear isomorphisms from a given linear space $\mathcal{V}$ to its dual $\mathcal{V}^*$. However, if no structure on $\mathcal{V}$ other than its structure as a linear space is given, none of these isomorphisms is natural (see the Remark at the end of Sect. 23). The specification of any one such isomorphism requires some capricious choice, such as the choice of a basis. By contrast, one can associate with each linear space $\mathcal{V}$ a natural isomorphism from $\mathcal{V}$ to its second dual, i.e. to the dual $\mathcal{V}^{**}$ of the dual $\mathcal{V}^*$ of $\mathcal{V}$. This isomorphism is an evaluation mapping as described in Sect. 04.

**Proposition 1:** Let $\mathcal{V}$ be a linear space. For each $v \in \mathcal{V}$, the evaluation $ev(v) : \mathcal{V}^* \to \mathbb{R}$, defined by

$$ev(v)(\lambda) := \lambda v \quad \text{for all} \quad \lambda \in \mathcal{V}^*, \quad (22.1)$$

is linear and hence a member of $\mathcal{V}^{**}$. The evaluation mapping $ev : \mathcal{V} \to \mathcal{V}^{**}$ defined in this way is a linear isomorphism.

**Proof:** The linearity of $ev(v) : \mathcal{V}^* \to \mathbb{R}$ merely reflects the fact that the linear-space operations in $\mathcal{V}^*$ are defined by value-wise applications of the operations in $\mathbb{R}$. The linearity of $ev : \mathcal{V} \to \mathcal{V}^{**}$ follows from the fact that each $\lambda \in \mathcal{V}^*$ is linear.

Put $\mathcal{U} := \text{Null} \ (ev)$ and let $v \in \mathcal{V}$ be given. By (22.1), we have

$$v \in \mathcal{U} \iff (\lambda v = 0 \quad \text{for all} \quad \lambda \in \mathcal{V}^*),$$

which means, by the definition of annihilator (see Sect. 21), that $\mathcal{V}^*$ coincides with the annihilator $\mathcal{U}^\perp$ of $\mathcal{U}$. Since $\dim \mathcal{V}^* = \dim \mathcal{V}$, we conclude from the Formula (21.15) for Dimension of Annihilators that $\dim \mathcal{U} = 0$ and hence $\text{Null} \ (ev) = \mathcal{U} = \{0\}$. Since $\dim \mathcal{V}^{**} = \dim \mathcal{V}$, we can use the Pigeonhole
22. THE SECOND DUAL SPACE

Principle for Linear Mappings (Sect. 17) to conclude that \( ev \) is invertible.

We use the natural isomorphism described by Prop.1 to identify \( \mathcal{V}^{**} \) with \( \mathcal{V} \):

\[
\mathcal{V}^{**} \cong \mathcal{V},
\]

and hence we use the same symbol for an element of \( \mathcal{V} \) and the corresponding element of \( \mathcal{V}^{**} \). Thus, (22.1) reduces to

\[
v\lambda = \lambda v \quad \text{for all} \quad \lambda \in \mathcal{V}^{*}, v \in \mathcal{V}, \tag{22.2}
\]

where the \( v \) on the left side is interpreted as an element of \( \mathcal{V}^{**} \).

The identification \( \mathcal{V}^{**} \cong \mathcal{V} \) induces identifications such as

\[
\text{Lin}(\mathcal{V}^{**}, \mathcal{W}^{**}) \cong \text{Lin}(\mathcal{V}, \mathcal{W})
\]

when \( \mathcal{V} \) and \( \mathcal{W} \) are given linear spaces. In particular, if \( H \in \text{Lin}(\mathcal{W}^{*}, \mathcal{V}^{*}) \), we will interpret \( H^\top \) as an element of \( \text{Lin}(\mathcal{V}, \mathcal{W}) \). In view of (22.2) and (21.3), \( H^\top \) is characterized by

\[
\mu H^\top v = (H\mu)v \quad \text{for all} \quad v \in \mathcal{V}, \mu \in \mathcal{W}^{*}. \tag{22.3}
\]

Using (21.3) and (22.3), we immediately obtain the following:

**Proposition 2:** Let \( \mathcal{V}, \mathcal{W} \) be linear spaces. Then the transposition \( (H \mapsto H^\top) : \text{Lin}(\mathcal{W}^{*}, \mathcal{V}^{*}) \to \text{Lin}(\mathcal{V}, \mathcal{W}) \) is the inverse of the transposition \( (L \mapsto L^\top) : \text{Lin}(\mathcal{V}, \mathcal{W}) \to \text{Lin}(\mathcal{W}^{*}, \mathcal{V}^{*}), \) so that

\[
(L^\top)^\top = L \quad \text{for all} \quad L \in \text{Lin}(\mathcal{V}, \mathcal{W}) \tag{22.4}
\]

The identification \( \mathcal{V}^{**} \cong \mathcal{V} \) also permits us to interpret the annihilator \( \mathcal{H}^\perp \) of a subset \( \mathcal{H} \) of \( \mathcal{V}^{*} \) as a subset of \( \mathcal{V} \).

**Proposition 3:** For every subspace \( \mathcal{U} \) of a given linear space \( \mathcal{V} \), we have

\[
(\mathcal{U}^{\perp})^\perp = \mathcal{U} \tag{22.5}
\]

**Proof:** Let \( u \in \mathcal{U} \) be given. By definition of \( \mathcal{U}^{\perp} \), we have \( \lambda u = 0 \) for all \( \lambda \in \mathcal{U}^{\perp} \). If we interpret \( u \) as an element of \( \mathcal{V}^{**} \) and use (22.2), this means that \( u\lambda = 0 \) for all \( \lambda \in \mathcal{U}^{\perp} \) and hence \( u_{\mathcal{U}^{\perp}}(\mathcal{U}^{\perp}) = \{0\} \). Therefore, we have \( u \in (\mathcal{U}^{\perp})^\perp \). Since \( u \in \mathcal{U} \) was arbitrary, it follows that \( \mathcal{U} \subset (\mathcal{U}^{\perp})^\perp \).

Applying the Formula (21.15) for Dimension of Annihilators to \( \mathcal{U} \) and \( \mathcal{U}^{\perp} \) and recalling \( \dim \mathcal{V} = \dim \mathcal{V}^{*} \), we find \( \dim \mathcal{U} = \dim (\mathcal{U}^{\perp})^\perp \). By Prop.2 of Sect.17, this is possible only when (22.5) holds.
Using (22.5) and (22.4) one obtains the following consequences of Props.3, 2, and 5 of Sect.21.

**Proposition 4:** For every subset $S$ of $V$, we have

$$\text{Lsp} S = (S^\perp)^\perp.$$  \hspace{1cm} (22.6)

If $U_1$ and $U_2$ are subspaces of $V$, then

$$U_1^\perp \subset U_2^\perp \implies U_2 \subset U_1$$

and

$$U_1^\perp + U_2^\perp = (U_1 \cap U_2)^\perp.$$  \hspace{1cm} (22.7)

Using (22.5) and (22.4) one also obtains the following consequence of the Theorem on Annihilators and Transposes (Sect.21).

**Proposition 5:** Let $V$ and $W$ be linear spaces and let $L \in \text{Lin}(V, W)$ be given. For every subspace $H$ of $W^*$, we then have

$$L^\top(H) = (L^<(H^\perp))^\perp.$$  \hspace{1cm} (22.8)

In particular, we have

$$\text{Rng} L^\top = (\text{Null} L)^\perp.$$  \hspace{1cm} (22.9)

**Notes 22**

(1) Our notation $\lambda v$ for the value of $\lambda \in V^* := \text{Lin}(V, R)$ at $v \in V$, as in (22.1), is in accord with the general notation for the values of linear mappings (see Sect.13). Very often, a more complicated notation, such as $\langle v, \lambda \rangle$ or $[v, \lambda]$, is used. I disagree with the claim of one author that this complication clarifies matters later on; I believe that it obscures them.

## 23 Dual Bases

We now consider the space $R^I$ of all families of real numbers indexed on a given finite set $I$. Let $\varepsilon := (ev_i \mid i \in I)$ be the evaluation family associated with $R^I$ (see Sect.04). As we already remarked in Sect.14, the evaluations $ev_i : R^I \to R$, $i \in I$, are linear, i.e. they are members of $(R^I)^*$. Thus, $\varepsilon$ is a family in $(R^I)^*$. The linear combination mapping $\text{ln}_{\varepsilon} : R^I \to (R^I)^*$ is defined by

$$\text{ln}_{\varepsilon} \lambda := \sum_{i \in I} \lambda_i ev_i \quad \text{for all} \quad \lambda \in R^I$$  \hspace{1cm} (23.1)
It follows from (23.1) that
\[
(\text{Inc}_\varepsilon \lambda)\mu = \sum_{i \in I} \lambda_i \text{ev}_i(\mu) = \sum_{i \in I} \lambda_i \mu_i
\] (23.2)
for all $\lambda, \mu \in \mathbb{R}^I$.

Let $\delta^I := (\delta^I_k | k \in I)$ be the standard basis of $\mathbb{R}^I$ (see Sect.16). Writing (23.2) with the choice $\mu := \delta^I_k$, $k \in I$, we obtain
\[
(\text{Inc}_\varepsilon \lambda)\delta^I_k = \lambda_k = \text{ev}_k \lambda
\] (23.3)
for all $\lambda \in \mathbb{R}^I$ and all $k \in I$. Given $\alpha \in (\mathbb{R}^I)^*$, it easily follows from (23.3) that $\lambda := (\alpha \delta^I_i | i \in I)$ is the unique solution of the equation
\[
?\lambda \in \mathbb{R}^I, \quad \text{Inc}_\varepsilon \lambda = \alpha.
\] Since $\alpha \in (\mathbb{R}^I)^*$ was arbitrary, we can conclude that Inc$_\varepsilon$ is invertible and hence a linear isomorphism.

It is evident from (23.2) that
\[
(\text{Inc}_\varepsilon \lambda)\mu = (\text{Inc}_\varepsilon \mu)\lambda \quad \text{for all} \quad \lambda, \mu \in \mathbb{R}^I.
\] (23.4)

Comparing this result with (22.3) and using the identification $(\mathbb{R}^I)^{**} \cong \mathbb{R}^I$, we obtain Inc$_\varepsilon^T$ = Inc$_\varepsilon$, where Inc$_\varepsilon^T \in \text{Lin}((\mathbb{R}^I)^{**}, (\mathbb{R}^I)^*)$ is identified with the corresponding element of $\text{Lin}(\mathbb{R}^I, (\mathbb{R}^I)^*)$. In other words, the isomorphism $(\text{Inc}_\varepsilon^T)^{-1}\text{Inc}_\varepsilon : \mathbb{R}^I \to (\mathbb{R}^I)^{**}$ coincides with the identification $\mathbb{R}^I \cong (\mathbb{R}^I)^{**}$ obtained from Prop.1 of Sect.22. Therefore, there is no conflict if we use Inc$_\varepsilon$ to identify $(\mathbb{R}^I)^*$ with $\mathbb{R}^I$.

From now on we shall use Inc$_\varepsilon$ to identify $(\mathbb{R}^I)^*$ with $\mathbb{R}^I$:
\[
(\mathbb{R}^I)^* \cong \mathbb{R}^I,
\]
except that, given $\lambda \in \mathbb{R}^I$, we shall write $\lambda := \text{Inc}_\varepsilon \lambda$ rather than merely $\lambda$ for the corresponding element in $(\mathbb{R}^I)^*$. Thus, (23.2) and (23.4) reduce to
\[
\mu \cdot \lambda = \lambda \cdot \mu = \sum_{i \in I} \lambda_i \mu_i \quad \text{for all} \quad \lambda, \mu \in \mathbb{R}^I.
\] (23.5)

The equations (23.5) and (23.3) yield
\[
\delta^I_k \cdot \lambda = \lambda \cdot \delta^I_k = \lambda_k = \text{ev}_k \lambda
\] (23.6)
for all $\lambda \in \mathbb{R}^I$ and all $k \in I$. It follows that ev$_k = \delta^I_k \cdot$ for all $k \in I$, i.e. that the standard basis $\delta^I$ becomes identified with the evaluation family $\varepsilon$. Since
\[
\text{ev}_k(\delta^I_i) = \delta_{k,i} \quad \text{for all} \quad i, k \in I
\]
we see that
\[
\delta^I_k \cdot \delta^I_i = \delta_{k,i} := \begin{cases} 
1 & \text{if } k = i \\
0 & \text{if } k \neq i
\end{cases}
\]  
(23.7)
holds for all \(i, k \in I\).

The following result is an easy consequence of (22.3), (23.5), and (16.2). It shows that the transpose of a matrix as defined in Sect.02 is the same as the transpose of the linear mapping identified with this matrix, and hence that there is no notational clash.

**Proposition 1:** Let \(I\) and \(J\) be finite index sets and let \(M \in \text{Lin}(\mathbb{R}^I, \mathbb{R}^J) \cong \mathbb{R}^{I \times J}\) be given. Then
\[M^\top \in \text{Lin}((\mathbb{R}^J)^*, (\mathbb{R}^I)^*) \cong \text{Lin}(\mathbb{R}^J, \mathbb{R}^I) \cong \mathbb{R}^{I \times J}\]

satisfies
\[(M^\top \mu) \cdot \lambda = \mu \cdot M \lambda \text{ for all } \mu \in \mathbb{R}^J, \lambda \in \mathbb{R}^I\]
(23.8)
and
\[(M^\top)_{i,j} = M_{j,i} \text{ for all } i \in I, j \in J.\]
(23.9)

Let \(V\) be a linear space, let \(b := (b_i | i \in I)\) be a basis of \(V\), and let \(\text{ln}c_b : \mathbb{R}^I \to V\) be the (invertible) linear combination mapping for \(b\) (see Sect.15). Using the identification \((\mathbb{R}^I)^* \cong \mathbb{R}^I\), we can regard \((\text{ln}c^{-1}_b)^\top\) as a mapping from \(\mathbb{R}^I\) to \(V^*\). Using the standard basis \(\delta^I\) of \(\mathbb{R}^I\), we define
\[b^*_i := (\text{ln}c^{-1}_b)^\top \delta^I_i \text{ for all } i \in I\]
(23.10)
and call the family \(b^* := (b^*_i | i \in I)\) in \(V^*\) the dual of the given basis \(b\). Since \((\text{ln}c^{-1}_b)^\top\) is invertible, it follows from Prop.2 of Sect.16 that the dual \(b^*\) is a basis of \(V^*\).

Using (23.10) and (21.7), we find that
\[
\text{ln}c^{-1}_b b^*_i = \delta^I_i = \text{ln}c^{-1}_b b_i \text{ for all } i \in I.
\]
(23.11)

The dual basis \(b^*\) can be used to evaluate the family of components of a given \(v \in V\) relative to the basis \(b\):

**Proposition 2:** Let \(b := (b_i | i \in I)\) be a basis of \(V\) and let \(b^*\) be its dual. For every \(v \in V\), we then have
\[(\text{ln}c^{-1}_b v)_i = b^*_i v \text{ for all } i \in I\]
(23.12)
and hence
\[v = \sum_{i \in I} (b^*_i v) b_i.\]
(23.13)
23. **DUAL BASES**

Proof: It follows from (23.10), (21.3) and (23.6) that

$$b^*_iv = ((\ln_c e^{-1}v) \cdot \delta^I_k) = (\ln_c e^{-1}v)_i$$

for all $i \in I$.

Using Prop. 2 and the formula (16.11) one easily obtains the following:

**Proposition 3:** Let $\mathcal{V}$ and $\mathcal{W}$ be linear spaces and let $b := (b_i \mid i \in I)$ and $c := (c_j \mid j \in J)$ be bases of $\mathcal{V}$ and $\mathcal{W}$, respectively. Then the matrix $M \in \mathbb{R}^{J \times I}$ of a given $L \in \text{Lin}(\mathcal{V}, \mathcal{W})$ relative to $b$ and $c$ can be obtained by the formula

$$M_{j,i} = c^*_j Lb_i$$

for all $i \in I, j \in J$. (23.14)

The following result gives the most useful characterization of the dual basis.

**Proposition 4:** Let $b := (b_i \mid i \in I)$ be a basis of $\mathcal{V}$ and let $\beta := (\beta_i \mid i \in I)$ be a family in $\mathcal{V}^*$. Then

$$\beta_k b_i = \delta_{k,i}$$

holds for all $i, k \in I$ if and only if $\beta$ coincides with the dual $b^*$ of $b$.

Proof: In view of (23.7) and (21.7), the relation (23.15) is valid if and only if

$$\beta_k b_i = \delta^I_k \cdot \delta^I_i = \delta^I_k \cdot \ln_c^{-1}(b_i) = ((\ln_c^{-1} \cdot \delta^I_k))b_i.$$

Therefore, since $b$ is a basis, it follows from the uniqueness assertion of Prop. 2 of Sect. 16 that (23.15) holds for all $i \in I$ if and only if $\beta_k = (\ln_c^{-1} \cdot \delta^I_k)$. The assertion now follows from the definition (23.10) of the dual basis.

The following result furnishes a useful criterion for linear independence.

**Proposition 5:** Let $f := (f_j \mid j \in J)$ be a family in $\mathcal{V}$ and $\varphi := (\varphi_j \mid j \in J)$ a family in $\mathcal{V}^*$ such that

$$\varphi_k f_j = \delta_{k,j}$$

for all $j, k \in J$. (23.16)

Then $f$ and $\varphi$ are both linearly independent.

Proof: By the definition (15.1) of $\ln_c$ it follows from (23.16) that

$$\varphi_k (\ln_c \lambda) = \sum_{j \in J} \lambda_j (\varphi_k f_j) = \lambda_k$$

for all $\lambda \in \mathbb{R}^J$ and all $k \in J$. Hence we can have $\lambda \in \text{Null } \ln_c$ i.e. $\ln_c \lambda = 0$, only if $\lambda_k = 0$ for all $k \in J$, i.e. only if $\lambda = 0$. It follows that $\text{Null } \ln_c = \{0\}$, which implies the linear independence of $f$ by Prop. 1 of
Sect.15. The linear independence of $\varphi$ follows by using the identification $V^{**} \cong V$ and by interchanging the roles of $f$ and $\varphi$.

If we apply Prop.4 to the case when $V$ is replaced by $V^*$ and $b$ by $b^*$ and use (22.2), we obtain:

**Proposition 6:** The dual $b^{**}$ of the dual $b^*$ of a basis $b$ of $V$ is identified with $b$ itself by the identification $V^{**} \cong V$, i.e. we have

$$b^{**} = b.$$  \hfill (23.17)

Using this result and Prop.3 we obtain:

**Proposition 7:** Let $V$, $W$, $b$ and $c$ be given as in Prop.3. If $M \in \mathbb{R}^{J \times I}$ is the matrix of a given $L \in \text{Lin}(V, W)$ relative to $b$ and $c$, then $M^\top \in \mathbb{R}^{I \times J}$ is the matrix of $L^\top \in \text{Lin}(W^*, V^*)$ relative to $c^*$ and $b^*$, i.e.

$$M_{ij} = b_i L^\top c_j^* = c_j L^\top b_i = M_{ji}$$  \hfill (23.18)

Let $\beta := (\beta_i \mid i \in I)$ be a family in $V^*$, so that $\beta \in V^*^I$. Using the identification $V^*^I = (\text{Lin}(V, \mathbb{R}))^I \cong \text{Lin}(V, \mathbb{R}^I)$ defined by termwise evaluation (see Sect.14), and the identification $(\mathbb{R}^I)^* \cong \mathbb{R}^I$ characterized by (23.5), we easily see that $\beta^\top \in \text{Lin}((\mathbb{R}^I)^*, V^*) \cong \text{Lin}(\mathbb{R}^I, V^*)$ is given by

$$\beta^\top = \text{lin}_c(\beta).$$  \hfill (23.19)

**Remark:** Let $b$ be a basis of $V$. The mapping $(\text{lin}^{-1}_b)(\text{lin}^{-1}_c) : V \to V^*$ is a linear isomorphism. In fact, by (23.10) and (23.11) it is the (unique) linear isomorphism that maps the basis $b$ termwise onto the dual basis $b^*$. This isomorphism is not a natural isomorphism because it depends on the capricious choice of the basis $b$. The natural isomorphism that is used for the identification $(\mathbb{R}^I)^* \cong \mathbb{R}^I$ exists because $\mathbb{R}^I$ has a natural basis, namely the standard basis. This basis gives $\mathbb{R}^I$ a structure beyond the mere linear-space structure. To use the metaphor mentioned in the Pitfall at the end of Sect.15, the bases in $\mathbb{R}^I$ do not form a “democracy”, as is the case in linear spaces without additional structure. Rather, they form a “monarchy” with the standard basis as king.

## 24 Bilinear Mappings

**Definition 1:** Let $V_1$, $V_2$, and $W$ be linear spaces. We say that the mapping $B : V_1 \times V_2 \to W$ is **bilinear** if $B(v_1, \cdot) := (v_2 \mapsto B(v_1, v_2)) : V_2 \to W$ is linear for all $v_1 \in V_1$ and $B(\cdot, v_2) := (v_1 \mapsto B(v_1, v_2)) : V_1 \to W$ is
24. BILINEAR MAPPINGS

linear for all $v_2 \in V_2$. The set of all bilinear mappings from $V_1 \times V_2$ to $W$ is denoted by $\text{Lin}_2(V_1 \times V_2, W)$.

Briefly, to say that $B$ is bilinear means that $B(v_1, \cdot) \in \text{Lin}(V_2, W)$ for all $v_1 \in V_1$ and $B(\cdot, v_2) \in \text{Lin}(V_1, W)$ for all $v_2 \in V_2$.

**Proposition 1:** $\text{Lin}_2(V_1 \times V_2, W)$ is a subspace of $\text{Map}(V_1 \times V_2, W)$.

**Proof:** $\text{Lin}_2(V_1 \times V_2, W)$ is not empty because the zero-mapping belongs to it. To show that $\text{Lin}_2(V_1 \times V_2, W)$ is stable under addition, consider two of its members $B$ and $C$. Using the definition of $B + C$ (see Sect. 14) we have

$$(B + C)(v_1, \cdot) = B(v_1, \cdot) + C(v_1, \cdot)$$

for all $v_1 \in V_1$. Since $\text{Lin}(V_2, W)$ is stable under addition, it follows that $(B + C)(v_1, \cdot) \in \text{Lin}(V_2, W)$ for all $v_1 \in V_1$. A similar argument shows that $(B + C)(\cdot, v_2) \in \text{Lin}(V_1, W)$ for all $v_2 \in V_2$ and hence that $B + C \in \text{Lin}_2(V_1 \times V_2, W)$. It is even easier to show that $\text{Lin}_2(V_1 \times V_2, W)$ is stable under scalar multiplication. ■

Pitfall: The space $\text{Lin}_2(V_1 \times V_2, W)$ of bilinear mappings has little connection with the space $\text{Lin}(V_1 \times V_2, W)$ of linear mappings from the product space $V_1 \times V_2$ to $W$. In fact, it is easily seen that as subspaces of $\text{Map}(V_1 \times V_2, W)$ the two are disjunct, i.e. they have only the zero mapping in common. ■

**Proposition 2:** Let $V_1$, $V_2$, and $W$ be linear spaces. For each $B \in \text{Lin}_2(V_1 \times V_2, W)$, the mapping

$$(v_1 \mapsto B(v_1, \cdot)) : V_1 \to \text{Lin}(V_2, W)$$

is linear. Moreover, the mapping

$$(B \mapsto (v_1 \mapsto B(v_1, \cdot))) : \text{Lin}_2(V_1 \times V_2, W) \to \text{Lin}(V_1, \text{Lin}(V_2, W))$$

is a linear isomorphism.

**Proof:** The first assertion follows from the definition of the linear-space operations in $\text{Lin}(V_2, W)$ and the linearity of $B(\cdot, v_2)$ for all $v_2 \in V_2$. The second assertion is an immediate consequence of the definitions of the spaces involved and of the definition of the linear-space operations in these spaces. ■

We use the natural isomorphism described in Prop. 2 to identify:

$$\text{Lin}_2(V_1 \times V_2, W) \cong \text{Lin}(V_1, \text{Lin}(V_2, W)). \quad (24.1)$$

This identification is expressed by

$$(Bv_1)v_2 = B(v_1, v_2) \quad \text{for all } v_1 \in V_1, v_2 \in V_2. \quad (24.2)$$
where on the left side $B$ is interpreted as an element of $\text{Lin}(V_1, \text{Lin}(V_2, W))$ and on the right side $B$ is interpreted as an element of $\text{Lin}_2(V_1 \times V_2, W)$. The identification given by (24.1) and (24.2) is consistent with the identification described by (04.28) and (04.29).

Using Prop.7 of Sect.17 and Prop.2 above, we obtain the following formula for the dimension of spaces of bilinear mappings:

$$\dim \text{Lin}_2(V_1 \times V_2, W) = (\dim V_1)(\dim V_2)(\dim W).$$  \hspace{1cm} (24.3) 

**Examples:**

1. The scalar multiplication $sm : \mathbb{R} \times V \to V$ of a linear space $V$ is bilinear, i.e.

   $$\text{Lin}_2(\mathbb{R} \times V, V) \cong \text{Lin}(\mathbb{R}, \text{Lin}(V, V)) = \text{Lin}(\mathbb{R}, \text{Lin}V).$$

   The corresponding linear mapping $sm \in \text{Lin}(\mathbb{R}, \text{Lin}V)$ is given by

   $$sm = (\xi \mapsto \xi 1_V) : \mathbb{R} \to \text{Lin}V.$$

   It is not only linear, but it also preserves products, i.e. $sm(\xi \eta) = (sm\xi)(sm\eta)$ holds for all $\xi, \eta \in \mathbb{R}$. In fact, $sm$ is an injective algebra-homomorphism from $\mathbb{R}$ to the algebra of lineons on $V$ (see Sect.18).

2. Let $V$ and $W$ be linear spaces. Then

   $$((L, v) \mapsto Lv) : \text{Lin}(V, W) \times V \to W$$

   is bilinear. The corresponding linear mapping is simply $1_{\text{Lin}(V, W)}$. In the special case $W := \mathbb{R}$, we obtain the bilinear mapping

   $$((\lambda, v) \mapsto \lambda v) : V^* \times V \to \mathbb{R}.$$

   The corresponding linear mapping is $1_{V^*}$.

3. Let $S$ be a set and let $V$ and $V'$ be linear spaces. It then follows from Props.1 and 3 of Sect.14 that

   $$((L, f) \mapsto Lf) : \text{Lin}(V, V') \times \text{Map}(S, V) \to \text{Map}(S, V')$$

   is bilinear.
4. Let $V$, $V'$, $V''$ be linear spaces. Then

$$( \langle M, L \rangle \mapsto ML ) : \text{Lin}(V', V'') \times \text{Lin}(V, V') \to \text{Lin}(V, V'')$$

defines a bilinear mapping. This follows from Prop. 1 of Sect. 13 and the result stated in the preceding example.

The following two results are immediate consequences of the definitions and Prop. 1 of Sect. 13.

**Proposition 3:** The composite of a bilinear mapping with a linear mapping is again bilinear. More precisely, if $V_1, V_2, W, W'$ are linear spaces and if $B \in \text{Lin}_2(V_1 \times V_2, W)$ and $L \in \text{Lin}(W, W')$, then $LB \in \text{Lin}_2(V_1 \times V_2, W')$.

**Proposition 4:** The composite of the cross-product of a pair of linear mappings with a bilinear mapping is again bilinear. More precisely, if $V_1, V_2, V'_1, V'_2$, and $W$ are linear spaces and if $L_1 \in \text{Lin}(V_1, V'_1)$, $L_2 \in \text{Lin}(V_2, V'_2)$ and $B \in \text{Lin}_2(V'_1 \times V'_2, W)$, then $B \circ (L_1 \times L_2) \in \text{Lin}_2(V_1 \times V_2, W)$.

With every $B \in \text{Lin}_2(V_1 \times V_2, W)$ we can associate a bilinear mapping $B^\sim \in \text{Lin}_2(V_2 \times V_1, W)$ defined by

$$B^\sim(v_2, v_1) := B(v_1, v_2) \text{ for all } v_1 \in V_1, v_2 \in V_2 \quad (24.4)$$

We call $B^\sim$ the **switch** of $B$. It is evident that

$$(B^\sim)^\sim = B \quad (24.5)$$

holds for all bilinear mappings $B$ and that the **switching**, defined by

$$(B \mapsto B^\sim) : \text{Lin}_2(V_1 \times V_2, W) \to \text{Lin}_2(V_2 \times V_1, W)$$

is a linear isomorphism.

**Definition 2:** Let $V$ and $W$ be linear spaces. We say that a bilinear mapping $B \in \text{Lin}_2(V^2, W)$ is **symmetric** if $B^\sim = B$, i.e. if

$$B(u, v) = B(v, u) \text{ for all } u, v \in V; \quad (24.6)$$

we say that it is **skew** if $B^\sim = -B$, i.e. if

$$B(u, v) = -B(v, u) \text{ for all } u, v \in V. \quad (24.7)$$

We use the notations

$$\text{Sym}_2(V^2, W) := \{ S \in \text{Lin}_2(V^2, W) | S^\sim = S \},$$
$$\text{Skew}_2(V^2, W) := \{ A \in \text{Lin}_2(V^2, W) | A^\sim = -A \}. $$
Proposition 5: A bilinear mapping \( A \in \text{Lin}_2(V^2, W) \) is skew if and only if
\[
A(u, u) = 0 \quad \text{for all} \quad u \in V. \tag{24.8}
\]

Proof: If \( A \) is skew, then, by (24.7), we have \( A(u, u) = -A(u, u) \) and hence \( A(u, u) = 0 \) for all \( u \in V \). If (24.8) holds, then
\[
0 = A(u + v, u + v) = A(u, u) + A(u, v) + A(v, u) + A(v, v) = A(u, v) + A(v, u)
\]
and hence \( A(u, v) = -A(v, u) \) for all \( u, v \in V \). \[\square\]

Proposition 6: To every \( B \in \text{Lin}_2(V^2, W) \) corresponds a unique pair \((S, A)\) with \( S \in \text{Sym}_2(V^2, W) \), \( A \in \text{Skew}_2(V^2, W) \) such that \( B = S + A \). In fact, \( S \) and \( A \) are given by
\[
S = \frac{1}{2}(B + B^\sim), \quad A = \frac{1}{2}(B - B^\sim). \tag{24.9}
\]

Proof: Assume that \( S \in \text{Sym}_2(V^2, W) \) and \( A \in \text{Skew}_2(V^2, W) \) are given such that \( B = S + A \). Since \( B \mapsto B^\sim \) is linear, it follows that \( B^\sim = S^\sim + A^\sim = S - A \). Therefore, we have \( B + B^\sim = (S + A) + (S - A) = 2S \) and \( B - B^\sim = (S + A) - (S - A) = 2A \), which shows that \( S \) and \( A \) must be given by (24.9) and hence are uniquely determined by \( B \). On the other hand, if we define \( S \) and \( A \) by (24.9), we can verify immediately that \( S \) is symmetric, that \( A \) is skew, and that \( B = S + A \). \[\square\]

In view of Prop.4 of Sect.12, Prop.6 has the following immediate consequence:

Proposition 7: \( \text{Sym}_2(V^2, W) \) and \( \text{Skew}_2(V^2, W) \) are supplementary subspaces of \( \text{Lin}_2(V^2, W) \).

Let \( V \) and \( W \) be linear spaces. We consider the identifications
\[
\text{Lin}_2(V \times W^*, \mathbb{R}) \cong \text{Lin}(V, W^{**}) \cong \text{Lin}(V, W)
\]
and
\[
\text{Lin}_2(W^* \times V, \mathbb{R}) \cong \text{Lin}(W^*, V^*).
\]
Hence if \( L \in \text{Lin}(V, W) \), we can not only form its transpose \( L^\top \in \text{Lin}(W^*, V^*) \) but, by interpreting \( L \) as an element of \( \text{Lin}_2(V \times W^*, \mathbb{R}) \), we can also form its switch \( L^\sim \in \text{Lin}_2(W^* \times V, \mathbb{R}) \). It is easily verified that \( L^\top \) and \( L^\sim \) correspond under the identification, i.e.
\[
L^\sim = L^\top \quad \text{for all} \quad L \in \text{Lin}(V, W). \tag{24.10}
\]
25. TENSOR PRODUCTS

**Pitfall:** For a bilinear mapping $B$ whose codomain is not $\mathbb{R}$, the linear mapping corresponding to the switch $B^\sim$ is not the same as the transpose $B^\top$ of the linear mapping corresponding to $B$. 

Let $V_1, V_2, W_1$, and $W_2$ be linear spaces. Let $L_1 \in \text{Lin}(V_1, W_1)$, $L_2 \in \text{Lin}(V_2, W_2)$ and $B \in \text{Lin}_2(W_1 \times W_2, \mathbb{R}) \cong \text{Lin}(W_1, W_2^\ast)$ be given. By Prop.4, we then have

$$B \circ (L_1 \times L_2) \in \text{Lin}_2(V_1 \times V_2, \mathbb{R}) \cong \text{Lin}(V_1, V_2^\ast).$$

Using these identifications, it is easily seen that

$$B \circ (L_1 \times L_2) = L_2^\top BL_1.$$  \hfill (24.11)

**Notes 24**

1. The terms “antisymmetric” and “skewsymmetric” are often used for what we call, simply, “skew”. A bilinear mapping that satisfies the condition (24.8) is often said to be “alternating”. In the case considered here, this term is synonymous with “skew”, but one obtains a different concept if one replaces $\mathbb{R}$ by a field of characteristic 2.

2. The pair $(S, A)$ associated with the bilinear mapping $B$ according to Prop.6 is sometimes called the “Cartesian decomposition” of $B$.

25 Tensor Products

For any linear space $V$ there is a natural isomorphism from $\text{Lin}(\mathbb{R}, V)$ onto $V$, given by $h \mapsto h(1)$. The inverse isomorphism associates with $v \in V$ the mapping $\xi \mapsto \xi v$ in $\text{Lin}(\mathbb{R}, V)$. We denote this mapping by $v \otimes$ (read “vee tensor”) so that

$$v \otimes \xi := \xi v \text{ for all } \xi \in \mathbb{R}. \quad \text{(25.1)}$$

In particular, there is a natural isomorphism from $\mathbb{R}$ onto $\mathbb{R}^* = \text{Lin}(\mathbb{R}, \mathbb{R})$. It associates with every number $\eta \in \mathbb{R}$ the operation of multiplication with that number, so that (25.1) reduces to $\eta \otimes \xi = \eta \xi$. We use this isomorphism to identify $\mathbb{R}^*$ with $\mathbb{R}$, i.e. we write $\eta = \eta \otimes$. However, when $V \neq \mathbb{R}$, we do not identify $\text{Lin}(\mathbb{R}, V)$ with $V$ because such an identification would conflict with the identification $V \cong V^{**}$ and lead to confusion.

If $\lambda \in V^* = \text{Lin}(V, \mathbb{R})$, we can consider $\lambda^\top \in \text{Lin}(\mathbb{R}^*, V^*) \cong \text{Lin}(\mathbb{R}, V^*)$. Using the identification $\mathbb{R}^* \cong \mathbb{R}$, it follows from (21.3) and (25.1) that $\lambda^\top \xi = \xi \lambda = \lambda \otimes \xi$ for all $\xi \in \mathbb{R}$, i.e. that $\lambda^\top = \lambda \otimes$. 

Definition 1: Let $V$ and $W$ be linear spaces. For every $w \in W$ and every $\lambda \in V^*$ the tensor product of $w$ and $\lambda$ is defined to be $w \otimes \lambda \in \text{Lin}(V,W)$, i.e. the composite of $\lambda$ with $w \otimes \in \text{Lin}(\mathbb{R}, W)$, so that

$$(w \otimes \lambda)v = (\lambda v)w \text{ for all } v \in V.$$  \hfill (25.2)$$

In view of Example 4 of Sect. 24, it is clear that the mapping

$$((w, \lambda) \mapsto w \otimes \lambda) : W \times V^* \rightarrow \text{Lin}(V, W)$$ \hfill (25.3)$$
is bilinear.

In the special case when $V := \mathbb{R}^I \cong (\mathbb{R}^I)^*$ and $W := \mathbb{R}^J$ for finite index sets $I$ and $J$, (16.4) and (25.2) show that the tensor product $\mu \otimes \lambda \in \text{Lin}(\mathbb{R}^I, \mathbb{R}^J) \cong \mathbb{R}^{J \times I}$ of $\mu \in \mathbb{R}^J$ and $\lambda \in \mathbb{R}^I$ has the components

$$(\mu \otimes \lambda)_{j,i} = \mu_j \lambda_i \text{ for all } (j, i) \in J \times I.$$ \hfill (25.4)$$

Using the identifications $V^{**} \cong V$ and $W^{**} \cong W$, we can form tensor products

$$w \otimes v \in \text{Lin}(V^*, W), \quad \mu \otimes v \in \text{Lin}(V^*, W^*),$$

$$w \otimes \lambda \in \text{Lin}(V, W), \quad \mu \otimes \lambda \in \text{Lin}(V, W^*)$$

for all $v \in V$, $w \in W$, $\lambda \in V^*$, $\mu \in W^*$. Also, using identifications such as

$$\text{Lin}(V, W) \cong \text{Lin}(V, W^{**}) \cong \text{Lin}_2(V \times W^*, \mathbb{R})$$

we can interpret any tensor product as a bilinear mapping to $\mathbb{R}$. For example, if $w \in W$ and $\lambda \in V^*$ we have

$$(w \otimes \lambda)(v, \mu) = (\lambda v)(w \mu) \text{ for all } v \in V, \mu \in W^*.$$ \hfill (25.5)$$

The following is an immediate consequence of (25.5), the definition (24.4) of a switch, and (24.10).

**Proposition 1:** For all $w \in W$ and $\lambda \in V^*$, we have

$$(w \otimes \lambda)^\sim = (w \otimes \lambda)^\top = \lambda \otimes w.$$  \hfill (25.6)$$

The following facts follow immediately from the definition of a tensor product.

**Proposition 2:** If $w \neq 0$ then

$$\text{Null } (w \otimes \lambda) = \text{Null } \lambda;$$ \hfill (25.7)$$
if $\lambda \neq 0$ then

$$\text{Rng}(w \otimes \lambda) = \mathbb{R}w.$$  \hfill (25.8)

**Proposition 3:** Let $V, V', W, W'$ be linear spaces. If $\lambda \in V^*$, $w \in W$, $L \in \text{Lin}(W, W')$, and $M \in \text{Lin}(V', V)$ then

$$L(w \otimes \lambda) = (Lw) \otimes \lambda$$  \hfill (25.9)

and

$$(w \otimes \lambda)M = w \otimes (\lambda M) = w \otimes (M^T \lambda).$$  \hfill (25.10)

If $v \in V$ and $\mu \in V'^*$, then

$$(w \otimes \lambda)(v \otimes \mu) = (\lambda v)(w \otimes \mu).$$  \hfill (25.11)

Tensor products can be used to construct bases of spaces of linear mappings:

**Proposition 4:** Let $V$ and $W$ be linear spaces, let $b := (b_i \mid i \in I)$ be a basis of $V$. Let $b^* := (b^*_i \mid i \in I)$ be the basis of $V^*$ dual to $b$ and let $c := (c_j \mid j \in J)$ be a basis of $W$. Then $(c_j \otimes b^*_i \mid (j, i) \in J \times I)$ is a basis of $\text{Lin}(V, W)$, and the matrix $M \in \mathbb{R}^{J \times I}$ of the components of $L$ relative to this basis is the same as the matrix $\text{ln}_c^{-1}L\text{ln}_b \in \text{Lin}(\mathbb{R}^I, \mathbb{R}^J) \cong \mathbb{R}^{J \times I}$, i.e. the matrix of $L$ relative to the bases $b$ and $c$ (see Sect. 16).

**Proof:** It is sufficient to prove that

$$L = \sum_{(j, i) \in J \times I} M_{j,i} (c_j \otimes b^*_i)$$  \hfill (25.12)

holds when $M$ is the matrix of $L$ relative to $b$ and $c$. It follows from (16.11) and from (25.2) and Prop.4 of Sect. 23 that the left and right sides of (25.12) give the same value when applied to the terms $b_k$ of the basis $b$. Using Prop.2 of Sect. 16, we conclude that (25.12) must hold. $\blacksquare$

Using (18.6) and (18.8), we obtain the following special case of (25.12):

**Proposition 5:** Let $V$ be a linear space and let $b := (b_i \mid i \in I)$ be a basis of $V$. For every lineon $L \in \text{Lin}V$, we then have

$$L = \sum_{(j, i) \in I \times I} ([L]_{b})_{j,i} (b_j \otimes b^*_i).$$  \hfill (25.13)

In particular, we have

$$1_V = \sum_{i \in I} b_i \otimes b^*_i.$$  \hfill (25.14)

Prop.4 has the following corollary:
Proposition 6: If $V$ and $W$ are linear spaces, then
\[ \text{Lin}(V, W) = \text{Lsp}\{ w \otimes \lambda \mid w \in W, \lambda \in V^* \}. \] (25.15)

Notes 25

(1) The term “dyadic product” and the notation $w \lambda$ is often used in the older literature for our “tensor product” $w \otimes \lambda$. We cannot use this older notation because it would lead to a clash with the evaluation notation described by (22.2).

(2) For other uses of the term “tensor product” see Note (1) to Sect. 26.

26 The Trace

Let $V$ and $W$ be linear spaces. Since the mapping (25.3) is bilinear, it follows from Prop.3 of Sect. 24 that for every linear form $\Omega$ on $\text{Lin}(V, W)$

\[ ((w, \lambda) \mapsto \Omega(w \otimes \lambda)) : W \times V^* \rightarrow \mathbb{R} \]

is a bilinear mapping. Using the identifications $\text{Lin}_2(V \times V^*, \mathbb{R}) \cong \text{Lin}(W, V^{**}) \cong \text{Lin}(W, V)$ (see Sects. 24 and 22) we see that there is a mapping

\[ \tau : (\text{Lin}(V, W))^* \rightarrow \text{Lin}(W, V) \] (26.1)

defined by

\[ \lambda(\tau(\Omega)w) = \Omega(w \otimes \lambda) \quad \text{for all} \quad \Omega \in (\text{Lin}(V, W))^*, \ w \in W, \ \lambda \in V^*. \] (26.2)

Lemma: The mapping (26.1) defined by (26.2) is a linear isomorphism.

Proof: The linearity of $\tau$ follows from the fact that every member of $\text{Lin}(V, W)$, and in particular $w \otimes \lambda$, can be identified with an element of $(\text{Lin}(V, W))^{**}$, i.e. a linear form on $(\text{Lin}(V, W))^*$.

If $\Omega \in \text{Null } \tau$ then $\tau(\Omega) = 0$ and hence, by (26.2), $\Omega(w \otimes \lambda) = 0$ for all $w \in W$ and all $\lambda \in V^*$. By Prop.6 of Sect. 25 this is possible only when $\Omega = 0$. We conclude that $\text{Null } \tau = \{0\}$. Since $\dim(\text{Lin}(V, W))^* = (\dim V)(\dim W) = \dim(\text{Lin}(W, V))$ (see (21.1) and (17.7) ) it follows from the Pigeonhole Principle for Linear Mappings (Sect. 17) that $\tau$ is invertible.

The following result shows that the algebra of lineons admits a natural linear form.
Characterization of the Trace: Let $\mathcal{V}$ be a linear space. There is exactly one linear form $\text{tr}_\mathcal{V}$ on $\text{Lin}\mathcal{V}$ that satisfies
\[
\text{tr}_\mathcal{V}(v \otimes \lambda) = \lambda v \quad \text{for all } v \in \mathcal{V}, \lambda \in \mathcal{V}^*.
\] (26.3)

This linear form $\text{tr}_\mathcal{V}$ is called the trace for $\mathcal{V}$.

Proof: Assume that $\text{tr}_\mathcal{V} \in (\text{Lin}\mathcal{V})^*$ satisfies (26.3). Using (26.2) with $\mathcal{W} := \mathcal{V}$ and the choice $\Omega := \text{tr}_\mathcal{V}$ we see that we must have $\tau(\text{tr}_\mathcal{V}) = 1_\mathcal{V}$. By the Lemma, it follows that $\text{tr}_\mathcal{V}$ is uniquely determined as $\text{tr}_\mathcal{V} = \tau^{-1}(1_\mathcal{V})$. On the other hand, if we define $\text{tr}_\mathcal{V} := \tau^{-1}(1_\mathcal{V})$ then (26.3) follows from (26.2).

If the context makes clear what $\mathcal{V}$ is, we often write tr for $\text{tr}_\mathcal{V}$.

Using (26.3), (25.9), and (25.10), it is easily seen that the definition (26.2) of the mapping $\tau$ is equivalent to the statement that
\[
\Omega L = \text{tr}_\mathcal{V}(\tau(\Omega)L) = \text{tr}_\mathcal{W}(L\tau(\Omega))
\] (26.4)
holds for all $\Omega \in (\text{Lin}(\mathcal{V}, \mathcal{W}))^*$ and all $L \in \text{Lin}(\mathcal{V}, \mathcal{W})$ that are tensor products, i.e. of the form $L = w \otimes \lambda$ for some $w \in \mathcal{W}, \lambda \in \mathcal{V}^*$. Since these tensor products span $\text{Lin}(\mathcal{V}, \mathcal{W})$ (Prop.6 of Sect. 25), it follows that the mapping $\tau$ can be characterized by the statement that (26.4) holds for all $L \in \text{Lin}(\mathcal{V}, \mathcal{W})$, whether $L$ is a tensor product or not. Using this fact and the Lemma we obtain the following two results.

Representation Theorem for Linear Forms on a Space of Linear Mappings: Let $\mathcal{V}$ and $\mathcal{W}$ be linear spaces. Every linear form $\Omega$ on $\text{Lin}(\mathcal{V}, \mathcal{W})$ is represented by a unique $M \in \text{Lin}(\mathcal{W}, \mathcal{V})$ in the sense that
\[
\Omega L = \text{tr}_\mathcal{V}(ML) \quad \text{for all } L \in \text{Lin}(\mathcal{V}, \mathcal{W}).
\] (26.5)

Proposition 1: For every $L \in \text{Lin}(\mathcal{V}, \mathcal{W})$ and every $M \in \text{Lin}(\mathcal{W}, \mathcal{V})$ we have
\[
\text{tr}_\mathcal{V}(ML) = \text{tr}_\mathcal{W}(LM).
\] (26.6)

Now let a linear space $\mathcal{V}$ be given.

Proposition 2: For every $L \in \text{Lin}\mathcal{V}$ we have
\[
\text{tr}_\mathcal{V}L^\top = \text{tr}_\mathcal{V}L.
\] (26.7)

Proof: By the Theorem on Characterization of Trace it suffices to show that the mapping $L \mapsto \text{tr}_\mathcal{V}L^\top$ from $\text{Lin}\mathcal{V}$ into $\mathbb{R}$ is linear and has the value $\lambda v$ when $L = v \otimes \lambda$. The first is immediate and the second follows from (25.6) and (26.3), applied to the case when $\mathcal{V}$ is replaced by $\mathcal{V}^*$. 

**Proposition 3:** Let $b := (b_i \mid i \in I)$ be a basis of $V$. For every lineon $L \in \text{Lin}V$, we have
\[
\text{tr}_V L = \sum_{i \in I} ([L]_b)_{i,i},
\]
where $[L]_b := \ln e^{-1}L\ln e_b$ is the matrix of $L$ relative to $b$.

**Proof:** It follows from (26.3) and Prop. 4 of Sect. 23 that
\[
\text{tr}_V (b_j \otimes b^*_i) = b^*_i b_j = \delta_{i,j} \quad \text{for all} \quad i, j \in I.
\]
Therefore, (26.8) is a consequence of (25.13) and the linearity of $\text{tr}_V$.

If we apply (26.8) to the case when $L := 1_V$ and hence $([L]_b)_{i,i} = 1$ for all $i \in I$ (see (18.8)), we obtain
\[
\text{tr}_V 1_V = \dim V.
\]

**Proposition 4:** Let $U$ be a subspace of $V$ and let $P : V \to U$ be a projection (see Sect. 19). Then
\[
\text{tr}_U (K) = \text{tr}_V (1_{U \subset V} KP)
\]
for all $K \in \text{Lin}U$.

**Proof:** It follows from Prop.1 that
\[
\text{tr}_V (1_{U \subset V} KP) = \text{tr}_U (P (1_{U \subset V} K)) = \text{tr}_U (P|_U K)
\]
for all $K \in \text{Lin}U$. By the definition of a projection (Sect. 19) we have $P|_U = 1_U$ and hence (26.10) holds.

**Proposition 5:** The trace of an idempotent lineon $E$ on $V$ is given by
\[
\text{tr}_V E = \dim \text{Rng} E
\]

**Proof:** Put $U := \text{Rng} E$ and $P := E|_U$. We then have $E = 1_{U \subset V} P$ and, by Prop.1 of Sect. 19, $P : V \to U$ is a projection. Applying (26.10) to the case when $K := 1_U$ we obtain $\text{tr}_U (1_U) = \text{tr}_V (E)$. The desired formula (26.11) now follows from (26.9).

**Proposition 6:** Let $V_1$, $V_2$, and $W$ be linear spaces. There is exactly one mapping
\[
\Lambda : \text{Lin}_2 (V_1 \times V_2, W) \to \text{Lin} (\text{Lin}(V_2^*, V_1), W)
\]
such that
\[
B(v_1, v_2) = \Lambda(B)(v_1 \otimes v_2) \quad \text{for all} \quad v_1 \in V_1, v_2 \in V_2
\]

\[
\text{(26.12)}
\]
26. THE TRACE

and all \( \mathbf{B} \in \text{Lin}_2(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{W}) \). Moreover, \( \Lambda \) is a linear isomorphism.

**Proof:** For every \( \mathbf{B} \in \text{Lin}_2(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{W}) \) and \( \mu \in \mathcal{W}^* \) we have \( \mu \mathbf{B} \in \text{Lin}_2(\mathcal{V}_1 \times \mathcal{V}_2, \mathbb{R}) \cong \text{Lin}(\mathcal{V}_1, \mathcal{V}_2^*) \) (see Prop. 3 of Sect. 24). Using the identification \( \mathcal{V}_2^{**} \cong \mathcal{V}_2 \), we see that (26.12) holds if and only if

\[
\mathbf{v}_2((\mu \mathbf{B}) \mathbf{v}_1) = (\mu \mathbf{B})(\mathbf{v}_1, \mathbf{v}_2) = (\mu \Lambda(\mathbf{B}))(\mathbf{v}_1 \otimes \mathbf{v}_2)
\]

(26.13) for all \( \mathbf{v}_1 \in \mathcal{V}_1, \mathbf{v}_2 \in \mathcal{V}_2 \) and all \( \mu \in \mathcal{W}^* \). Comparing (26.13) with (26.2), we conclude that (26.12) is equivalent to

\[
\tau(\mu \Lambda(\mathbf{B})) = \mu \mathbf{B} \quad \text{for all} \quad \mu \in \mathcal{W}^*
\]

where \( \tau \) is defined according to (26.1) and (26.2), in which \( \mathcal{V} \) and \( \mathcal{W} \) must be replaced by \( \mathcal{V}_1 \) and \( \mathcal{V}_2^* \), respectively. Since \( \tau \) is invertible by the Lemma, we conclude that (26.12) is equivalent to

\[
\mu \Lambda(\mathbf{B}) = \tau^{-1}(\mu \mathbf{B}) \quad \text{for all} \quad \mu \in \mathcal{W}^*
\]

(26.14)

Since \( \mathcal{W} \) is isomorphic to \( \mathcal{W}^{**} \), the uniqueness and existence of \( \Lambda \) follow from the equivalence of (26.14) with (26.12). The linearity of \( \Lambda \) follows from the linearity of \( \tau^{-1} \). Also it is clear from (26.14) that \( \Lambda \) has an inverse \( \Lambda \), which is characterized by

\[
\mu \Lambda^{-1}(\Phi) = \tau(\mu \Phi) \quad \text{for all} \quad \mu \in \mathcal{W}^*
\]

(26.15)

and all \( \Phi \in \text{Lin}(\text{Lin}(\mathcal{V}_2^*, \mathcal{V}_1), \mathcal{W}) \).

**Remark:** Prop. 6 shows that every bilinear mapping \( \mathbf{B} \) on \( \mathcal{V}_1 \times \mathcal{V}_2 \) is the composite of the tensor product mapping from \( \mathcal{V}_1 \times \mathcal{V}_2 \) into \( \text{Lin}(\mathcal{V}_2^*, \mathcal{V}_1) \) with a linear mapping \( \Lambda(\mathbf{B}) \). In a setting more abstract than the one used here, the term tensor product mapping is often employed for any bilinear mapping \( \otimes \) on \( \mathcal{V}_1 \times \mathcal{V}_2 \) with the following universal factorization property: Every bilinear mapping on \( \mathcal{V}_1 \times \mathcal{V}_2 \) is the composite of \( \otimes \) with a unique linear mapping. The codomain of \( \otimes \) is then called a tensor product space and is denoted by \( \mathcal{V}_1 \otimes \mathcal{V}_2 \). For the specific tensor product used here, we have \( \mathcal{V}_1 \otimes \mathcal{V}_2 := \text{Lin}(\mathcal{V}_2^*, \mathcal{V}_1) \).

Notes 26

(1) There are many ways of constructing a tensor-product space in the sense of the Remark above from given linear spaces \( \mathcal{V}_1 \) and \( \mathcal{V}_2 \). The notation \( \mathcal{V}_1 \otimes \mathcal{V}_2 \) for such a space is therefore ambiguous and we will not use it. One can associate with the construction of tensor-product spaces a concept of “tensor product” of linear mappings which generalizes the concept of tensor product introduced in Sect. 25. Tensor-product spaces are of little practical value in the present context, even though they give important insights in abstract algebra.
27 Bilinear Forms and Quadratic Forms

We assume that a linear space $\mathcal{V}$ is given. The bilinear mappings from $\mathcal{V}^2$ to $\mathbb{R}$, i.e. the members of $\text{Lin}_2(\mathcal{V}^2, \mathbb{R})$ are called bilinear forms on $\mathcal{V}$. The identification $\text{Lin}_2(\mathcal{V}^2, \mathbb{R}) \cong \text{Lin}(\mathcal{V}, \mathcal{V}^*)$—a special case of (24.1)—enables us to interpret bilinear forms as mappings from $\mathcal{V}$ to its dual $\mathcal{V}^*$. We use the notation $\text{Sym}(\mathcal{V}, \mathcal{V}^*)$ and $\text{Skew}(\mathcal{V}, \mathcal{V}^*)$ for the subspaces of $\text{Lin}(\mathcal{V}, \mathcal{V}^*)$ that correspond to the subspaces $\text{Sym}_2(\mathcal{V}^2, \mathbb{R})$ and $\text{Skew}_2(\mathcal{V}^2, \mathbb{R})$ of $\text{Lin}_2(\mathcal{V}^2, \mathbb{R})$ (see Def.2 in Sect. 24). In view of (24.10), we have

$$\text{Sym}(\mathcal{V}, \mathcal{V}^*) = \{ S \in \text{Lin}(\mathcal{V}, \mathcal{V}^*) \mid S^\top = S \}, \quad (27.1)$$

$$\text{Skew}(\mathcal{V}, \mathcal{V}^*) = \{ A \in \text{Lin}(\mathcal{V}, \mathcal{V}^*) \mid A^\top = -A \}. \quad (27.2)$$

By Prop.7 of Sect. 24, $\text{Sym}(\mathcal{V}, \mathcal{V}^*)$ and $\text{Skew}(\mathcal{V}, \mathcal{V}^*)$ are supplementary subspaces of $\text{Lin}(\mathcal{V}, \mathcal{V}^*)$.

Let $b := (b_i \mid i \in I)$ be a basis of $\mathcal{V}$ and let $B \in \text{Lin}_2(\mathcal{V}^2, \mathbb{R}) \cong \text{Lin}(\mathcal{V}, \mathcal{V}^*)$ be a bilinear form. The matrix $M \in \mathbb{R}^{I \times I}$ defined by

$$M_{i,j} := B(b_i, b_j) \quad \text{for all } i, j \in I \quad (27.3)$$

is called the matrix of $B$ relative to $b$. It is easily seen that $M$ coincides with the matrix of $B$ when regarded as an element of $\text{Lin}(\mathcal{V}, \mathcal{V}^*)$, relative to the bases $b$ and $b^*$ in the sense of the definition in Sect. 16. Using the fact that $b^{**} = b$, (see (23.17)) it follows from Prop.4 of Sect. 25 that $M$ is also the matrix of the components of $B$ relative to the basis $(b_i^* \otimes b_j^* \mid (i, j) \in I \times I)$ of $\text{Lin}_2(\mathcal{V}^2, \mathbb{R}) \cong \text{Lin}(\mathcal{V}, \mathcal{V}^*)$, i.e. that

$$B = \sum_{(i,j) \in I \times I} M_{i,j} (b_i^* \otimes b_j^*) \quad (27.4)$$

holds if and only if $M$ is given by (27.3). It is clear from (27.3) that $B$ is symmetric, i.e. $B^\top = B^\sim = B$, if and only if

$$M_{i,j} = M_{j,i} \quad \text{for all } i, j \in I, \quad (27.5)$$

and that $B$ is skew, i.e. $B^\top = B^\sim = -B$, if and only if

$$M_{i,j} = -M_{j,i} \quad \text{for all } i, j \in I \quad (27.6)$$

**Proposition 1:** If $n := \dim \mathcal{V}$ then

$$\dim \text{Sym}_2(\mathcal{V}^2, \mathbb{R}) = \dim \text{Sym}(\mathcal{V}, \mathcal{V}^*) = \frac{n(n + 1)}{2} \quad (27.7)$$

and
dim Skew$_2(V^2, \mathbb{R}) = \dim \text{Skew}(V, V^*) = \frac{n(n-1)}{2}$. \quad (27.8)

**Proof:** We choose a list basis $b := (b_i \mid i \in n^1)$ of $V$. Let $A \in \text{Skew}_2(V^2, \mathbb{R})$ be given. We claim that

$$A = \sum (K_{i,j} (b_i^* \otimes b_j^* - b_j^* \otimes b_i^* \mid (i,j) \in n^1 \times n^1, i < j) \quad (27.9)$$

holds if and only if

$$K_{i,j} = A(b_i, b_j) \quad \text{for all} \quad (i,j) \in n^1 \times n^1 \quad \text{with} \quad i < j. \quad (27.10)$$

Indeed, since $(b_i^* \otimes b_j^* \mid (i,j) \in n^1 \times n^1)$ is a basis of Lin$_2(V^2, \mathbb{R})$, we find, by comparing (27.9) with (27.4) and (27.10) with (27.3), that (27.9) can hold only if (27.10) holds. On the other hand, if we define $(K_{i,j} \mid (i,j) \in n^1 \times n^1, i < j)$ by (27.10), it follows from (27.6) that the matrix $M$ of $A$ is given by

$$M_{i,j} = \begin{cases} K_{i,j} & \text{if} \quad i < j \\ -K_{j,i} & \text{if} \quad j < i \\ 0 & \text{if} \quad i = j \end{cases} \quad \text{for all} \quad i,j \in I.$$ 

Therefore, (27.4) reduces to (27.9).

It follows from (25.6) that the terms of the family

$$b^* \wedge b^* := (b_i^* \otimes b_j^* - b_j^* \otimes b_i^* \mid (i,j) \in n^1 \times n^1, i < j) \quad (27.11)$$

all belong to Skew$_2(V^2, \mathbb{R})$. Therefore, the equivalence of (27.9) and (27.10) proves that the family (27.11) is a basis of Skew$_2(V^2, \mathbb{R})$. This basis has $\sharp\{(i,j) \in n^1 \times n^1 \mid i < j\} = \frac{n(n-1)}{2}$ terms, and hence (27.8) holds.

Since Sym$_2(V^2, \mathbb{R})$ is a supplement of Skew$_2(V^2, \mathbb{R})$ in Lin$_2(V^2, \mathbb{R})$ and since dim Lin$_2(V^2, \mathbb{R}) = n^2$ by (24.3), we can use Prop. 5 of Sect. 17 and (27.8) to obtain

$$\dim \text{Sym}_2(V^2, \mathbb{R}) = \dim \text{Lin}_2(V^2, \mathbb{R}) - \dim \text{Skew}_2(V^2, \mathbb{R}) = n^2 - \frac{n(n-1)}{2} = \frac{n(n+1)}{2}.$$ 

We note that if $S \in \text{Sym}_2(V^2, \mathbb{R})$, then $S \circ (1_V, 1_V) = (u \mapsto S(u,u))$ is a real-valued function on $V$ (see Sect. 04). Thus, one can consider the mapping

$$(S \mapsto S \circ (1_V, 1_V)) : \text{Sym}_2(V^2, \mathbb{R}) \to \text{Map}(V, \mathbb{R}). \quad (27.12)$$
CHAPTER 2. DUALITY, BILINEARITY

**Proposition 2:** The mapping (27.12) is linear and injective.

**Proof:** The linearity follows from Prop.1 of Sect. 14. To show that (27.12) is injective, it suffices to show that its nullspace contains only the zero of Sym$_2$(V$^2$, R). If S belongs to this nullspace, then S ◦ (1$_V$, 1$_V$) = 0, i.e. S(u, u) = 0 for all u ∈ V. By Props.5 and 7 of Sect. 24 we conclude that S must also be skew and hence that S = 0.

**Definition 1:** The range space of the mapping (27.12) will be denoted by Qu(V). Its members are called **quadratic forms** on V. We write

$$S := S ◦ (1_V, 1_V) \text{ when } S \in \text{Sym}_2(V^2, R),$$

(27.13)

and the inverse of the linear isomorphism

$$(S \mapsto \overline{S}) : \text{Sym}_2(V^2, R) \rightarrow Qu(V)$$

will be denoted by

$$(Q \mapsto \overline{Q} ) : Qu(V) \rightarrow \text{Sym}_2(V^2, R) \cong \text{Sym}(V, V^\ast),$$

so that $Q = \overline{Q} ◦ (1_V, 1_V)$, i.e.,

$$Q(u) = \overline{Q}(u, u) = (\overline{Q} u)u \text{ for all } u \in V.$$  

(27.14)

We say that $Q \in Qu(V)$ is **non-degenerate** if $\overline{Q} \in \text{Sym}(V, V^\ast)$ is injective, **positive** [**negative**] if $\text{Rng} Q \subset P$ [$\text{Rng} Q \subset -P$], **strictly positive** [**strictly negative**] if $Q > 0(V^\ast) \subset P^\ast$ [$Q > 0(V^\ast) \subset -P^\ast$]. We say that Q is **single-signed** if it is either positive or negative, **double-signed** otherwise.

The same adjectives will be used for the corresponding bilinear form $\overline{Q}$.

**Example:** Let V be a linear space. The mapping

$$(L \mapsto \text{tr}_V(L^2)) : \text{Lin}V \rightarrow \mathbb{R}$$

(27.15)

is a quadratic form on the algebra of lineons LinV, i.e. a member of Qu(LinV). The corresponding element of Sym$_2((\text{Lin}V)^2, \mathbb{R})$ is given by $(L, M) \mapsto \text{tr}_V(LM)$. It follows from the Representation Theorem for Linear Forms on LinV (Sect. 26) that the quadratic form (27.15) is non-degenerate.

Using the symmetry and bilinearity of $\overline{Q}$, one obtains the formulas

$$Q(\xi u) = \xi^2 Q(u) \text{ for all } \xi \in \mathbb{R}, \ u \in V,$$

(27.16)

$$Q(u + v) = Q(u) + 2 \overline{Q}(u, v) + Q(v) \text{ for all } u, v \in V.$$  

(27.17)
As a consequence of (27.17), we find
\[
Q(u, v) = \frac{1}{2} (Q(u + v) - Q(u) - Q(v))
\]
\[= \frac{1}{2} (Q(u) + Q(v) - Q(u - v))
\]
\[= \frac{1}{4} (Q(u + v) - Q(u - v)),
\]
which give various ways of expressing \(\overline{Q}\) explicitly in terms of \(Q\).

The following result follows from Prop. 4 of Sect. 24, from (24.11), and from Prop. 1 of Sect. 14.

**Proposition 3:** Let \(V\) and \(W\) be linear spaces and let \(L \in \text{Lin}(V, W)\). For every \(Q \in \text{Qu}(W)\) we have \(Q \circ L \in \text{Qu}(V)\) and
\[
\overline{Q \circ L} = \overline{Q} \circ (L \times L) = L^\top \overline{Q} L.
\] (27.19)

The mapping
\[
(Q \mapsto \overline{Q \circ L}) : \text{Qu}(W) \rightarrow \text{Qu}(V)
\] (27.20)
is linear; it is invertible if and only if \(L\) is invertible.

If \(U\) is a subspace of \(V\) and \(Q \in \text{Qu}(V)\), then \(Q|_U \in \text{Qu}(U)\) and \(Q|_U = \overline{Q}|_{U \times U}\). If \(Q\) is positive, strictly positive, negative, or strictly negative, so is \(Q|_U\). However, \(Q|_U\) need not be non-degenerate even if \(Q\) is.

**Proposition 4:** A positive [negative] quadratic form is strictly positive [strictly negative] if and only if it is non-degenerate.

**Proof:** Let \(Q \in \text{Qu}(V)\) and \(u \in \text{Null} \overline{Q}\), so that \(\overline{Q} u = 0\). Then
\[
0 = (\overline{Q} u) u = \overline{Q} (u, u) = Q(u).
\] (27.21)

Now, if \(Q\) is strictly positive then (27.21) can hold only if \(u = 0\), which implies that \(\text{Null} \overline{Q} = \{0\}\) and hence that \(\overline{Q}\) is injective, i.e. that \(Q\) is positive and non-degenerate. Assume that \(Q\) is positive and non-degenerate and that \(Q(v) = 0\) for a given \(v \in V\). Using (27.18) we find that
\[
\xi \overline{Q} (u, v) = \overline{Q} (u, \xi v) = \frac{1}{2} (Q(u) + Q(\xi v) - Q(u - \xi v)),
\]
and hence, since \(Q(u - \xi v) \geq 0\) and \(Q(\xi v) = \xi^2 Q(v) = 0\),
\[
\xi \overline{Q} (u, v) \leq \frac{1}{2} Q(u)
\]
for all \(v \in V\) and all \(\xi \in \mathbb{R}^\times\). It is easily seen that this is possible only when \(0 = \overline{Q} (u, v) = (\overline{Q} u)v\) for all \(v \in V\), i.e. if \(\overline{Q} u = 0\). Since \(\overline{Q}\) is injective,
it follows that \( u = 0 \). We have shown that \( Q(u) = 0 \) implies \( u = 0 \), and hence, since \( Q \) was assumed to be positive, that \( Q \) is strictly positive. \( \blacksquare \)

Notes 27

(1) Many people use the clumsy terms “positive semidefinite” or “non-negative” when we speak of a “positive” quadratic or bilinear form. They then use “positive definite” or “positive” when we use “strictly positive”.

(2) The terms “single-signed” and “double-signed” for quadratic and bilinear forms are used here for the first time. They are clearer than the terms “definite” and “indefinite” found in the literature, sometimes with somewhat different meanings.

28 Problems for Chapter 2

1. Let \( \mathcal{V} \) be a linear space and let \( L \in \text{Lin}\mathcal{V} \) and \( \lambda, \mu \in \mathbb{R} \) with \( \lambda \neq \mu \) be given. Prove that

\[
\text{Null } (L^T - \mu 1_{\mathcal{V}^*}) \subset (\text{Null } (L - \lambda 1_{\mathcal{V}}))^\perp.
\]  

(\text{P2.1})

2. Let \( n \in \mathbb{N} \) be given and let the linear space \( \mathcal{P}_n \) be defined as in Problem 4 in Chap.1. For each \( k \in n! \), let \( \beta_k : \mathcal{P}_n \to \mathbb{R} \) be defined by

\[
\beta_k(f) := f^{(k)}(0) \quad \text{for all } f \in \mathcal{P}_n,
\]  

(\text{P2.2})

where \( f^{(k)} \) denotes the \( k \)’th derivative of \( f \) (see Sect. 08). Note that, for each \( k \in n! \), \( \beta_k \) is linear, so that \( \beta_k \in \mathcal{P}_n^* \).

(a) Determine a list \( h := (h_k \mid k \in n!) \) in \( \mathcal{P}_n \) such that

\[
\beta_k h_j = \delta_{k,j} \quad \text{for all } k, j \in n!.
\]  

(\text{P2.3})

(b) Show that the list \( h \) determined in (a) is a basis of \( \mathcal{P}_n \) and that the dual of \( h \) is given by

\[
h^* = \beta := (\beta_k \mid k \in n!).
\]  

(\text{P2.4})

(Hint: Apply Props.4 and 5 of Sect. 23).
28. PROBLEMS FOR CHAPTER 2

For each $t \in \mathbb{R}$, let the evaluation $ev_t : \mathcal{P}_n \to \mathbb{R}$ be defined by

$$ev_t(f) := f(t) \text{ for all } f \in \mathcal{P}_n \quad (P2.5)$$

(see Sect. 04) and note that $ev_t$ is linear, so that $ev_t \in \mathcal{P}_n^\ast$.

(c) Let $t \in \mathbb{R}$ be given. Determine $\lambda := \ln c_\beta^{-1}ev_t \in \mathbb{R}^n$, where $\beta$ is the basis of $\mathcal{P}_n^\ast$ defined by (P2.4) and (P2.2), so that

$$ev_t = \ln c_\beta \lambda = \sum_{k \in n} \lambda_k \beta_k. \quad (P2.6)$$

3. Let $n \in \mathbb{N}$ be given and let the linear space $\mathcal{P}_n$ be defined as in Problem 4 of Chap.1. Also, let a subset $F$ of $\mathbb{R}$ with $n$ elements be given, so that $n = \#F$.

(a) Determine a family $g := (g_s | s \in F)$ in $\mathcal{P}_n$ such that

$$ev_t(g_s) = \delta_{t,s} \text{ for all } t, s \in F, \quad (P2.7)$$

where $ev_t \in \mathcal{P}_n^\ast$ is the evaluation given by (P2.5).

(b) Show that the family $g$ determined in (a) is a basis of $\mathcal{P}_n$ and that the dual of $g$ is given by

$$g^* = (ev_t | t \in F). \quad (P2.8)$$

(Hint: Apply Props.4 and 5 of Sect. 23).

4. Let $\mathcal{V}$ be a linear space. Let the mappings $\Lambda$ and $\Upsilon$ from $\text{Lin}\mathcal{V} \times \text{Lin}\mathcal{V}$ to $\text{Lin}\mathcal{V}$ be defined by

$$\Lambda(L, M) := LM - ML \quad (P2.9)$$

and

$$\Upsilon(L, M) := LM + ML \quad (P2.10)$$

for all $L, M \in \text{Lin}\mathcal{V}$.

(a) Show that $\Lambda$ is bilinear and skew, so that $\Lambda \in \text{Skew}_2((\text{Lin}\mathcal{V})^2, \text{Lin}\mathcal{V})$. Also, show that...
\[ \sum_{\gamma \in C_3} \Lambda(\Lambda(L_{\gamma(1)}, L_{\gamma(2)}), L_{\gamma(3)}) = 0 \quad (P2.11) \]

for every triple \((L_1, L_2, L_3)\) in \(\text{Lin}\mathcal{V}\), where \(C_3\) denotes the set of cyclic permutations of 3, i.e.

\[ C_3 := \left\{ 13, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix} \right\}. \]

(b) Show that \(\Upsilon\) is bilinear and symmetric, so that \(\Upsilon \in \text{Sym}_2((\text{Lin}\mathcal{V})^2, \text{Lin}\mathcal{V})\). Also, show that

\[ \Upsilon(L, \Upsilon(\Upsilon(L, L), M)) = \Upsilon(\Upsilon(L, L), \Upsilon(L, M)) \quad (P2.12) \]

for all \(L, M \in \text{Lin}\mathcal{V}\).

Remark: The mapping \(\Lambda\) endows \(\text{Lin}\mathcal{V}\) with the structure of a “Lie-algebra” and the mapping \(\Upsilon\) endows \(\text{Lin}\mathcal{V}\) with the structure of a “Jordan-algebra”. The theory of these is an important topic in abstract algebra.

5. Let \(D, M \in \text{Lin} C^\infty(\mathbb{R})\) and \(\mathcal{P}_n\), with \(n \in \mathbb{N}\), be defined as in Problems 4 and 5 of Chapt.1.

(a) Show that \(\mathcal{P}_n\) is \(D\)-subspace, a \((MD)\)-subspace, and a \((DM)\)-subspace of \(C^\infty(\mathbb{R})\). Calculate \(\text{tr}\mathcal{P}_n, \text{tr}(MD)\mathcal{P}_n, \text{and} \text{tr}(DM)\mathcal{P}_n\).

(b) Prove: If \(\mathcal{V}\) is a linear space of finite and non-zero dimension, there do not exist \(L, K \in \text{Lin}\mathcal{V}\) such that \(LK - KL = \mathbf{1}_\mathcal{V}\). (Compare this assertion with the one of Problem 5 of Chap.1.)

6. Let \(\mathcal{V}\) be a linear space.

(a) Prove that \(\Omega \in (\text{Lin}\mathcal{V})^*\) satisfies

\[ \Omega(LM) = \Omega(ML) \quad \text{for all} \quad L, M \in \text{Lin}\mathcal{V} \quad (P2.13) \]

if and only if \(\Omega = \xi \text{tr}\mathcal{V}\) for some \(\xi \in \mathbb{R}\).

(Hint: Consider the case when \(L\) and \(M\) in \((P2.13)\) are tensor products.)

(b) Consider the left-multiplication mapping \(\text{Le}_L \in \text{Lin(Lin}\mathcal{V})\) defined by \((P1.1)\) for each \(L \in \text{Lin}\mathcal{V}\). Prove that

\[ \text{tr}_{\text{Lin}\mathcal{V}}(\text{Le}_L) = (\dim \mathcal{V})\text{tr}_\mathcal{V}L \quad \text{for all} \quad L \in \text{Lin}\mathcal{V} \quad (P2.14) \]
7. Let $U_1$ and $U_2$ be supplementary subspaces of the given linear space $V$ and let $P_1, P_2$ be the projections associated with $U_1, U_2$ according to Prop.4 of Sect. 19. Prove that, for every $L \in \text{Lin}V$, we have
\[
\text{tr}_V L = \text{tr}_{U_1}(P_1 L|_{U_1}) + \text{tr}_{U_2}(P_2 L|_{U_2}).
\] (P2.15)

(Hint: Use part (a)).

8. Let $V$ be a linear space and put $n := \dim V$.

(a) Given a list basis $b := (b_i | i \in n)$ of $V$ construct a basis of $\text{Sym}_2(V^2, \mathbb{R}) \cong \text{Sym}(V, V^*)$ by a procedure analogous to the one described in the proof of Prop.1 of Sect. 27.

(b) Show that, for every $\lambda \in V^*$, the function $\lambda^2 : V \to \mathbb{R}$ obtained from $\lambda$ by value-wise squaring (i.e. by $\lambda^2(v) := (\lambda v)^2$ for all $v \in V$), is a quadratic form on $V$, i.e. that $\lambda^2 \in \text{Qu}(V)$. (See Def.1 in Sect. 27)

(c) Prove that
\[
\text{Lsp}\{\lambda^2 | \lambda \in V^*\} = \text{Qu}(V).
\] (P2.16)
Chapter 3

Flat Spaces

In this chapter, the term “linear space” will be used as a shorthand for “linear space over the real field $\mathbb{R}$”. (Actually, many definitions remain meaningful and many results remain valid when $\mathbb{R}$ is replaced by an arbitrary field. The interested reader will be able to decide for himself when this is the case.)

31 Actions of Groups

Let $\mathcal{E}$ be a set and consider the permutation group $\text{Perm}\, \mathcal{E}$, which consists of all invertible mappings from $\mathcal{E}$ onto itself. $\text{Perm}\, \mathcal{E}$ is a group under composition, and the identity mapping $1_\mathcal{E}$ is the neutral of this group (see Sect. 06).

**Definition 1:** Let $\mathcal{E}$ be a non-empty set and $\mathcal{G}$ a group. By an **action of $\mathcal{G}$ on $\mathcal{E}$** we mean a group homomorphism $\tau : \mathcal{G} \rightarrow \text{Perm}\, \mathcal{E}$.

We write $\tau_g$ for the value of $\tau$ at $g \in \mathcal{G}$. To say that $\tau$ is an action of $\mathcal{G}$ on $\mathcal{E}$ means that $\tau_{\text{cmb}(g,h)} = \tau_g \circ \tau_h$ for all $g, h \in \mathcal{G}$, where cmb denotes the combination of $\mathcal{G}$ (see Sect. 06). If $n$ is the neutral of $\mathcal{G}$, then $\tau_n = 1_\mathcal{E}$, and if $\text{rev}(g)$ is the group-reverse of $g \in \mathcal{G}$, then $\tau_{\text{rev}(g)} = \tau_g^{-1}$.

**Examples:**

1. If $\mathcal{G}$ is any subgroup of $\text{Perm}\, \mathcal{E}$, then the inclusion mapping $1_{\mathcal{G} \subseteq \text{Perm}\, \mathcal{E}}$ is an action of $\mathcal{G}$ on $\mathcal{E}$. For subgroups of $\text{Perm}\, \mathcal{E}$ it is always understood that the action is this inclusion.

2. If $\mathcal{E}$ is a set with a prescribed structure, then the automorphisms of $\mathcal{E}$ form a subgroup of $\text{Perm}\, \mathcal{E}$ which acts on $\mathcal{E}$ by inclusion in $\text{Perm}\, \mathcal{E}$ as
explained in Example 1. For instance, if $V$ is a linear space, then the
elementary group $\text{Lis} V$ acts on $V$.

3. If $V$ is a linear space, define $\tau : V \to \text{Perm} V$ by letting $\tau_v$ be the
operation of adding $v \in V$, so that $\tau_v(u) := u + v$ for all $u \in V$. It is
easily seen that $\tau$ is a homomorphism of the additive group of $V$ into
the permutation group $\text{Perm} V$, i.e., $\tau$ is an action of $V$ on itself.

4. If $V$ is a linear space, define $\mu_l : P^x \to \text{Perm} V$ by letting $\mu_l(\lambda) :=
\lambda 1_V$ for all $\lambda \in P$ and $v \in V$. It is evident that $\mu_l$ is a homomorphism of the
multiplicative group $P^x$ into $\text{Perm} V$ and hence an action of $P^x$ on $V$.

$\mu_l(P^x)$ is a subgroup of $\text{Lis} V$ that is isomorphic to $P^x$; namely,
$\mu_l(P^x) = P^x 1_V$.

Let $\tau$ be an action of $G$ on $E$. We define a relation $\sim_{\tau}$ on $E$ by

$$x \sim_{\tau} y :\iff y = \tau_g(x) \text{ for some } g \in G. \quad (31.1)$$

It is easily seen that $\sim_{\tau}$ is an equivalence relation. The equivalence
classes are called the orbits in $E$ under the action $\tau$ of $G$. If $x \in E$, then
$\{\tau_g(x) | g \in G\}$ is the orbit to which $x$ belongs, also called the orbit of $x$.

**Definition 2:** An action $\tau$ of a group $G$ on a non-empty set $E$ is said
to be **transitive** if for all $x, y \in E$ there is $g \in G$ such that $\tau_g(x) = y$.
The action is said to be **free** if $\tau_g$, with $g \in G$, can have a fixed point (see
Sect. 03) only if $g = n$, the neutral of $G$.

To say that an action is transitive means that all of $E$ is the only orbit
under the action. Of course, since $\tau_n = 1_E$, all points in $E$ are fixed points
of $\tau_n$.

If the action $\tau$ of $G$ on $E$ is free, one easily sees that $\tau : G \to \text{Perm} E$
must be injective, and hence that $\tau_g(G)$ is a subgroup of $\text{Perm} E$ that is an
isomorphic image of $G$.

**Proposition 1:** An action $\tau$ of a group $G$ on a non-empty set $E$ is both
transitive and free if and only if for all $x, y \in E$ there is exactly one $g \in G$
such that $\tau_g(x) = y$.

**Proof:** Assume the action is both transitive and free and that $x, y \in E$
are given. If $\tau_g(x) = y$ and $\tau_h(x) = y$, then $x = (\tau_g)^{-1}(y) = \tau_{\text{rev}(g)}(y) =
\tau_{\text{rev}(g)}(\tau_h(x)) = (\tau_{\text{rev}(g)} \circ \tau_h)(x) = \tau_{\text{camb}(\text{rev}(g), h)}(x)$. Since the action is free,
it follows that $\text{camb}(\text{rev}(g), h) = n$ and hence $g = h$. Thus there can be at
most one $g \in G$ such that $\tau_g(x) = y$. The transitivity of the action assures
the existence of such $g$. 
Assume now that the condition is satisfied. The action is then transitive. Let \( \tau_g(x) = x \) for some \( x \in \mathcal{E} \). Since \( \tau_n(x) = 1_{\mathcal{E}}(x) = x \) it follows from the uniqueness that \( g = n \). Hence the action is free. ■

32 Flat Spaces and Flats

Definition 1: A flat space is a non-empty set \( \mathcal{E} \) endowed with structure by the prescription of

(i) a commutative subgroup \( V \) of \( \text{Perm} \mathcal{E} \) whose action is transitive.

(ii) a mapping \( \text{sm} : \mathbb{R} \times V \to V \) which makes \( V \) a linear space when composition is taken as the addition and \( \text{sm} \) as the scalar multiplication in \( V \).

The linear space \( V \) is then called the translation space of \( \mathcal{E} \).

The elements of \( \mathcal{E} \) are called points and the elements of \( V \) translations or vectors. If \( \xi \in \mathbb{R} \) and \( v \in V \) we write, as usual, \( \xi v \) for \( \text{sm}(\xi, v) \). In \( V \), we use additive notation for composition, i.e., we write \( u + v \) for \( v \circ u \), \( -v \) for \( v^- \), and \( 0 \) for \( 1_{\mathcal{E}} \).

Proposition 1: The action of the translation space \( V \) on the flat space \( \mathcal{E} \) is free.

Proof: Suppose that \( v \in V \) satisfies \( v(x) = x \) for some \( x \in \mathcal{E} \). Let \( y \in \mathcal{E} \) be given. Since the action of \( V \) is transitive, we may choose \( u \in V \) such that \( y = u(x) \). Using the commutativity of the group \( V \), we find that

\[
 v(y) = v(u(x)) = (v \circ u)(x) = (u \circ v)(x) = u(v(x)) = u(x) = y.
\]

Since \( y \in \mathcal{E} \) was arbitrary, it follows that \( v = 1_{\mathcal{E}} \). ■

This Prop.1 and Prop.1 of Sect.31 have the following immediate consequence:

Proposition 2: There is exactly one mapping

\[
 \text{diff} : \mathcal{E} \times \mathcal{E} \to V
\]

such that

\[
 (\text{diff}(x, y))(y) = x \quad \text{for all} \quad x, y \in \mathcal{E}. \tag{32.1}
\]

We call \( \text{diff}(x, y) \in V \) the point-difference between \( x \) and \( y \) and use the following simplified notations:

\[
 x - y := \text{diff}(x, y) \quad \text{when} \quad x, y \in \mathcal{E},
\]

\[
 x + v := v(x) \quad \text{when} \quad x \in \mathcal{E}, v \in V.
\]
CHAPTER 3. FLAT SPACES

The following rules are valid for all \( x, y, z \in E \) and for all \( u, v \in V \).

\[
\begin{align*}
    x - x &= 0, \\
    x - y &= -(y - x), \\
    (x - y) + (y - z) &= x - z, \\
    (x + u) + v &= x + (u + v), \\
    x + v = y &\iff v = y - x.
\end{align*}
\]

In short, an equation involving points and vectors is valid if it is valid according to the rules of ordinary algebra, provided that all expressions occurring in the equation make sense. Of course, an expression such as \( x + y \) does not make sense when \( x, y \in E \).

The following notations, which are similar to notations introduced in Sect.06., are very suggestive and useful. They apply when \( x \in E, v \in V, \mathcal{H}, \mathcal{Y} \in \text{Sub}E, S \in \text{Sub}V \).

\[
\begin{align*}
    x + S &:= \{x + s \mid s \in S\}, \\
    \mathcal{H} + v &:= v_\to(\mathcal{H}) = \{x + v \mid x \in \mathcal{H}\}, \\
    \mathcal{H} + S &:= \bigcup_{s \in S} s_\to(\mathcal{H}) = \{x + s \mid x \in \mathcal{H}, s \in S\}, \\
    \mathcal{H} - \mathcal{Y} &:= \{x - y \mid x \in \mathcal{H}, y \in \mathcal{Y}\}.
\end{align*}
\]

The first three of the above are subsets of \( E \), the last is a subset of \( V \). We have \( V = E - E \) and we sometimes use \( E - E \) as a notation for the translation space of \( E \).

Flat spaces serve as mathematical models for physical planes and spaces. The vectors, i.e. the mappings in \( V \) (except the identity mapping \( 0 = 1_E \)), are to be interpreted as parallel shifts (hence the term “translation”). If a vector \( v \in V \) is given, we can connect points \( x, y, \ldots \), with their images \( x + v, y + v, \ldots \), by drawing arrows as shown in Figure 1. In this sense, vectors can be represented pictorially by arrows. The commutativity of \( V \) is illustrated in Figure 2. The assumption that the action of \( V \) is transitive corresponds to the fact that given any two points \( x \) and \( y \), there exists a vector that carries \( x \) to \( y \).
It often happens that a set $E$, a linear space $V$, and an action of the additive group of $V$ on $E$ are given. If this action is transitive and injective, then $E$ acquires the structure of a flat space whose translation space is the isomorphic image of $V$ in $\text{Perm} E$ under the given action. Under such circumstances, we identify $V$ with its isomorphic image and, using poetic license, call $V$ itself the translation space of $E$. If we wish to emphasize the fact that $V$ becomes the translation space only after such an identification, we say that $V$ acts as an external translation space of $E$.

Let $V$ be a linear space. The action $\tau : V \to \text{Perm} V$ described in Example 3 of the previous section is easily seen to be transitive and injective. Hence we have the following trivial but important result:

**Proposition 3:** Every linear space $V$ has the natural structure of a flat space. The space $V$ becomes its own external translation space by associating with each $v \in V$ the mapping $u \mapsto u + v$ from $V$ to itself.

The linear-space structure of $V$ embodies more information than its flat-space structure. As a linear space, $V$ has a distinguished element, $0$, but as a flat space it is homogeneous in the sense that all of its elements are of equal standing. Roughly, the flat structure of $V$ is obtained from its linear structure by forgetting where $0$ is.

Note that the operations involving points and vectors introduced earlier, when applied to the flat structure of a linear space, reduce to ordinary linear-space operations. For example, point-differences reduce to ordinary differences.

Of course, the set $\mathbb{R}$ of reals has the natural structure of a flat space whose (external) translation space is $\mathbb{R}$ itself, regarded as a linear space.

We now assume that a flat space $E$ with translation space $V$ is given. A subset $F$ of $E$ will inherit from $E$ the structure of a flat space if the translations of $E$ that leave $F$ invariant can serve, after adjustment (see Sect.03), as translations of $F$. The precise definition is this:
Definition 2: A non-empty subset $\mathcal{F}$ of $\mathcal{E}$ is called flat subspace of $\mathcal{E}$, or simply a flat in $\mathcal{E}$, if the set

$$U := \{u \in V \mid F + u \subset F\}$$

is a linear subspace of $V$ whose additive group acts transitively on $F$ under the action which associates with every $u \in U$ the mapping $u_F$ of $F$ onto itself.

The action of $U$ on $F$ described in this definition endows $F$ with the natural structure of a flat space whose (external) translation space is $U$. We say that $U$ is the direction space of $F$. The following result is immediate from Def.2.

Proposition 4: Let $U$ be a subspace of $V$. A non-empty subset $F$ of $\mathcal{E}$ is a flat with direction space $U$ if and only if

$$F - F \subset U \quad \text{and} \quad F + U \subset F.$$  \hfill (32.3)

Let $U$ be a subspace of $V$. The inclusion of the additive group of $U$ in the group Perm $\mathcal{E}$ is an action of $U$ on $\mathcal{E}$. This action is transitive only when $U = V$. The orbits in $\mathcal{E}$ under the action of $U$ are exactly the flats with direction space $U$. Since $x + U$ is the orbit of $x \in \mathcal{E}$ we obtain:

Proposition 5: Let $U$ be a subspace of $V$. A subset $F$ of $\mathcal{E}$ is a flat with direction space $U$ if and only if it is of the form

$$F = x + U$$ \hfill (32.4)

for some $x \in \mathcal{E}$.

We say that two flats are parallel if the direction space of one of them is included in the direction space of the other. For example, two flats with the same direction space are parallel; if they are distinct, then their intersection is empty.

The following result is an immediate consequence of Prop.4.

Proposition 6: The intersection of any collection of flats in $\mathcal{E}$ is either empty or a flat in $\mathcal{E}$. If it is not empty, then the direction space of the intersection is the intersection of the direction spaces of the members of the collection.

In view of the remarks on span-mappings made in Sect.03, we have the following result.

Proposition 7: Given any non-empty subset $S$ of $\mathcal{E}$, there is a unique smallest flat that includes $S$. More precisely: There is a unique flat that includes $S$ and is included in every flat that includes $S$. It is called the flat
span of $S$ and is denoted by $\text{Fsp } S$. We have $\text{Fsp } S = S$ if and only if $S$ is a flat.

If $x$ and $y$ are two distinct points in $\mathcal{E}$, then

$$\overrightarrow{xy} := \text{Fsp}\{x, y\}$$

is called the line passing through $x$ and $y$.

It is easily seen that, if $S \subset \mathcal{E}$ and if $q \in S$, then $\text{Lsp}(S - q)$ is the direction space of $\text{Fsp } S$ and hence, by Prop. 5,

$$\text{Fsp } S = q + \text{Lsp}(S - q).$$

Let $\mathcal{E}_1, \mathcal{E}_2$ be flat spaces with translation spaces $V_1, V_2$. The set-product $\mathcal{E}_1 \times \mathcal{E}_2$ then has the natural structure of a flat space as follows: The (external) translation space of $\mathcal{E}_1 \times \mathcal{E}_2$ is the product space $V_1 \times V_2$. The action of $V_1 \times V_2$ on $\mathcal{E}_1 \times \mathcal{E}_2$ is defined by associating with each $(v_1, v_2) \in V_1 \times V_2$ the mapping $(x_1, x_2) \mapsto (x_1 + v_1, x_2 + v_2)$ of $\mathcal{E}_1 \times \mathcal{E}_2$ onto itself. More generally, if $(\mathcal{E}_i \mid i \in I)$ is any family of flat spaces, then $\times (\mathcal{E}_i \mid i \in I)$ has the natural structure of a flat space whose (external) translation space is the product-space $\times (V_i \mid i \in I)$, where $V_i$ is the translation space of $\mathcal{E}_i$ for each $i \in I$. The action of $\times (V_i \mid i \in I)$ on $\times (\mathcal{E}_i \mid i \in I)$ is defined by

$$(x_i \mid i \in I) + (v_i \mid i \in I) := (x_i + v_i \mid i \in I).$$

We say that a flat space $\mathcal{E}$ is finite-dimensional if its translation space is finite-dimensional. If this is the case, we define the dimension of $\mathcal{E}$ by

$$\dim \mathcal{E} := \dim (\mathcal{E} - \mathcal{E}).$$

Let $\mathcal{E}$ be any flat space. The only flat in $\mathcal{E}$ whose direction space is $V := \mathcal{E} - \mathcal{E}$ is $\mathcal{E}$ itself. The singleton subsets of $\mathcal{E}$ are the zero-dimensional flats, their direction space is the zero-subspace $\{0\}$ of $V$. The one-dimensional flats are called lines, the two-dimensional flats are called planes, and, if $n := \dim \mathcal{E} \in \mathbb{N}^\times$, then the $(n - 1)$-dimensional flats are called hyperplanes. The only n-dimensional flat is $\mathcal{E}$ itself.

Let $V$ be a linear space, which is its own translation space when regarded as a flat space (see Prop. 3). A subset $\mathcal{U}$ of $V$ is then a subspace of $V$ if and only if it is a flat in $V$ and contains $0$. 
Notes 32

(1) The traditional term for our “flat space” is “affine space”. I believe that the term “flat space” is closer to the intuitive content of the concept.

(2) A fairly rigorous definition of the concept of a flat space, close to the one given here, was given 65 years ago by H. Weyl (Raum, Zeit, Materie; Springer, 1918). After all these years, the concept is still not used as much as it should be. One finds explanations (often bad ones) of the concept in some recent geometry and abstract algebra textbooks, but the concept is almost never introduced in analysis. One of my reasons for writing this book is to remedy this situation, for I believe that flat spaces furnish the most appropriate conceptual background for analysis.

(3) Strictly speaking, the term “point” for an element of a flat space and the term “vector” for an element of its translation space are appropriate only if the flat space is used as a primitive structure to describe some geometrical-physical reality. Sometimes a flat space may come up as a derived structure, and then the terms “point” and “vector” may not be appropriate to physical reality. Nevertheless, we will use these terms when dealing with the general theory.

(4) What we call a “flat” or a “flat subspace” is sometimes called a “linear manifold”, “translated subspace”, or “affine subset”, especially when it is a subset of a linear space.

(5) What we call “flat span” is sometimes called “affine hull” and the notation aff$S$ is then used instead of Fsp$S$.

33 Flat Mappings

Let $E, E'$ be flat spaces with translation spaces $V, V'$. We say that a mapping $\alpha : E \rightarrow E'$ is flat if, roughly, it preserves translations and scalar multiples of translations. This means that if any two points in $E$ are related by a given translation $v$, their $\alpha$-values must be related by a corresponding translation $v'$, and if $v$ is replaced by $\xi v$ then $v'$ can be replaced by $\xi v'$. The precise statement is this:

**Definition 1:** A mapping $\alpha : E \rightarrow E'$ is called a flat mapping if for every $v \in V$, there is a $v' \in V'$ such that

$$\alpha \circ (\xi v) = (\xi v') \circ \alpha \quad \text{for all } \xi \in \mathbb{R}. \quad (33.1)$$

If we evaluate (33.1) at $x \in E$ and take $\xi := 1$, we see that $v' = \alpha(x + v) - \alpha(x)$ is uniquely determined by $v$. Thus, the following result is immediate from the definition.

**Proposition 1:** A mapping $\alpha : E \rightarrow E'$ is flat if and only if there is a mapping $\nabla \alpha : V \rightarrow V'$ such that

$$(\nabla \alpha)(v) = \alpha(x + v) - \alpha(x) \quad \text{for all } x \in E \text{ and } v \in V, \quad (33.2)$$
and

\[(\nabla \alpha)(\xi v) = \xi(\nabla \alpha(v)) \quad \text{for all} \quad v \in \mathcal{V} \quad \text{and} \quad \xi \in \mathbb{R}. \quad (33.3)\]

The mapping \(\nabla \alpha\) is uniquely determined by \(\alpha\) and is called the gradient of \(\alpha\).

Note that condition (33.2) is equivalent to

\[\alpha(x) - \alpha(y) = (\nabla \alpha)(x - y) \quad \text{for all} \quad x, y \in \mathcal{E}. \quad (33.4)\]

**Proposition 2:** The gradient \(\nabla \alpha\) of a flat mapping \(\alpha\) is linear, i.e., \(\nabla \alpha \in \text{Lin}(\mathcal{V}, \mathcal{V}')\).

**Proof:** Choose \(x \in \mathcal{E}\). Using condition (33.2) three times, we see that

\[
(\nabla \alpha)(v + u) = \alpha(x + (v + u)) - \alpha(x) = (\alpha((x + v) + u) - \alpha(x + v)) + (\alpha(x + v) - \alpha(x)) = (\nabla \alpha)(u) + (\nabla \alpha)(v)
\]

is valid for all \(u, v \in \mathcal{V}\). Hence \(\nabla \alpha\) preserves addition. The condition (33.3) states that \(\nabla \alpha\) preserves scalar multiplication. ■

**Theorem on Specification of Flat Mappings:** Let \(q \in \mathcal{E}, q' \in \mathcal{E}'\) and \(L \in \text{Lin}(\mathcal{V}, \mathcal{V}')\) be given. Then there is exactly one flat mapping \(\alpha : \mathcal{E} \to \mathcal{E}'\) such that \(\alpha(q) = q'\) and \(\nabla \alpha = L\). It is given by

\[\alpha(x) := q' + L(x - q) \quad \text{for all} \quad x \in \mathcal{E}. \quad (33.5)\]

**Proof:** Using (33.4) with \(y\) replaced by \(q\) we see that (33.5) must be valid when \(\alpha\) satisfies the required conditions. Hence \(\alpha\) is uniquely determined. Conversely, if \(\alpha : \mathcal{E} \to \mathcal{E}'\) is defined by (33.5), one easily verifies that it is flat and satisfies the conditions. ■

**Examples:**

1. All constant mappings from one flat space into another are flat. A flat mapping is constant if and only if its gradient is zero.

2. If \(F\) is a flat in \(\mathcal{E}\), then the inclusion \(1_{F \subset \mathcal{E}}\) is flat. Its gradient is the inclusion \(1_{\mathcal{U} \subset \mathcal{V}}\) of the direction space \(\mathcal{U}\) of \(F\) in the translation space \(\mathcal{V}\) of \(\mathcal{E}\).

3. A function \(a : \mathbb{R} \to \mathbb{R}\) is flat if and only if it is of the form \(a = \xi t + \eta\) with \(\xi, \eta \in \mathbb{R}\). The gradient of \(a\) is \(\xi \in \text{Lin}\mathbb{R}\), which one identifies with the number \(\xi \in \mathbb{R}\) (see Sect.25). The graph of \(a\) is a straight line with slope \(\xi\). This explains our use of the term “gradient” (gradient means slope or inclination).
4. The translations $\mathbf{v} \in \mathcal{V}$ of a flat space $\mathcal{E}$ are flat mappings of $\mathcal{E}$ into itself. They all have the same gradient, namely $1_\mathcal{V}$. In fact, a flat mapping from $\mathcal{E}$ into itself is a translation if and only if its gradient is $1_\mathcal{V}$.

5. Let $\mathcal{V}$ and $\mathcal{V}'$ be linear spaces, regarded as flat spaces (see Prop.3 of Sect.32). A mapping $L : \mathcal{V} \to \mathcal{V}'$ is linear if and only if it is flat and preserves zero, i.e. $L0 = 0'$. If this is the case, then $L$ is its own gradient.

6. Let $(\mathcal{E}_i \mid i \in I)$ be a family of flat spaces and let $(\mathcal{V}_i \mid i \in I)$ be the family of their translation spaces. Given $j \in I$, the evaluation mapping $\text{ev}_{\mathcal{E}_j} : \times (\mathcal{E}_i \mid i \in I) \to \mathcal{E}_j$ (see Sect.04) is a flat mapping when $\times (\mathcal{E}_i \mid i \in I)$ is endowed with the natural flat-space structure described in the previous section. The gradient of $\text{ev}_{\mathcal{E}_j}$ is the evaluation mapping $\text{ev}_{\mathcal{V}_j} : \times (\mathcal{V}_i \mid i \in I) \to \mathcal{V}_j$ (see Sect.14).

7. If $\mathcal{E}$ is a flat space and $\mathcal{V} := \mathcal{E} - \mathcal{E}$, then $((x, \mathbf{v}) \mapsto x + \mathbf{v}) : \mathcal{E} \times \mathcal{V} \to \mathcal{E}$ is a flat mapping. Its gradient is the vector-addition $\mathcal{V} \times \mathcal{V} \to \mathcal{V}$.

8. The point-difference mapping defined by (32.1) is flat. Its gradient is the vector-difference mapping $((u, \mathbf{v}) \mapsto (u - \mathbf{v})) : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$.

**Proposition 3:** Let $\alpha : \mathcal{E} \to \mathcal{E}'$ be a flat mapping, $\mathcal{F}$ a flat in $\mathcal{E}$ with direction space $\mathcal{U}$, and $\mathcal{F}'$ a flat in $\mathcal{E}'$ with direction space $\mathcal{U}'$. Then:

(i) $\alpha>\mathcal{F}$ is a flat in $\mathcal{E}'$ with direction space $(\nabla \alpha)>\mathcal{U}$.

(ii) $\alpha<\mathcal{F}'$ is either empty or a flat in $\mathcal{E}$ with direction space $(\nabla \alpha)<\mathcal{U}'$.

**Proof:** Using (33.4) and (33.2) one easily verifies that the conditions of Prop.4 of Sect.32 are verified in the situations described in (i) and (ii).

If $\mathcal{F}' := \{x'\}$ is a singleton in Prop.3, then $\mathcal{U}' = \{0\}$, and (ii) states that $\alpha<\mathcal{F}'$ is either empty or a flat in $\mathcal{E}$ whose direction space is $\text{Null} \nabla \alpha$. If $\mathcal{F} := \mathcal{E}$, then $\mathcal{U} = \mathcal{V}$ and (i) states that $\text{Rng} \alpha$ is a flat in $\mathcal{E}'$ whose direction space is $\text{Rng} \nabla \alpha$. Therefore, the following result is an immediate consequence of Prop.3.

**Proposition 4:** A flat mapping is injective [surjective] if and only if its gradient is injective [surjective].
The following two rules describe the behavior of flat mappings with respect to composition and inversion. They follow immediately from Prop.1.

**Chain Rule for Flat Mappings:** The composite of two flat mappings is again flat, and the gradient of their composite is the composite of their gradients. More precisely: If $\mathcal{E}, \mathcal{E}', \mathcal{E}''$ are flat spaces and $\alpha : \mathcal{E} \to \mathcal{E}'$ and $\beta : \mathcal{E}' \to \mathcal{E}''$ are flat mappings, then $\beta \circ \alpha : \mathcal{E} \to \mathcal{E}''$ is again a flat mapping and

$$\nabla (\beta \circ \alpha) = \nabla \beta \nabla \alpha.$$  \hspace{1cm} (33.6)

**Inversion Rule for Flat Mappings:** A flat mapping $\alpha : \mathcal{E} \to \mathcal{E}'$ is invertible if and only if its gradient $\nabla \alpha \in \text{Lin}(\mathcal{V}, \mathcal{V}')$ is invertible. If this is the case, then the inverse $\alpha^{-1} : \mathcal{E}' \to \mathcal{E}$ is also flat and

$$\nabla (\alpha^{-1}) = (\nabla \alpha)^{-1}.$$  \hspace{1cm} (33.7)

Hence every invertible flat mapping is a flat isomorphism.

The set of all flat automorphisms of flat space $\mathcal{E}$, i.e. all flat isomorphism from $\mathcal{E}$ to itself, is denoted by $\text{Fis} \mathcal{E}$. It is a subgroup of $\text{Perm} \mathcal{E}$. Let $\mathcal{V}$ be the translation space of $\mathcal{E}$ and let $h : \text{Fis} \mathcal{E} \to \text{Lis} \mathcal{V}$ be defined by $h(\alpha) := \nabla \alpha$ for all $\alpha \in \text{Fis} \mathcal{E}$. It follows from the Chain Rule and Inversion Rule for Flat Mappings that $h$ is a group-homomorphism. It is easily seen that $h$ is surjective and that the kernel of $h$ is the translation group $\mathcal{V}$ (see Sect.06).

Notes 33

1. The traditional terms for our “flat mapping” are “affine mapping” or “affine transformation”.

2. In one recent textbook, the notation $\alpha^\dagger$ and the unfortunate term “trace” are used for what we call the “gradient” $\nabla \alpha$ of a flat mapping $\alpha$.

34 Charge Distributions, Barycenters, Mass-Points

Let $\mathcal{E}$ be a flat space with translation space $\mathcal{V}$. By a **charge distribution** $\gamma$ on $\mathcal{E}$ we mean a family in $\mathbb{R}$, indexed on $\mathcal{E}$ with finite support, i.e., $\gamma \in \mathbb{R}^{(\mathcal{E})}$. We may think of the term $\gamma_x$ of $\gamma$ at $x$ as an electric charge placed at the point $x$. The total charge of the distribution $\gamma$ is its sum (see (15.2))

$$\sum_{x \in \mathcal{E}} \gamma_x = \sum_{x \in \text{Supt} \gamma} \gamma_x.$$  \hspace{1cm} (34.1)
CHAPTER 3. FLAT SPACES

**Definition 1:** We say that the charge distributions \(\gamma, \beta \in \mathbb{R}(E)\) are equivalent, and we write \(\gamma \sim \beta\), if \(\gamma\) and \(\beta\) have the same total charge, i.e. \(\sum_{x \in E} \gamma(x) = \sum_{y \in E} \beta(y)\), and if
\[
\sum_{x \in E} \gamma(x)(x-q) = \sum_{y \in E} \beta(y)(y-q) \quad \text{for all } q \in E. \tag{34.2}
\]

It is easily verified that the relation \(\sim\) thus defined is indeed an equivalence relation on \(\mathbb{R}(E)\). Moreover, we have \(\gamma \sim \beta\) if and only if \(\gamma - \beta \sim 0\).

Also, one easily proves the following result:

**Proposition 1:** If \(\gamma, \beta \in \mathbb{R}(E)\) have the same total charge, i.e. \(\sum_{x \in E} \gamma(x) = \sum_{y \in E} \beta(y)\), and if (34.2) holds for some \(q \in E\), then \(\gamma, \beta\) are equivalent, i.e., (34.2) holds for all \(q \in E\).

Given any \(x \in E\) we define \(\delta_x \in \mathbb{R}(E)\) by
\[
(\delta_x)_y = \begin{cases} 0 & \text{if } x \neq y \\ 1 & \text{if } x = y \end{cases}, \tag{34.3}
\]
in words, the distribution \(\delta_x\) places a unit charge at \(x\) and no charge anywhere else. Actually, \(\delta_x\) is the same as the \(x\)-term \(\delta^E_x\) of the standard basis \(\delta^E\) of \(\mathbb{R}(E)\) as defined in Sect.16.

The following result is an immediate consequence of Prop.1.

**Proposition 2:** Given \(\gamma \in \mathbb{R}(E)\), \(b \in E\) and \(\sigma \in \mathbb{R}\), we have \(\gamma \sim \sigma \delta_b\) if and only if
\[
\sigma = \sum_{x \in E} \gamma(x) \quad \text{and} \quad \sum_{x \in E} \gamma(x)(x-b) = 0.
\]

The next result states that every charge distribution with a non-zero total charge is equivalent to a single charge.

**Theorem on Unique Existence of Barycenters:** Let \(\gamma \in \mathbb{R}(E)\) be given so that \(\sigma := \sum_{x \in E} \gamma(x) \neq 0\). Then there is exactly one \(b \in E\) such that \(\gamma \sim \sigma \delta_b\). We have
\[
b = q + \frac{1}{\sigma} \sum_{x \in E} \gamma(x)(x-q) \quad \text{for all } q \in E. \tag{34.4}
\]
The point \(b\) is called the **barycenter** of the distribution \(\gamma\).

**Proof:** Assume that \(b \in E\) is such that \(\gamma \sim \sigma \delta_b\). Using (34.2) with \(\beta := \sigma \delta_b\) we get \(\sum_{x \in E} (\gamma(x)(x-q) \mid x \in E) = \sigma(b-q)\) for all \(q \in E\). Since \(\sigma \neq 0\), it follows that (34.4) must hold for each \(q \in E\). This proves the uniqueness of \(b \in E\). To prove existence, we choose \(q \in E\) arbitrarily and define \(b \in E\) by (34.4). Using Prop.1, it follows immediately that \(\gamma \sim \sigma \delta_b\).
Proposition 3: The barycenter of a charge distribution with non-zero total charge belongs to the flat span of the support of the distribution. More precisely, if $\gamma \in \mathbb{R}^E$, $b \in E$ and $\sigma \in \mathbb{R}^\times$ are such that $\gamma \sim \sigma \delta_b$, then $b \in \text{Fsp}(\text{Supt}\ \gamma)$.

Proof: Let $F$ be any flat such that $\text{Supt}\ \gamma \subset F$. Then $\gamma|_F \in \mathbb{R}^F$ is a charge distribution on $F$ and has a barycenter in $F$. But this is also the barycenter $b$ in $E$ and hence we have $b \in F$. Since $F$, with $\text{Supt}\ \gamma \subset F$, was arbitrary, it follows that $b \in \text{Fsp}(\text{Supt}\ \gamma)$.

We call a pair $(\mu, p) \in \mathbb{P} \times E$ a mass-point. We may think of it as describing a particle of mass $\mu$ placed at $p$. By the barycenter $b$ of a non-empty finite family $((\mu_i, p_i) \mid i \in I)$ of mass-points we mean the barycenter of the distribution $\sum(\mu_i \delta_{p_i} \mid i \in I)$. It is characterized by

$$\sum_{i \in I} \mu_i (p_i - b) = 0$$

and satisfies

$$b = q + \frac{1}{\sum \mu} \sum_{i \in I} \mu_i (p_i - q)$$

for all $q \in E$.

Proposition 4: Let $\Pi$ be a partition of a given non-empty finite set $I$ and let $((\mu_i, p_i) \mid i \in I)$ be a family of mass-points. For every $J \in \Pi$, let $\lambda_J := \sum_i \mu_i | J_i$ and let $q_J$ be the barycenter of the family $((\mu_j, p_j) \mid j \in J)$. Then $((\mu_i, p_i) \mid i \in I)$ and $((\lambda_J, q_J) \mid J \in \Pi)$ have the same barycenter.

Proof: Using (34.6) with $I$ replaced by $J$, $b$ by $q_J$ and $q$ by the barycenter $b$ of $((\mu_i, p_i) \mid i \in I)$, we find that

$$\lambda_J (q_J - b) = \sum_{j \in J} \mu_j (p_j - b)$$

holds for all $J \in \Pi$. Using the summation rule ($07.5$), we obtain

$$\sum_{J \in \Pi} \lambda_J (q_J - b) = \sum_{i \in I} \mu_i (p_i - b) = 0,$$

which proves that $b$ is the barycenter of $((\lambda_J, q_J) \mid J \in \Pi)$.

By the centroid of a non-empty finite family $(p_i \mid i \in I)$ of points we mean the barycenter of the family $((1, p_i) \mid i \in I)$.

Examples and Remarks:

1. If $z$ is the barycenter of the pair $((\mu, x), (\lambda, y)) \in (\mathbb{P}^\times \times E)^2$ i.e. of the distribution $\mu \delta_x + \lambda \delta_y$, we say that $z$ divides $(x, y)$ in the ratio $\lambda : \mu$. 

By (34.5) \( z \) is characterized by \( \mu(x - z) + \lambda(y - z) = 0 \) (see Figure 1).

If \( \lambda = \mu \), we say that \( z \) is the **midpoint** of \( (x, y) \).

Thus, the midpoint of \( (x, y) \) is just the centroid of \( (x, y) \).

---

2. Consider a triple \( (x_1, x_2, x_3) \) of points and let \( c \) be centroid of this triple. If we apply Prop. 4 to the partition \( \{\{1\}, \{2, 3\}\} \) of \( 3 \) and to the case when \( \mu_i = 1 \) for all \( i \in 3 \), we find that \( c \) is also the barycenter of \( ((1, x_1), (2, z)) \), where \( z \) is the midpoint of \( (x_2, x_3) \). In other words, \( c \) divides \( (x_1, z) \) in the ratio 2:1. (see Figure 2). We thus obtain the following well known theorem of elementary geometry: The three medians of a triangle all meet at the centroid, which divides each in the ratio 2:1.

---

3. Let \( c \) be the centroid of a quadruple \( (x_1, x_2, x_3, x_4) \) of points. If we apply Prop. 4 to the partition \( \{\{1, 2\}, \{3, 4\}\} \) of \( 4 \) and to the case when \( \mu_i = 1 \) for all \( i \in 4 \), we find that \( c \) is also the midpoint of \( (y, z) \), where \( y \) is the midpoint of \( (x_1, x_2) \) and \( z \) the midpoint of \( (x_3, x_4) \) (see Figure 3). We thus obtain the following geometrical theorem: The four line-segments that join the midpoints of opposite edges of a tetrahedron all meet at the centroid of the tetrahedron, which is the midpoint of all four.
If Prop.4 is applied to the partition \{\{1, 2, 3\}, \{4\}\} of 4\] and to the case when \(\mu_i = 1\) for all \(i \in 4\]\, one obtains the following geometric theorem: The four line segments that join the vertices of a tetrahedron to the centroids of the opposite faces all meet at the centroid of the tetrahedron, which divides each in the ratio 3:1.

4. If a non-empty finite family \((\mu_i, p_i) | i \in I\) in \(\mathbb{P}^\times \times \mathcal{E}\) is interpreted physically as a system of point-particles then the barycenter of the family is the center of mass of the system. If the system moves, then the places \(p_i\) will change in time, and so will the center of mass. Newtonian mechanics teaches that the center of mass moves on a straight line with constant speed if no external forces act on the system.

5. Let \(\mathcal{E}\), with \(\text{dim } \mathcal{E} = 2\), be a mathematical model for a rigid, plane, horizontal plate. A distribution \(\gamma \in \mathbb{R}^{(\mathcal{E})}\) can then be interpreted as a system of vertical forces. The term \(\gamma_x\) with \(x \in \text{Supt } \gamma\) gives the magnitude and direction of a force acting at \(x\). We take \(\gamma_x\) as positive if the force acts downwards, negative if it acts upwards. If \(\sigma := \text{sum}_x \gamma\) is not zero and if \(b\) is the barycenter of \(\gamma\), then the force system is statically equivalent to a single resultant force \(\sigma\) applied at \(b\); i.e., the plate can be held in equilibrium by applying on opposite force \(-\sigma\) at \(b\).

6. We can define an “addition” \(\text{add}: (\mathbb{P}^\times \times \mathcal{E})^2 \to \mathbb{P}^\times \times \mathcal{E}\) on \(\mathbb{P}^\times \times \mathcal{E}\) by

\[
\text{add}((\mu, x), (\lambda, y)) := (\mu + \lambda, z),
\]

where \(z\) is the barycenter of the pair \((\mu, x), (\lambda, y)\), i.e., the point that divides \((x, y)\) in the ratio \(\lambda: \mu\). It is obvious that this addition is
CHAPTER 3. FLAT SPACES

commutative. Application of Prop. 4 to triples easily shows that this addition is also associative and hence endows $\mathbb{P}^\times \times \mathcal{E}$ with the structure of a commutative pre-monoid (see Sect.06). We use additive notation, i.e. we write

$$(\mu, x) + (\lambda, y) := \text{add}(\langle \mu, x \rangle, \langle \lambda, y \rangle).$$

For any non-empty finite family $((\mu_i, p_i) \mid i \in I)$ in $\mathbb{P}^\times \times \mathcal{E}$ we then have

$$\sum_{i \in I} (\mu_i, p_i) = (\text{sum}_I \mu, b),$$

where $b$ is the barycenter of the family. It is easily shown that the pre-monoid $\mathbb{P}^\times \times \mathcal{E}$ is cancellative.

**Pitfall:** The statements “$z$ divides $(x, y)$ in the ratio $\lambda : \mu$” and “$z$ is the midpoint of $(x, y)$” should not be interpreted as statements about the distances from $z$ to $x$ and $y$. There is no concept of distance in a flat space unless it is endowed with additional structure as in Chap.4. ■

Notes 34

(1) The terms “weight” or “mass” are often used instead of our “charge”. The trouble with “weight” and “mass” is that they lead one to assume that their values are positive.

(2) There is no agreement in the literature about the terms “barycenter” and “centroid”. Sometimes “centroid” is used for what we call “barycenter” and sometimes “barycenter” for what we call “centroid”.

35 Flat Combinations

We assume that a flat space $\mathcal{E}$ with translation space $\mathcal{V}$ and a non-empty index set $I$ are given. Recall that the summation mapping $\text{sum}_I : \mathbb{R}^{(I)} \to \mathbb{R}$ defined by (15.2) is linear and hence flat. For each $\nu \in \mathbb{R}$, we write

$$\left(\mathbb{R}^{(I)}\right)_\nu := \text{sum}_I^\leq (\{\nu\}) = \{\lambda \in \mathbb{R}^{(I)} \mid \text{sum}_I \lambda = \nu\}. \quad (35.1)$$

It follows from Prop. 3 of Sect. 33 that $\left(\mathbb{R}^{(I)}\right)_\nu$ is a flat in $\mathbb{R}^{(I)}$ whose direction space is $\left(\mathbb{R}^{(I)}\right)_0$.

**Definition 1:** The mapping

$$\text{flc}_p : \left(\mathbb{R}^{(I)}\right)_1 \to \mathcal{E},$$
defined so that $\text{flc}_p(\lambda)$ is the barycenter of the distribution $\sum (\lambda_i \delta_{p_i} \mid i \in I)$, is called the flat-combination mapping for $p$. The value $\text{flc}_p(\lambda)$ is called the flat combination of $p$ with coefficient family $\lambda$.

By the Theorem on the Unique Existence of Barycenters of Sect.34, the mapping $\text{flc}_p$ is characterized by

$$
\sum_{i \in I} \lambda_i (p_i - \text{flc}_p(\lambda)) = 0 \quad \text{for all } \lambda \in \mathbb{R}^{(I)}_1
$$

and we have, for every $q \in \mathcal{E}$,

$$
\text{flc}_p(\lambda) = q + \sum_{i \in I} \lambda_i (p_i - q) \quad \text{for all } \lambda \in \mathbb{R}^{(I)}_1.
$$

If $\mathcal{F}$ is a flat in $\mathcal{E}$ and if $\text{Rng} p \subset \mathcal{F}$, one can apply Def.1 to $\mathcal{F}$ instead of $\mathcal{E}$ and obtain the flat-combination mapping $\text{flc}_p^{\mathcal{F}}$ for $p$ relative to $\mathcal{F}$. It is easily seen that $\text{flc}_p^{\mathcal{F}}$ coincides with the adjustment $\text{flc}_p|_{\mathcal{F}}$.

**Proposition 1:** The flat-combination mapping $\text{flc}_p$ is a flat mapping. Its gradient is the restriction to $\mathbb{R}^{(I)}_0$ of the linear-combination mapping $\text{lnC}_{p-q}$ for the family $p - q := (p_i - q \mid i \in I)$ in $\mathcal{V}$, no matter how $q \in \mathcal{E}$ is chosen.

**Proof:** It follows from (35.3) that

$$
\text{flc}_p(\lambda) - \text{flc}_p(\mu) = \sum_{i \in I} (\lambda_i - \mu_i)(p_i - q) = \text{lnC}_{p-q}(\lambda - \mu)
$$

holds for all $\lambda, \mu \in \mathbb{R}^{(I)}_1$. Comparison of this result with (33.4) gives the desired conclusion.

If $\mathcal{S}$ is a non-empty subset of $\mathcal{E}$, we identify $\mathcal{S}$ with the family $(x \mid x \in \mathcal{S})$ indexed on $\mathcal{S}$ itself (see Sect.02). Thus, the flat combination mapping $\text{flc}_\mathcal{S} : \mathbb{R}^{(\mathcal{S})}_1 \to \mathcal{E}$ has the following interpretation: For every $\gamma \in \mathbb{R}^{(\mathcal{S})}_1$, which can be viewed as a charge distribution with total charge 1 and support included in $\mathcal{S}$, $\text{flc}_\mathcal{S}(\gamma)$ is the barycenter of $\gamma$.

**Flat Span Theorem:** The set of all flat combinations of a non-empty family $p$ of points in $\mathcal{E}$ is the flat span of the range of $p$, i.e. $\text{Rng} \text{flc}_p = \text{Fsp} \text{Rng} p$. In particular, if $\mathcal{S}$ is a non-empty subset of $\mathcal{E}$, then $\text{Rng} \text{flc}_\mathcal{S} = \text{Fsp} \mathcal{S}$.

**Proof:** Noting that $\text{Rng} \text{flc}_p = \text{Rng} \text{flc}_\mathcal{S}$ if $\mathcal{S} = \text{Rng} p$, we see that it is sufficient to prove that $\text{Rng} \text{flc}_\mathcal{S} = \text{Fsp} \mathcal{S}$ for every non-empty $\mathcal{S} \in \text{Sub} \mathcal{E}$.

Let $x \in \mathcal{S}$. Then $\delta_x \in \mathbb{R}^{(\mathcal{S})}_1$ and $\text{flc}_\mathcal{S}(\delta_x) = x$. Hence $x \in \text{Rng} \text{flc}_\mathcal{S}$. Since $x \in \mathcal{S}$ was arbitrary, it follows that $\mathcal{S} \subset \text{Rng} \text{flc}_\mathcal{S}$. Since $\text{flc}_\mathcal{S}$ is a flat
mapping by Prop.1, it follows, by Prop.3 of Sect.33, that \( \text{Rng flc}_S \) is a flat in \( \mathcal{E} \) and hence that \( S \subset \text{Fsp} \subset \text{Rng flc}_S \). On the other hand, by Prop.3 of Sect.34, we have \( \text{flc}_S(\gamma) \in \text{Fsp}(\text{Supt} \gamma) \subset \text{Fsp} \) for every \( \gamma \in (\mathbb{R}(S))_1 \) and hence \( \text{Rng flc}_S \subset \text{Fsp} \).

Applying the Flat Span Theorem to the case when \( p := (x, y) \) is a pair of distinct points in \( \mathcal{E} \), we find that the line \( xy \) passing through \( x \) and \( y \) (see (32.5)) is just the set of all flat combinations of \( (x, y) \):

\[
xy = \{ \text{flc}_{(x,y)}(\lambda, \mu) \mid \lambda, \mu \in \mathbb{R}, \lambda + \mu = 1 \}.
\]

(35.4)

If \( V \) is a linear space, regarded as a flat space that is its own translation space, and if \( f := (f_i \mid i \in I) \) is a family of elements of \( V \), then the flat combination mapping \( \text{flc}_f \) for \( f \) is nothing but the restriction to \( (\mathbb{R}(I))_1 \) of the linear combination mapping for \( f \), i.e. \( \text{flc}_f = \text{lnc}_f \mid (\mathbb{R}(I))_1 \). In other words we have

\[
\sum_{i \in I} \lambda_i f_i = \text{flc}_f(\lambda) \quad \text{for all} \quad \lambda \in (\mathbb{R}(I))_1.
\]

If \( \mathcal{E} \) is a flat space that does not carry the structure of a linear space, and if \( p := (p_i \mid i \in I) \) is a family in \( \mathcal{E} \), we often use the **symbolic notation**

\[
\sum_{i \in I} \lambda_i p_i := \text{flc}_p(\lambda) \quad \text{for all} \quad \lambda \in (\mathbb{R}(I))_1,
\]

(35.5)
even though the terms \( \lambda_i p_i \) do not make sense by themselves. If \( p := (p_1, p_2) \) is a pair of points, we write

\[
\lambda_1 p_1 + \lambda_2 p_2 := \text{flc}_{(p_1, p_2)}(\lambda_1, \lambda_2)
\]

(35.6)
for all \( (\lambda_1, \lambda_2) \in (\mathbb{R}^2)_1 \), i.e. for all \( \lambda_1, \lambda_2 \in \mathbb{R} \) with \( \lambda_1 + \lambda_2 = 1 \). Similar notations are used for other lists with a small number of terms.

If \( p \) is a family of points in \( \mathcal{E} \) and \( f \) a family of vectors in \( V \), both with the same index set \( I \), then

\[
p + f := (p_i + f_i \mid i \in I)
\]

(35.7)
is also a family of points in \( \mathcal{E} \).

**Definition 2:** A non-empty family \( p \) of points in \( \mathcal{E} \) is said to be **flatly independent**, **flatly spanning**, or a **flat basis** of \( \mathcal{E} \) if the flat-combination mapping \( \text{flc}_p \) is injective, surjective, or invertible, respectively.

By the Flat Span Theorem, the family \( p \) is flatly spanning if and only if the flat span of its range is all of \( \mathcal{E} \).
Proposition 2: Let \( p = (p_i \mid i \in I) \) be a non-empty family of points in \( \mathcal{E} \) and let \( k \in I \). Then \( p \) is flatly independent, flatly spanning, or a flat basis of \( \mathcal{E} \) according as the family \( f := (p_i - p_k \mid i \in I \setminus \{k\}) \) of vectors is linearly independent, spanning, or a basis of \( V \), respectively.

Proof: Using (35.3) with the choice \( q := p_k \) we see that

\[
\text{flc}_p(\lambda) = p_k + \text{lnc}_f(\lambda_{|I \setminus \{k\}})
\]

for all \( \lambda \in (\mathbb{R}^I)_1 \). (35.8)

Since the mapping \( (\lambda \mapsto \lambda_{|I \setminus \{k\}}) : (\mathbb{R}^I)_1 \to \mathbb{R}^{I \setminus \{k\}} \) is easily seen to be invertible, it follows from (35.8) that \( \text{flc}_p \) is injective, surjective, or invertible according as \( \text{lnc}_f \) is injective, surjective, or invertible, respectively. In view of Def.2 of Sect.15, this is the desired result. □

The following result follows from Prop.2 and the Theorem on Characterization of Dimension (Sect.17).

Proposition 3: Let \( \mathcal{E} \) be a finite-dimensional flat space and let \( p := (p_i \mid i \in I) \) be a family of points in \( \mathcal{E} \).

(a) If \( p \) is flatly independent then \( I \) is finite and \( \sharp I \leq \dim \mathcal{E} + 1 \), with equality if and only if \( p \) is a flat basis.

(b) If \( p \) is flatly spanning, then \( \sharp I \geq \dim \mathcal{E} + 1 \), with equality if and only if \( p \) is a flat basis.

Let \( p := (p_i \mid i \in I) \) be a flat basis of \( \mathcal{E} \). Then, by Def.2, for every \( x \in \mathcal{E} \) there is a unique \( \lambda \in (\mathbb{R}^I)_1 \) such that \( x = \text{flc}_p(\lambda) \). The family \( \lambda \) is called the family of barycentric coordinates of \( x \) relative to the flat basis \( p \).

The following result states that the flat mappings are exactly those that “preserve” flat combinations. We leave its proof to the reader (see Problem 7 below).

Proposition 4: Let \( \mathcal{E}, \mathcal{E}' \) be flat spaces. A mapping \( \alpha : \mathcal{E} \to \mathcal{E}' \) is flat if and only if for every non-empty family \( p \) of points in \( \mathcal{E} \) we have

\[
\alpha \circ \text{flc}_p = \text{flc}_{\alpha p}.
\]

The following result is an analogue, and a consequence, of Prop.2 of Sect.16.

Proposition 5: Let \( \mathcal{E}, \mathcal{E}' \) be flat spaces. Let \( p := (p_i \mid i \in I) \) be a flat basis of \( \mathcal{E} \) and let \( p' := (p'_i \mid i \in I) \) be a family of points in \( \mathcal{E}' \). Then there is a unique flat mapping \( \alpha : \mathcal{E} \to \mathcal{E}' \) such that \( \alpha \circ p = p' \). This \( \alpha \) is injective, surjective, or invertible depending on whether \( p' \) is flatly independent, flatly spanning, or a flat basis, respectively.
CHAPTER 3. FLAT SPACES

Notes 35

(1) In the case when \( p \) is a list, the flat-combination \( \text{flc}_p(\lambda) \) is sometimes called the "average" of \( p \) with "weights" \( \lambda \).

(2) The term "frame" is sometimes used for what we call a "flat basis".

36 Flat Functions

In this section, we assume that \( E \) is a finite-dimensional flat space with translation space \( \mathcal{V} \). We denote the set of all flat mappings from \( E \) into \( \mathbb{R} \) by \( \text{Flf} \mathcal{E} \) and call its members flat functions. Given any \( q \in E \), we use the notation

\[
\text{Flf}_q \mathcal{E} := \{ a \in \text{Flf} \mathcal{E} \mid a(q) = 0 \} \quad (36.1)
\]

for the set of all flat functions that have the value 0 at the point \( q \). The set of all real-valued constants with domain \( \mathcal{E} \) will be denoted by \( \mathbb{R} \mathcal{E} \). Note that the gradient \( \nabla a \) of a flat function \( a \) belongs to the dual \( \mathcal{V}^* := \text{Lin}(\mathcal{V}, \mathbb{R}) \) of the translation space \( \mathcal{V} \).

The following two basic facts are easily verified using Prop. 1 of Sect. 33.

**Proposition 1:** \( \text{Flf} \mathcal{E} \) is a subspace of \( \text{Map} (\mathcal{E}, \mathbb{R}) \). The set \( \mathbb{R} \mathcal{E} \) is a one-dimensional subspace of \( \text{Flf} \mathcal{E} \) and, for each \( q \in \mathcal{E} \), \( \text{Flf}_q \mathcal{E} \) is a supplement of \( \mathbb{R} \mathcal{E} \) in \( \text{Flf} \mathcal{E} \), so that \( \mathbb{R} \mathcal{E} \cap \text{Flf}_q \mathcal{E} = \{0\} \) and

\[
\text{Flf} \mathcal{E} = \mathbb{R} \mathcal{E} + \text{Flf}_q \mathcal{E}. \quad (36.2)
\]

**Proposition 2:** The mapping \( G_{\mathcal{E}} : \text{Flf} \mathcal{E} \to \mathcal{V}^* \) defined by \( G_{\mathcal{E}} a := \nabla a \) is linear and surjective and has the nullspace \( \text{Null} G_{\mathcal{E}} = \mathbb{R} \mathcal{E} \).

The next result follows from Prop. 2 by applying the Theorem on Dimension of Range and Nullspace ( Sect. 14) and (21.1).

**Proposition 3:** We have

\[
\dim \mathcal{E} = \dim(\text{Flf}_q \mathcal{E}) = \dim(\text{Flf} \mathcal{E}) - 1 \quad (36.3)
\]

for all \( q \in \mathcal{E} \).

The following result states, among other things, that every flat function defined on a flat in \( \mathcal{E} \) can be extended to a flat function on all of \( \mathcal{E} \).

**Proposition 4:** Let \( \mathcal{F} \) be a flat with direction space \( \mathcal{U} \). The restriction mapping \( (a \mapsto a|_{\mathcal{F}}) : \text{Flf} \mathcal{E} \to \text{Flf} \mathcal{F} \) is linear and surjective. Its nullspace \( \mathcal{N} := \{ a \in \text{Flf} \mathcal{E} \mid a|_{\mathcal{F}} = 0 \} \) has the following properties:

(i) For every \( q \in \mathcal{F} \) we have \( \mathcal{N} = \{ a \in \text{Flf}_q \mathcal{E} \mid \nabla a \in \mathcal{U}^\perp \} \).
(ii) For every $q \in E \setminus F$ there is an $a \in \mathcal{N}$ such that $a(q) = 1$.

**Proof:** The linearity of $a \mapsto a|_F$ is evident. Now let $q \in F$ be given. If $b \in \text{Flf} F$ is given we can choose, by Prop.6 of Sect.21, $\lambda \in \mathcal{V}^*$ such that $\lambda|_U = \nabla b$. By the Theorem on Specification of Flat Mappings (Sect.33), we can define $a \in \text{Flf} E$ by $a(x) := b(q) + \lambda(x - q)$ for all $x \in E$. We then have $a|_F = b$. Since $b \in \text{Flf} F$ was arbitrary, this shows the surjectivity of $a \mapsto a|_F$.

The property (i) of $\mathcal{N}$ is an immediate consequence of Prop.1 of Sect.33.

Assume now that $q \in E \setminus F$ is given and choose $z \in F$. Then $v := q - z \notin U$ and hence $U \subseteq U + Rv$. Using Prop.4 of Sect.22, it follows that $(U + Rv)^\perp \subseteq U^\perp$. Hence we can choose $\lambda \in U^\perp$ such that $\lambda v \neq 0$.

By the Theorem on Specification of Flat Mappings, we can define $a \in \text{Flf} E$ by

$$a(x) := \frac{1}{\lambda v} \lambda(x - z) \quad \text{for all } x \in E.$$ 

It is evident that $a \in \mathcal{N}$ and $a(q) = 1$. ■

Consider now a non-constant flat function $a : E \to \mathbb{R}$. As stated in Example 1 of Sect.33, we then have $\nabla a \neq 0$. It follows from Prop.3 of Sect.33 that the direction space of $a^\perp(\{0\})$ is $\{\nabla a\}^\perp = (\mathbb{R}\nabla a)^\perp$. Since $\dim(\mathbb{R}\nabla a) = 1$, it follows from the Formula for Dimension of Annihilators (21.15) that $\dim a^\perp(\{0\}) = \dim(\nabla a)^\perp = n - 1$, i.e. that $a^\perp(\{0\})$ is a hyperplane. We say that $a^\perp(\{0\})$ is the hyperplane determined by $a$. Given any hyperplane $F$ in $E$, it follows from Prop.4, (ii) that there is a non-constant $a \in \text{Flf} E$ such that $F$ is the hyperplane determined by $a$. Also, Prop.4, (ii), and Prop.6 of Sect.32 have the following immediate consequence.

**Proposition 5:** Every flat in $E$ other than $E$ itself is the intersection of all hyperplanes that include it.

**Proposition 6:** The evaluation mapping $e : E \to (\text{Flf} E)^*$ defined by

$$ev(x)a := a(x) \quad \text{for all } a \in \text{Flf} E, \ x \in E \quad (36.4)$$

is flat and injective. Its gradient $\nabla ev \in \text{Lin}(\mathcal{V}, (\text{Flf} E)^*)$ coincides with the transpose of the linear mapping $G_E \in \text{Lin}(\text{Flf} E, \mathcal{V}^*)$ defined in Prop.2.

**Proof:** Using (33.2) and (36.4) we see that

$$(\nabla a)v = a(x + v) - a(x) = (ev(x + v) - ev(x))a$$

holds for all $a \in \text{Flf} E, x \in E$ and $v \in \mathcal{V}$. Hence, if we define $L : \mathcal{V} \to (\text{Flf} E)^*$ by

$$L(v)a := (\nabla a)v \quad \text{for all } a \in \text{Flf} E, v \in \mathcal{V}, \quad (36.5)$$

then $L = G_E$. Therefore, $ev$ is injective.
we obtain

\[ L(v) = ev(x + v) - ev(x) \quad \text{for all} \quad x \in \mathcal{E}, \; v \in \mathcal{V}. \]

It is clear from (36.5) that \( L(\xi v) = \xi L(v) \) for all \( v \in \mathcal{V}, \xi \in \mathbb{R} \). Thus, \( ev \) satisfies the two conditions of Prop.1 of Sect.33 and \( L \) is the gradient of \( ev \), i.e. \( \nabla ev = L \).

The equation (36.5) states that

\[ ((\nabla ev)v) a = (G_{\mathcal{E}}a)v \quad \text{for all} \quad a \in \text{Flf} \mathcal{E}, \; v \in \mathcal{V} \quad (36.6) \]

In view of the characterization (22.3) of the transpose, this means that \( \nabla ev = G_{\mathcal{E}}^{\top} \). By the Theorem on Annihilators and Transposes, (21.13), and by Prop.2, it follows that \( \text{Null} \nabla ev = (\text{Rng} G_{\mathcal{E}})^\perp = \mathcal{V}^* \perp = \{0\} \). Therefore \( \nabla ev \) is injective and so is \( ev \) (see Prop.4 of Sect.33).

If we write

\[ \tilde{x} := ev(x), \quad \tilde{v} := (\nabla ev)v \quad (36.7) \]

when \( x \in \mathcal{E}, v \in \mathcal{V} \), we can, by Prop.6, regard \( x \mapsto \tilde{x} \) as a flat isomorphism from \( \mathcal{E} \) onto the flat \( \tilde{\mathcal{E}} := ev>(\mathcal{E}) \) in \( (\text{Flf} \mathcal{E})^* \) and \( v \mapsto \tilde{v} \) as a linear isomorphism from \( \mathcal{V} \) onto the subspace \( \tilde{\mathcal{V}} := (\nabla ev)>\mathcal{V} \) of \( (\text{Flf} \mathcal{E})^* \). This subspace \( \tilde{\mathcal{V}} \) is the direction space of the flat \( \tilde{\mathcal{E}} \). For all \( x, y \in \mathcal{E}, v \in \mathcal{V} \), we have

\[ x - y = v \iff \tilde{x} - \tilde{y} = \tilde{v} \]

and

\[ x + v = y \iff \tilde{x} + \tilde{v} = \tilde{y}. \]

Hence, if points in \( \mathcal{E} \) and translations in \( \mathcal{V} \) are replaced by their images in \( \tilde{\mathcal{E}} \) and \( \tilde{\mathcal{V}} \), a point-difference becomes an ordinary difference in \( (\text{Flf} \mathcal{E})^* \) and the symbolic sum of a point and a translation becomes an ordinary sum in \( (\text{Flf} \mathcal{E})^* \). If points in \( \mathcal{E} \) are replaced by their images in \( \tilde{\mathcal{E}} \), then flat combinations become linear combinations. In other words, the image in \( \tilde{\mathcal{E}} \) of a symbolic sum of the form (35.5) is the ordinary sum \( \sum_{i \in I} \lambda_i \tilde{p}_i \).

It is not hard to show that every element of \( (\text{Flf} \mathcal{E})^* \) is either of the form \( \xi \tilde{x} \) with \( x \in \mathcal{E}, \xi \in \mathbb{R}^\times \), or of the form \( \tilde{v} \) with \( v \in \mathcal{V} \).

\[
\text{Notes 36}
\]

1. In accord with Note 1 to Sect.32, the usual term for our “flat function” is “affine function”. However, in very elementary texts (especially those for high-school use), the term “linear function” is often found. Of course, such usage clashes with our (and the standard) use of “linear”.

\[
\textit{CHAPTER 3. FLAT SPACES}
\]
37 Convex Sets

Let $E$ be a flat space with translation space $V$. Given any points $x, y \in E$ we define the **segment joining $x$ and $y$** to be the set of all flat combinations of $(x, y)$ with positive coefficients and denote it by

$$[x, y] := \{\lambda x + \mu y \mid \lambda, \mu \in P, \lambda + \mu = 1\}, \quad (37.1)$$

where the symbolic-sum notation (35.6) is used. Comparing this definition with (35.4), we see that, if $x \neq y$, the segment joining $x$ and $y$ is a subset of the line passing through $x$ and $y$, as one would expect.

**Definition 1:** A subset $C$ of $E$ is said to be **convex** if the segment joining any two points in $C$ is included in $C$, i.e. if $x, y \in C$ implies $[x, y] \subset C$.

It is evident that the empty set $\emptyset$ and all flats in $E$, and in particular $E$ itself, are convex. Also, all segments are convex. If $\mathbb{R}$ is regarded as a one-dimensional flat space, then its convex sets are exactly the intervals (see Sect.08). If $s, t \in \mathbb{R}$ and $s \leq t$, the notation (37.1) is consistent with the notation $[s, t] := \{r \in \mathbb{R} \mid s \leq r \leq t\}$ (see (08.4)).

The following result is an immediate consequence of Def.1.

**Proposition 1:** The intersection of any collection of convex sets is again a convex set.

In view of the remarks on span-mappings made in Sect.03 we have the following result.

**Proposition 2:** Given any subset $S$ of $E$, there is a unique smallest convex set that includes $S$. More precisely, there is a unique convex set that includes $S$ and is included in every convex set that includes $S$. It is called the **convex hull of $S$** and is denoted by $C_{\text{cxh}}S$. We have $S = C_{\text{cxh}}S$ if and only if $S$ is convex.

It is clear that the convex hull of two points is just the segment joining the two points; more precisely $C_{\text{cxh}}\{x, y\} = [x, y]$ for all $x, y \in E$.

The following result follows from the fact (Prop.4 of Sect.35) that flat mappings preserve flat combinations.

**Proposition 3:** Images and pre-images of convex sets under flat mappings are again convex. In particular, if $x, y \in E$ and if $\alpha$ is a flat mapping with domain $E$, then

$$\alpha([x, y]) = [\alpha(x), \alpha(y)]. \quad (37.2)$$

Using (35.7) for pairs $p := (x, y)$ and $f := (u, v)$ we immediately obtain the following result.

**Proposition 4:** Let $C$ be a convex subset of $E$, and $B$ a convex subset of $V$. Then $C + B$ is a convex subset of $E$. In particular, $C + v$ is convex for all $v \in V$ and $x + B$ is convex for all $x \in E$. 

CHAPTER 3. FLAT SPACES

Given any non-empty set $I$, we use the notation

$$(\mathbb{P}(I))_1 := (\mathbb{R}(I))_1 \cap \mathbb{P}(I)$$

for the set of all families of positive numbers that have finite support and add up to 1. It is easily seen that $(\mathbb{P}(I))_1$ is a convex subset of the flat space $(\mathbb{R}(I))_1$.

**Definition 3:** Let $p := (p_i \mid i \in I)$ be a non-empty family of points in $\mathcal{E}$. The restriction of the flat-combination mapping for $p$ to $(\mathbb{P}(I))_1$ is called the convex-combination mapping for $p$ and is denoted by

$$cxc_p := \text{flc}_p \mid_{(\mathbb{P}(I))_1}.$$  \hfill (37.3)

**Convex Hull Theorem:** For every non-empty family $p$ of points in $\mathcal{E}$, we have

$$\text{Rng } cxc_p = \text{Cxh}(\text{Rng } p).$$  \hfill (37.4)

In particular, a subset $C$ of $\mathcal{E}$ is convex if and only if $\text{Rng } cxc_C = C$.

**Proof:** We have, for all $j \in I$, $\delta_j \in (\mathbb{P}(I))_1$ and hence $p_j = \text{flc}_p(\delta_j) = cxc_p(\delta_j)$. Since $j \in I$ was arbitrary, it follows that $\text{Rng } p \subset \text{Rng } cxc_p$. Since $(\mathbb{P}(I))_1$ is convex and $\text{flc}_p$ flat, it follows from Prop.3 that $\text{Rng } cxc_p = (\text{flc}_p)_*( (\mathbb{P}(I))_1 )$ is a convex subset of $\mathcal{E}$ and so $\text{Cxh}(\text{Rng } p) \subset \text{Rng } cxc_p$.

To prove the reverse inclusion $\text{Rng } cxc_p \subset \text{Cxh}(\text{Rng } p)$, we must show that $cxc_p(\lambda) \in \text{Cxh}(\text{Rng } p)$ for all $\lambda \in (\mathbb{P}(I))_1$. We do so by induction over $\sharp \text{Supt } \lambda$. If $\sharp \text{Supt } \lambda = 1$, then

$$\lambda = \delta_j \quad \text{for some } j \in I \quad \text{and } cxc_p(\lambda) = p_j \in \text{Rng } p \subset \text{Cxh}(\text{Rng } p).$$

Assume, then, that $\lambda \in (\mathbb{P}(I))_1$ with $\sharp \text{Supt } \lambda > 1$ is given, and that $cxc_p(\mu) \in \text{Cxh}(\text{Rng } p)$ holds for all $\mu \in (\mathbb{P}(I))_1$ with $\sharp \text{Supt } \mu < \sharp \text{Supt } \lambda$. We may and do choose $j \in \text{Supt } \lambda$ and we put $\sigma := \sum (\lambda_i \mid i \in I \setminus \{j\}) \in ]0,1[$. We define $\mu \in (\mathbb{P}(I))_1$ by

$$\mu := \left\{ \begin{array}{ll} \frac{1}{\sigma} \lambda_i & \text{if } i \in I \setminus \{j\} \\ 0 & \text{if } i = j \end{array} \right\}$$

Clearly, $\sharp \text{Supt } \mu = \sharp \text{Supt } \lambda - 1$ and $\lambda = \sigma \mu + \lambda_j \delta_j$. Since $\sigma + \lambda_j = \sum I \lambda = 1$, this states that $\lambda \in [\mu, \delta_j]$ in $(\mathbb{P}(I))_1$. Since $\text{flc}_p$ is flat, it follows by (37.2) that

$$cxc_p(\lambda) = \text{flc}_p(\lambda) \in [\text{flc}_p(\mu), \text{flc}_p(\delta_j)] = [cxc_p(\mu), p_j]$$

By the induction hypothesis, $cxc_p(\mu) \in \text{Cxh}(\text{Rng } p)$. Since $p_j \in \text{Rng } p \subset \text{Cxh}(\text{Rng } p)$ and since $\text{Cxh}(\text{Rng } p)$ is convex, it follows by Def.1 that $cxc_p(\lambda) \in \text{Cxh}(\text{Rng } p)$. \hfill \blacksquare
Since flat mappings preserve flat combinations (Prop. 4 of Sect. 35) and hence convex combinations, we immediately obtain the following consequence of the Convex Hull Theorem.

**Proposition 5:** If \( S \) is a subset of \( E \) and \( \alpha \) a flat mapping with domain \( E \), then

\[
\alpha_>(\text{C}x\text{h}S) = \text{C}x\text{h}(\alpha_>(S)). \tag{37.5}
\]

The following is a refinement of the Convex Hull Theorem.

**Strong Convex Hull Theorem:** Let \( p \) be a non-empty family of points in \( E \) and let \( x \in \text{C}x\text{h}(\text{Rng} p) \) be given. Then there is a \( \lambda \in (\mathbb{P}(I))_1 \) such that \( x = cxc_p(\lambda) \) and such that \( p|\text{Supt} \lambda \) is flatly independent.

**Proof:** By the Convex Hull Theorem, \( cxc_p^\leq(\{x\}) \) is not empty. We choose a \( \lambda \in cxc_p^\leq(\{x\}) \) whose support has minimum cardinal. Put \( J := \text{Supt} \lambda \).

Assume that \( p' := p|_J \) is flatly dependent, i.e. that \( \text{flc}_{p'} \) is not injective. Then, by Prop. 4 of Sect. 33, the gradient of \( \text{flc}_{p'} \) is not injective. Hence we can choose \( \nu \in ((\mathbb{R}(I))(0)^\times \) such that \( \text{Supt} \nu \subset J \) and

\[
0 = (\text{grad}_{\text{flc}_{p'}})\nu|_J = (\text{grad}_{\text{flc}_p})\nu \tag{37.6}
\]

We now choose \( k \in J \) such that

\[
\frac{\nu_k}{\lambda_k} = \max\{\frac{\nu_i}{\lambda_i} \mid i \in J\}.
\]

Since \( \sum_{J} \nu|_J = 0 \) and \( \nu|_J \neq 0 \) and \( \lambda_i > 0 \) for all \( i \in J \), we must have \( \nu_k > 0 \) and we conclude that

\[
\lambda_i \geq \frac{\nu_i}{\nu_k} \lambda_k \quad \text{for all} \quad i \in J.
\]

Hence we have \( \lambda' := \lambda - \frac{\lambda_k}{\nu_k} \nu \in (\mathbb{P}(I))_1 \), so that \( cxc_p \lambda' \) is meaningful. Using (37.6) we obtain

\[
cxc_p \lambda' = \text{flc}_p(\lambda - \frac{\lambda_k}{\nu_k} \nu) = \text{flc}_p \lambda - \frac{\lambda}{\nu_k} (\text{grad}_{\text{flc}_p})\nu
\]

\[
= \text{flc}_p \lambda = cxc_p \lambda = x,
\]

which means that \( \lambda' \in cxc_p^\leq(\{x\}) \). On the other hand, we have \( \text{Supt} \lambda' \not\subset \text{Supt} \lambda \) because \( \lambda'_k = 0 \). It follows that \( \not\exists \text{Supt} \lambda' < \not\exists \text{Supt} \lambda \) which contradicts the assumption that \( \text{Supt} \lambda \) has minimum cardinal. Hence \( p|_J \) cannot be flatly dependent. □

The following consequence of the Strong Convex Hull Theorem is obtained by using Prop. 3, (a) of Sect. 35.
**Corollary:** Let $S$ be a subset of a finite-dimensional flat space $E$. Then, for every $x \in \text{C}xhS$, there is a finite subset $I$ of $S$ such that $x \in \text{C}xhI$ and $\#I \leq (\dim E) + 1$.

**Notes 37**

(1) Many other notations, for example Conv$S$ and $\hat{S}$, can be found for our $\text{C}xhS$.

(2) The Corollary to the Strong Convex Hull Theorem is often called “Carathéodory’s Theorem”.

### 38 Half-Spaces

Let $E$ be a flat space with translation space $V$. Consider a non-constant flat function $a : E \to \mathbb{R}$, i.e. a flat function such that Rng $a$ is not a singleton. Since the only flats in $\mathbb{R}$ are the singletons and $\mathbb{R}$ itself, it follows by Prop.3 of Sect.33 that Rng $a = \mathbb{R}$ and hence that $a^<(\{0\})$, $a^<(P)$, and $a^<(P^\times)$ are all non-empty. As we have seen in Sect.36, $a^<(\{0\})$ is the hyperplane determined by $a$ if $E$ is finite-dimensional. By Prop.3 of Sect.37, $a^<(P)$ and $a^<(P^\times)$ are (non-empty) convex subsets of $E$, called the **half-space** and the **open-half-space** determined by $a$. The hyperplane $a^<(\{0\})$ is called the **boundary** of these half spaces. Of course, the half-space $a^<(P)$ is the union of its boundary $a^<(\{0\})$ and the open-half-space $a^<(P^\times)$, and these two are disjoint. If $\xi \in P^\times$, then $a$ and $\xi a$ determine the same hyperplane and half-spaces.

If $z \in E$ and $a \in (\text{Flf}_zE)^\times$, then $a$ is not constant and $z$ belongs to the boundary of the half-space $a^<(P)$.

**Half-Space Inclusion Theorem:** Let $E$ be a finite-dimensional flat space, $C$ a non-empty convex subset of $E$ and $q \in E \setminus C$. Then there is a half-space that includes $C$ and has $q$ on its boundary. In other words, there is an $a \in (\text{Flf}_qE)^\times$ such that $a|_C \geq 0$.

The proof will be based on the following preliminary result:

**Lemma:** Let $C$ be a convex subset of $E$ and let $b \in \text{Flf}E$ be given. If both $C \cap b^<(P^\times)$ and $C \cap b^<(-P^\times)$ are non-empty, so is $C \cap b^<(\{0\})$. If $c \in \text{Flf}E$ satisfies $c|_{C \cap b^<(\{0\})} \geq 0$, then

$$\frac{c(x)}{b(x)} \geq \frac{c(y)}{b(y)} \quad \text{for all} \quad x \in C \cap b^<(P^\times), \; y \in C \cap b^<(-P^\times). \quad (38.1)$$
Proof: Let $x \in C \cap b^\leq(\mathbb{P}^\times)$, $y \in C \cap b^\leq(-\mathbb{P}^\times)$ be given, so that $b(x) > 0$, $b(y) < 0$, and hence $b(x) - b(y) > 0$. We define

$$
\lambda := -\frac{b(y)}{b(x) - b(y)}, \quad \mu := \frac{b(x)}{b(x) - b(y)}.
$$

(38.2)

It is clear that $\lambda, \mu \in \mathbb{P}^\times$, $\lambda + \mu = 1$. Hence $z := \lambda x + \mu y$ belongs to $[x, y]$. Since $b$ preserves convex combinations, we have, by (38.2),

$$
b(z) = \lambda b(x) + \mu b(y) = 0
$$

and hence $z \in b^\leq(\{0\})$. Since $C$ is convex, we also have $z \in C$ and hence $z \in C \cap b^\leq(\{0\})$. The situation is illustrated in Fig.1.

Figure 1.

If $C \cap b^\leq(\mathbb{P}^\times)$ and $C \cap b^\leq(-\mathbb{P}^\times)$ are both non-empty, we may choose a point in each and obtain a point in $C \cap b^\leq\{0\}$ by the above method, showing that $C \cap b^\leq\{0\}$ is not empty.

Now let $c \in \text{Flf} \mathcal{E}$ be such that $c|_{C \cap b^\leq\{0\}} \geq 0$. Then, if $x$ and $y$ are given as above and $z$ constructed accordingly, we have

$$
0 \leq c(z) = \lambda c(x) + \mu c(y).
$$

Substituting (38.2) into this inequality and multiplying the result by $b(x) - b(y) \in \mathbb{P}^\times$ yields $-b(y)c(x) + b(x)c(y) \geq 0$ which is equivalent to (38.1).

Proof of Theorem: We proceed by induction. The assertion is vacuously valid if $\dim \mathcal{E} = 0$. Assume then, that $\dim \mathcal{E} > 0$, that a non-empty convex $C \in \text{Sub} \mathcal{E}$ and a $q \in \mathcal{E} \setminus C$ are given and that the assertion is valid when $\mathcal{E}$ is replaced by a hyperplane $\mathcal{F}$ in $\mathcal{E}$.

CHAPTER 3. FLAT SPACES

Since, by Prop. 3 of Sect. 36, \( \dim(\text{Fl}_q \mathcal{E}) = \dim \mathcal{E} > 0 \), we may and do choose \( b \in (\text{Fl}_q \mathcal{E})^\times \) and put \( \mathcal{F} := b^< \{0\} \). If \( \mathcal{C} \cap b^<(\mathbb{P}^\times) \) or \( \mathcal{C} \cap b^<(\mathbb{P}^\times) \) is empty then \( a := -b \) or \( a := b \), respectively, fulfills the requirement of the theorem. Therefore, we may assume that both \( \mathcal{C} \cap b^<(\mathbb{P}^\times) \) and \( \mathcal{C} \cap b^<(\mathbb{P}^\times) \) are not empty. By the Lemma, \( \mathcal{C} \cap \mathcal{F} \) is then also non-empty. Since \( \mathcal{C} \cap \mathcal{F} \) is a convex subset of \( \mathcal{F} \) and \( q \in \mathcal{F} \setminus (\mathcal{F} \cap \mathcal{C}) \) we may use the induction hypothesis and choose \( d \in (\text{Fl}_q \mathcal{F})^\times \) such that \( d|_{\mathcal{C} \cap \mathcal{F}} \geq 0 \). In view of Prop. 4 of Sect. 36, we may and do choose a flat extension \( c \) of \( d \) to \( \mathcal{E} \), so that \( c \in (\text{Fl}_q \mathcal{E})^\times \) and \( c|_{\mathcal{C} \cap \mathcal{F}} \geq 0 \). We define subsets \( S \) and \( T \) of \( \mathbb{R} \) by

\[
S := \left\{ \frac{c(x)}{b(x)} \mid x \in \mathcal{C} \cap b^<(\mathbb{P}^\times) \right\},
\]

\[
T := \left\{ \frac{c(y)}{b(y)} \mid y \in \mathcal{C} \cap b^<(\mathbb{P}^\times) \right\}.
\]

Both \( S \) and \( T \) are non-empty. Applying the second statement of the Lemma, we see that every number in \( T \) is less than every number in \( S \). It follows that \( -\infty < \sup T \leq \inf S \leq \infty \). We choose \( \xi \in [\sup T, \inf S] \) and put \( a := c - \xi b \). If \( x \in \mathcal{C} \cap \mathcal{F} = \mathcal{C} \cap b^<\{0\} \) we have \( a(x) = c(x) \geq 0 \) since \( c|_{\mathcal{C} \cap \mathcal{F}} \geq 0 \). If \( x \in \mathcal{C} \cap b^<(\mathbb{P}^\times) \) we have \( \frac{c(x)}{b(x)} \geq \inf S \geq \xi \), \( b(x) > 0 \), and hence \( a(x) = c(x) - \xi b(x) \geq 0 \). Finally, if \( x \in \mathcal{C} \cap b^<(\mathbb{P}^\times) \), we have \( \frac{c(x)}{b(x)} \leq \sup T \leq \xi \), \( b(x) < 0 \), and hence \( a(x) = c(x) - \xi b(x) \geq 0 \). Therefore, since \( \mathcal{E} = b^<\{0\} \cup b^<(\mathbb{P}^\times) \cup b^<(\mathbb{P}^\times) \), we have \( a(x) \geq 0 \) for all \( x \in \mathcal{C} \). Since \( \mathcal{F} = b^<\{0\} \) we have \( a|_{\mathcal{F}} = c|_{\mathcal{F}} = d \neq 0 \) and hence \( a \neq 0 \). Hence \( a \) fulfills the requirement of the theorem.

Remark 1: The requirement, in the Theorem, that \( \mathcal{C} \) be non-empty can obviously be omitted when \( \dim \mathcal{E} > 0 \). This requirement is needed only to make the assertion vacuously valid when \( \dim \mathcal{E} = 0 \) and thus to simplify the induction proof.

Remark 2: In general, there are many half-spaces that meet the requirements of the Theorem. However, if \( \mathcal{C} := a^<(\mathbb{P}^\times) \cup \{z\} \), where \( z \in \mathcal{E} \) and \( a \in (\text{Fl}_z \mathcal{E})^\times \), and if \( q \in a^<(\{0\}) \setminus \{z\} \), then \( a^<(\mathbb{P}) \) is the only half-space that meets the requirements (see Fig. 2).

![Figure 2](image-url)
Remark 3: The Half-Space Inclusion Theorem depends strongly on the assumption that the translation space of \( \mathcal{E} \) is a linear space over the real field \( \mathbb{R} \). Its conclusion is no longer valid if \( \mathbb{R} \) is replaced by \( \mathbb{Q} \) (see Problem 10, below).

Notes 38

(1) What we call the “Half-Space Inclusion Theorem” is at the root of a number of results, often called “Separation Theorems”, in convex analysis. Some of these results are stated in Sect. 54.

39 Problems for Chapter 3

(1) Let \( \mathcal{S} \) be a subset of a linear space \( \mathcal{V} \). Show that

\[
\text{Lsp\,}\mathcal{S} = \text{Fsp}(\mathcal{S} \cup \{0\}). \tag{P3.1}
\]

(2) Let \( \mathcal{E}, \mathcal{E}' \) be flat spaces. Show that a mapping \( \alpha : \mathcal{E} \to \mathcal{E}' \) is flat if and only if its graph \( \text{Gr}(\alpha) \) is a flat in \( \mathcal{E} \times \mathcal{E}' \), and, if this is the case, show that the direction space of \( \text{Gr}(\alpha) \) is \( \text{Gr}(\nabla \alpha) \). (Hint: Use Problem 1 of Chap. 1)

(3) Let \( \mathcal{E} \) be a flat space and let \( \varepsilon \) be a flat mapping from \( \mathcal{E} \) to itself that is idempotent in the sense that

\[
\varepsilon \circ \varepsilon = \varepsilon. \tag{P3.2}
\]

Show:

(a) We have

\[
\varepsilon|_{\text{Rng}\varepsilon} = I_{\text{Rng}\varepsilon \subset \mathcal{E}}. \tag{P3.3}
\]

(b) \( \nabla \varepsilon \) is an idempotent linéon on \( \mathcal{V} := \mathcal{E} - \mathcal{E} \).

(c) We have

\[
\text{Rng}\varepsilon + \text{Null} \nabla \varepsilon = \mathcal{E}. \tag{P3.4}
\]

(d) Let \( x \in \text{Rng}\varepsilon \) be given. Then \( \varepsilon^<(\{x\}) \) is a flat with direction space \( \text{Null} \nabla \varepsilon \) and we have
CHAPTER 3. FLAT SPACES

\[ \varepsilon^\leq(\{x\}) \cap \text{Rng} \varepsilon = \{x\}, \quad (P3.5) \]

\[ \varepsilon^\leq(\{x\}) + \text{Rng} (\nabla \varepsilon) = \mathcal{E}. \quad (P3.6) \]

Moreover, if \( \mathcal{E} \) is finite-dimensional, we have

\[ \dim \varepsilon^\leq(\{x\}) + \dim \text{Rng} \varepsilon = \dim \mathcal{E}. \quad (P3.7) \]

(4) Let \( \mathcal{E} \) be a flat space and let \( \gamma \) be a charge distribution on \( \mathcal{E} \) with total charge zero. Show: Given any \( p \in \mathcal{E} \) and \( \sigma \in \mathbb{P}^\times \) there is exactly one \( q \in \mathcal{E} \) such that \( \gamma \sim \sigma \delta_p - \sigma \delta_q \).

**Note:** If \( p \neq q \), the distribution \( \sigma \delta_p - \sigma \delta_q \) is called a **dipole**.

(5) (a) Let \((x_1, x_2, x_3, x_4)\) be a quadruple of points in a flat space. Show that the following are equivalent:

(i) The midpoint of \((x_1, x_3)\) coincides with the midpoint of \((x_2, x_4)\),

(ii) \( x_2 - x_1 = x_3 - x_4 \),

(iii) \( x_3 - x_2 = x_4 - x_1 \).

**Remark:** If the quadruple is injective and if any and hence all of these conditions are satisfied, then the points are the vertices of a parallelogram. In fact, one may take this to be the definition of “parallelogram”.

(b) Show that the midpoints of the sides of a quadrilateral (not necessarily plane) in a flat space form a parallelogram.

(6) Consider a flatly independent triple \((x_1, x_2, x_3)\) in a flat space \( \mathcal{E} \). Assume that \( y_1 \) divides \((x_2, x_3)\) in the ratio \( \lambda_1 : \mu_1 \) and that \( y_2 \) divides \((x_1, x_3)\) in the ratio \( \lambda_2 : \mu_2 \). Determine the ratios in which the point \( c \) of intersection of \( \overrightarrow{x_1y_1} \) and \( \overrightarrow{x_2y_2} \) divides \((x_1, y_1)\) and \((x_2, y_2)\) (see Figure).
39. PROBLEMS FOR CHAPTER 3

(7) Let $\mathcal{E}, \mathcal{E}'$ be flat spaces and let $\alpha : \mathcal{E} \to \mathcal{E}'$ be a mapping which preserves flat combinations of pairs of points, i.e. which satisfies $\alpha \circ \text{flc}_p = \text{flc}_{\alpha \circ p}$ for all pairs $p := (p_1, p_2) \in \mathcal{E}^2$. Prove that $\alpha$ must be a flat mapping.

(8) Let $\mathcal{E}$ be a flat space and consider the pre-monoid $\mathbb{P}^\times \times \mathcal{E}$ described in Example 6 of Sect.34.

(a) Prove that the pre-monoid $\mathbb{P}^\times \times \mathcal{E}$ is cancellative (see Sect.06).

(b) Show that $\mathbb{P}^\times \times \mathcal{E}$ does not contain an element that satisfies the neutrality law (06.2) and hence is not monoidable.

(9) Let $\mathcal{E}$ be a flat space and put $\mathcal{V} := \mathcal{E} - \mathcal{E}$. Let $\lambda \in \mathcal{V}^\times$ and $a, b \in \mathcal{V}^\times$ be given such that $\lambda a = 1, \lambda b = 0$. Put

$$ E := 1_\mathcal{V} - (a \otimes \lambda), \quad N := b \otimes \lambda. \quad (P3.8) $$

(a) Show that $E + \xi N$ is idempotent for each $\xi \in \mathbb{R}$ and determine $\text{Null} (E + \xi N)$ and $\text{Rng} (E + \xi N)$.

Now let $q \in \mathcal{E}$ be given and define, for each $v \in \mathcal{V}$, $\varphi_v : \mathcal{E} \to \mathcal{E}$ by

$$ \varphi_v(x) := x + (E + (\lambda(x - q))N)v \quad \text{for all} \quad x \in \mathcal{E}. \quad (P3.9) $$

(b) Show that, for each $v \in \mathcal{V}$, $\varphi_v$ is flat and determine $\nabla \varphi_v$; also, show that $\varphi_v$ is invertible, so that $\varphi_v \in \text{Fis} \mathcal{E}$.

(c) Show that the mapping

$$ \varphi := (v \mapsto \varphi_v) : \mathcal{V} \to \text{Perm} \mathcal{E} \quad (P3.10) $$
is an injective homomorphism from the additive group of $\mathcal{V}$ to $\text{Perm} \mathcal{E}$. (Hence, in view of (b), $\varphi_>(\mathcal{V})$ is a subgroup of $\text{Fis} \mathcal{E}$ that is isomorphic to—but different from—the additive group $\mathcal{V}$.)

(d) Is the action of $\mathcal{V}$ on $\mathcal{E}$ defined by (P3.10) free? Is it transitive?

(10) Note that all the definitions of Chap.3 remain meaningful if the field $\mathbb{R}$ of real numbers is replaced by the field $\mathbb{Q}$ of rational numbers. Give an example of a finite-dimensional flat space $\mathcal{E}$ over $\mathbb{Q}$, a non-empty convex subset $\mathcal{C}$ of $\mathcal{E}$ and a point $q \in \mathcal{E} \setminus \mathcal{C}$ such that $a_>(\mathcal{C}) \not\subseteq \mathcal{P}$ for all $a \in (\text{Flf}_q \mathcal{E})^\times$, where $\text{Flf} \mathcal{E}$ denotes the set of all flat functions on $\mathcal{E}$ with values in $\mathbb{Q}$. (Thus, the Half-Space Inclusion Theorem does not extend to the case when $\mathbb{R}$ is replaced by $\mathbb{Q}$.)
Chapter 4

Inner-Product Spaces, Euclidean Spaces

As in Chap. 2, the term "linear space" will be used as a shorthand for "finite dimensional linear space over \( \mathbb{R} \). However, the definitions of an inner-product space and a Euclidean space do not really require finite-dimensionality. Many of the results, for example the Inner-Product Inequality and the Theorem on Subadditivity of Magnitude, remain valid for infinite-dimensional spaces. Other results extend to infinite-dimensional spaces after suitable modification.

41 Inner-Product Spaces

**Definition 1:** An inner-product space is a linear space \( V \) endowed with additional structure by the prescription of a non-degenerate quadratic form \( \text{sq} \in \text{Qu}(V) \) (see Sect. 27). The form \( \text{sq} \) is then called the inner square of \( V \) and the corresponding symmetric bilinear form

\[
\text{ip} := \text{sq} \in \text{Sym}_2(V^2, \mathbb{R}) \cong \text{Sym}(V, V^*)
\]

the inner product of \( V \).

We say that the inner-product space \( V \) is genuine if \( \text{sq} \) is strictly positive.

It is customary to use the following simplified notations:

\[
\begin{align*}
\mathbf{v}^2 & := \text{sq}(\mathbf{v}) \quad \text{when} \quad \mathbf{v} \in V, \\
\mathbf{u} \cdot \mathbf{v} & := \text{ip}(\mathbf{u}, \mathbf{v}) = (\text{ip} \mathbf{u})\mathbf{v} \quad \text{when} \quad \mathbf{u}, \mathbf{v} \in V.
\end{align*}
\]

(41.1)  (41.2)
The symmetry and bilinearity of $ip$ is then reflected in the following rules, valid for all $u, v, w \in V$ and $\xi \in \mathbb{R}$

$$u \cdot v = v \cdot u,$$  

(41.3)

$$w \cdot (u + v) = w \cdot u + w \cdot v,$$  

(41.4)

$$u \cdot (\xi v) = \xi (u \cdot v) = (\xi u) \cdot v$$  

(41.5)

We say that $u \in V$ is orthogonal to $v \in V$ if $u \cdot v = 0$. The assumption that the inner product is non-degenerate is expressed by the statement that, given $u \in V$,

$$(u \cdot v = 0 \text{ for all } v \in V) \implies u = 0. \quad (41.6)$$

In words, the zero of $V$ is the only member of $V$ that is orthogonal to every member of $V$.

Since $ip \in \text{Lin}(V, V^*)$ is injective and since $\dim V = \dim V^*$, it follows from the Pigeonhole Principle for Linear Mappings that $ip$ is a linear isomorphism. It induces the linear isomorphism $ip^\top : V^{**} \rightarrow V^*$. Since $ip$ is symmetric, we have $ip^\top = ip$ when $V^{**}$ is identified with $V$ as explained in Sect.22. Thus, this identification is the same as the isomorphism $(ip^\top)^{-1}ip : V \rightarrow V^{**}$ induced by $ip$. Therefore, there is no conflict if we use $ip$ to identify $V$ with $V^*$.

From now on we shall identify $V \cong V^*$ by means of $ip$ except that, given $v \in V$, we shall write $v \cdot := ip v$ for the corresponding element in $V^*$ so as to be consistent with the notation (41.2).

Every space $\mathbb{R}^I$ of families of numbers, indexed on a given finite set $I$, carries the natural structure of an inner-product space whose inner square is given by

$$\text{sq}(\lambda) = \lambda^2 := \sum_{i \in I} \lambda_i^2$$  

(41.7)

for all $\lambda \in \mathbb{R}^I$. The corresponding inner product is given by

$$(ip \lambda)\mu = \lambda \cdot \mu = \sum_{i \in I} \lambda_i \mu_i$$  

(41.8)

for all $\lambda, \mu \in \mathbb{R}^I$. The identification $(\mathbb{R}^I)^* \cong \mathbb{R}^I$ resulting from this inner product is the same as the one described in Sect.23.

Let $V$ and $W$ be inner-product spaces. The identifications $V^* \cong V$ and $W^* \cong W$ give rise to the further identifications such as

$$\text{Lin}(V, W) \cong \text{Lin}(V, W^*) \cong \text{Lin}_2(V \times W, \mathbb{R}),$$
\( \text{Lin}(\mathcal{W}, \mathcal{V}) \cong \text{Lin}(\mathcal{W}^*, \mathcal{V}^*) \).

Thus \( \mathbf{L} \in \text{Lin}(\mathcal{V}, \mathcal{W}) \) becomes identified with the bilinear form \( \mathbf{L} \in \text{Lin}_2(\mathcal{V} \times \mathcal{W}, \mathbb{R}) \) whose value at \((\mathbf{v}, \mathbf{w}) \in \mathcal{V} \times \mathcal{W}\) satisfies

\[
\mathbf{L}(\mathbf{v}, \mathbf{w}) = (\mathbf{L}\mathbf{v}) \cdot \mathbf{w},
\]

and \( \mathbf{L}^\top \in \text{Lin}(\mathcal{W}^*, \mathcal{V}^*) \) becomes identified with \( \mathbf{L}^\top \in \text{Lin}(\mathcal{W}, \mathcal{V}) \cong \text{Lin}_2(\mathcal{W} \times \mathcal{V}, \mathbb{R}) \) in such a way that

\[
(\mathbf{L}^\top \mathbf{w}) \cdot \mathbf{v} = \mathbf{L}^\top(\mathbf{w}, \mathbf{v}) = \mathbf{L}(\mathbf{v}, \mathbf{w}) = (\mathbf{L}\mathbf{v}) \cdot \mathbf{w}
\]

for all \( \mathbf{v} \in \mathcal{V}, \mathbf{w} \in \mathcal{W} \).

We identify the supplementary subspaces \( \text{Sym}_2(\mathcal{V}^2, \mathbb{R}) \) and \( \text{Skew}_2(\mathcal{V}^2, \mathbb{R}) \) of \( \text{Lin}_2(\mathcal{V}^2, \mathbb{R}) \cong \text{Lin}\mathcal{V} \) (see Prop.7 of Sect.24) with the supplementary subspaces

\[
\text{Sym}\mathcal{V} := \{ \mathbf{S} \in \text{Lin}\mathcal{V} \mid \mathbf{S} = \mathbf{S}^\top \},
\]

\[
\text{Skew}\mathcal{V} := \{ \mathbf{A} \in \text{Lin}\mathcal{V} \mid \mathbf{A}^\top = -\mathbf{A} \}
\]

of \( \text{Lin}\mathcal{V} \). The members of \( \text{Sym}\mathcal{V} \) are called \textit{symmetric lineons} and the members of \( \text{Skew}\mathcal{V} \) \textit{skew lineons}.

Given \( \mathbf{L} \in \text{Lin}(\mathcal{V}, \mathcal{W}) \) and hence \( \mathbf{L}^\top \in \text{Lin}(\mathcal{W}, \mathcal{V}) \) we have \( \mathbf{L}^\top \mathbf{L} \in \text{Sym}\mathcal{V} \) and \( \mathbf{L}\mathbf{L}^\top \in \text{Sym}\mathcal{W} \). Clearly, if \( \mathbf{L}^\top \mathbf{L} \) is invertible (and hence injective), then \( \mathbf{L} \) must also be injective. Also, if \( \mathbf{L} \) is invertible, so is \( \mathbf{L}^\top \mathbf{L} \). Applying these observations to the linear-combination mapping of a family \( \mathbf{f} := (\mathbf{f}_i \mid i \in I) \) in \( \mathcal{V} \) and noting the identification

\[
\text{Gr}_\mathcal{V} := \text{Lin}\mathcal{V} \cong \mathbb{R}^{|I|^2} \]

we obtain the following results.

**Proposition 1:** Let \( \mathbf{f} := (\mathbf{f}_i \mid i \in I) \) be a family in \( \mathcal{V} \). Then \( \mathbf{f} \) is linearly independent if the matrix \( \text{Gr}_\mathcal{V} \) given by (41.13) is invertible.

**Proposition 2:** A family \( \mathbf{b} := (\mathbf{b}_i \mid i \in I) \) in \( \mathcal{V} \) is a basis if and only if the matrix \( \text{Gr}_\mathcal{V} \) is invertible and \( |I| = \dim \mathcal{V} \).

Let \( \mathbf{b} := (\mathbf{b}_i \mid i \in I) \) be a basis of \( \mathcal{V} \). The identification \( \mathcal{V}^* \cong \mathcal{V} \) identifies the dual of the basis \( \mathbf{b} \) with another basis \( \mathbf{b}^* \) of \( \mathcal{V} \) in such a way that

\[
\mathbf{b}_i^* \cdot \mathbf{b}_k = \delta_{i,k} \quad \text{for all} \quad i, k \in I,
\]

where \( \delta_{i,k} \) is defined by (16.2). Using the notation (41.13) we have

\[
\mathbf{b}_k = \sum_{i \in I} (G_{\mathbf{b}})_{k,i} \mathbf{b}_i^*
\]
and
\[ G_{b^*} = G_b^{-1}. \] (41.16)
Moreover, for each \( v \in \mathcal{V} \), we have
\[ \ln v^{-1} = b^* \cdot v := (b_i^* \cdot v \mid i \in I) \] (41.17)
and hence
\[ v = \ln b (b^* \cdot v) = \ln b^* (b \cdot v). \] (41.18)
For all \( u, v \in \mathcal{V} \), we have
\[ u \cdot v = (b \cdot u) \cdot (b^* \cdot v) = \sum_{i \in I} (b_i \cdot u)(b_i^* \cdot v). \] (41.19)

We say that a family \( e := (e_i \mid i \in I) \) in \( \mathcal{V} \) is orthonormal if
\[ e_i \cdot e_k = \begin{cases} 1 & \text{or } -1 \text{ if } i = k \\ 0 & \text{if } i \neq k \end{cases} \text{ for all } i, k \in I \] (41.20)
We say that \( e \) is genuinely orthonormal if, in addition, \( e_i^2 = 1 \) for all \( i \in I \), which—using the notation (16.2)—means that
\[ e_i \cdot e_k = \delta_{i,k} \text{ for all } i, k \in I. \] (41.21)
To say that \( e \) is orthonormal means that the matrix \( G_e \), as defined by (41.13), is diagonal and each of its diagonal terms is 1 or \(-1\). It is clear from Prop.1 that every orthonormal family is linearly independent and from Prop.2 that a given orthonormal family is a basis if and only if \( \mathcal{I} = \dim \mathcal{V} \).

The standard basis \( \delta^I := (\delta_i^I \mid i \in I) \) (see Sect.16) is a genuinely orthonormal basis of \( \mathbb{R}^I \).

Let \( v \in \mathcal{V}, w \in \mathcal{W} \) be given. Then \( w \otimes v \in \text{Lin}(\mathcal{V}^*, \mathcal{W}) \) becomes identified with the element \( w \otimes v \) of \( \text{Lin}(\mathcal{V}, \mathcal{W}) \) whose values are given by
\[ (w \otimes v)u = (v \cdot u)w \text{ for all } u \in \mathcal{V}. \] (41.22)

We say that a subspace \( \mathcal{U} \) of a given inner-product space \( \mathcal{V} \) is regular if the restriction \( \sq|_{\mathcal{U}} \) of the inner square to \( \mathcal{U} \) is non-degenerate, i.e., if for each \( u \in \mathcal{U} \),
\[ u \cdot v = 0 \text{ for all } v \in \mathcal{U} \implies u = 0. \] (41.23)
If \( \mathcal{U} \) is regular, then \( \text{sq}_{\mathcal{U}} \) endows \( \mathcal{U} \) with the structure of an inner-product space. If \( \mathcal{U} \) is not regular, it does not have a natural structure of an inner-product space.

The identification \( \mathcal{V}^* \cong \mathcal{V} \) identifies the annihilator \( S^\perp \) of a given subset \( S \) of \( \mathcal{V} \) with the subspace

\[
S^\perp = \{ v \in \mathcal{V} \mid v \cdot u = 0 \text{ for all } u \in S \}
\]

of \( \mathcal{V} \). Thus, \( S^\perp \) consists of all elements of \( \mathcal{V} \) that are orthogonal to every element of \( S \). The following result is very easy to prove with the use of the Formula for Dimension of Annihilators (Sect. 21) and of Prop. 5 of Sect. 17.

**Characterization of Regular Subspaces:** Let \( \mathcal{U} \) be a subspace of \( \mathcal{V} \). Then the following are equivalent:

(i) \( \mathcal{U} \) is regular,

(ii) \( \mathcal{U} \cap \mathcal{U}^\perp = \{0\} \),

(iii) \( \mathcal{U} + \mathcal{U}^\perp = \mathcal{V} \),

(iv) \( \mathcal{U} \) and \( \mathcal{U}^\perp \) are supplementary.

Moreover, if \( \mathcal{U} \) is regular, so is \( \mathcal{U}^\perp \).

If \( \mathcal{U} \) is regular, then its annihilator \( \mathcal{U}^\perp \) is also called the **orthogonal supplement** of \( \mathcal{U} \). The following two results exhibit a natural one-to-one correspondence between the regular subspaces of \( \mathcal{V} \) and the symmetric idempotents in \( \text{Lin} \mathcal{V} \), i.e., the lineons \( E \in \text{Lin} \mathcal{V} \) that satisfy \( E = E^\top \) and \( E^2 = E \).

**Proposition 3:** Let \( E \in \text{Lin} \mathcal{V} \) be an idempotent. Then \( E \) is symmetric if and only if \( \text{Null} E = (\text{Rng} E)^\perp \).

**Proof:** If \( E = E^\top \), then \( \text{Null} E = (\text{Rng} E)^\perp \) follows from (21.13).

Assume that \( \text{Null} E = (\text{Rng} E)^\perp \). It follows from (21.13) that \( \text{Null} E = \text{Null} E^\top \) and hence from (22.9) that \( \text{Rng} E^\top = \text{Rng} E \). Since, by (21.6), \( (E^\top)^2 = (E^2)^\top = E^\top \), it follows that \( E^\top \) is also an idempotent. The assertion of uniqueness in Prop. 4 of Sect. 19 shows that \( E = E^\top \).

**Proposition 4:** If \( E \in \text{Lin} \mathcal{V} \) is symmetric and idempotent, then \( \text{Rng} E \) is a regular subspace of \( \mathcal{V} \) and \( \text{Null} E \) is its orthogonal supplement. Conversely, if \( \mathcal{U} \) is a regular subspace of \( \mathcal{V} \), then there is exactly one symmetric idempotent \( E \in \text{Lin} \mathcal{V} \) such that \( \mathcal{U} = \text{Rng} E \).

**Proof:** Assume that \( E \in \text{Lin} \mathcal{V} \) is symmetric and idempotent. It follows from Prop. 3 that \( \text{Null} E = (\text{Rng} E)^\perp \) and hence by the implication \( (v) \Rightarrow (i) \) of Prop. 4 of Sect. 19 that \( \text{Rng} E \) and \( (\text{Rng} E)^\perp \) are supplementary. Hence,

---

**41. INNER-PRODUCT SPACES**

137
by the implication (iv) ⇒ (i) of the Theorem on the Characterization of
Regular Subspaces, applied to \( \mathcal{U} := \text{Rng } E \), it follows that \( \text{Rng } E \) is regular.

Assume now that \( \mathcal{U} \) is a regular subspace of \( \mathcal{V} \). By the implication
(i) ⇒ (iv) of the Theorem just mentioned, \( \mathcal{U}^\perp \) is then a supplement of \( \mathcal{U} \).
By Prop.3, an idempotent \( E \) with \( \text{Rng } E = \mathcal{U} \) is symmetric if and only if
\( \text{Null } E = \mathcal{U}^\perp \). By Prop.4 of Sect.19 there is exactly one such idempotent.

Notes 41

(1) Most textbooks deal only with what we call “genuine” inner-product spaces and use
the term “inner-product space” only in this restricted sense. The terms “Euclidean
vector space” or even “Euclidean space” are also often used in this sense.

(2) The terms “scalar product” or “dot product” are sometimes used instead of “inner
product”.

(3) Other notations for the value \( u \cdot v \) of the inner product are \( \langle u, v \rangle \), \( (u|v) \), \( \langle u|v \rangle \), and
\( (u, v) \). The last is very common, despite the fact that it clashes with the notation
for the pair with terms \( u \) and \( v \).

(4) Some people use simply \( v^2 \) instead of \( v^2 \) for an inner square. In fact, I have
done this many times myself. We use \( v^2 \) because omitting the dot would lead
to confusion of the compositional square with the inner square of a lineon (see
Sect.44).

(5) The terms of \( b^* \cdot v \) are often called the “contravariant components” of \( v \) relative
to the basis \( b \) and are denoted by \( v^i \). The terms of \( b \cdot v \) are called “covariant
components” and are denoted by \( v_i \).

(6) In most of the literature, the term “orthonormal” is used for what we call “genuinely
orthonormal”.

(7) Some people use the term “non-isotropic” or “non-singular” when we speak of a
“regular” subspace.

(8) Some authors use the term “perpendicular projection” or the term “orthogonal
projection” to mean the same as our “symmetric idempotent”. See also Note 1 to
Sect.19.

(9) The dual \( b^* \) of a basis \( b \) of an inner-product space \( \mathcal{V} \), when regarded as a basis of
\( \mathcal{V} \) rather than \( \mathcal{V}^* \), is often called the “reciprocal” of \( b \).

42 Genuine Inner-Product Spaces

In this section, a genuine inner-product space \( \mathcal{V} \) is assumed to be given. We
then have

\[ v^2 > 0 \quad \text{for all } v \in \mathcal{V}^\times. \] (42.1)

Of course, if \( \mathcal{U} \) is any subspace of \( \mathcal{V} \), then the restriction \( \text{sq|}_\mathcal{U} \) of the inner
square of \( \mathcal{V} \) to \( \mathcal{U} \) is again strictly positive and hence endows \( \mathcal{U} \) with the
natural structure of a genuine inner-product space. Thus every subspace \( U \) of \( V \) is regular and hence has an orthogonal supplement \( U^\perp \).

**Definition 1:** For every \( v \in V \), the number \( |v| \in \mathbb{P} \) defined by
\[
|v| := \sqrt{v \cdot 2}
\] (42.2)
is called the magnitude of \( v \). We say that \( u \in V \) is a unit vector if
\[
|u| = 1.
\] In the case when \( V \) is \( \mathbb{R} \) the magnitude turns out to be the absolute value; hence the notation (42.2) is consistent with the usual notation for absolute values.

It follows from (42.1) that for all \( v \in V \),
\[
|v| = 0 \iff v = 0.
\] (42.3)
The following formula follows directly from (42.2) and (41.5):
\[
|\xi v| = |\xi||v| \quad \text{for all } v \in V, \xi \in \mathbb{R}.
\] (42.4)

**Inner-Product Inequality:** Every pair \((u, v)\) in a genuine inner-product space \( V \) satisfies
\[
|u \cdot v| \leq |u||v|;
\] (42.5)
equality holds if and only if \((u, v)\) is linearly dependent.

**Proof:** Assume, first, that \((u, v)\) is linearly independent. Then, by (42.3), \( |u| \neq 0, |v| \neq 0 \). Moreover, if we put \( w := |v|^2u - (u \cdot v)v \), then \( w \) cannot be zero because it is a linear combination of \((u, v)\) with at least one coefficient, namely \( |v|^2 \), not zero. Hence, by (42.1), we have
\[
0 < w^2 = (|v|^2)^2u^2 - 2|v|^2(u \cdot v)(u \cdot v) + (u \cdot v)^2v^2
= |v|^2(|v|^2|u|^2 - (u \cdot v)^2),
\]
which is equivalent to (42.5) with equality excluded.

Assume, now, that \((u, v)\) is linearly dependent. Then one of \( u \) and \( v \) is a scalar multiple of the other. Without loss we may assume, for example, that \( u = \xi v \) for some \( \xi \in \mathbb{R} \). By (42.4), we then have
\[
|u \cdot v| = |u \cdot (\xi u)| = |\xi u^2| = |\xi||u||u| = |v||u|,
\]
which shows that (42.5) holds with equality.

**Subadditivity of Magnitude:** For every \( u, v \in V \) we have
\[
|u + v| \leq |u| + |v|.
\] (42.6)
**Proof:** Using (42.5) we find
\[ |u + v|^2 = (u + v)^2 = u^2 + 2u \cdot v + v^2 \leq |u|^2 + 2|u||v| + |v|^2 = (|u| + |v|)^2, \]
which is equivalent to (42.6). 

**Proposition 1:** An inner-product space \( V \) is genuine if and only if it has some genuinely orthonormal basis.

**Proof:** Assume that \( V \) has a genuinely orthonormal basis \( e := (e_i \mid i \in I) \) and let \( v \in V^* \) be given. Since \( e = e^* \), we infer from (41.19) that
\[ v^2 = v \cdot v = \sum_{i \in I} (e_i \cdot v)^2, \]
which is strictly positive. Since \( v \in V^* \) was arbitrary, it follows that \( V \) is genuine.

Assume, now, that \( V \) is genuine. Let \( e \) be an orthonormal set in \( V \). We have seen in the previous section that \( e \) is linearly independent. Hence it is a basis if and only if \( \text{Lsp} e = V \). Now, if \( \text{Lsp} e \) is a proper subset of \( V \), then its orthogonal supplement \( (\text{Lsp} e)^\perp \) is not the zero-space. Hence we may choose \( u \in (\text{Lsp} e)^\perp \) and put \( f := \frac{1}{|u|} u \), so that \( f^2 = 1 \). It is clear that \( e \cup \{f\} \) is an orthonormal set that has \( e \) as a proper subset. It follows that every maximal orthonormal set in \( V \) is a basis. Such sets exist because the empty set \( \emptyset \) is orthonormal and because no orthonormal set in \( V \) can have more than \( \dim V \) members.

**Definition 2:** The set of all elements of \( V \) with magnitude strictly less than 1 is called the **unit ball** of \( V \) and is denoted by
\[ \text{Ubl} V := \{ v \in V \mid |v| < 1 \} = \text{sq}^<([0, 1[). \tag{42.7} \]

The set of all elements of \( V \) with magnitude less than 1 is called the **closed unit ball** of \( V \) and is denoted by
\[ \overline{\text{Ubl}} V := \{ v \in V \mid |v| \leq 1 \} = \text{sq}^<([0, 1]). \tag{42.8} \]

The set of all unit vectors in \( V \) is called the **unit sphere** of \( V \) and is denoted by
\[ \text{Usph} V := \{ v \in V \mid |v| = 1 \} = \text{sq}^<\{1\} = \overline{\text{Ubl}} V \setminus \text{Ubl} V. \tag{42.9} \]
43. ORTHOGONAL MAPPINGs

Notes 42

(1) The magnitude |v| is often called “Euclidean norm”, “norm”, or “length”.

(2) Many people use ||v|| instead of |v| for the magnitude. I consider this a waste of a good symbol that could be employed for something else. In fact, in this book we reserve the use of double bars for operator-norms only (see Sect.52) and thus have an easy notational distinction between the magnitude |L| and the operator-norm ||L|| of a linear L (see Example 3 in Sect.52).

(3) The Inner-Product Inequality is often called “Cauchy’s inequality” (particularly in France), “Schwarz’s inequality” (particularly in Germany), or “Bunyakovsky’s inequality” (particularly in Russia). Various combinations of these three names are also sometimes used.

43 Orthogonal Mappings

Let V and V′ be inner-product spaces.

**Definition 1:** A mapping \( R : V \rightarrow V' \) is called an **orthogonal mapping** if it is linear and preserves inner squares, i.e. if \( R \in \text{Lin}(V, V') \) and

\[
(Rv)^2 = v^2 \quad \text{for all} \quad v \in V.
\]  

(43.1)

The set of all orthogonal mappings from V to V′ will be denoted by \( \text{Orth}(V, V') \). An orthogonal mapping that has an orthogonal inverse will be called an **orthogonal isomorphism**. We say that V and V′ are orthogonally isomorphic if there exists an orthogonal isomorphism from V to V′. Of course, if V and V′ are orthogonally isomorphic, they are also linearly isomorphic; but they may be linearly isomorphic without being orthogonally isomorphic (see Prop.1 of Sect.47 below). The following is evident from Def.1 and from Props.1, 2 of Sect.13.

**Proposition 1:** The identity mapping of an inner-product space is orthogonal. The composite of two orthogonal mappings is orthogonal. If an orthogonal mapping is invertible, its inverse is again orthogonal, and hence it is an orthogonal isomorphism.

The following is a direct consequence of Def.1 and of Prop.3 of Sect.27, applied to \( Q := \text{sq} \).

**Proposition 2:** \( R \in \text{Lin}(V, V') \) is orthogonal if and only if

\[
R^\top R = 1_V,
\]

(43.2)
or, equivalently, R preserves inner products.

As an immediate consequence of Prop.2 and the Pigeonhole Principle for Linear Mappings we obtain the following result.
Proposition 3: For every orthogonal mapping \( R \in \text{Orth}(V, V') \) we have: \( R \) is injective, \( \text{Rng} R \) is a regular subspace of \( V' \), and \( R|_{\text{Rng} R} \) is an orthogonal isomorphism from \( V \) to \( \text{Rng} R \). Moreover, \( R \) is an orthogonal isomorphism if and only if \( R \) is surjective or, equivalently,

\[
RR^\top = 1_{V'}.
\] (43.3)

This is the case if and only if \( \dim V = \dim V' \).

We write \( \text{Orth} V := \text{Orth}(V, V) \) and call its members orthogonal lineons. In view of Prop.1 and Prop.3, \( \text{Orth} V \) is a subgroup of \( \text{Lis} V \) and consists of all orthogonal automorphisms of \( V \). The group \( \text{Orth} V \) is called the orthogonal group of the inner product space \( V \).

Proposition 4: Let \( b := (b_i \mid i \in I) \) be a basis of the inner-product space \( V \), let \( V' \) be an inner-product space and \( R \in \text{Lin}(V, V') \). Then \( R \) is orthogonal if and only if

\[
R b_i \cdot R b_k = b_i \cdot b_k \quad \text{for all } i, k \in I.
\] (43.4)

In particular, if \( e := (e_i \mid i \in I) \) is a genuinely orthonormal basis, then \( R \) is orthogonal if and only if \( R e := (R e_i \mid i \in I) \) is a genuinely orthonormal family.

Proof: Since \( R b_i \cdot R b_k = (R^\top R b_i) \cdot b_k \) for all \( i, k \in I \), (43.4) states that for each \( i \in I \), \( (R^\top R b_i) \in V^* \) and \( b_i \in V^* \) agree on a basis. Hence, by Prop.2 of Sect.16, (43.4) holds if and only if \( R^\top R b_i = b_i \) for all \( i \in I \), which, in turn, is the case if and only if \( R^\top R = 1_V \). The assertion follows then from Prop.2.

Proposition 5: Let \( V, V' \) be genuine inner-product spaces. Then \( V \) and \( V' \) are orthogonally isomorphic if and only if \( \dim V = \dim V' \).

Proof: The “only if” part follows from Cor.2 to the Theorem on Characterization of Dimension of Sect.17.

Assume that \( \dim V = \dim V' =: n \). By Prop.1 of Sect.42 and by Cor.1 to the Theorem on Characterization of Dimension, we may choose genuinely orthonormal bases \( e := (e_i \mid i \in n) \) and \( e' := (e'_i \mid i \in n) \) of \( V \) and \( V' \), respectively. By Prop.2 of Sect.16, there is an invertible linear mapping \( R : V \to V' \) such that \( R e_i = e'_i \) for all \( i \in n \). By Prop.4, \( R \) is an orthogonal isomorphism.

Proposition 6: Let \( V, V' \) be inner-product spaces. Assume that \( R : V \to V' \) preserves inner products, i.e., that

\[
R(\mathbf{u}) \cdot R(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v} \quad \text{for all } \mathbf{u}, \mathbf{v} \in V,
\] (43.5)
and that LspRng R is a regular subspace of V\\prime. Then R is linear and hence orthogonal.

**Proof:** To show that R preserves sums, let u, v \in V be given. It follows from (43.5) that
\[
R(u + v) \cdot R(w) = (u + v) \cdot w = u \cdot w + v \cdot w
\]
and hence
\[
(R(u + v) - R(u) - R(v)) \cdot R(w) = 0
\]
for all w \in V. This means that
\[
R(u + v) - R(u) - R(v) \in (\text{Rng } R)^\perp.
\]
Since LspRng R is regular we have LspRng R \cap (\text{Rng } R)^\perp = \{0\}. It follows that
\[
R(u + v) - R(u) - R(v) = 0.
\]
Since u, v \in V were arbitrary, we conclude that R preserves sums. A similar argument shows that R preserves scalar multiples and hence is linear.

**Pitfall:** If R : \mathcal{V} \rightarrow \mathcal{V}\\prime preserves inner products but LspRng R is not regular, then R need not be linear. For example, if \mathcal{V} = \{0\} and n \in \mathcal{V}\\prime is such that n^2 = 0 but n \neq 0, then the mapping R : \mathcal{V} \rightarrow \mathcal{V}\\prime defined by R(0) = n preserves inner products but is not linear.

In view of Prop.3, Prop.6 has the following corollary.

**Proposition 7:** A surjective mapping that preserves inner products is necessarily an orthogonal isomorphism.

Notes 43

(1) Orthogonal mappings are commonly called “orthogonal transformations”, but the definition is often restricted to the case in which the domain and codomain coincide (i.e. when we use “orthogonal lineon”) and the spaces involved are genuine inner-product spaces.

(2) The notations O_n, O_n(\mathbb{R}), or O(n, \mathbb{R}) are often used for the orthogonal group Orth \mathbb{R}^n.

(3) If the inner-product is double-signed and if its index is 1 (see Sect.47), an orthogonal lineon is often called a “Lorentz transformation”, especially in the context of the theory of relativity. The orthogonal group is then called the “Lorentz group”.

44 Induced Inner Products

Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be inner-product spaces. Then $\mathcal{V}_1 \times \mathcal{V}_2$ carries a natural induced structure of a linear space, as explained in Sect. 14. We endow $\mathcal{V}_1 \times \mathcal{V}_2$ also with the structure of an inner-product space by prescribing its inner square by the rule

$$(v_1, v_2)^2 := v_1^2 + v_2^2 \quad \text{for all} \quad v_1 \in \mathcal{V}_1, \; v_2 \in \mathcal{V}_2. \tag{44.1}$$

The inner product in $\mathcal{V}_1 \times \mathcal{V}_2$ is given by

$$(u_1, u_2) \cdot (v_1, v_2) = u_1 \cdot v_1 + u_2 \cdot v_2$$

for all $(u_1, u_2), (v_1, v_2) \in \mathcal{V}_1 \times \mathcal{V}_2$. More generally, if $(\mathcal{V}_i \mid i \in I)$ is a finite family of inner-product spaces, we endow the product-space $\prod_{i \in I} (\mathcal{V}_i \mid i \in I)$ with the natural structure of an inner-product space by prescribing the inner square by the rule

$$v^2 := \sum_{i \in I} v_i^2 \quad \text{for all} \quad v \in \prod_{i \in I} \mathcal{V}_i. \tag{44.2}$$

It is clear that $\prod_{i \in I} (\mathcal{V}_i \mid i \in I)$ is genuine if all the $\mathcal{V}_i, i \in I$, are genuine.

**Remark:** The inner products on $\mathcal{V}_1$ and $\mathcal{V}_2$ induce on $\mathcal{V}_1 \times \mathcal{V}_2$, in a natural manner, non-degenerate quadratic forms other than the inner square given by (44.1). For example, one such quadratic form $Q : \mathcal{V}_1 \times \mathcal{V}_2 \to \mathbb{R}$ is given by

$$Q(v_1, v_2) := v_1^2 - v_2^2 \tag{44.3}$$

This form is double-signed (see Sect. 27) when $\mathcal{V}_1$ and $\mathcal{V}_2$ are genuine inner-product spaces of dimension greater than zero.

**Theorem on the Induced Inner Product:** Let $\mathcal{V}$ and $\mathcal{W}$ be inner-product spaces. Then $\text{Lin}(\mathcal{V}, \mathcal{W})$ has the natural structure of an inner product space, obtained by prescribing its inner square by the rule

$$L^2 := \text{tr}_\mathcal{V}(L^\top L) \quad \text{for all} \quad L \in \text{Lin}(\mathcal{V}, \mathcal{W}). \tag{44.4}$$

The corresponding inner product is given by

$$L \cdot M := \text{tr}_\mathcal{V}(L^\top M) = \text{tr}_\mathcal{W}(ML^\top) \tag{44.5}$$

for all $L, M \in \text{Lin}(\mathcal{V}, \mathcal{W})$.

If $\mathcal{V}$ and $\mathcal{W}$ are genuine inner-product spaces, so is $\text{Lin}(\mathcal{V}, \mathcal{W})$. 
44. **INDUCED INNER PRODUCTS**

**Proof:** The bilinearity of the mapping

\[
((L, M) \mapsto \text{tr}_V(L^\top M)) : (\text{Lin}(V, W))^2 \to \mathbb{R} \tag{44.6}
\]

is a consequence of the linearity of transposition, the bilinearity of composition, and the linearity of the trace (see Chap.2). To prove the symmetry of the mapping (44.6) we use (26.7), (21.6), and (22.4) to obtain

\[
M \cdot L = \text{tr}_V(M^\top L) = \text{tr}_V((M^\top L)^\top) = \text{tr}_V(L^\top M^\top) = \text{tr}_V(L^\top M) = L \cdot M
\]

for all \(L, M \in \text{Lin}(V, W)\).

The Representation Theorem for Linear Forms on \(\text{Lin}(V, W)\) of Sect.26 states that the mapping that associates with \(H \in \text{Lin}(W, V)\) the linear form \((M \mapsto \text{tr}_V(HM)) \in (\text{Lin}(V, W))^*\) is a linear isomorphism. The transposition \((L \mapsto L^\top) : \text{Lin}(V, W) \to \text{Lin}(W, V)\) is also a linear isomorphism. It follows that their composite, the linear mapping from \(\text{Lin}(V, W)\) into \((\text{Lin}(V, W))^*\) associated with the bilinear mapping (44.6), is an isomorphism and hence injective. This means that (44.4) indeed defines a non-degenerate quadratic form whose associated bilinear form is given by (44.6).

The last equality of (44.5) is an immediate consequence of (26.6).

The last statement of the Theorem is an immediate consequence of Prop.1 of Sect.42 and the following Lemma.

**Lemma:** If \(e := (e_i | i \in I)\) is a genuinely orthonormal basis of \(V\) and \(L \in \text{Lin}(V, W)\), then

\[
L^2 = \text{tr}_W(LL^\top) = \sum_{i \in I}(Le_i)^2. \tag{44.7}
\]

**Proof:** Since \(e\) coincides with its dual, we have, by (25.14),

\[
e_\perp = \sum_{i \in I} e_i \otimes e_i.
\]

Using Prop.3 of Sect.25, we obtain

\[
LL^\top = L_\perp L^\top = \sum_{i \in I} L(e_i \otimes e_i) L^\top = \sum_{i \in I} (Le_i) \otimes (Le_i).
\]

Since \(\text{tr}_W((Le_i) \otimes (Le_i)) = (Le_i) \cdot (Le_i) = (Le_i)^2\) by (26.3), and since the trace is linear, the last of the equalities (44.7) follows. The first is a consequence of (44.5). \(\blacksquare\)
Let $I$ and $J$ be finite sets. As we have seen in Sect. 41, the spaces $\mathbb{R}^I$, $\mathbb{R}^J$ and $\mathbb{R}^{J \times I}$ can all be regarded as inner-product spaces. Therefore, the Theorem on the Induced Inner Product shows that $\text{Lin}(\mathbb{R}^I, \mathbb{R}^J)$ carries the natural structure of an inner product space. We claim that this structure is compatible with the identification $\text{Lin}(\mathbb{R}^I, \mathbb{R}^J) \cong \mathbb{R}^{J \times I}$ (see Sect. 16). Indeed, given $M \in \text{Lin}(\mathbb{R}^I, \mathbb{R}^J) \cong \mathbb{R}^{J \times I}$ we have, by (16.7) and (23.9)

$$(M^T M)_{i,k} = \sum_{j \in I} (M^T)_{i,j} M_{j,k} = \sum_{j \in I} M_{j,i} M_{j,k}$$

for all $i, k \in I$ and hence, by (26.8),

$$\text{tr}(M^T M) = \sum_{i \in I} (M^T M)_{i,i} = \sum_{(j,i) \in J \times I} (M_{j,i})^2. \quad (44.8)$$

We now assume that inner-product spaces $V$ and $W$ are given. The following formulas, valid for all $L \in \text{Lin}(V, W), v \in V,$ and $w \in W$ follow easily from the definitions (44.4) and (44.5):

$$L \cdot (w \otimes v) = w \cdot Lv = L(v, w), \quad (44.9)$$
$$\left(w \otimes v\right)^2 = (v^2)(w^2). \quad (44.10)$$

In view of (26.9) we have

$$1^2_V = \dim V. \quad (44.11)$$

**Proposition 1:** The transposition $(L \mapsto L^\top) : \text{Lin}(V, W) \rightarrow \text{Lin}(W, V)$ is an orthogonal isomorphism, i.e. we have

$$L \cdot M = L^\top \cdot M^\top \quad \text{for all } L, M \in \text{Lin}(V, W). \quad (44.12)$$

**Proof:** Given $L, M \in \text{Lin}(V, W)$, we may use (44.5) to obtain $L^\top \cdot M^\top = \text{tr}_W(L^\top T M^\top) = \text{tr}_V(L M^\top) = M \cdot L$ and hence (44.12).

**Proposition 2:** The subspaces $\text{Sym} V$ and $\text{Skew} V$ of $\text{Lin} V$ are orthogonal supplements of each other (and hence both regular).

**Proof:** By (44.5) we have, for all $S \in \text{Sym} V$ and all $A \in \text{Skew} V$,

$$S \cdot A = \text{tr}(S^\top A) = \text{tr}(SA) = -\text{tr}(SA^\top) = -A \cdot S$$

and hence $S \cdot A = 0$. It follows that $\text{Sym} V \subset (\text{Skew} V)^\perp$ and $\text{Skew} V \subset (\text{Sym} V)^\perp$. We already know (Prop. 7 of Sect. 24) that $\text{Sym} V$ and $\text{Skew} V$ are supplementary.
From now on we assume that both $V$ and $W$ are genuine inner-product spaces. By the last assertion of the Theorem on the Induced Inner Product, $\text{Lin}(V, W)$ is then genuine, too. Thus, according to the Definition of Sect.42, every $L \in \text{Lin}(V, W)$ has a magnitude

$$|L| := \sqrt{L^2} = \sqrt{\text{tr}_V(L^\top L)} = \sqrt{\text{tr}_W(LL^\top)}. \quad (44.13)$$

In addition to the rules concerning magnitudes in general, the magnitude in $\text{Lin}(V, W)$ satisfies the following rules.

**Proposition 3:** For all $v \in V$, $w \in W$, and $L \in \text{Lin}(V, W)$ we have

$$|w \otimes v| = |w||v|, \quad (44.14)$$

$$|Lv| \leq |L||v|. \quad (44.15)$$

**Proof:** (44.14) is an immediate consequence of (44.10). To prove (44.15), we use (44.9), the Inner-Product Inequality of Sect.42 as applied to $\text{Lin}(V, W)$, and (44.14) to show that

$$|w \cdot Lv| = |L \cdot (w \otimes v)| \leq |L||w \otimes v| = |L||w||v|. $$

Putting $w := Lv$ we get $|Lv|^2 \leq |Lv||L||v|$, which is equivalent to (44.15).

In view of (44.11), we have

$$|1_V| = \sqrt{\dim(V)}. \quad (44.16)$$

**Proposition 4:** Let $V$, $V'$ and $V''$ be genuine inner-product spaces and let $L \in \text{Lin}(V, V')$, $M \in \text{Lin}(V', V'')$. Then

$$|ML| \leq |M||L|. \quad (44.17)$$

**Proof:** Choose a genuinely orthonormal basis $e := (e_i \mid i \in I)$ of $V$. By the Lemma, (44.7), we have

$$|ML|^2 = (ML)^2 = \sum_{i \in I}(MLe_i)^2 = \sum_{i \in I}M(Le_i)^2.$$

Applying (44.15) to each term, we get

$$|ML|^2 \leq \sum_{i \in I}|M|^2|Le_i|^2 = |M|^2 \sum_{i \in I}(Le_i)^2.$$

Applying the Lemma, (44.7), again, we obtain $|ML|^2 \leq |M|^2|L|^2$, which is equivalent to (44.17).
45 Euclidean Spaces

**Definition 1:** A Euclidean space is a finite-dimensional flat space \( E \) endowed with additional structure by the prescription of a separation function

\[
\text{sep} : E \times E \rightarrow \mathbb{R}
\]

such that

(a) \( \text{sep} \) is translation invariant, i.e.

\[
\text{sep} \circ (v \times v) = \text{sep}
\]

for all \( v \) in the translation space \( V \) and \( E \),

(b) For some \( x \in E \), the mapping

\[
(v \mapsto \text{sep}(x, x + v)) : V \rightarrow \mathbb{R}
\]

is a non-degenerate quadratic form on \( V \).

It follows from (a) that the mapping (45.2) from \( V \) to \( \mathbb{R} \) does not depend on the choice of \( x \). Hence the translation space \( V \) of a Euclidean space \( E \) carries the natural structure of an inner-product space whose inner square \( v \mapsto v^2 \) satisfies

\[
\text{sep}(x, y) = (y - x)^2 \quad \text{for all} \quad x, y \in E.
\]

Conversely, if \( E \) is a finite-dimensional flat space and if its translation space \( V \) is endowed with additional structure as in Def.1 of Sect.41 so as to make it an inner-product space, then \( E \) acquires the natural structure of a Euclidean space when we use (45.3) to define the separation function.

Every inner-product space \( V \) has the natural structure of a Euclidean space that is its own (external) translation space. Indeed, the natural structure of \( V \) as a flat space is the one described by Prop.3 of Sect.32, and the inner-product of \( V \) then gives \( V \) the natural structure of a Euclidean space as remarked above.

We say that a flat \( F \) in a Euclidean space \( E \) is regular if its direction space is a regular subspace of \( V \). If \( F \) is regular, then \( \text{sep}|_{F \times F} \) endows \( F \) with the structure of a Euclidean space; but if \( F \) is not regular, it does not have a natural structure of a Euclidean space.

**Definition 2:** Let \( E, E' \) be Euclidean spaces. We say that the mapping \( \alpha : E \rightarrow E' \) is Euclidean if it is flat and preserves separation, so that

\[
\text{sep}' \circ (\alpha \times \alpha) = \text{sep},
\]
where \( \text{sep} \) and \( \text{sep}' \) are the separation functions of \( \mathcal{E} \) and \( \mathcal{E}' \), respectively.

A Euclidean mapping that has a Euclidean inverse is called a Euclidean isomorphism.

The following is evident from Def. 2, (45.3), (33.4), and Def. 1 in Sect. 43.

**Proposition 1:** The mapping \( \alpha : \mathcal{E} \rightarrow \mathcal{E}' \) is Euclidean if and only if it is flat and has an orthogonal gradient.

Using this Proposition and Prop. 1 of Sect. 43, we obtain:

**Proposition 2:** The identity mapping of a Euclidean space is a Euclidean mapping. The composite of two Euclidean mappings is Euclidean. If a Euclidean mapping is invertible, then its inverse is again Euclidean, and hence it is a Euclidean isomorphism.

In view of Prop. 3 of Sect. 43, we have:

**Proposition 3:** Every Euclidean mapping \( \alpha : \mathcal{E} \rightarrow \mathcal{E}' \) is injective, its range \( \text{Rng} \alpha \) is a regular subspace of \( \mathcal{E}' \), and \( \alpha|_{\text{Rng} \alpha} \) is a Euclidean isomorphism from \( \mathcal{E} \) onto \( \text{Rng} \alpha \). Moreover, \( \alpha \) is a Euclidean isomorphism if and only if \( \alpha \) is surjective, and this is the case if and only if \( \dim \mathcal{E} = \dim \mathcal{E}' \).

The set of all Euclidean automorphisms of a Euclidean space \( \mathcal{E} \) will be denoted by \( \text{EisE} \). It is a subgroup of the group \( \text{FisE} \) of all flat automorphisms of \( \mathcal{E} \). The mapping \( g : \text{EisE} \rightarrow \text{OrthV} \) defined by \( g(\alpha) := \nabla \alpha \) is a surjective group-homomorphism whose kernel is the translation group \( V \). In fact, \( g \) is obtained from the homomorphism \( h : \text{FisE} \rightarrow \text{LisV} \) described at the end of Sect. 33 by the adjustment \( g := h|_{\text{EisE}} \).

**Proposition 4:** Let \( \mathcal{E}, \mathcal{E}' \) be Euclidean spaces, let \( \alpha : \mathcal{E} \rightarrow \mathcal{E}' \) be a flat mapping, and let \( p := (p_i \mid i \in I) \) be a flat basis of \( \mathcal{E} \) such that

\[
\text{sep}(p_i, p_j) = \text{sep}'(\alpha(p_i), \alpha(p_j)) \quad \text{for all} \quad i, j \in I. \tag{45.5}
\]

Then \( \alpha \) is a Euclidean mapping.

**Proof:** Since \( \alpha \) is flat we have, by (33.4),

\[
\alpha(p_i) - \alpha(p_j) = \nabla \alpha(p_i - p_j) \quad \text{for all} \quad i, j \in I. \tag{45.6}
\]

We now choose \( k \in I \) and put \( I' := I \setminus \{k\} \) and \( u_i := p_i - p_k \) for all \( i \in I' \). Then (45.6) gives

\[
\alpha(p_i) - \alpha(p_k) = (\nabla \alpha) u_i \quad \text{for all} \quad i \in I'
\]

and

\[
\alpha(p_i) - \alpha(p_j) = (\nabla \alpha)(u_i - u_j) \quad \text{for all} \quad i, j \in I'.
\]

In view of (45.3), it follows from (45.5) that

\[
((\nabla \alpha) u_i)^2 = u_i^2 \quad \text{for all} \quad i \in I'. \tag{45.7}
\]
and
\[ ((\nabla \alpha)(u_i - u_j))^2 = (u_i - u_j)^2 \]
for all \( i, j \in I' \),
which is equivalent to
\[ ((\nabla \alpha)u_i)^2 - 2((\nabla \alpha)u_i) \cdot ((\nabla \alpha)u_j) + ((\nabla \alpha)u_j)^2 = u_i^2 - 2u_i \cdot u_j + u_j^2. \]
Using (45.7) this gives
\[ (\nabla \alpha)u_i \cdot (\nabla \alpha)u_j = u_i \cdot u_j \]
for all \( i, j \in I' \).

Since \((u_i | i \in I')\) is a basis of \( V \) by Prop.2 of Sect.35, it follows from Prop.4 of Sect.43 that \( \nabla \alpha \) is orthogonal. Hence \( \alpha \) is Euclidean by Prop.1.

The following result shows that, in Def.2, the requirement that \( \alpha \) be flat can in many cases be omitted.

**Proposition 5:** Let \( E, E' \) be Euclidean spaces and let \( \alpha : E \to E' \) be a mapping that preserves separation, i.e. satisfies (45.4). Then \( \alpha \) is flat and hence Euclidean provided that the flat span of \( \text{Rng} \alpha \) is regular.

**Proof:** Let \( V \) and \( V' \) be the translation spaces of \( E \) and \( E' \), respectively.
We choose \( q \in E \) and define \( R : V \to V' \) by
\[ R(v) := \alpha(q + v) - \alpha(q) \quad \text{for all} \quad v \in V. \quad (45.8) \]
We have, for all \( u, v \in V \)
\[ R(u) - R(v) = \alpha(q + u) - \alpha(q + v). \]
Since \( \alpha \) preserves separation, it follows that
\[ (R(u) - R(v))^2 = (u - v)^2 \quad \text{for all} \quad u, v \in V. \]
By an argument similar to one given in the proof of Prop.4, we conclude that
\[ R(u) \cdot R(v) = u \cdot v \quad \text{for all} \quad u, v \in V. \]
Now, (45.8) shows that \( \text{Rng} R = (\text{Rng} \alpha) - \alpha(q) \). Hence, by (32.6), \( \text{LspRng} R \)
is the direction space of \( \text{Fsp Rng} \alpha \). If \( \text{Fsp Rng} \alpha \) is regular, so is \( \text{LspRng} R \), and we can apply Prop.6 of Sect.43 to conclude that \( R \) is orthogonal. It follows from (45.8) that \( \alpha \) is flat and that \( R = \nabla \alpha. \]

The following corollary to Prop.5 shows that the prescription of the separation function alone is sufficient to determine the structure of a Euclidean space on a set.
Proposition 6: Let a set \( \mathcal{E} \) and a function \( \text{sep} : \mathcal{E} \times \mathcal{E} \to \mathbb{R} \) be given. Then \( \mathcal{E} \) can be endowed with at most one structure of a Euclidean space whose separation function is \( \text{sep} \).

Proof: Let two such structures be given with \( V \) and \( V' \) as the corresponding translation spaces. The identity \( 1_\mathcal{E} \) preserves separation and hence Prop. 5 may be applied to it when \( \mathcal{E} \) as domain of \( 1_\mathcal{E} \) is considered as endowed with the first structure and as codomain of \( 1_\mathcal{E} \) with the second. Let \( R \in \text{Orth}(V, V') \) be the gradient of \( 1_\mathcal{E} \) when interpreted in this manner. In view of (33.1), it follows that \( 1_\mathcal{E} \circ v = (Rv) \circ 1_\mathcal{E} \) and hence \( v = Rv \) for all \( v \in V \), which means that \( V \subset V' \) and \( R = 1_{V \subset V'} \). Reversing the roles of \( V \) and \( V' \) we get \( V' \subset V \) and hence \( V = V' \) and \( R = 1_V \). Since \( R \) is an orthogonal isomorphism, it follows that \( V \) and \( V' \) coincide not only as subgroups of \( \text{Perm} \mathcal{E} \) but also as inner-product spaces.

Proposition 7: Let \( \mathcal{E} \) be a Euclidean space and \( p := (p_i \mid i \in I) \) a flatly spanning family of points in \( \mathcal{E} \). Then for all \( x, y \in \mathcal{E} \)

\[
(\text{sep}(x, p_i) = \text{sep}(y, p_i) \text{ for all } i \in I) \implies x = y.
\]

Proof: Let \( x, y \in \mathcal{E} \) be given and let \( z \) be the midpoint of \( (x, y) \) (see Example 1 of Sect. 34). This means that \( x = z + u \) and \( y = z - u \) for a suitable \( u \in V \). The family \( b := ((p_i - z) \mid i \in I) \) spans \( V \) and we have

\[
\begin{align*}
\text{sep}(x, p_i) &= (p_i - x)^2 = (b_i - u)^2 \\
\text{sep}(y, p_i) &= (p_i - y)^2 = (b_i + u)^2
\end{align*}
\]

for all \( i \in I \). Hence, if \( \text{sep}(x, p_i) = \text{sep}(y, p_i) \) for all \( i \in I \) we have

\[
b_i \cdot u = \frac{1}{4}((b_i + u)^2 - (b_i - u)^2) = 0
\]

for all \( i \in I \). Since \( b \) is spanning, it follows that \( u \in (\text{Rng} b)^\perp = V^\perp = \{0\} \) and hence \( u = 0 \), which means that \( x = z = y \).

Notes 45

(1) What we call a “genuine Euclidean space” (see Def. 1 of Sect. 46) is usually just called a “Euclidean space”, and non-genuine Euclidean spaces are often referred to as “pseudo-Euclidean spaces”. I believe it is more convenient to have a term that covers both cases.

The term “Euclidean space” is often misused to mean a genuine inner-product space or even \( \mathbb{R}^n \).

(2) In the past, the term “separation” has been used only in the special case when the space \( \mathcal{E} \) is a model of the event-world in the theory of relativity.
Genuine Euclidean Spaces, Congruences

Definition 1: We say that a Euclidean space $E$ is genuine if its separation function $\text{sep}$ has only positive values, i.e. if $\text{Rng} \text{sep} \subseteq \mathbb{P}$. The distance function

$$\text{dst} : E \times E \to \mathbb{P}$$

is then defined by

$$\text{dst}(x, y) := \sqrt{\text{sep}(x, y)} \quad \text{for all} \quad x, y \in E. \quad (46.1)$$

It is clear from (45.3) that a Euclidean space $E$ is genuine if and only if its translation space $V$ is a genuine inner-product space. Hence all flats in a genuine Euclidean space are regular and have natural structures of genuine Euclidean spaces. Comparing (46.1) with (42.2), we see that (45.3) is equivalent to

$$\text{dst}(x, y) = |y - x| \quad \text{for all} \quad x, y \in E. \quad (46.2)$$

Using (42.4) with $\xi := -1$ and (42.6) and (42.3), we obtain:

**Proposition 1:** The distance function $\text{dst}$ of a genuine Euclidean space $E$ has the following properties, valid for all $x, y, z \in E$:

$$\text{dst}(x, y) = \text{dst}(y, x), \quad (46.3)$$
$$\text{dst}(x, y) + \text{dst}(y, z) \geq \text{dst}(x, z), \quad (46.4)$$
$$\text{dst}(x, y) = 0 \iff x = y. \quad (46.5)$$

Let $q \in E$ and $\rho \in \mathbb{P}^\times$ be given. Then

$$\text{Ball}_{q, \rho}(E) := \{ x \in E \mid \text{dst}(x, q) < \rho \} \quad (46.6)$$

is called the **ball of radius** $\rho$ **centered at** $q$,

$$\overline{\text{Ball}}_{q, \rho}(E) := \{ x \in E \mid \text{dst}(x, q) \leq \rho \} \quad (46.7)$$

is called the **closed ball** of radius $\rho$ centered at $q$, and

$$\text{Sph}_{q, \rho}(E) := \{ x \in E \mid \text{dst}(x, q) = \rho \} \quad (46.8)$$

$$= \overline{\text{Ball}}_{q, \rho}(E) \setminus \text{Ball}_{q, \rho}(E)$$

is called the **sphere** of radius $\rho$ centered at $q$. If $\dim E = 2$, the term “disc” is often used instead of “ball”, and the term “circle” instead of “sphere”.

46. GENUINE EUCLIDEAN SPACES, CONGRUENCES

In view of Def.2 of Sect.42, we have
\[
\text{Ball}_{q,\rho}(\mathcal{E}) = q + \rho U_{bl} V, \quad (46.9)
\]
\[
\text{Ball}_{q,\rho}(\mathcal{E}) = q + \rho U_{bl} V, \quad (46.10)
\]
\[
\text{Sph}_{q,\rho}(\mathcal{E}) = q + \rho U_{sp} V. \quad (46.11)
\]

We now assume that a genuine Euclidean space \( \mathcal{E} \) is given.

**Definition 2:** We say that the families \( p := (p_i \mid i \in I) \) and \( p' := (p'_i \mid i \in I) \) of points in \( \mathcal{E} \) are congruent in \( \mathcal{E} \) if there is a Euclidean automorphism \( \alpha \) of \( \mathcal{E} \) such that \( \alpha(p_i) = p'_i \) for all \( i \in I \).

**Congruence Theorem:** The families \( p := (p_i \mid i \in I) \) and \( p' := (p'_i \mid i \in I) \) are congruent if and only if
\[
\text{dst}(p_i, p_j) = \text{dst}(p'_i, p'_j) \quad \text{for all} \quad i, j \in I. \quad (46.12)
\]

**Proof:** The “only if” part follows from the definitions.

Assume that (46.12) holds. Put \( \mathcal{F} := \text{Fsp Rng } p \) and choose a subset \( K \) of \( I \) such that \( p|_K = (p_k \mid k \in K) \) is a flat basis of \( \mathcal{F} \). Put \( \mathcal{F}' := \text{Fsp Rng } p' \).

By Prop.5 of Sect.35, there is a unique flat mapping \( \beta : \mathcal{F} \to \mathcal{F}' \) such that \( \beta(p_k) = p'_k \) for all \( k \in K \). By Prop.4 of Sect.45, we can conclude from (46.12) that \( \beta \) must be Euclidean and hence injective (see Prop.3 of Sect.45).

It follows that \( \dim \mathcal{F} \leq \dim \mathcal{F}' \). Reversing the roles of \( p \) and \( p' \), we also conclude that \( \dim \mathcal{F}' \leq \dim \mathcal{F} \). Therefore, we have \( \dim \mathcal{F} = \dim \mathcal{F}' \) and \( \beta \) is a Euclidean isomorphism. Hence \( p'|_K = \beta_\ast(p|_K) \) is a flat basis of \( \mathcal{F}' \). Since \( \beta \) preserves distance, we have for all \( x \in \mathcal{F} \)
\[
\text{dst}(x, p_k) = \text{dst}(\beta(x), \beta(p_k)) = \text{dst}(\beta(x), p'_k)
\]
for all \( k \in K \). Thus, using (46.12) we get
\[
\text{dst}(\beta(p_i), p'_k) = \text{dst}(p_i, p_k) = \text{dst}(p'_i, p'_k)
\]
for all \( k \in K, i \in I \). Using Prop.7 of Sect.45, we conclude that
\[
\beta(p_i) = p'_i \quad \text{for all} \quad i \in I. \quad (46.13)
\]

Let \( \mathcal{U} \) and \( \mathcal{U}' \) be the direction spaces of \( \mathcal{F} \) and \( \mathcal{F}' \). Since
\[
\dim \mathcal{U} = \dim \mathcal{F} = \dim \mathcal{F}' = \dim \mathcal{U}',
\]
it follows that \( \dim \mathcal{U}^\perp = \dim \mathcal{U}'^\perp \). Therefore, in view of Prop.5 of Sect.43, we may choose an orthogonal isomorphism \( R : \mathcal{U}^\perp \to \mathcal{U}'^\perp \). Since \( \mathcal{E} = \)}
CHAPTER 4. INNER-PRODUCT SPACES, EUCLIDEAN SPACES

\[ \mathcal{F} + \mathcal{U}^\perp = \mathcal{F}' + \mathcal{U}'^\perp, \]

there is a unique Euclidean isomorphism \( \alpha : \mathcal{E} \to \mathcal{E} \)
such that

\[ \alpha(x + w) = \beta(x) + Rw \]

for all \( x \in \mathcal{F}, \ w \in \mathcal{U}^\perp. \)

By (46.13) we have \( \alpha(p_i) = p'_i \) for all \( i \in I, \) showing that \( p \) and \( p' \) are congruent.

The Congruence Theorem shows that two families of points in \( \mathcal{E} \) are congruent in \( \mathcal{E} \) if and only if they are congruent in any flat \( \mathcal{F} \) that includes the ranges of both families. It is for this reason that we may simply say “congruent” rather than “congruent in \( \mathcal{E} \)”.

**Definition 3:** Let \( S \) and \( S' \) be two subsets of \( \mathcal{E} \). We say that a mapping \( \gamma : S \to S' \) is a congruence if \( (x \mid x \in S) \) and \( (\gamma(x) \mid x \in S) \) are congruent families of points. The set of all congruences from a given set \( S \) to itself is called the symmetry-group of \( S \) and is denoted by \( \text{Cong}_S \).

The Congruence Theorem has the following immediate consequence:

**Corollary:** The mapping \( \gamma : S \to S' \) is a congruence if and only if it is surjective and preserves distance.

All congruences are invertible mappings and \( \text{Cong}_S \) is a subgroup of the permutation group \( \text{Perm}_S \) of \( S \). The group \( \text{Cong}_S \) describes the internal symmetry of \( S \). For example, if \( \#S = 3 \), then \( S \) can be viewed as the set of vertices of a triangle. If the triangle is equilateral, then \( \text{Cong}_S = \text{Perm}_S \).

If the triangle is isosceles but not equilateral, then \( \text{Cong}_S \) is a two-element subgroup of \( \text{Perm}_S \) that contains \( 1_S \) and one switch. If the triangle is scalene, then \( \text{Cong}_S \) is the identity-subgroup \( \{1_S\} \) of \( \text{Perm}_S \).

**Notes 46**

1. In view of Prop.5 of Sect.45, it turns out that a mapping between genuine Euclidean spaces is a Euclidean isomorphism if and only if it is an isometry, i.e. an invertible mapping that preserves distances. It is for this reason that many people say “isometry” in place of “Euclidean isomorphism”.

47 Double-Signed Inner-Product Spaces

We assume that an inner-product space \( \mathcal{V} \) with inner square \( \text{sq} \) is given. If \( \text{sq} \) is strictly positive, i.e. if the space is a genuine inner-product space, then the results of Sect.42 apply. If \( \text{sq} \) is strictly negative, then a change from \( \text{sq} \) to \(-\text{sq}\) converts the space into a genuine inner-product space, and the results of Sect.42, with suitable adjustments of sign, still apply. In particular, two
inner-product spaces with strictly negative inner squares are isomorphic if they have the same dimension.

The results of this section are of interest only if the inner product is double-signed, i.e. if sq is neither strictly positive nor strictly negative. Recall that a subspace $\mathcal{U}$ of $\mathcal{V}$ is called regular if $\text{sq}|_{\mathcal{U}}$ is non-degenerate (see Sect. 41). We say that $\mathcal{U}$ is

- singular if it is not regular,
- totally singular if $\text{sq}|_{\mathcal{U}} = 0$,
- positive-regular if $\text{sq}|_{\mathcal{U}}$ is strictly positive,
- negative-regular if $\text{sq}|_{\mathcal{U}}$ is strictly negative.

**Definition 1:** Let $\mathcal{V}$ be an inner product space. The greatest among the dimensions of all positive-regular [negative-regular] subspaces of $\mathcal{V}$ will be denoted by $\text{sig}^+\mathcal{V}$ [$\text{sig}^-\mathcal{V}$]. The pair $(\text{sig}^+\mathcal{V}, \text{sig}^-\mathcal{V})$ is called the signature of $\mathcal{V}$.

The structure of double-signed inner-product spaces is described by the following result.

**Inner-Product Signature Theorem:** Let $\mathcal{V}$ be an inner-product space. If a positive-regular [negative-regular] subspace $\mathcal{U}$ of $\mathcal{V}$ satisfies one of the following conditions, then it satisfies all three:

(a) $\dim \mathcal{U} = \text{sig}^+\mathcal{V}$ [$\dim \mathcal{U} = \text{sig}^-\mathcal{V}$],

(b) $\mathcal{U}$ is maximal among the positive-regular [negative-regular] subspaces of $\mathcal{V}$,

(c) $\mathcal{U}^\perp$ is negative-regular [positive-regular].

**Proof:** It is sufficient to consider the case when $\mathcal{U}$ is positive-regular, because a change from sq to $-\text{sq}$ will then take care of the other case. We choose a positive-regular subspace $\mathcal{W}$ such that $\dim \mathcal{W} = \text{sig}^+\mathcal{V}$; then $\dim \mathcal{U} \leq \dim \mathcal{W}$.

(a) $\Rightarrow$ (b): This is trivial.

(b) $\Rightarrow$ (c): Suppose that $\mathcal{U}^\perp$ is not negative-regular. Since $\mathcal{U}^\perp$ is regular by the last assertion of the Theorem on the Characterization of Regular Subspaces of Sect. 41, $\text{sq}|_{\mathcal{U}^\perp}$ cannot be negative by Prop. 4 of Sect. 27. Hence we may choose $w \in \mathcal{U}^\perp$ such that $w^2 > 0$. Then $\mathcal{U} + \mathbb{R}w$ strictly includes $\mathcal{U}$ and is still positive-regular. Hence $\mathcal{U}$ is not maximal.

(c) $\Rightarrow$ (a): Since $\mathcal{U}$ is regular, $\mathcal{U}^\perp$ is a supplement of $\mathcal{U}$ by the Theorem on the Characterization of Regular Subspaces. Hence, by Prop. 4 of Sect. 19, there is a projection $\mathbf{P} : \mathcal{V} \to \mathcal{U}$ such that $\mathcal{U}^\perp = \text{Null} \mathbf{P}$. Let $w \in \mathcal{W} \cap \mathcal{U}^\perp$ be given. Since $\mathcal{W}$ is positive-regular and $\mathcal{U}^\perp$ is negative regular, we would
have \( w^2 > 0 \) and \( w^2 < 0 \) if \( w \) were not zero. Since this cannot be, it follows that

\[
\text{Null } P|_W = W \cap U^\perp = \{0\}
\]

and hence that \( P|_W \) is injective. By the Pigeonhole Principle for Linear Mappings, it follows that \( \dim W \leq \dim U = \dim W = \text{sig}^+ V \).

Using the equivalence (a) \( \iff \) (c) twice, once for \( U \) and once when \( U \) is replaced by \( U^\perp \), we obtain the following:

**Corollary:** There exist positive-regular subspaces \( U \) such that \( U^\perp \) is negative-regular. If \( U \) is such a subspace, then

\[
\dim U = \text{sig}^+ V, \quad \dim U^\perp = \text{sig}^- V. \tag{47.1}
\]

We have

\[
\text{sig}^+ V + \text{sig}^- V = \dim V. \tag{47.2}
\]

Of course, spaces that are orthogonally isomorphic have the same signature. The converse is also true:

**Proposition 1:** Two inner-product spaces are orthogonally isomorphic if (and only if) they have the same signature.

**Proof:** Assume that \( V \) and \( V' \) are inner-product spaces with the same signature. In view of the Corollary above, we may choose positive-regular subspaces \( U \) and \( U' \) of \( V \) and \( V' \), respectively, such that \( U^\perp \) and \( U'^\perp \) are negative regular. Since \( V \) and \( V' \) have the same signature, we have, by (47.1),

\[
\dim U = \dim U', \quad \dim U^\perp = \dim U'^\perp.
\]

Therefore, we may apply Prop.5 of Sect.43 and choose orthogonal isomorphisms \( R_1 : U \to U' \) and \( R_2 : U^\perp \to U'^\perp \). Since \( U \) and \( U^\perp \) are supplementary, there is a unique linear mapping \( R : V \to V' \) such that \( R|_U = R_1|_V \) and \( R|_{U^\perp} = R_2|_{U'^\perp} \) (See Prop.5 of Sect.19). It is easily seen that \( R \) is an orthogonal isomorphism.

**Proposition 2:** Every inner-product space \( V \) has orthonormal bases.

Moreover, if \( e := (e_i | i \in I) \) is an orthonormal basis of \( V \), then the signature of \( V \) is given by

\[
\text{sig}^+ V = \sharp\{i \in I | e_i^2 = 1\}, \tag{47.3}
\]

\[
\text{sig}^- V = \sharp\{i \in I | e_i^2 = -1\}. \tag{47.4}
\]

**Proof:** In view of the Corollary above, we may choose a positive-regular subspace \( U \) of \( V \) such that \( U^\perp \) is negative regular. By Prop.1 of Sect.42, we
may choose a genuinely orthonormal set basis $b$ of $U$ and an orthonormal set basis $c$ of $U^\perp$ such that $e^2 = -1$ for all $e \in c$. Then $b \cup c$ is an orthonormal set basis of $V$.

Let an orthonormal basis $e := (e_i \mid i \in I)$ of $V$ be given. We put

$$U_1 := \text{Lsp}\{e_i \mid i \in I, e_i^2 = 1\}, \quad U_2 := \text{Lsp}\{e_i \mid i \in I, e_i^2 = -1\}.$$ 

It is clear that $U_1$ is positive-regular and that $U_2$ is negative regular and that the right sides of (47.3) and (47.4) are the dimensions of $U_1$ and $U_2$, respectively. Hence, by Def.1, $\dim U_1 \leq \text{sig}^+ V$, $\dim U_2 \leq \text{sig}^- V$. On the other hand, since $e$ is a basis, $U_1$ and $U_2$ are supplementary, in $V$ and hence by Prop.5 of Sect.17, $\dim U_1 + \dim U_2 = \dim V$. This is compatible with (47.2) only if $\dim U_1 = \text{sig}^+ V$ and $\dim U_2 = \text{sig}^- V$.

The greatest among the dimensions of all totally singular subspaces of a given inner product space $V$ is called the \textbf{index} of $V$ and is denoted by $\text{ind} V$.

**Inner-Product Index Theorem:** The index of an inner-product space $V$ is given by

$$\text{ind} V = \min\{ \text{sig}^+ V, \text{sig}^- V\} \quad (47.5)$$

and all maximal totally singular subspaces of $V$ have dimension $\text{ind} V$.

**Proof:** By the Corollary above, we may choose a positive-regular subspace $U$ of $V$ such that $U^\perp$ is negative-regular. Let $P_1 : V \to U$ and $P_2 : V \to U^\perp$ be the projections for which $\text{Null} P_1 = U^\perp$ and $\text{Null} P_2 = U$ (see Prop.4 of Sect.19).

Let a totally singular subspace $N$ of $V$ be given. Since $U^\perp$ is negative-regular, every non-zero element of $U^\perp$ has a strictly negative inner square and hence cannot belong to $N$. It follows that

$$\text{Null} P_1|_N \cap U^\perp = \{0\}.$$ 

and hence that $P_1|_N$ is injective. By the Pigeonhole Principle for Linear Mappings, it follows that

$$\dim N = \dim(\text{Rng} P_1|_N) \leq \dim U = \text{sig}^+ V. \quad (47.6)$$

A similar argument shows that

$$\dim N = \dim(\text{Rng} P_2|_N) \leq \dim U^\perp = \text{sig}^- V. \quad (47.7)$$

If the inequalities in (47.6) and (47.7) are both strict, then the orthogonal supplements of $\text{Rng} P_1|_N$ and $\text{Rng} P_2|_N$ relative to $U$ and $U^\perp$, respectively,
are both non-zero. Hence we may choose \( u \in U \) and \( w \in U^\perp \) such that \( u^2 = 1 \), \( w^2 = -1 \), and \( u \cdot P_1 n = w \cdot P_2 n = 0 \) for all \( n \in N \). It is easily verified that \( N + \mathbb{R}(u + w) \) is a totally singular subspace of \( V \) that includes \( N \) as a proper subspace. We conclude that \( N \) is maximal among the totally singular subspaces of \( V \) if and only if at least one of the inequalities (47.6) and (47.7) reduces to an equality, i.e., if and only if \( \dim N = \min \{ \text{sig}^+ V, \text{sig}^- V \} \).

**Remark:** The mathematical model for space-time in the Theory of Special Relativity has the structure of a (non-genuine) Euclidean space \( E \) whose translation space \( V \) has index 1. It is customary, by convention, to assume

\[ 1 = \text{ind} V = \text{sig}^- V \leq \text{sig}^+ V. \]

(However, many physicists now use the only other possible convention: \( 1 = \text{ind} V = \text{sig}^+ V \leq \text{sig}^- V \)). The elements of \( V \) are called *world-vectors*. A non-zero world-vector \( v \) is said to be *time-like*, *space-like*, or *signal-like*, depending on whether \( v^2 < 0 \), \( v^2 > 0 \), or \( v^2 = 0 \), respectively.

Notes 47

(1) The terms “isotropic” and “totally isotropic” are sometimes used for our “singular” and “totally singular”.

(2) Some people use the term “signature” to mean an appropriate list of + signs and – signs rather than just a pair of numbers in \( \mathbb{N} \).

(3) The Inner-Product Signature Theorem, or the closely related Prop.2, is often referred to as “the Law of Inertia” or “Sylvester’s Law of Inertia”.

48 Problems for Chapter 4

(1) Let \( V \) be an inner-product space.

(a) Prove: If \( e := (e_i \mid i \in I) \) is an orthonormal family in \( V \) and if \( v \in V \), then

\[ w := v - \sum_{i \in I} \text{sgn}(e_i^2)(e_i \cdot v)e_i \quad (P4.1) \]

satisfies \( w \cdot e_j = 0 \) for all \( j \in I \). (For the definition of sgn see (08.13)).

(b) Assume that \( V \) is genuine. Prove: If \( b := (b_i \mid i \in n^l) \) is a list-basis of \( V \) then there is exactly one orthonormal list basis \( e := (e_i \mid i \in n^l) \) of \( V \) such that \( b_k \cdot e_k > 0 \) and
48. PROBLEMS FOR CHAPTER 4

\[ \text{Lsp}\{e_i \mid i \in k\} = \text{Lsp}\{b_i \mid i \in k\} \quad \text{(P4.2)} \]

for all \( k \in \mathbb{N} \). (Hint: Use induction and Part (a).)

**Note:** The procedure for obtaining the orthonormal basis \( e \) of (b) is often called "Gram-Schmidt orthogonalization". I consider it of far less importance than some mathematicians do.

(2) Let \( \mathcal{V} \) be a genuine inner-product space of dimension 2.

(a) Show that the set \( \text{Skew} \cap \text{Orth} \) has exactly two members; moreover, if \( J \) is one of them then \( -J \) is the other, and we have \( J^2 = -1\mathcal{V} \).

(b) Show that \( L - L^\top = -\text{tr}(LJ)J \) for all \( L \in \text{Lin} \mathcal{V} \).

(3) Let \( \mathcal{V} \) be a genuine inner-product space.

(a) Prove: If \( A \) is a skew lineon on \( \mathcal{V} \), then \( 1\mathcal{V} - A \) is invertible, so that we may define \( \Phi : \text{Skew} \mathcal{V} \rightarrow \text{Lin} \mathcal{V} \) by

\[ \Phi(A) := (1\mathcal{V} + A)(1\mathcal{V} - A)^{-1} \quad \text{for all} \quad A \in \text{Skew} \mathcal{V} \quad \text{(P4.3)} \]

(b) Show that \( \Phi(A) \) is orthogonal when \( A \in \text{Skew} \mathcal{V} \).

(c) Show that \( \Phi \) is injective and that

\[ \text{Rng} \Phi = \{ R \in \text{Orth} \mathcal{V} \mid R + 1\mathcal{V} \text{ is invertible} \} \quad \text{(P4.4)} \]

(4) Let \( \mathcal{V} \) be a genuine inner-product space.

(a) Show that, for each \( a \in \mathcal{V}^\times \), there is exactly one orthogonal lineon \( R \) on \( \mathcal{V} \) such that \( Ra = -a \) and \( Ru = u \) for all \( u \in \{ a \}^\perp \). Show that, for all \( v \in \mathcal{V} \),

\[ Rv = -v \quad \iff \quad v \in \mathbb{R}a. \quad \text{(P4.5)} \]

(b) Let \( L \in \text{Lin} \mathcal{V} \) be given. Show that \( L \) commutes with every orthogonal lineon on \( \mathcal{V} \) if and only if \( L \in \mathbb{R}1\mathcal{V} \). (Hint: use Part (a).)
(5) Let $\mathcal{E}$ be a genuine Euclidean space. Prove that the sum of the squares of the lengths of the sides of a parallelogram (see Problem 5 of Chapt. 3) in $\mathcal{E}$ equals the sum of the squares of the lengths of its diagonals. (Hint: Reduce the problem to an identity involving magnitudes of vectors in $\mathcal{V} := \mathcal{E} - \mathcal{E}$.)

**Note:** This result is often called the “parallelogram law”.

(6) Let $\mathcal{E}$ be a genuine Euclidean space with $\dim \mathcal{E} > 0$ and let $p$ be a flat basis of $\mathcal{E}$. Show that there is exactly one sphere that includes $\text{Rng } p$, i.e. there is exactly one $q \in \mathcal{E}$ and one $\rho \in \mathbb{P}^\times$ such that $\text{Rng } p \subset Sph_{q,\rho} \mathcal{E}$ (see (46.8)). (Hint: Use induction over $\dim \mathcal{E}$.)

(7) Let $p := (p_i \mid i \in I)$ and $p' := (p'_i \mid i \in I)$ be families in a genuine Euclidean space $\mathcal{E}$ and assume that there is a $k \in I$ such that

$$(p_i - p_k) \cdot (p_j - p_k) = (p'_i - p'_k) \cdot (p'_j - p'_k) \quad \text{for all } i, j \in I. \quad (P4.6)$$

Show that $p$ and $p'$ are congruent.

(8) Let $\mathcal{V}$ and $\mathcal{V}'$ be inner-product spaces and consider $\text{Lin}(\mathcal{V}, \mathcal{V}')$ with its natural inner-product space structure characterized by (44.4) and (44.5). Let $(p, n)$ and $(p', n')$ be the signatures of $\mathcal{V}$ and $\mathcal{V}'$, respectively (see Sect.47). Show that the signature of $\text{Lin}(\mathcal{V}, \mathcal{V}')$ is given by $(pp' + nn', pn' + np')$. (Hint: Use Prop.4 of Sect.25 and Prop.2 of Sect.47.)

(9) Let $\mathcal{V}$ be an inner-product space with $\text{sig }^+ \mathcal{V} = 1$ and $\dim \mathcal{V} > 1$. Put

$$\mathcal{V}_+ := \{v \in \mathcal{V} \mid v^2 > 0\}. \quad (P4.7)$$

(a) Prove: For all $u, v \in \mathcal{V}_+$ we have

$$(u \cdot v)^2 \geq u^2 v^2; \quad (P4.8)$$

equality holds if and only if $(u, v)$ is linearly dependent. (Hint: Use a trick similar to the one used in the proof of the Inner-Product Inequality, Sect.42, and use the Inner-Product Signature Theorem.)

(b) Prove: For all $u, v, w \in \mathcal{V}_+$ we have

$$(u \cdot v)(v \cdot w)(w \cdot u) > 0. \quad (P4.9)$$
(Hint: Consider $z := (v \cdot w)u - (u \cdot w)v$ and use the Inner-Product Signature Theorem).

(c) Define the relation $\sim$ on $V_+$ by

$$u \sim v \iff u \cdot v > 0.$$  \hfill (P4.10)

Show that $\sim$ is an equivalence relation on $V_+$ and that the corresponding partition of $V_+$ (see Sect.01) has exactly two members; moreover, if $C$ is one of them then $-C$ is the other.

**Note:** In the application to the theory of relativity, one of the two equivalence classes of Part (b), call it $C$, is singled out. The elements of $C$ are then called "future-directed" world-vectors while the members of $-C$ are called "past-directed" world vectors. ■
Chapter 6

Differential Calculus

In this chapter, it is assumed that all linear spaces and flat spaces under consideration are finite-dimensional.

61 Differentiation of Processes

Let $E$ be a flat space with translation space $V$. A mapping $p : I \to E$ from some interval $I \in \text{Sub} \mathbb{R}$ to $E$ will be called a process. It is useful to think of the value $p(t) \in E$ as describing the state of some physical system at time $t$. In special cases, the mapping $p$ describes the motion of a particle and $p(t)$ is the place of the particle at time $t$. The concept of differentiability for real-valued functions (see Sect.08) extends without difficulty to processes as follows:

Definition 1: The process $p : I \to E$ is said to be differentiable at $t \in I$ if the limit
$$\partial p := \lim_{s \to 0} \frac{1}{s} (p(t + s) - p(t))$$
exists. Its value $\partial p \in V$ is then called the derivative of $p$ at $t$. We say that $p$ is differentiable if it is differentiable at all $t \in I$. In that case, the mapping $\partial p : I \to V$ defined by $(\partial p)(t) := \partial p$ for all $t \in I$ is called the derivative of $p$.

Given $n \in \mathbb{N}^\times$, we say that $p$ is $n$ times differentiable if $\partial^n p : I \to V$ can be defined by the recursion
$$\partial^1 p := \partial p, \quad \partial^{k+1} p := \partial(\partial^k p) \quad \text{for all } k \in (n - 1)^\mathbb{N}.$$  

We say that $p$ is of class $C^n$ if it is $n$ times differentiable and $\partial^n p$ is continuous. We say that $p$ is of class $C^\infty$ if it is of class $C^n$ for all $n \in \mathbb{N}^\times$.  

209
As for a real-valued function, it is easily seen that a process \( p \) is continuous at \( t \in \text{Dom} \ p \) if it is differentiable at \( t \). Hence \( p \) is continuous if it is differentiable, but it may also be continuous without being differentiable.

In analogy to (08.34) and (08.35), we also use the notation
\[
p^{(k)} := \partial^k p \quad \text{for all} \quad k \in (n - 1)
\]
when \( p \) is an \( n \)-times differentiable process, and we use
\[
p \cdot := p^{(1)} = \partial p, \quad p \cdot \cdot := p^{(2)} = \partial^2 p, \quad p \cdot \cdot \cdot := p^{(3)} = \partial^3 p,
\]
if meaningful.

We use the term “process” also for a mapping \( p : I \to \mathcal{D} \) from some interval \( I \) into a subset \( \mathcal{D} \) of the flat space \( \mathcal{E} \). In that case, we use poetic license and ascribe to \( p \) any of the properties defined above if \( p|\mathcal{E} \) has that property. Also, we write \( \partial p \) instead of \( \partial(p|\mathcal{E}) \) if \( p \) is differentiable, etc. (If \( \mathcal{D} \) is included in some flat \( \mathcal{F} \), then one can take the direction space of \( \mathcal{F} \) rather than all of \( \mathcal{V} \) as the codomain of \( \partial p \). This ambiguity will usually not cause any difficulty.)

The following facts are immediate consequences of Def.1, and Prop.5 of Sect.56 and Prop.6 of Sect.57.

**Proposition 1:** The process \( p : I \to \mathcal{E} \) is differentiable at \( t \in I \) if and only if, for each \( \lambda \) in some basis of \( \mathcal{V}^* \), the function \( \lambda(p - p(t)) : I \to \mathbb{R} \) is differentiable at \( t \).

The process \( p \) is differentiable if and only if, for every flat function \( a \in \text{Flf} \mathcal{E} \), the function \( a \circ p \) is differentiable.

**Proposition 2:** Let \( \mathcal{E}, \mathcal{E}' \) be flat spaces and \( \alpha : \mathcal{E} \to \mathcal{E}' \) a flat mapping. If \( p : I \to \mathcal{E} \) is a process that is differentiable at \( t \in I \), then \( \alpha \circ p : I \to \mathcal{E}' \) is also differentiable at \( t \in I \) and
\[
\partial t(\alpha \circ p) = (\nabla \alpha)(\partial t p).
\]
If \( p \) is differentiable then (61.5) holds for all \( t \in I \) and we get
\[
\partial(\alpha \circ p) = (\nabla \alpha)\partial p.
\]

Let \( p : I \to \mathcal{E} \) and \( q : I \to \mathcal{E}' \) be processes having the same domain \( I \). Then \( (p, q) : I \to \mathcal{E} \times \mathcal{E}' \), defined by value-wise pair formation, (see (04.13)) is another process. It is easily seen that \( p \) and \( q \) are both differentiable at \( t \in I \) if and only if \( (p, q) \) is differentiable at \( t \). If this is the case we have
\[
\partial t(p, q) = (\partial t p, \partial t q).
\]
Both \( p \) and \( q \) are differentiable if and only if \((p, q)\) is, and, in that case,
\[
\partial(p, q) = (\partial p, \partial q). \tag{61.8}
\]

Let \( p \) and \( q \) be processes having the same domain \( I \) and the same codomain \( E \). Since the point-difference \((x, y) \mapsto x - y\) is a flat mapping from \( E \times E \) into \( V \) whose gradient is the vector-difference \((u, v) \mapsto u - v\) from \( V \times V \) into \( V \) we can apply Prop.2 to obtain

**Proposition 3:** If \( p : I \to E \) and \( q : I \to E \) are both differentiable at \( t \in I \), so is the value-wise difference \( p - q : I \to V \), and
\[
\partial_t(p - q) = (\partial_t p) - (\partial_t q). \tag{61.9}
\]

If \( p \) and \( q \) are both differentiable, then (61.9) holds for all \( t \in I \) and we get
\[
\partial(p - q) = \partial p - \partial q. \tag{61.10}
\]

The following result generalizes the Difference-Quotient Theorem stated in Sect.08.

**Difference-Quotient Theorem:** Let \( p : I \to E \) be a process and let \( t_1, t_2 \in I \) with \( t_1 < t_2 \). If \( p|_{t_1, t_2} \) is continuous and if \( p \) is differentiable at each \( t \in ]t_1, t_2[ \) then
\[
\frac{p(t_2) - p(t_1)}{t_2 - t_1} \in \overline{\text{Clo Cxh}}\{\partial_t p \mid t \in ]t_1, t_2[\}. \tag{61.11}
\]

**Proof:** Let \( a \in \text{Flf } E \) be given. Then \((a \circ p)|_{t_1, t_2}\) is continuous and, by Prop.1, \( a \circ p \) is differentiable at each \( t \in ]t_1, t_2[ \). By the elementary Difference-Quotient Theorem (see Sect.08) we have
\[
\frac{(a \circ p)(t_2) - (a \circ p)(t_2)}{t_2 - t_1} \in \{\partial_t (a \circ p) \mid t \in ]t_1, t_2[\}.
\]

Using (61.5) and (33.4), we obtain
\[
\nabla a \left( \frac{p(t_2) - p(t_1)}{t_2 - t_1} \right) \in (\nabla a)_{>(S)}, \tag{61.12}
\]
where
\[
S := \{\partial_t p \mid t \in ]t_1, t_2[\}.
\]

Since (61.12) holds for all \( a \in \text{Flf } E \) we can conclude that
\[
\hat{b}(\frac{p(t_2) - p(t_1)}{t_2 - t_1}) \geq 0
\]
holds for all those \( b \in \text{Flf } V \) that satisfy \( b_{>(S)} \subset \mathbb{P} \). Using the Half-Space Intersection Theorem of Sect.54, we obtain the desired result (61.11). □

**Notes 61**

(1) See Note (8) to Sect.08 concerning notations such as \( \partial_t p \), \( \partial p \), \( \partial^n p \), \( p^n \), etc.
62 Small and Confined Mappings

Let $V$ and $V'$ be linear spaces of strictly positive dimension. Consider a mapping $n$ from a neighborhood of zero in $V$ to a neighborhood of zero in $V'$. If $n(0) = 0$ and if $n$ is continuous at $0$, then we can say, intuitively, that $n(v)$ approaches $0$ in $V'$ as $v$ approaches $0$ in $V$. We wish to make precise the idea that $n$ is small near $0 \in V$ in the sense that $n(v)$ approaches $0 \in V'$ faster than $v$ approaches $0 \in V$.

**Definition 1:** We say that a mapping $n$ from a neighborhood of $0 \in V$ to a neighborhood of $0 \in V'$ is small near $0 \in V$ if $n(0) = 0$ and, for all norms $\nu$ and $\nu'$ on $V$ and $V'$, respectively, we have

$$\lim_{u \to 0} \frac{\nu'(n(u))}{\nu(u)} = 0.$$ (62.1)

The set of all such small mappings will be denoted by $\text{Small}(V, V')$.

**Proposition 1:** Let $n$ be a mapping from a neighborhood of $0 \in V$ to a neighborhood of $0 \in V'$ is small near $0 \in V$. Then the following conditions are equivalent:

(i) $n \in \text{Small}(V, V')$.

(ii) $n(0) = 0$ and the limit-relation (62.1) holds for some norm $\nu$ on $V$ and some norm $\nu'$ on $V'$.

(iii) For every bounded subset $S$ of $V$ and every $N' \in \text{Nhd}_0(V')$ there is a $\delta \in \mathbb{P}^\times$ such that

$$n(sv) \in sN' \quad \text{for all} \quad s \in [−\delta, \delta[ \quad (62.2)$$

and all $v \in S$ such that $sv \in \text{Dom} n$.

**Proof:** (i) $\Rightarrow$ (ii): This implication is trivial.

(ii) $\Rightarrow$ (iii): Assume that (ii) is valid. Let $N' \in \text{Nhd}_0(V')$ and a bounded subset $S$ of $V$ be given. By Cor.1 to the Cell-Inclusion Theorem of Sect.52, we can choose $b \in \mathbb{P}^\times$ such that

$$\nu(v) \leq b \quad \text{for all} \quad v \in S.$$ (62.3)

By Prop.3 of Sect.53 we can choose $\varepsilon \in \mathbb{P}^\times$ such that

$$\varepsilon b \text{Ce}(\nu') \subset N'.$$ (62.4)

Applying Prop.4 of Sect.57 to the assumption (ii) we obtain $\delta \in \mathbb{P}^\times$ such that, for all $u \in \text{Dom} n$,

$$\nu'(n(u)) < \varepsilon \nu(u) \quad \text{if} \quad 0 < \nu(u) < \delta b.$$ (62.5)
Now let \( v \in S \) be given. Then \( \nu(sv) = |s|\nu(v) \leq |s|\delta b \) for all \( s \in [-\delta, \delta] \) such that \( sv \in \text{Dom } n \). Therefore, by (62.5), we have

\[
\nu'(n(sv)) < \varepsilon \nu(sv) = \varepsilon |s|\nu(v) \leq |s|\varepsilon b
\]

if \( sv \neq 0 \), and hence

\[
n(sv) \in s \varepsilon b \text{Ce}(\nu')
\]

for all \( s \in [-\delta, \delta] \) such that \( sv \in \text{Dom } n \). The desired conclusion (62.2) now follows from (62.4).

**(iii) ⇒ (i):** Assume that (iii) is valid. Let a norm \( \nu \) on \( V \), a norm \( \nu' \) on \( V' \), and \( \varepsilon \in \mathbb{P}^\times \) be given. We apply (iii) to the choices \( S := \text{Ce}(\nu) \), \( N' := \varepsilon \text{Ce}(\nu') \) and determine \( \delta \in \mathbb{P}^\times \) such that (62.2) holds. If we put \( s := 0 \) in (62.2) we obtain \( n(0) = 0 \). Now let \( u \in \text{Dom } n \) be given such that \( 0 < \nu(u) < \delta \). If we apply (62.2) with the choices \( s := \nu(u) \) and \( v := \frac{1}{s}u \), we see that \( n(u) \in \nu(u)\varepsilon \text{Ce}(\nu') \), which yields

\[
\frac{\nu'(n(u))}{\nu(u)} < \varepsilon.
\]

The assertion follows by applying Prop. 4 of Sect. 57.

The condition (iii) of Prop. 1 states that

\[
\lim_{s \to 0} \frac{1}{s} n(sv) = 0 \quad (62.6)
\]

for all \( v \in V \) and, roughly, that the limit is approached uniformly as \( v \) varies in an arbitrary bounded set.

We also wish to make precise the intuitive idea that a mapping \( h \) from a neighborhood of \( 0 \) in \( V \) to a neighborhood of \( 0 \) in \( V' \) is confined near zero in the sense that \( h(v) \) approaches \( 0 \in V' \) not more slowly than \( v \) approaches \( 0 \in V \).

**Definition 2:** A mapping \( h \) from a neighborhood of \( 0 \) in \( V \) to a neighborhood of \( 0 \) in \( V' \) is said to be **confined near 0** if for every norm \( \nu \) on \( V \) and every norm \( \nu' \) on \( V' \) there is \( N \in \text{Nhdo}(V) \) and \( \kappa \in \mathbb{P}^\times \) such that

\[
\nu'(h(u)) \leq \kappa \nu(u) \quad \text{for all } u \in N \cap \text{Dom } h. \quad (62.7)
\]

The set of all such confined mappings will be denoted by \( \text{Conf}(V, V') \).

**Proposition 2:** Let \( h \) be a mapping from a neighborhood of \( 0 \) in \( V \) to a neighborhood of \( 0 \) in \( V' \). Then the following are equivalent:

(i) \( h \in \text{Conf}(V, V') \).
CHAPTER 6. DIFFERENTIAL CALCULUS

(ii) There exists a norm \( \nu \) on \( V \), a norm \( \nu' \) on \( V' \), a neighborhood \( N \) of \( 0 \) in \( V \), and \( \kappa \in \mathbb{P}^\times \) such that (62.7) holds.

(iii) For every bounded subset \( S \) of \( V \) there is \( \delta \in \mathbb{P}^\times \) and a bounded subset \( S' \) of \( V' \) such that \[
\mathbf{h}(sv) \in sS' \quad \text{for all} \quad s \in [-\delta, \delta]\]
and all \( v \in S \) such that \( sv \in \text{Dom} \, \mathbf{h} \).

Proof: (i) \( \Rightarrow \) (ii): This is trivial.

(ii) \( \Rightarrow \) (iii): Assume that (ii) holds. Let a bounded subset \( S \) of \( V \) be given. By Cor.1 to the Cell-Inclusion Theorem of Sect.52, we can choose \( b \in \mathbb{P}^\times \) such that \( \nu(v) \leq b \) for all \( v \in S \).

By Prop.3 of Sect.53, we can determine \( \delta \in \mathbb{P}^\times \) such that \( \delta b \subseteq \nu \) \( \subset N \).

Hence, by (62.7), we have for all \( u \in \text{Dom} \, \mathbf{h} \) \[
\nu'(\mathbf{h}(u)) \leq \kappa \nu(u) \quad \text{if} \quad \nu(u) < \delta b. \quad (62.9)
\]

Now let \( v \in S \) be given. Then \( \nu(sv) = |s|\nu(v) \leq |s|b < \delta b \) for all \( s \in [-\delta, \delta] \) such that \( sv \in \text{Dom} \, \mathbf{h} \). Therefore, by (62.9) we have \[
\nu'(\mathbf{h}(sv)) \leq \kappa \nu(sv) = \kappa |s|\nu(v) < |s|\kappa b
\]

and hence \[
\mathbf{h}(sv) \in s\kappa b \nu'(v)
\]
for all \( s \in [-\delta, \delta] \) such that \( sv \in \text{Dom} \, \mathbf{h} \). If we put \( S' := \kappa b \nu'(v) \), we obtain the desired conclusion (62.8).

(iii) \( \Rightarrow \) (i): Assume that (iii) is valid. Let a norm \( \nu \) on \( V \) and a norm \( \nu' \) on \( V' \) be given. We apply (iii) to the choice \( S := \nu(v) \) and determine \( S' \) and \( \delta \in \mathbb{P}^\times \) such that (62.8) holds. Since \( S' \) is bounded, we can apply the Cell-Inclusion Theorem of Sect.52 and determine \( \kappa \in \mathbb{P}^\times \) such that \( S' \subseteq \kappa \nu'(v) \).

We put \( \mathcal{N} := \delta \nu'(v) \cap \text{Dom} \, \mathbf{h} \), which belongs to \( \text{Nh}(V) \). If we put \( s := 0 \) in (62.8) we obtain \( \mathbf{h}(0) = 0 \), which shows that (62.7) holds for \( u := 0 \).

Now let \( u \in \mathcal{N}^\times \) be given, so that \( 0 < \nu(u) < \delta \). If we apply (62.8) with the choices \( s := \nu(u) \) and \( v := \frac{1}{s} u \), we see that \[
\mathbf{h}(u) \in \nu(u)S' \subseteq \nu(u)\kappa \nu'(v),
\]
which yields the assertion (62.7). ■

The following results are immediate consequences of the definition and of the properties on linear mappings discussed in Sect.52:
62. SMALL AND CONFINED MAPPINGS

(I) Value-wise sums and value-wise scalar multiples of mappings that are small [confined] near zero are again small [confined] near zero.

(II) Every mapping that is small near zero is also confined near zero, i.e.

$$\text{Small}(\mathcal{V}, \mathcal{V}') \subset \text{Conf}(\mathcal{V}, \mathcal{V}')$$

(III) If $h \in \text{Conf}(\mathcal{V}, \mathcal{V}')$, then $h(0) = 0$ and $h$ is continuous at $0$.

(IV) Every linear mapping is confined near zero, i.e.

$$\text{Lin}(\mathcal{V}, \mathcal{V}') \subset \text{Conf}(\mathcal{V}, \mathcal{V}')$$

(V) The only linear mapping that is small near zero is the zero-mapping, i.e.

$$\text{Lin}(\mathcal{V}, \mathcal{V}') \cap \text{Small}(\mathcal{V}, \mathcal{V}') = \{0\}$$

Proposition 3: Let $\mathcal{V}, \mathcal{V}', \mathcal{V}''$ be linear spaces and let $h \in \text{Conf}(\mathcal{V}, \mathcal{V}')$ and $k \in \text{Conf}(\mathcal{V}', \mathcal{V}'')$ be such that $\text{Dom} k = \text{Cod} h$. Then $k \circ h \in \text{Conf}(\mathcal{V}, \mathcal{V}'')$. Moreover, if one of $k$ or $h$ is small near zero so is $k \circ h$.

Proof: Let norms $\nu, \nu', \nu''$ on $\mathcal{V}, \mathcal{V}', \mathcal{V}''$, respectively, be given. Since $h$ and $k$ are confined we can find $\kappa, \kappa' \in \mathbb{P} \times \mathbb{N}$ and $N \in \text{Nhd}_0(\mathcal{V}), N' \in \text{Nhd}_0(\mathcal{V}')$ such that

$$\nu''((k \circ h)(u)) \leq \kappa' \nu'(h(u)) \leq \kappa \kappa' \nu(u) \quad (62.10)$$

for all $u \in N \cap \text{Dom} h$ such that $h(u) \in N' \cap \text{Dom} k$, i.e. for all $u \in N \cap h^<(N' \cap \text{Dom} k)$. Since $h$ is continuous at $0 \in \mathcal{V}$, we have $h^< (N' \cap \text{Dom} k) \in \text{Nhd}_0(\mathcal{V})$ and hence $N \cap h^<(N' \cap \text{Dom} k) \in \text{Nhd}_0(\mathcal{V})$. Thus, (62.7) remains satisfied when we replace $h, \kappa$ and $N$ by $k \circ h, \kappa' \kappa$, and $N \cap h^<(N' \cap \text{Dom} k)$, respectively, which shows that $k \circ h \in \text{Conf}(\mathcal{V}, \mathcal{V}'')$.

Assume now, that one of $k$ and $h$, say $h$, is small. Let $\varepsilon \in \mathbb{P} \times $ be given. Then we can choose $N \in \text{Nhd}_0(\mathcal{V})$ such that $\nu'(h(u)) \leq \kappa \nu(u)$ holds for all $u \in N \cap \text{Dom} h$ with $\kappa := \frac{\varepsilon}{\nu}$. Therefore, (62.10) gives $\nu''((k \circ h)(u)) \leq \varepsilon \nu(u)$ for all $u \in N \cap h^<(N' \cap \text{Dom} k)$. Since $\varepsilon \in \mathbb{P} \times$ was arbitrary this proves that

$$\lim_{u \to 0} \frac{\nu''((k \circ h)(u))}{\nu(u)} = 0,$$

i.e. that $k \circ h$ is small near zero. ■

Now let $\mathcal{E}$ and $\mathcal{E}'$ be flat spaces with translation spaces $\mathcal{V}$ and $\mathcal{V}'$, respectively.
Definition 3: Let $x \in \mathcal{E}$ be given. We say that a mapping $\sigma$ from a neighborhood of $x \in \mathcal{E}$ to a neighborhood of $0 \in \mathcal{V}'$ is small near $x$ if the mapping $v \mapsto \sigma(x + v)$ from $(\text{Dom } \sigma) - x$ to $\text{Cod } \sigma$ is small near $0$. The set of all such small mappings will be denoted by $\text{Small}_x(\mathcal{E}, \mathcal{V}')$.

We say that a mapping $\varphi$ from a neighborhood of $x \in \mathcal{E}$ to a neighborhood of $\varphi(x) \in \mathcal{E}'$ is confined near $x$ if the mapping $v \mapsto \varphi(x + v) - \varphi(x)$ from $(\text{Dom } \varphi) - x$ to $(\text{Cod } \varphi) - \varphi(x)$ is confined near zero.

The following characterization is immediate.

Proposition 4: The mapping $\varphi$ is confined near $x \in \mathcal{E}$ if and only if for every norm $\nu$ on $\mathcal{V}$ and every norm $\nu'$ on $\mathcal{V}'$ there is $N \in \text{Nhd}_x(\mathcal{E})$ and $\kappa \in \mathbb{P}^\times$ such that

$$
\nu'(\varphi(y) - \varphi(x)) \leq \kappa \nu(y - x) \quad \text{for all } y \in N. 
$$

We now state a few facts that are direct consequences of the definitions, the results (I)–(V) stated above, and Prop. 3:

(VI) Value-wise sums and differences of mappings that are small [confined] near $x$ are again small [confined] near $x$. Here, “sum” can mean either the sum of two vectors or sum of a point and a vector, while “difference” can mean either the difference of two vectors or the difference of two points.

(VII) Every $\sigma \in \text{Small}_x(\mathcal{E}, \mathcal{V}')$ is confined near $x$.

(VIII) If a mapping is confined near $x$ it is continuous at $x$.

(IX) A flat mapping $\alpha : \mathcal{E} \to \mathcal{E}'$ is confined near every $x \in \mathcal{E}$.

(X) The only flat mapping $\beta : \mathcal{E} \to \mathcal{V}'$ that is small near some $x \in \mathcal{E}$ is the constant $0_{\mathcal{E} \to \mathcal{V}'}$.

(XI) If $\varphi$ is confined near $x \in \mathcal{E}$ and if $\psi$ is a mapping with $\text{Dom } \psi = \text{Cod } \varphi$ that is confined near $\varphi(x)$ then $\psi \circ \varphi$ is confined near $x$.

(XII) If $\sigma \in \text{Small}_x(\mathcal{E}, \mathcal{V}')$ and $h \in \text{Conf}(\mathcal{V}', \mathcal{V}'')$ with $\text{Cod } \sigma = \text{Dom } h$, then $h \circ \sigma \in \text{Small}_x(\mathcal{E}, \mathcal{V}'')$.

(XIII) If $\varphi$ is confined near $x \in \mathcal{E}$ and if $\sigma$ is a mapping with $\text{Dom } \sigma = \text{Cod } \varphi$ that is small near $\varphi(x)$ then $\sigma \circ \varphi \in \text{Small}_x(\mathcal{E}, \mathcal{V}'')$, where $\mathcal{V}''$ is the linear space for which $\text{Cod } \sigma \in \text{Nhd}_0(\mathcal{V}'')$.

(XIV) An adjustment of a mapping that is small [confined] near $x$ is again small [confined] near $x$, provided only that the concept small [confined] near $x$ remains meaningful after the adjustment.
63. GRADIENTS, CHAIN RULE

Notes 62

(1) In the conventional treatments, the norms $\nu$ and $\nu'$ in defs. 1 and 2 are assumed to be prescribed and fixed. The notation $n = o(\nu)$, and the phrase "$n$ is small oh of $\nu$", are often used to express the assertion that $n \in \text{Small}(\mathcal{V}, \mathcal{V}')$. The notation $h = O(\nu)$, and the phrase "$h$ is big oh of $\nu$", are often used to express the assertion that $h \in \text{Conf}(\mathcal{V}, \mathcal{V}')$. I am introducing the terms "small" and "confined" here for the first time because I believe that the conventional terminology is intolerably awkward and involves a misuse of the $=$ sign.

63 Gradients, Chain Rule

Let $I$ be an open interval in $\mathbb{R}$. One learns in elementary calculus that if a function $f : I \to \mathbb{R}$ is differentiable at a point $t \in I$, then the graph of $f$ has a tangent at $(t, f(t))$. This tangent is the graph of a flat function $a \in \text{Flf}(\mathbb{R})$. Using poetic license, we refer to this function itself as the tangent to $f$ at $t \in I$. In this sense, the tangent $a$ is given by $a(r) := f(t) + (\partial_t f)(r - t)$ for all $r \in \mathbb{R}$.

If we put $\sigma := f - a|_I$, then $\sigma(r) = f(r) - f(t) - (\partial_t f)(r - t)$ for all $r \in I$. We have $\lim_{s \to 0} \frac{\sigma(t + s)}{s} = 0$, from which it follows that $\sigma \in \text{Small}_t(\mathbb{R}, \mathbb{R})$. One can use the existence of a tangent to define differentiability at $t$. Such a definition generalizes directly to mappings involving flat spaces.

Let $\mathcal{E}, \mathcal{E}'$ be flat spaces with translation spaces $\mathcal{V}, \mathcal{V}'$, respectively. We consider a mapping $\varphi : D \to D'$ from an open subset $D$ of $\mathcal{E}$ into an open subset $D'$ of $\mathcal{E}'$.

**Proposition 1:** Given $x \in D$, there can be at most one flat mapping $\alpha : \mathcal{E} \to \mathcal{E}'$ such that the value-wise difference $\varphi - \alpha|_D : D \to \mathcal{V}'$ is small near $x$.

**Proof:** If the flat mappings $\alpha_1, \alpha_2$ both have this property, then the value-wise difference $(\alpha_2 - \alpha_1)|_D = (\varphi - \alpha_1|_D) - (\varphi - \alpha_2|_D)$ is small near
\( x \in \mathcal{E} \). Since \( \alpha_2 - \alpha_1 \) is flat, it follows from (X) and (XIV) of Sect. 62 that \( \alpha_2 - \alpha_1 \) is the zero mapping and hence that \( \alpha_1 = \alpha_2 \). 

**Definition 1:** The mapping \( \varphi : D \to D' \) is said to be **differentiable** at \( x \in D \) if there is a flat mapping \( \alpha : \mathcal{E} \to \mathcal{E}' \) such that

\[
\varphi - \alpha|_D \in \text{Small}_x(\mathcal{E}, \mathcal{V}'). \tag{63.1}
\]

This (unique) flat mapping \( \alpha \) is then called the **tangent to** \( \varphi \) at \( x \). The **gradient** of \( \varphi \) at \( x \) is defined to be the gradient of \( \alpha \) and is denoted by

\[
\nabla_x \varphi := \nabla \alpha. \tag{63.2}
\]

We say that \( \varphi \) is **differentiable** if it is differentiable at all \( x \in D \). If this is the case, the mapping

\[
\nabla \varphi : D \to \text{Lin}(\mathcal{V}, \mathcal{V}') \tag{63.3}
\]

defined by

\[
(\nabla \varphi)(x) := \nabla_x \varphi \quad \text{for all} \quad x \in D \tag{63.4}
\]

is called the **gradient** of \( \varphi \). We say that \( \varphi \) is of **class** \( \mathbf{C}^1 \) if it is differentiable and if its gradient \( \nabla \varphi \) is continuous. We say that \( \varphi \) is **twice differentiable** if it is differentiable and if its gradient \( \nabla \varphi \) is also differentiable. The gradient of \( \nabla \varphi \) is then called the **second gradient** of \( \varphi \) and is denoted by

\[
\nabla^{(2)} \varphi := \nabla (\nabla \varphi) : D \to \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \mathcal{V}')) \cong \text{Lin}_2(\mathcal{V}^2, \mathcal{V}'). \tag{63.5}
\]

We say that \( \varphi \) is of **class** \( \mathbf{C}^2 \) if it is twice differentiable and if \( \nabla^{(2)} \varphi \) is continuous.

If the subsets \( D \) and \( D' \) are arbitrary, not necessarily open, and if \( x \in \text{Int} D \), we say that \( \varphi : D \to D' \) is differentiable at \( x \) if \( \varphi|_{\text{Int} D} \) is differentiable at \( x \) and we write \( \nabla_x \varphi \) for \( \nabla_x (\varphi|_{\text{Int} D}) \).

The differentiability properties of a mapping \( \varphi \) remain unchanged if the codomain of \( \varphi \) is changed to any open subset of \( \mathcal{E}' \) that includes \( \text{Rng} \varphi \). The gradient of \( \varphi \) remains unaltered. If \( \text{Rng} \varphi \) is included in some flat \( \mathcal{F}' \) in \( \mathcal{E}' \), one may change the codomain to a subset that is open in \( \mathcal{F}' \). in that case, the gradient of \( \varphi \) at a point \( x \in D \) must be replaced by the adjustment \( \nabla_x \varphi|_{\mathcal{U}'} \) of \( \nabla_x \varphi \), where \( \mathcal{U}' \) is the direction space of \( \mathcal{F}' \).

The differentiability and the gradient of a mapping at a point depend only on the values of the mapping near that point. To be more precise, let \( \varphi_1 \) and \( \varphi_2 \) be two mappings whose domains are neighborhoods of a given \( x \in \mathcal{E} \) and whose codomains are open subsets of \( \mathcal{E}' \). Assume that \( \varphi_1 \) and \( \varphi_2 \) agree
on some neighborhood of \( x \), i.e. that \( \varphi_1|_{N'} = \varphi_2|_{N'} \) for some \( N \in \text{Nhd}_x(\mathcal{E}) \). Then \( \varphi_1 \) is differentiable at \( x \) if and only if \( \varphi_2 \) is differentiable at \( x \). If this is the case, we have \( \nabla_x \varphi_1 = \nabla_x \varphi_2 \).

Every flat mapping \( \alpha : \mathcal{E} \to \mathcal{E}' \) is differentiable. The tangent of \( \alpha \) at every point \( x \in \mathcal{E} \) is \( \alpha \) itself. The gradient of \( \alpha \) as defined in this section is the constant \( (\nabla \alpha)_{\mathcal{E} \to \text{Lin}(\mathcal{V}, \mathcal{V}')} \) whose value is the gradient \( \nabla \alpha \) of \( \alpha \) in the sense of Sect.33. The gradient of a linear mapping is the constant whose value is this linear mapping itself.

In the case when \( \mathcal{E} := \mathcal{V} := \mathbb{R} \) and when \( \mathcal{D} := I \) is an open interval, differentiability at \( t \in I \) of a process \( p : I \to \mathcal{D}' \) in the sense of Def.1 above reduces to differentiability of \( p \) at \( t \) in the sense of Def.1 of Sect.1. The gradient \( \nabla_t p \in \text{Lin}(\mathbb{R}, \mathcal{V}') \) becomes associated with the derivative \( \partial_t p \in \mathcal{V}' \) by the natural isomorphism from \( \mathcal{V}' \) to \( \text{Lin}(\mathbb{R}, \mathcal{V}') \), so that \( \nabla_t p = (\partial_t p) \otimes \) and \( \partial_t p = (\nabla_t p)1 \) (see Sect.25). If \( \mathcal{J} \) is an interval that is not open and if \( t \) is an endpoint of it, then the derivative of \( p : I \to \mathcal{D}' \) at \( t \), if it exists, cannot be associated with a gradient.

If \( \varphi \) is differentiable at \( x \) then it is confined and hence continuous at \( x \). This follows from the fact that its tangent \( \alpha \), being flat, is confined near \( x \) and that the difference \( \varphi - \alpha|_D \), being small near \( x \), is confined near \( x \). The converse is not true. For example, it is easily seen that the absolute-value function \( (t \mapsto |t|) : \mathbb{R} \to \mathbb{R} \) is confined near \( 0 \in \mathbb{R} \) but not differentiable at 0.

The following criterion is immediate from the definition:

**Characterization of Gradients:** The mapping \( \varphi : \mathcal{D} \to \mathcal{D}' \) is differentiable at \( x \in \mathcal{D} \) if and only if there is an \( \mathcal{L} \in \text{Lin}(\mathcal{V}, \mathcal{V}') \) such that \( \mathcal{N} : (\mathcal{D} - x) \to \mathcal{V} \), defined by

\[
\mathcal{N}(\mathcal{V}) := (\varphi(x + \mathcal{V}) - \varphi(x)) - \mathcal{L} \mathcal{V} \quad \text{for all} \quad \mathcal{V} \in \mathcal{D} - x,
\]

is small near \( 0 \) in \( \mathcal{V} \). If this is the case, then \( \nabla_x \varphi = \mathcal{L} \).

Let \( \mathcal{D}, \mathcal{D}_1, \mathcal{D}_2 \) be open subsets of flat spaces \( \mathcal{E}, \mathcal{E}_1, \mathcal{E}_2 \) with translation spaces \( \mathcal{V}, \mathcal{V}_1, \mathcal{V}_2 \), respectively. The following result follows immediately from the definitions if we use the term-wise evaluations (04.13) and (14.12).

**Proposition 2:** The mapping \( (\varphi_1, \varphi_2) : \mathcal{D} \to \mathcal{D}_1 \times \mathcal{D}_2 \) is differentiable at \( x \in \mathcal{D} \) if and only if both \( \varphi_1 \) and \( \varphi_2 \) are differentiable at \( x \). If this is the case, then

\[
\nabla_x (\varphi_1, \varphi_2) = (\nabla_x \varphi_1, \nabla_x \varphi_2) \in \text{Lin}(\mathcal{V}, \mathcal{V}_1) \times \text{Lin}(\mathcal{V}, \mathcal{V}_2) \cong \text{Lin}(\mathcal{V}, \mathcal{V}_1 \times \mathcal{V}_2)
\]

**General Chain Rule:** Let \( \mathcal{D}, \mathcal{D}', \mathcal{D}'' \) be open subsets of flat spaces \( \mathcal{E}, \mathcal{E}', \mathcal{E}'' \) with translation spaces \( \mathcal{V}, \mathcal{V}', \mathcal{V}'' \), respectively. If \( \varphi : \mathcal{D} \to \mathcal{D}' \) is
differentiable at \( x \in \mathcal{D} \) and if \( \psi : \mathcal{D}' \to \mathcal{D}'' \) is differentiable at \( \varphi(x) \), then the composite \( \psi \circ \varphi : \mathcal{D} \to \mathcal{D}'' \) is differentiable at \( x \). The tangent to the composite \( \psi \circ \varphi \) at \( x \) is the composite of the tangent to \( \varphi \) at \( x \) and the tangent to \( \psi \) at \( \varphi(x) \) and we have

\[
\nabla_x (\psi \circ \varphi) = (\nabla_{\varphi(x)} \psi)(\nabla_x \varphi). \tag{63.8}
\]

If \( \varphi \) and \( \psi \) are both differentiable, so is \( \psi \circ \varphi \), and we have

\[
\nabla (\psi \circ \varphi) = (\nabla \psi \circ \varphi)(\nabla \varphi), \tag{63.9}
\]

where the product on the right is understood as value-wise composition.

**Proof:** Let \( \alpha \) be the tangent to \( \varphi \) at \( x \) and \( \beta \) the tangent to \( \psi \) at \( \varphi(x) \). Then

\[
\sigma := \varphi - \alpha|_{\mathcal{D}} \in \text{Small}_x(\mathcal{E}, \mathcal{V}'),
\]

\[
\tau := \psi - \beta|_{\mathcal{D}'} \in \text{Small}_{\varphi(x)}(\mathcal{E}', \mathcal{V}'').
\]

We have

\[
\psi \circ \varphi = (\beta + \tau) \circ (\alpha + \sigma)
\]

\[
= \beta \circ (\alpha + \sigma) + \tau \circ (\alpha + \sigma)
\]

\[
= \beta \circ \alpha + (\nabla \beta) \circ \sigma + \tau \circ (\alpha + \sigma)
\]

where domain restriction symbols have been omitted to avoid clutter. It follows from (VI), (IX), (XII), and (XIII) of Sect. 62 that \( (\nabla \beta) \circ \sigma + \tau \circ (\alpha + \sigma) \in \text{Small}_x(\mathcal{E}, \mathcal{V}'') \), which means that

\[
\psi \circ \varphi - \beta \circ \alpha \in \text{Small}_x(\mathcal{E}, \mathcal{V}'').
\]

If follows that \( \psi \circ \varphi \) is differentiable at \( x \) with tangent \( \beta \circ \alpha \). The assertion (63.8) follows from the Chain Rule for Flat Mappings of Sect.33.

Let \( \varphi : \mathcal{D} \to \mathcal{E}' \) and \( \psi : \mathcal{D} \to \mathcal{E}' \) both be differentiable at \( x \in \mathcal{D} \). Then the value-wise difference \( \varphi - \psi : \mathcal{D} \to \mathcal{V}' \) is differentiable at \( x \) and

\[
\nabla_x (\varphi - \psi) = \nabla_x \varphi - \nabla_x \psi. \tag{63.10}
\]

This follows from Prop.2, the fact that the point-difference \( (x', y') \mapsto (x' - y') \) is a flat mapping from \( \mathcal{E}' \times \mathcal{E}' \) into \( \mathcal{V}' \), and the General Chain Rule.

When the General Chain Rule is applied to the composite of a vector-valued mapping with a linear mapping it yields
Proposition 3: If \( h : \mathcal{D} \to \mathcal{V}' \) is differentiable at \( x \in \mathcal{D} \) and if \( L \in \text{Lin}(\mathcal{V}', \mathcal{W}) \), where \( \mathcal{W} \) is some linear space, then \( \mathbf{L}h : \mathcal{D} \to \mathcal{W} \) is differentiable at \( x \in \mathcal{D} \) and
\[
\nabla_x (\mathbf{L}h) = \mathbf{L}(\nabla_x h).
\] (63.11)

Using the fact that vector-addition, transpositions of linear mappings, and the trace operations are all linear operations, we obtain the following special cases of Prop.3:

(I) Let \( h : \mathcal{D} \to \mathcal{V}' \) and \( k : \mathcal{D} \to \mathcal{V}' \) both be differentiable at \( x \in \mathcal{D} \). Then the value-wise sum \( h + k : \mathcal{D} \to \mathcal{V}' \) is differentiable at \( x \) and
\[
\nabla_x (h + k) = \nabla_x h + \nabla_x k.
\] (63.12)

(II) Let \( \mathcal{W} \) and \( \mathcal{Z} \) be linear spaces and let \( F : \mathcal{D} \to \text{Lin}(\mathcal{W}, \mathcal{Z}) \) be differentiable at \( x \). If \( F^\top : \mathcal{D} \to \text{Lin}(\mathcal{Z}^*, \mathcal{W}^*) \) is defined by value-wise transposition, then \( F^\top \) is differentiable at \( x \in \mathcal{D} \) and
\[
\nabla_x (F^\top) v = ((\nabla_x F) v)^\top \quad \text{for all} \quad v \in \mathcal{V}.
\] (63.13)

In particular, if \( I \) is an open interval and if \( F : I \to \text{Lin}(\mathcal{W}, \mathcal{Z}) \) is differentiable, so is \( F^\top : I \to \text{Lin}(\mathcal{Z}^*, \mathcal{W}^*) \), and
\[
(F^\top)^* = (F^*)^\top.
\] (63.14)

(III) Let \( \mathcal{W} \) be a linear space and let \( F : \mathcal{D} \to \text{Lin}(\mathcal{W}) \) be differentiable at \( x \). If \( \text{tr}F : \mathcal{D} \to \mathbb{R} \) is the value-wise trace of \( F \), then \( \text{tr}F \) is differentiable at \( x \) and
\[
\nabla_x (\text{tr}F) v = \text{tr}((\nabla_x F) v) \quad \text{for all} \quad v \in \mathcal{V}.
\] (63.15)

In particular, if \( I \) is an open interval and if \( F : I \to \text{Lin}(\mathcal{W}) \) is differentiable, so is \( \text{tr}F : I \to \mathbb{R} \), and
\[
(\text{tr}F)^* = \text{tr}(F^*).
\] (63.16)

We note three special cases of the General Chain Rule: Let \( I \) be an open interval and let \( \mathcal{D} \) and \( \mathcal{D}' \) be open subsets of \( \mathcal{E} \) and \( \mathcal{E}' \), respectively. If \( p : I \to \mathcal{D} \) and \( \varphi : \mathcal{D} \to \mathcal{D}' \) are differentiable, so is \( \varphi \circ p : I \to \mathcal{D}' \), and
\[
(\varphi \circ p)^* = ((\nabla \varphi) \circ p)p^*.
\] (63.17)
CHAPTER 6. DIFFERENTIAL CALCULUS

If $\varphi : D \to D'$ and $f : D' \to \mathbb{R}$ are differentiable, so is $f \circ \varphi : D \to \mathbb{R}$, and

$$\nabla (f \circ \varphi) = (\nabla \varphi)^\top ((\nabla f) \circ \varphi)$$  \hspace{1cm} (63.18)

(see (21.3)). If $f : D \to I$ and $p : I \to D'$ are differentiable, so is $p \circ f$, and

$$\nabla (p \circ f) = (p^* \circ f) \otimes \nabla f.$$  \hspace{1cm} (63.19)

Notes 63

(1) Other common terms for the concept of “gradient” of Def.1 are “differential”, “Fréchet differential”, “derivative”, and “Fréchet derivative”. Some authors make an artificial distinction between “gradient” and “differential”. We cannot use “derivative” because, for processes, “gradient” and “derivative” are distinct though related concepts.

(2) The conventional definitions of gradient depend, at first view, on the prescription of a norm. Many texts never even mention the fact that the gradient is a norm-invariant concept. In some contexts, as when one deals with genuine Euclidean spaces, this norm-invariance is perhaps not very important. However, when one deals with mathematical models for space-time in the theory of relativity, the norm-invariance is crucial because it shows that the concepts of differential calculus have a “Lorentz-invariant” meaning.

(3) I am introducing the notation $\nabla_x \varphi$ for the gradient of $\varphi$ at $x$ because the more conventional notation $\nabla \varphi(x)$ suggests, incorrectly, that $\nabla \varphi(x)$ is necessarily the value at $x$ of a gradient-mapping $\nabla \varphi$. In fact, one cannot define the gradient-mapping $\nabla \varphi$ without first having a notation for the gradient at a point (see 63.4).

(4) Other notations for $\nabla_x \varphi$ in the literature are $d\varphi(x)$, $D\varphi(x)$, and $\varphi'(x)$.

(5) I conjecture that the “Chain” of “Chain Rule” comes from an old terminology that used “chaining” (in the sense of “concatenation”) for “composition”. The term “Composition Rule” would need less explanation, but I retained “Chain Rule” because it is more traditional and almost as good.

64 Constricted Mappings

In this section, $D$ and $D'$ denote arbitrary subsets of flat spaces $\mathcal{E}$ and $\mathcal{E}'$ with translation spaces $\mathcal{V}$ and $\mathcal{V}'$, respectively.

Definition 1: We say that the mapping $\varphi : D \to D'$ is constricted if for every norm $\nu$ on $\mathcal{V}$ and every norm $\nu'$ on $\mathcal{V}'$ there is $\kappa \in \mathbb{R}^+$ such that

$$\nu'((\varphi(y) - \varphi(x)) \leq \kappa \nu(y - x)$$  \hspace{1cm} (64.1)
holds for all $x, y \in D$. The infimum of the set of all $\kappa \in \mathbb{P}^\times$ for which (64.1) holds for all $x, y \in D$ is called the \textbf{striction} of $\varphi$ relative to $\nu$ and $\nu'$; it is denoted by $\text{str}(\varphi; \nu, \nu')$.

It is clear that

$$\text{str}(\varphi; \nu, \nu') = \sup \left\{ \frac{\nu'(y) - \varphi(x)}{\nu(y - x)} \mid x, y \in D, x \neq y \right\}. \quad (64.2)$$

\textbf{Proposition 1:} For every mapping $\varphi : D \to D'$, the following are equivalent:

(i) $\varphi$ is constricted.

(ii) There exist norms $\nu$ and $\nu'$ on $V$ and $V'$, respectively, and $\kappa \in \mathbb{P}^\times$ such that (64.1) holds for all $x, y \in D$.

(iii) For every bounded subset $C$ of $V$ and every $N' \in \text{NhD}_0(V')$ there is $\rho \in \mathbb{P}^\times$ such that

$$x - y \in sC \implies \varphi(x) - \varphi(y) \in s\rho N' \quad (64.3)$$

for all $x, y \in D$ and $s \in \mathbb{P}^\times$.

\textbf{Proof:} (i) $\implies$ (ii): This implication is trivial.

(ii) $\implies$ (iii): Assume that (ii) holds, and let a bounded subset $C$ of $V$ and $N' \in \text{NhD}_0(V')$ be given. By Cor.1 of the Cell-Inclusion Theorem of Sect.52, we can choose $b \in \mathbb{P}^\times$ such that

$$\nu(u) \leq b \text{ for all } u \in C. \quad (64.4)$$

By Prop.3 of Sect.53, we can choose $\sigma \in \mathbb{P}^\times$ such that $\sigma \text{Ce}(\nu') \subset N'$. Now let $x, y \in D$ and $s \in \mathbb{P}^\times$ be given and assume that $x - y \in sC$. Then $\frac{1}{s}(x - y) \in C$ and hence, by (64.4), we have $\nu(\frac{1}{s}(x - y)) \leq b$, which gives $\nu(x - y) \leq sb$. Using (64.1) we obtain $\nu'(\varphi(y) - \varphi(x)) \leq skb$. If we put $\rho := \frac{sk}{\sigma}$, this means that

$$\varphi(y) - \varphi(x) \in s\rho \sigma \text{Ce}(\nu') \subset s\rho N',$$

i.e. that (64.3) holds.

(iii) $\implies$ (i): Assume that (iii) holds and let a norm $\nu$ on $V$ and a norm $\nu'$ on $V'$ be given. We apply (iii) with the choices $C := \text{Bdy Ce}(\nu)$ and $N' := \text{Ce}(\nu')$. Let $x, y \in D$ with $x \neq y$ be given. If we put $s := \nu(x - y)$
then \( x - y \in s \text{Bdy } Ce(\nu) \) and hence, by (64.3), \( \varphi(x) - \varphi(y) \in s \rho Ce(\nu') \), which implies
\[
\nu'(\varphi(x) - \varphi(y)) < \rho \varphi(x - y).
\]
Thus, if we put \( \kappa := \rho \), then (64.1) holds for all \( x, y \in D \).

If \( D \) and \( D' \) are open sets and if \( \varphi : D \to D' \) is constricted, it is confined near every \( x \in D \), as is evident from Prop.4 of Sect.62. The converse is not true. For example, one can show that the function \( f : [-1,1[ \to \mathbb{R} \) defined by
\[
f(t) := \begin{cases} 
t \sin(\frac{1}{t}) & \text{if } t \in ]0,1[ \\
0 & \text{if } t \in ]-1,0[ 
\end{cases}
\]
(64.5) is not constricted, but is confined near every \( t \in ]-1,1[ \).

Every flat mapping \( \alpha : E \to E' \) is constricted and
\[
\text{str}(\alpha; \nu, \nu') = ||\nabla \alpha||_{\nu, \nu'},
\]
where \( ||.||_{\nu, \nu'} \) is the operator norm on \( \text{Lin}(V, V') \) corresponding to \( \nu \) and \( \nu' \) (see Sect.52).

Constrictedness and strictions remain unaffected by a change of codomain. If the domain of a constricted mapping is restricted, then it remains constricted and the striction of the restriction is less than or equal to the striction of the original mapping. (Pardon the puns.)

**Proposition 2:** If \( \varphi : D \to D' \) is constricted then it is uniformly continuous.

**Proof:** We use condition (iii) of Prop.1. Let \( \mathcal{N} \in \text{Nhdo}(\nu') \) be given. We choose a bounded neighborhood \( \mathcal{C} \) of \( 0 \) in \( V \) and determine \( \rho \in \mathbb{P}^x \) according to (iii). Putting \( \mathcal{N} := \frac{1}{\rho} \mathcal{C} \in \text{Nhdo}(\nu) \) and \( s := \frac{1}{\rho} \), we see that (64.3) gives
\[
x - y \in \mathcal{N} \implies \varphi(x) - \varphi(y) \in \mathcal{N}' \quad \text{for all } x, y \in D.
\]

**Pitfall:** The converse of Prop.2 is not valid. A counterexample is the square-root function \( \sqrt{\cdot} : \mathbb{P} \to \mathbb{P} \), which is uniformly continuous but not constricted. Another counterexample is the function defined by (64.5).

The following result is the most useful criterion for showing that a given mapping is constricted.

**Striction Estimate for Differentiable Mappings:** Assume that \( D \) is an open convex subset of \( E \), that \( \varphi : D \to D' \) is differentiable and that the gradient \( \nabla \varphi : D \to \text{Lin}(V, V') \) has a bounded range. Then \( \varphi \) is constricted and
\[
\text{str}(\varphi; \nu, \nu') \leq \sup \{ ||\nabla z \varphi||_{\nu, \nu'} | z \in D \}
\]
(64.6)
for all norms $\nu, \nu'$ on $\mathcal{V}, \mathcal{V}'$, respectively.

**Proof:** Let $x, y \in \mathcal{D}$. Since $\mathcal{D}$ is convex, we have $tx + (1 - t)y \in \mathcal{D}$ for all $t \in [0, 1]$ and hence we can define a process

$$p : [0, 1] \to \mathcal{D}' \quad \text{by} \quad p(t) := \varphi(tx + (1 - t)y).$$

By the Chain Rule, $p$ is differentiable at $t$ when $t \in ]0, 1[$ and we have

$$\partial_t p = (\nabla_z \varphi)(x - y) \quad \text{with} \quad z := tx + (1 - t)y \in \mathcal{D}.$$

Applying the Difference-Quotient Theorem of Sect.61 to $p$, we obtain

$$\varphi(x) - \varphi(y) = p(1) - p(0) \in \text{Clo Cxh}\{(\nabla_z \varphi)(x - y) \mid z \in \mathcal{D}\}. \quad (64.7)$$

If $\nu, \nu'$ are norms on $\mathcal{V}, \mathcal{V}'$ then, by (52.7),

$$\nu'(\nabla_z \varphi)(x - y) \leq ||\nabla_z \varphi||_{\nu, \nu'} \nu(x - y) \quad (64.8)$$

for all $z \in \mathcal{D}$. To say that $\nabla \varphi$ has a bounded range is equivalent, by Cor.1 to the Cell-Inclusion Theorem of Sect.52, to

$$\kappa := \sup\{||\nabla_z \varphi||_{\nu, \nu'} \mid z \in \mathcal{D}\} < \infty. \quad (64.9)$$

It follows from (64.8) that $\nu'(\nabla_z \varphi)(x - y) \leq \kappa \nu(x - y)$ for all $z \in \mathcal{D}$, which can be expressed in the form

$$(\nabla_z \varphi)(x - y) \in \kappa \nu(x - y)\text{Ce}(\nu') \quad \text{for all} \quad z \in \mathcal{D}.$$

Since the set on the right is closed and convex we get

$$\text{Clo Cxh}\{(\nabla_z \varphi)(y - x) \mid z \in \mathcal{D}\} \subset \kappa \nu(x - y)\text{Ce}(\nu')$$

and hence, by (64.7), $\varphi(x) - \varphi(y) \in \kappa \nu(x - y)\text{Ce}(\nu')$. This may be expressed in the form

$$\nu'(\varphi(x) - \varphi(y)) \leq \kappa \nu(x - y).$$

Since $x, y \in \mathcal{D}$ were arbitrary it follows that $\varphi$ is constricted. The definition (64.9) shows that (64.6) holds.

**Remark:** It is not hard to prove that the inequality in (64.6) is actually an equality.

**Proposition 3:** If $\mathcal{D}$ is a non-empty open convex set and if $\varphi : \mathcal{D} \to \mathcal{D}'$ is differentiable with gradient zero, then $\varphi$ is constant.

**Proof:** Choose norms $\nu, \nu'$ on $\mathcal{V}, \mathcal{V}'$, respectively. The assumption $\nabla \varphi = 0$ gives $||\nabla_z \varphi||_{\nu, \nu'} = 0$ for all $z \in \mathcal{D}$. Hence, by (64.6), we have $\text{str}(\varphi, \nu, \nu') = 0$.
0. Using (64.2), we conclude that \( \nu'(\varphi(y) - \varphi(x)) = 0 \) and hence \( \varphi(x) = \varphi(y) \) for all \( x, y \in D \).

**Remark:** In Prop. 3, the condition that \( D \) be convex can be replaced by the weaker one that \( D \) be “connected”. This means, intuitively, that every two points in \( D \) can be connected by a continuous curve entirely within \( D \).

**Proposition 4:** Assume that \( D \) and \( D' \) are open subsets and that \( \varphi: D \to D' \) is of class \( C^1 \). Let \( \mathfrak{k} \) be a compact subset of \( D \). For every norm \( \nu \) on \( \mathcal{V} \), every norm \( \nu' \) on \( \mathcal{V}' \), and every \( \varepsilon \in \mathbb{P} \times \mathbb{P} \), there exists \( \delta \in \mathbb{P} \times \mathbb{P} \) such that \( \mathfrak{k} + \delta \text{Ce}(\nu) \subset D \) and such that, for every \( x \in \mathfrak{k} \), the function \( n_x: \delta \text{Ce}(\nu) \to \mathcal{V}' \) defined by

\[
\nu_x(v) := \varphi(x + v) - \varphi(x) - (\nabla_x \varphi)v
\]

is constricted with

\[
\text{str}(n_x; \nu, \nu') \leq \varepsilon.
\]

**Proof:** Let a norm \( \nu \) on \( \mathcal{V} \) be given. By Prop. 6 of Sect. 58, we can obtain \( \delta_1 \in \mathbb{P} \times \mathbb{P} \) such that \( \mathfrak{k} + \delta_1 \text{Ce}(\nu) \subset D \). For each \( x \in \mathfrak{k} \), we define \( m_x: \delta_1 \text{Ce}(\nu) \to \mathcal{V}' \) by

\[
m_x(v) := \varphi(x + v) - \varphi(x) - (\nabla_x \varphi)v.
\]

Differentiation gives

\[
\nabla_v m_x = \nabla \varphi(x + v) - \nabla \varphi(x)
\]

for all \( x \in \mathfrak{k} \) and all \( v \in \delta_1 \text{Ce}(\nu) \). Since \( \mathfrak{k} + \delta_1 \text{Ce}(\nu) \) is compact by Prop. 6 of Sect. 58 and since \( \nabla \varphi \) is continuous, it follows by the Uniform Continuity Theorem of Sect. 58 that \( \nabla \varphi|_{\mathfrak{k} + \delta_1 \text{Ce}(\nu)} \) is uniformly continuous. Now let a norm \( \nu' \) on \( \mathcal{V}' \) and \( \varepsilon \in \mathbb{P} \times \mathbb{P} \) be given. By Prop. 4 of Sect. 56, we can then determine \( \delta_2 \in \mathbb{P} \times \mathbb{P} \) such that

\[
\nu(y - x) < \delta_2 \implies ||\nabla \varphi(y) - \nabla \varphi(x)||_{\nu, \nu'} < \varepsilon
\]

for all \( x, y \in \mathfrak{k} + \delta_1 \text{Ce}(\nu) \). In view of (64.13), it follows that

\[
v \in \delta_2 \text{Ce}(\nu) \implies ||\nabla_v m_x||_{\nu, \nu'} < \varepsilon
\]

for all \( x \in \mathfrak{k} \) and all \( v \in \delta_1 \text{Ce}(\nu) \). If we put \( \delta := \min\{\delta_1, \delta_2\} \) and if we define \( n_x := m_x|_{\delta \text{Ce}(\nu)} \) for every \( x \in \mathfrak{k} \), we see that \( \{||\nabla_v n_x||_{\nu, \nu'} : v \in \delta \text{Ce}(\nu)\} \) is bounded by \( \varepsilon \) for all \( x \in \mathfrak{k} \). By (64.12) and the Striction Estimate for Differentiable Mappings, the desired result follows.
**Definition 2:** Let \( \varphi : \mathcal{D} \to \mathcal{D} \) be a constricted mapping from a set \( \mathcal{D} \) into itself. Then

\[
\text{str}(\varphi) := \inf \{ \text{str}(\varphi; \nu, \nu) \mid \nu \text{ a norm on } \mathcal{V} \}
\]

is called the absolute striction of \( \varphi \). If \( \text{str}(\varphi) < 1 \) we say that \( \varphi \) is a contraction.

**Contraction Fixed Point Theorem:** Every contraction has at most one fixed point and, if its domain is closed and not empty, it has exactly one fixed point.

**Proof:** Let \( \varphi : \mathcal{D} \to \mathcal{D} \) be a contraction. We choose a norm \( \nu \) on \( \mathcal{V} \) such that \( \kappa := \text{str}(\varphi; \nu, \nu) < 1 \) and hence, by (64.1),

\[
\nu(\varphi(x) - \varphi(y)) \leq \kappa \nu(x - y) \quad \text{for all } x, y \in \mathcal{D}.
\]

If \( x \) and \( y \) are fixed points of \( \varphi \), so that \( \varphi(x) = x \), \( \varphi(y) = y \), then (64.15) gives \( \nu(x - y)(1 - \kappa) \leq 0 \). Since \( 1 - \kappa > 0 \) this is possible only when \( \nu(x - y) = 0 \) and hence \( x = y \). Therefore, \( \varphi \) can have at most one fixed point.

We now assume \( \mathcal{D} \neq \emptyset \), choose \( s_0 \in \mathcal{D} \) arbitrarily, and define

\[
s_n := \varphi^n(s_0) \quad \text{for all } n \in \mathbb{N}^x
\]

(see Sect.03). It follows from (64.15) that

\[
\nu(s_{m+1} - s_m) = \nu(\varphi(s_m) - \varphi(s_{m-1})) \leq \kappa \nu(s_m - s_{m-1})
\]

for all \( m \in \mathbb{N}^x \). Using induction, one concludes that

\[
\nu(s_{m+1} - s_m) \leq \kappa^m \nu(s_1 - s_0) \quad \text{for all } m \in \mathbb{N}.
\]

Now, if \( n \in \mathbb{N} \) and \( r \in \mathbb{N} \), then

\[
s_{n+r} - s_n = \sum_{k \in r} (s_{n+k+1} - s_{n+k})
\]

and hence

\[
\nu(s_{n+r} - s_n) \leq \sum_{k \in r} \kappa^{n+k} \nu(s_1 - s_0) \leq \frac{\kappa^n}{1 - \kappa} \nu(s_1 - s_0).
\]

Since \( \kappa < 1 \), we have \( \lim_{n \to \infty} \kappa^n = 0 \), and it follows that for every \( \varepsilon \in \mathbb{P}^x \) there is an \( m \in \mathbb{N} \) such that \( \nu(s_{n+r} - s_n) < \varepsilon \) whenever \( n \in m + \mathbb{N}, r \in \mathbb{N} \). By
the Basic Convergence Criterion of Sect.55 it follows that \( s := (s_n \mid n \in \mathbb{N}) \) converges. Since \( \varphi(s_n) = s_{n+1} \) for all \( n \in \mathbb{N} \), we have \( \varphi \circ s = (s_{n+1} \mid n \in \mathbb{N}) \), which converges to the same limit as \( s \). We put

\[
x := \lim s = \lim (\varphi \circ s).
\]

Now assume that \( \mathcal{D} \) is closed. By Prop.6 of Sect.55 it follows that \( x \in \mathcal{D} \). Since \( \varphi \) is continuous, we can apply Prop.2 of Sect.56 to conclude that \( \lim (\varphi \circ s) = \varphi(\lim s) \), i.e. that \( \varphi(x) = x \).

Notes 64

(1) The traditional terms for “constricted” and “striction” are “Lipschitzian” and “Lipschitz number (or constant)”, respectively. I am introducing the terms “constricted” and “striction” here because they are much more descriptive.

(2) It turns out that the absolute striction of a line coincides with what is often called its “spectral radius”.

(3) The Contraction Fixed Point Theorem is often called the “Contraction Mapping Theorem” or the “Banach Fixed Point Theorem”.

65 Partial Gradients, Directional Derivatives

Let \( \mathcal{E}_1, \mathcal{E}_2 \) be flat spaces with translation spaces \( \mathcal{V}_1, \mathcal{V}_2 \). As we have seen in Sect.33, the set-product \( \mathcal{E} := \mathcal{E}_1 \times \mathcal{E}_2 \) is then a flat space with translation space \( \mathcal{V} := \mathcal{V}_1 \times \mathcal{V}_2 \). We consider a mapping \( \varphi : \mathcal{D} \to \mathcal{D}' \) from an open subset \( \mathcal{D} \) of \( \mathcal{E} \) into an open subset \( \mathcal{D}' \) of a flat space \( \mathcal{E}' \) with translation space \( \mathcal{V}' \). Given any \( x_2 \in \mathcal{E}_2 \), we define \( (\cdot, x_2) : \mathcal{E}_1 \to \mathcal{E} \) according to (04.21), put

\[
\mathcal{D}_{(\cdot,x_2)} := (\cdot, x_2)^\circ(\mathcal{D}) = \{ z \in \mathcal{E}_1 \mid (z, x_2) \in \mathcal{D} \},
\]

which is an open subset of \( \mathcal{E}_1 \) because \( (\cdot, x_2) \) is flat and hence continuous, and define \( \varphi_{(\cdot, x_2)} : \mathcal{D}_{(\cdot,x_2)} \to \mathcal{D}' \) according to (04.22). If \( \varphi_{(\cdot, x_2)} \) is differentiable at \( x_1 \) for all \( (x_1, x_2) \in \mathcal{D} \) we define the partial 1-gradient \( \nabla_{(1)} \varphi : \mathcal{D} \to \text{Lin}(\mathcal{V}_1, \mathcal{V}') \) of \( \varphi \) by

\[
\nabla_{(1)} \varphi(x_1, x_2) := \nabla_{x_1} \varphi_{(\cdot, x_2)} \quad \text{for all} \quad (x_1, x_2) \in \mathcal{D}.
\]

In the special case when \( \mathcal{E}_1 := \mathbb{R} \), we have \( \mathcal{V}_1 = \mathbb{R} \), and the partial 1-derivative \( \varphi_{,1} : \mathcal{D} \to \mathcal{V}' \), defined by

\[
\varphi_{,1}(t, x_2) := (\partial \varphi_{(\cdot, x_2)})(t) \quad \text{for all} \quad (t, x_2) \in \mathcal{D},
\]
is related to $\nabla (1) \varphi$ by $\nabla (1) \varphi = \varphi, 1 \otimes$ and $\varphi, 1 = (\nabla (1) \varphi) 1$ (value-wise).

Similar definitions are employed to define $D_{(x, \cdot)}$, the mapping $\varphi(x, \cdot) : D_{(x, \cdot)} \rightarrow D'$, the partial 2-gradient $\nabla (2) \varphi : D \rightarrow \text{Lin} (V_2, V')$ and the partial 2-derivative $\varphi, 2 : D \rightarrow V'$.

The following result follows immediately from the definitions.

**Proposition 1:** If $\varphi : D \rightarrow D'$ is differentiable at $x := (x_1, x_2) \in D$, then the gradients $\nabla x_1 \varphi(x, \cdot)$ and $\nabla x_2 \varphi(x, \cdot)$ exist and

$$(\nabla x \varphi) v = (\nabla x_1 \varphi(x, \cdot)) v_1 + (\nabla x_2 \varphi(x, \cdot)) v_2 \quad (65.4)$$

for all $v := (v_1, v_2) \in V$.

If $\varphi$ is differentiable, then the partial gradients $\nabla (1) \varphi$ and $\nabla (2) \varphi$ exist and we have

$$\nabla \varphi = \nabla (1) \varphi \oplus \nabla (2) \varphi \quad (65.5)$$

where the operation $\oplus$ on the right is understood as value-wise application of (14.13).

**Pitfall:** The converse of Prop. 1 is not true: A mapping can have partial gradients without being differentiable. For example, the mapping $\varphi : R \times R \rightarrow R$, defined by

$$\varphi(s, t) := \begin{cases} \frac{st}{s^2 + t^2} & \text{if } (s, t) \neq (0, 0) \\ 0 & \text{if } (s, t) = (0, 0) \end{cases},$$

has partial derivatives at $(0, 0)$ since both $\varphi(\cdot, 0)$ and $\varphi(0, \cdot)$ are equal to the constant 0. Since $\varphi(t, t) = \frac{1}{2}$ for $t \in R^{\times}$ but $\varphi(0, 0) = 0$, it is clear that $\varphi$ is not even continuous at $(0, 0)$, let alone differentiable.

The second assertion of Prop. 1 shows that if $\varphi$ is of class $C^1$, then $\nabla (1) \varphi$ and $\nabla (2) \varphi$ exist and are continuous. The converse of this statement is true, but the proof is highly nontrivial:

**Proposition 2:** Let $E_1, E_2, E'$ be flat spaces. Let $D$ be an open subset of $E_1 \times E_2$ and $D'$ an open subset of $E'$. A mapping $\varphi : D \rightarrow D'$ is of class $C^1$ if (and only if) the partial gradients $\nabla (1) \varphi$ and $\nabla (2) \varphi$ exist and are continuous.

**Proof:** Let $(x_1, x_2) \in D$. Since $D - (x_1, x_2)$ is a neighborhood of $(0, 0)$ in $V_1 \times V_2$, we may choose $M_1 \in \text{Nhd}_0 (V_1)$ and $M_2 \in \text{Nhd}_0 (V_2)$ such that

$$M := M_1 \times M_2 \subset (D - (x_1, x_2)).$$

We define $m : M \rightarrow V'$ by

$$m(v_1, v_2) := \varphi((x_1, x_2) + (v_1, v_2)) - \varphi(x_1, x_2) - (\nabla (1) \varphi)(x_1, x_2)v_1 + (\nabla (2) \varphi)(x_1, x_2)v_2.$$
It suffices to show that \( m \in \text{Small}(V_1 \times V_2, V') \), for if this is the case, then the Characterization of Gradients of Sect.63 tells us that \( \varphi \) is differentiable at \( (x_1, x_2) \) and that its gradient at \( (x_1, x_2) \) is given by (65.4). In addition, since \( (x_1, x_2) \in D \) is arbitrary, we can then conclude that (65.5) holds and, since both \( \nabla_{(1)} \varphi \) and \( \nabla_{(2)} \varphi \) are continuous, that \( \nabla \varphi \) is continuous.

We note that
\[
\text{Small}(\varphi) := \{ h : M \to V' \}.
\]

We now generalize Prop.2 to the case when \( E \in \mathbb{P}^\alpha \), realizing that \( \nabla_{(1)} \varphi \) and \( \nabla_{(2)} \varphi \) are continuous at \( (x_1, x_2) \), and that \( \nabla \varphi \) is continuous.

Remark: In examining the proof above one observes that one needs merely the existence of the partial gradients and the continuity of one of them in order to conclude that the mapping is differentiable.

We now generalize Prop.2 to the case when \( \mathcal{E} := \bigotimes_{i \in I} (\mathcal{E}_i | i \in I) \) is the set-product of a finite family \( (\mathcal{E}_i | i \in I) \) of flat spaces. This product \( \mathcal{E} \) is a
PARTIAL GRADIENTS, DIRECTIONAL DERIVATIVES

flat space whose translation space is the set-product \( V := \bigtimes (V_i \mid i \in I) \) of the translation spaces \( V_i \) of the \( E_i \), \( i \in I \). Given any \( x \in E \) and \( j \in I \), we define the mapping \((x.j) : E_j \to E\) according to (04.24).

We consider a mapping \( \varphi : D \to D' \) from an open subset \( D \) of \( E \) into an open subset \( D' \) of \( E' \). Given any \( x \in D \) and \( j \in I \), we put \( D(x.j) := (x.j)^{(D)} \subset E_j \) and define \( \varphi(x.j) : D(x.j) \to D' \) according to (04.25). If \( \varphi(x.j) \) is differentiable at \( x \) for all \( x \in D \), we define the partial j-gradient \( \nabla_j \varphi : D \to \text{Lin}(V_j, V') \) of \( \varphi \) by \( \nabla_j \varphi(x) := \nabla_{x.j} \varphi(x.j) \) for all \( x \in D \).

In the special case when \( E_j := \mathbb{R} \) for some \( j \in I \), the partial j-derivative \( \varphi_j : D \to V' \), defined by
\[
\varphi_j(x) := (\partial \varphi(x.j))(x_j) \quad \text{for all} \quad x \in D,
\]
is related to \( \nabla_j \varphi \) by \( \nabla_j \varphi = \varphi_j \otimes \) and \( \varphi_j = (\nabla_j \varphi)1 \).

In the case when \( I := \{1, 2\} \), these notations and concepts reduce to the ones explained in the beginning (see (04.21)).

The following result generalizes Prop.1.

**Proposition 3:** If \( \varphi : D \to D' \) is differentiable at \( x = (x_i \mid i \in I) \in D \), then the gradients \( \nabla_{x.j} \varphi(x.j) \) exist for all \( j \in I \) and
\[
(\nabla x \varphi) v = \sum_{j \in I} (\nabla_{x.j} \varphi(x.j)) v_j
\]
for all \( v = (v_i \mid i \in I) \in V \).

If \( \varphi \) is differentiable, then the partial gradients \( \nabla_j \varphi \) exist for all \( j \in I \) and we have
\[
\nabla \varphi = \bigoplus_{j \in I}(\nabla_j \varphi),
\]
where the operation \( \bigoplus \) on the right is understood as value-wise application of (14.18).

In the case when \( E := \mathbb{R}^I \), we can put \( E_i := \mathbb{R} \) for all \( i \in I \) and (65.10) reduces to
\[
\nabla \varphi = \text{ln}(\varphi, i \in I),
\]
where the value at \( x \in D \) of the right side is understood to be the linear-combination mapping of the family \( (\varphi_i(x) \mid i \in I) \) in \( V' \). If moreover, \( E' \) is also a space of the form \( E' := \mathbb{R}^K \) with some finite index set \( K \), then
ln\((\varphi, i | i \in I)\) can be identified with the function which assigns to each \(x \in D\) the matrix
\[
(\varphi_{k,i}(x) | (k, i) \in K \times I) \in \mathbb{R}^{K \times I} \cong \text{Lin}(\mathbb{R}^I, \mathbb{R}^K).
\]

Hence \(\nabla \varphi\) can be identified with the matrix
\[
\nabla \varphi = (\varphi_{k,i} | (k, i) \in K \times I).
\] (65.12)

of the partial derivatives \(\varphi_{k,i} : D \to \mathbb{R}\) of the component functions. As we have seen, the mere existence of these partial derivatives is not sufficient to insure the existence of \(\nabla \varphi\). Only if \(\nabla \varphi\) is known a priori to exist is it possible to use the identification (65.12).

Using the natural isomorphism between \(\bigotimes (E_i | i \in I)\) and \(E_j \times \bigotimes (E_i | i \in I \setminus \{j\})\) and Prop.2, one easily proves, by induction, the following generalization.

**Partial Gradient Theorem:** Let \(I\) be a finite index set and let \(E_i, i \in I\) and \(E'\) be flat spaces. Let \(D\) be an open subset of \(E := \bigotimes (E_i | i \in I)\) and \(D'\) an open subset of \(E'\). A mapping \(\varphi : D \to D'\) is of class \(C^1\) if and only if the partial gradients \(\nabla_{(i)} \varphi\) exist and are continuous for all \(i \in I\).

The following result is an immediate corollary:

**Proposition 4:** Let \(I\) and \(K\) be finite index sets and let \(D\) and \(D'\) be open subsets of \(\mathbb{R}^I\) and \(\mathbb{R}^K\), respectively. A mapping \(\varphi : D \to D'\) is of class \(C^1\) if and only if the partial derivatives \(\varphi_{k,i} : D \to \mathbb{R}\) exist and are continuous for all \(k \in K\) and all \(i \in I\).

Let \(E, E'\) be flat spaces with translation spaces \(V\) and \(V'\), respectively. Let \(D, D'\) be open subsets of \(E\) and \(E'\), respectively, and consider a mapping \(\varphi : D \to D'\). Given any \(x \in D\) and \(v \in V^x\), the range of the mapping \((s \mapsto (x + sv)) : \mathbb{R} \to E\) is a line through \(x\) whose direction space is \(\mathbb{R}v\).

Let \(S_{x,v} := \{s \in \mathbb{R} | x + sv \in D\}\) be the pre-image of \(D\) under this mapping and \(x + 1_{S_{x,v}}v : S_{x,v} \to D\) a corresponding adjustment of the mapping. Since \(D\) is open, \(S_{x,v}\) is an open neighborhood of zero in \(\mathbb{R}\) and \(\varphi \circ (x + 1_{S_{x,v}}v) : S_{x,v} \to D'\) is a process. The derivative at \(0 \in \mathbb{R}\) of this process, if it exists, is called the **directional derivative of \(\varphi\) at \(x\)** and is denoted by

\[
(dd_{v} \varphi)(x) := \partial_0(\varphi \circ (x + 1_{S_{x,v}}v)) = \lim_{s \to 0} \frac{\varphi(x + sv) - \varphi(x)}{s}. \tag{65.13}
\]
If this directional derivative exists for all \( x \in D \), it defines a function
\[
\text{dd}_v \varphi : D \to V'
\]
which is called the **directional v-derivative** of \( \varphi \). The following result is immediate from the General Chain Rule:

**Proposition 5:** If \( \varphi : D \to D' \) is differentiable at \( x \in D \) then the directional v-derivative of \( \varphi \) at \( x \) exists for all \( v \in V \) and is given by
\[
(\text{dd}_v \varphi)(x) = (\nabla_x \varphi)v.
\]

**Pitfall:** The converse of this Proposition is false. In fact, a mapping can have directional derivatives in all directions at all points in its domain without being differentiable. For example, the mapping \( \varphi : \mathbb{R}^2 \to \mathbb{R} \) defined by
\[
\varphi(s,t) := \begin{cases} 
  s^2t & \text{if } (s,t) \neq (0,0) \\
  0 & \text{if } (s,t) = (0,0) 
\end{cases}
\]
has the directional derivatives
\[
(\text{dd}_{(\alpha,\beta)} \varphi)(0,0) = \begin{cases} 
  \frac{\alpha^2}{\beta} & \text{if } \beta \neq 0 \\
  0 & \text{if } \beta = 0 
\end{cases}
\]
at \((0,0)\). Using Prop.4 one easily shows that \( \varphi|_{\mathbb{R}^2 \setminus \{(0,0)\}} \) is of class \( C^1 \) and hence, by Prop.5, \( \varphi \) has directional derivatives in all directions at all \((s,t) \in \mathbb{R}^2\). Nevertheless, since \( \varphi(s,s^2) = \frac{1}{2} \) for all \( s \in \mathbb{R} \), \( \varphi \) is not even continuous at \((0,0)\), let alone differentiable.

**Proposition 6:** Let \( b \) be a set basis of \( V \). Then \( \varphi : D \to D' \) is of class \( C^1 \) if and only if the directional \( b \)-derivatives \( \text{dd}_b \varphi : D \to V' \) exist and are continuous for all \( b \in b \).

**Proof:** We choose \( q \in D \) and define \( \alpha : \mathbb{R}^b \to E \) by \( \alpha(\lambda) := q + \sum_{b \in b} \lambda_b b \) for all \( \lambda \in \mathbb{R}^b \). Since \( b \) is a basis, \( \alpha \) is a flat isomorphism. If we define \( \overline{D} := \alpha^{-1}(D) \subset \mathbb{R}^b \) and \( \overline{\varphi} := \varphi \circ \alpha|_{\overline{D}} : \overline{D} \to D' \), we see that the directional \( b \)-derivatives of \( \varphi \) correspond to the partial derivatives of \( \overline{\varphi} \). The assertion follows from the Partial Gradient Theorem, applied to the case when \( I := b \) and \( E_i := \mathbb{R} \) for all \( i \in I \).

Combining Prop.5 and Prop.6, we obtain

**Proposition 7:** Assume that \( S \in \text{Sub} V \) spans \( V \). The mapping \( \varphi : D \to D' \) is of class \( C^1 \) if and only if the directional derivatives \( \text{dd}_v \varphi : D \to V' \) exist and are continuous for all \( v \in S \).
(1) The notation $\nabla_i$ for the partial $i$-gradient is introduced here for the first time. I am not aware of any other notation in the literature.

(2) The notation $\varphi, i$ for the partial $i$-derivative of $\varphi$ is the only one that occurs frequently in the literature and is not objectionable. Frequentiy seen notations such as $\partial \varphi/\partial x_i$ or $\varphi_x$, for $\varphi_x$, are poison to me because they contain dangling dummies (see Part D of the Introduction).

(3) Some people use the notation $\nabla_v$ for the directional derivative $dd_v$. I am introducing $dd_v$ because (by Def.1 of Sect.63) $\nabla_v$ means something else, namely the gradient at $v$.

66 The General Product Rule

The following result shows that bilinear mappings (see Sect.24) are of class $C^1$ and hence continuous. (They are not uniformly continuous except when zero.)

**Proposition 1:** Let $\mathcal{V}_1, \mathcal{V}_2$ and $\mathcal{W}$ be linear spaces. Every bilinear mapping $B : \mathcal{V}_1 \times \mathcal{V}_2 \rightarrow \mathcal{W}$ is of class $C^1$ and its gradient $\nabla B : \mathcal{V}_1 \times \mathcal{V}_2 \rightarrow \text{Lin}(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{W})$ is the linear mapping given by

$$\nabla B(v_1, v_2) = (B^\sim v_2) \oplus (Bv_1)$$

(66.1)

for all $v_1 \in \mathcal{V}_1, v_2 \in \mathcal{V}_2$.

**Proof:** Since $B(v_1, \cdot) = B v_1 : \mathcal{V}_2 \rightarrow \mathcal{W}$ (see (24.2)) is linear for each $v_1 \in \mathcal{V}$, the partial 2-gradient of $B$ exists and is given by $\nabla(2)B(v_1, v_2) = B v_1$ for all $(v_1, v_2) \in \mathcal{V}_1 \times \mathcal{V}_2$. It is evident that $\nabla(2)B : \mathcal{V}_1 \times \mathcal{V}_2 \rightarrow \text{Lin}(\mathcal{V}_2, \mathcal{W})$ is linear and hence continuous. A similar argument shows that $\nabla(1)B$ is given by $\nabla(1)B(v_1, v_2) = B^\sim v_2$ for all $(v_1, v_2) \in \mathcal{V}_1 \times \mathcal{V}_2$ and hence is also linear and continuous. By the Partial Gradient Theorem of Sect.65, it follows that $B$ is of class $C^1$. The formula (66.1) is a consequence of (65.5).

Let $\mathcal{D}$ be an open set in a flat space $\mathcal{E}$ with translation space $\mathcal{V}$. If $h_1 : \mathcal{D} \rightarrow \mathcal{V}_1$, $h_2 : \mathcal{D} \rightarrow \mathcal{V}_2$ and $B \in \text{Lin}(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{W})$ we write

$$B(h_1, h_2) : \mathcal{D} \rightarrow \mathcal{W}$$

so that

$$B(h_1, h_2)(x) = B(h_1(x), h_2(x)) \quad \text{for all } x \in \mathcal{D}.$$
The following theorem, which follows directly from Prop.1 and the General Chain Rule of Sect.63, is called “Product Rule” because the term "product" is often used for the bilinear forms to which the theorem is applied.

**General Product Rule:** Let \( V_1, V_2 \) and \( W \) be linear spaces and let \( B \in \text{Lin}_2(V_1 \times V_2, W) \) be given. Let \( D \) be an open subset of a flat space and let mappings \( h_1 : D \to V_1 \) and \( h_2 : D \to V_2 \) be given. If \( h_1 \) and \( h_2 \) are both differentiable at \( x \in D \) so is \( B(h_1, h_2) \), and we have

\[
\nabla_x B(h_1, h_2) = (Bh_1(x))\nabla_x h_2 + (B^\sim h_2(x))\nabla_x h_1.
\]

(66.2)

If \( h_1 \) and \( h_2 \) are both differentiable [of class \( C^1 \)], so is \( B(h_1, h_2) \), and we have

\[
\nabla B(h_1, h_2) = (Bh_1)\nabla h_2 + (B^\sim h_2)\nabla h_1,
\]

(66.3)

where the products on the right are understood as value-wise compositions.

We now apply the General Product Rule to special bilinear mappings, namely to the ordinary product in \( \mathbb{R} \) and to the four examples given in Sect.24. In the following list of results \( D \) denotes an open subset of a flat space having \( V \) as its translation space; \( W, W' \) and \( W'' \) denote linear spaces.

(I) If \( f : D \to \mathbb{R} \) and \( g : D \to \mathbb{R} \) are differentiable [of class \( C^1 \)], so is the value-wise product \( fg : D \to \mathbb{R} \) and we have

\[
\nabla (fg) = f\nabla g + g\nabla f.
\]

(66.4)

(II) If \( f : D \to \mathbb{R} \) and \( h : D \to W \) are differentiable [of class \( C^1 \)], so is the value-wise scalar multiple \( fh : D \to W \), and we have

\[
\nabla (fh) = h \odot \nabla f + f \nabla h.
\]

(66.5)

(III) If \( h : D \to W \) and \( \eta : D \to W^* \) are differentiable [of class \( C^1 \)], so is the function \( \eta h : D \to \mathbb{R} \) defined by \( \eta h(x) := \eta(x)h(x) \) for all \( x \in D \), and we have

\[
\nabla (\eta h) = (\nabla h)^\top \eta + (\nabla \eta)^\top h.
\]

(66.6)

(IV) If \( F : D \to \text{Lin}(W,W') \) and \( h : D \to W \) are differentiable at \( x \in D \), so is \( Fh \) (defined by \( (Fh)(y) := F(y)h(y) \) for all \( y \in D \)) and we have

\[
\nabla_x (Fh)v = ((\nabla_x F)v)h(x) + (F(x)\nabla_x h)v
\]

(66.7)

for all \( v \in V \). If \( F \) and \( h \) are differentiable [of class \( C^1 \)], so is \( Fh \) and (66.7) holds for all \( x \in D, v \in V \).
If $F : D \rightarrow \text{Lin}(W, W')$ and $G : D \rightarrow \text{Lin}(W', W'')$ are differentiable at $x \in D$, so is $GF$, defined by value-wise composition, and we have
\[
\nabla_x (GF)v = ((\nabla_x G)v)F(x) + G(x)(\nabla_x F)v
\]
for all $v \in V$. If $F$ and $G$ are differentiable \([\text{of class } C^1]\), so is $GF$, and (66.8) holds for all $x \in D, v \in V$.

If $W$ is an inner-product space, then $W^*$ can be identified with $W$, (see Sect.41) and (III) reduces to the following result.

(VI) If $h, k : D \rightarrow W$ are differentiable \([\text{of class } C^1]\), so is their value-wise inner product $h \cdot k$ and
\[
\nabla(h \cdot k) = (\nabla h)^\top k + (\nabla k)^\top h.
\]

In the case when $D$ reduces to an interval in $\mathbb{R}$, (66.4) becomes the Product Rule (08.32) of elementary calculus. The following formulas apply to differentiable processes $f, h, \eta, F$, and $G$ with values in $\mathbb{R}, W, W^*, \text{Lin}(W, W')$, and $\text{Lin}(W', W'')$, respectively:
\[
(fh)^\bullet = f^\bullet h + fh^\bullet,
\]
\[
(\eta h)^\bullet = \eta^\bullet h + \eta h^\bullet,
\]
\[
(Fh)^\bullet = F^\bullet h + Fh^\bullet,
\]
\[
(GF)^\bullet = G^\bullet F + GF^\bullet.
\]
If $h$ and $k$ are differentiable processes with values in an inner product space, then
\[
(h \cdot k)^\bullet = h^\bullet \cdot k + h \cdot k^\bullet.
\]

Proposition 2: Let $W$ be an inner-product space and let $R : D \rightarrow \text{Lin}W$ be given. If $R$ is differentiable at $x \in D$ and if $\text{Rng } R \subset \text{Orth}W$ then
\[
R(x)^\top ((\nabla_x R)v) \in \text{Skew}W \quad \text{for all } v \in V.
\]
Conversely, if $R$ is differentiable, if $D$ is convex, if (66.15) holds for all $x \in D$, and if $R(q) \in \text{Orth}W$ for some $q \in D$ then $\text{Rng } R \subset \text{Orth}W$.

Proof: Assume that $R$ is differentiable at $x \in D$. Using (66.8) with the choice $F := R$ and $G := R^\top$ we find that
\[
\nabla_x (R^\top R)v = ((\nabla_x R^\top)v)R(x) + R^\top(x)((\nabla_x R)v)
\]
holds for all \( v \in V \). By (63.13) we obtain

\[
\nabla_x (R^\top R)v = W^\top + W \quad \text{with} \quad W := R^\top (x)((\nabla_x R)v)
\]

(66.16)

for all \( v \in V \). Now, if \( \text{Rng } R \subset \text{Orth} V \) then \( R^\top R = 1_W \) is constant (see Prop.2 of Sect.43). Hence the left side of (66.16) is zero and \( W \) must be skew for all \( v \in V \), i.e. (66.15) holds.

Conversely, if (66.15) holds for all \( x \in D \), then, by (66.16) \( \nabla_x (R^\top R) = 0 \) for all \( x \in D \). Since \( D \) is convex, we can apply Prop.3 of Sect.64 to conclude that \( R^\top R \) must be constant. Hence, if \( R(q) \in \text{Orth} W \), i.e. if \( (R^\top R)(q) = 1_W \), then \( R^\top R \) is the constant \( 1_W \), i.e. \( \text{Rng } R \subset \text{Orth} W \).

**Remark:** In the second part of Prop.2, the condition that \( D \) be convex can be replaced by the one that \( D \) be “connected” as explained in the Remark after Prop.3 of Sect.64.

**Corollary:** Let \( I \) be an open interval and let \( R : I \to \text{Lin} W \) be a differentiable process. Then \( \text{Rng } R \subset \text{Orth} W \) if and only if \( R(t) \in \text{Orth} W \) for some \( t \in I \) and \( \text{Rng } (R^\top R) \subset \text{Skew} W \).

The following result shows that quadratic forms (see Sect.27) are of class \( C^1 \) and hence continuous.

**Proposition 3:** Let \( V \) be a linear space. Every quadratic form \( Q : V \to \mathbb{R} \) is of class \( C^1 \) and its gradient \( \nabla Q : V \to V^\ast \) is the linear mapping

\[
\nabla Q = 2 \overline{Q},
\]

(66.17)

where \( \overline{Q} \) is identified with the symmetric bilinear form \( \overline{Q} \in \text{Sym}_2(V^2, \mathbb{R}) \cong \text{Sym}(V, V^\ast) \) associated with \( Q \) (see (27.14)).

**Proof:** By (27.14), we have

\[
Q(u) = \overline{Q}(u, u) = (\overline{Q}(1_V, 1_V))(u)
\]

for all \( u \in V \). Hence, if we apply the General Product Rule with the choices \( B := \overline{Q} \) and \( h_1 := h_2 := 1_V \), we obtain \( \nabla Q = \overline{Q} + \overline{Q}^\sim \). Since \( \overline{Q} \) is symmetric, this reduces to (66.17).

Let \( V \) be a linear space. The lineonic \( n \)th power \( \text{pow}_n : \text{Lin} V \to \text{Lin} V \) on the algebra \( \text{Lin} V \) of lineons (see Sect.18) is defined by

\[
\text{pow}_n(L) := L^n \quad \text{for all} \quad n \in \mathbb{N}.
\]

(66.18)

The following result is a generalization of the familiar differentiation rule \( (u^n)' = nu^{n-1} \).
Proposition 4: For every \( n \in \mathbb{N} \) the lineonic nth power \( \text{pow}_n \) on \( \text{LinV} \) defined by \( (66.18) \) is of class \( C^1 \) and, for each \( L \in \text{LinV} \), its gradient \( \nabla_L \text{pow}_n \in \text{Lin} (\text{LinV}) \) at \( L \in \text{LinV} \) is given by

\[
(\nabla_L \text{pow}_n) M = \sum_{k \in \mathbb{N}} L^{k-1} ML^{n-k} \quad \text{for all } M \in \text{LinV}.
\]  

(66.19)

Proof: For \( n = 0 \) the assertion is trivial. For \( n = 1 \), we have \( \text{pow}_1 = 1_{\text{LinV}} \) and hence \( \nabla_L \text{pow}_1 = 1_{\text{LinV}} \) for all \( L \in \text{LinV} \), which is consistent with \( (66.19) \). Also, since \( \nabla \text{pow}_1 \) is constant, \( \text{pow}_1 \) is of class \( C^1 \). Assume, then, that the assertion is valid for a given \( n \in \mathbb{R} \). Since \( \text{pow}_{n+1} = \text{pow}_1 \text{pow}_n \) holds in terms of value-wise composition, we can apply the result (V) above to the case when \( F := \text{pow}_n \), \( G := \text{pow}_1 \) in order to conclude that \( \text{pow}_{n+1} \) is of class \( C^1 \) and that

\[
(\nabla_L \text{pow}_{n+1}) M = (\nabla_L \text{pow}_1) M \text{pow}_n(L) + \text{pow}_1(L)(\nabla_L \text{pow}_n) M
\]

for all \( M \in \text{LinV} \). Using \( (66.18) \) and \( (66.19) \) we get

\[
(\nabla_L \text{pow}_{n+1}) M = ML^n + L\left( \sum_{k \in \mathbb{N}} L^{k-1} ML^{n-k} \right),
\]

which shows that \( (66.19) \) remains valid when \( n \) is replaced by \( n + 1 \). The desired result follows by induction. \( \blacksquare \)

If \( M \) commutes with \( L \), then \( (66.19) \) reduces to

\[
(\nabla_L \text{pow}_n) M = (nL^{n-1}) M.
\]

Using Prop.4 and the form \((63.17)\) of the Chain Rule, we obtain

Proposition 5: Let \( I \) be an interval and let \( F : I \to \text{LinV} \) be a process. If \( F \) is differentiable \( [ \text{of class } C^1 ] \), so is its value-wise nth power \( F^n : I \to \text{LinV} \) and we have

\[
(F^n) \cdot = \sum_{k \in \mathbb{N}} F^{k-1} F \cdot F^{n-k} \quad \text{for all } n \in \mathbb{N}.
\]

(66.20)

67 Divergence, Laplacian

In this section, \( D \) denotes an open subset of a flat space \( \mathcal{E} \) with translation space \( \mathcal{V} \), and \( \mathcal{W} \) denotes a linear space.
If \( h : D \rightarrow V \) is differentiable at \( x \in D \), we can form the trace (see Sect.26) of the gradient \( \nabla_x h \in \text{Lin}V \). The result

\[
\text{div}_x h := \text{tr}(\nabla_x h) \tag{67.1}
\]

is called the divergence of \( h \) at \( x \). If \( h \) is differentiable we can define the divergence \( \text{div} h : D \rightarrow \mathbb{R} \) of \( h \) by

\[
(\text{div} h)(x) := \text{div}_x h \quad \text{for all} \quad x \in D. \tag{67.2}
\]

Using the product rule (66.5) and (26.3) we obtain

**Proposition 1:** Let \( h : D \rightarrow V \) and \( f : D \rightarrow \mathbb{R} \) be differentiable. Then the divergence of \( fh : D \rightarrow V \) is given by

\[
(\text{div} fh)(x) = (\nabla f)h + f(\text{div} h). \tag{67.3}
\]

Consider now a mapping \( H : D \rightarrow \text{Lin}(V^*, W) \) that is differentiable at \( x \in D \). For every \( \omega \in W^* \) we can form the value-wise composite \( \omega H : D \rightarrow \text{Lin}(V^*, \mathbb{R}) = V^{**} \cong V \) (see Sect.22). Since \( \omega H \) is differentiable at \( x \) for every \( \omega \in W^* \) (see Prop.3 of Sect.63) we may consider the mapping

\[
(\omega \mapsto \text{tr}(\nabla_x (\omega H))) : W^* \rightarrow \mathbb{R}.
\]

It is clear that this mapping is linear and hence an element of \( W^{**} \cong W \).

**Definition 1:** Let \( H : D \rightarrow \text{Lin}(V^*, W) \) be differentiable at \( x \in D \). Then the divergence of \( H \) at \( x \) is defined to be the (unique) element \( \text{div}_x H \) of \( W \) which satisfies

\[
\omega(\text{div}_x H) = \text{tr}(\nabla_x (\omega H)) \quad \text{for all} \quad \omega \in W^*. \tag{67.4}
\]

If \( H \) is differentiable, then its divergence \( \text{div} H : D \rightarrow W \) is defined by

\[
(\text{div} H)(x) := \text{div}_x H \quad \text{for all} \quad x \in D. \tag{67.5}
\]

In the case when \( W := \mathbb{R} \), we also have \( W^* \cong \mathbb{R} \) and \( \text{Lin}(V^*, W) = V^{**} \cong V \). Thus, using (67.4) with \( \omega := 1 \in \mathbb{R} \), we see that the definition just given is consistent with (67.1) and (67.2).

If we replace \( h \) and \( L \) in Prop.3 of Sect.63 by \( H \) and

\[
(K \mapsto \omega K) \in \text{Lin}(\text{Lin}(V^*, W), V),
\]

respectively, we obtain

\[
\nabla_x (\omega H) = \omega \nabla_x H \quad \text{for all} \quad \omega \in W^*, \tag{67.6}
\]
where the right side must be interpreted as the composite of $\nabla_x H \in \text{Lin}_2(\mathcal{V} \times \mathcal{V}^*, \mathcal{W})$ with $\omega \in \text{Lin}(\mathcal{W}, \mathbb{R})$. This composite, being an element of $\text{Lin}_2(\mathcal{V} \times \mathcal{V}^*, \mathbb{R})$, must be reinterpreted as an element of $\text{Lin}\mathcal{V}$ via the identifications $\text{Lin}_2(\mathcal{V} \times \mathcal{V}^*, \mathbb{R}) \cong \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}^*, \mathbb{R})) \cong \text{Lin}\mathcal{V}$ to make (67.6) meaningful. Using (67.6), (67.4), and (67.1) we see that $\text{div} H$ satisfies

$$\omega(\text{div}_x H) = \text{div}_x (\omega H) = \text{tr}(\omega \nabla_x H)$$

(67.7)

for all $\omega \in \mathcal{W}^*$.

The following results generalize Prop.1.

**Proposition 2:** Let $H : D \to \text{Lin}(\mathcal{V}^*, \mathcal{W})$ and $\rho : D \to \mathcal{W}^*$ be differentiable. Then the divergence of the value-wise composite $\rho H : D \to \text{Lin}(\mathcal{V}^*, \mathbb{R}) \cong \mathcal{V}$ is given by

$$\text{div} (\rho H) = \rho \text{div} H + \text{tr}(H^\top \nabla \rho),$$

(67.8)

where value-wise evaluation and composition are understood.

**Proof:** Let $x \in D$ and $v \in \mathcal{V}$ be given. Using (66.8) with the choices $G := \rho$ and $F := H$ we obtain

$$\nabla_x (\rho H)v = ((\nabla_x \rho)v)H(x) + \rho(x)((\nabla_x H)v).$$

(67.9)

Since $(\nabla_x \rho)v \in \mathcal{W}^*$, (21.3) gives

$$((\nabla_x \rho)v)H(x) = (H(x)^\top \nabla_x \rho)v.$$

On the other hand, we have

$$\rho(x)(\nabla_x H)v = (\rho(x)(\nabla_x H))v$$

if we interpret $\nabla_x H$ on the right as an element of $\text{Lin}_2(\mathcal{V} \times \mathcal{V}^*, \mathcal{W})$. Hence, since $v \in \mathcal{V}$ was arbitrary, (67.9) gives

$$\nabla_x (\rho H) = \rho(x)\nabla_x H + H(x)^\top \nabla_x \rho.$$

Taking the trace, using (67.1), and using (67.7) with $\omega := \rho(x)$, we get

$$\text{div}_x (\rho H) = \rho(x)\text{div}_x H + \text{tr}(H(x)^\top \nabla_x \rho).$$

Since $x \in D$ was arbitrary, the desired result (67.8) follows.

**Proposition 3:** Let $h : D \to \mathcal{V}$ and $k : D \to \mathcal{W}$ be differentiable. The divergence of $k \otimes h : D \to \text{Lin}(\mathcal{V}^*, \mathcal{W})$, defined by taking the value-wise tensor product (see Sect.25), is then given by

$$\text{div} (k \otimes h) = (\nabla k)h + (\text{div} h)k.$$

(67.10)
67. DIVERGENCE, LAPLACIAN

**Proof:** By (67.7) we have
\[ \omega \text{div}_{x} (k \otimes h) = \text{div}_{x} ((\omega k)h) = \text{div}_{x} (\omega (k \otimes h)) \]
for all \( x \in D \) and all \( \omega \in \mathcal{W}^* \). Hence, by Prop.1,
\[ \omega \text{div} (k \otimes h) = (\nabla (\omega k))h + (\omega k) \text{div} h = \omega ((\nabla k)h + k \text{div} h) \]
holds for all \( \omega \in \mathcal{W}^* \), and (67.10) follows.

**Proposition 4:** Let \( H : D \to \text{Lin}(\mathcal{V}^*, \mathcal{W}) \) and \( f : D \to \mathbb{R} \) be differentiable. The divergence of \( fH : D \to \text{Lin}(\mathcal{V}^*, \mathcal{W}) \) is then given by
\[ \text{div} (fH) = f \text{div} H + H(\nabla f). \quad (67.11) \]

**Proof:** Let \( \omega \in \mathcal{W}^* \) be given. If we apply Prop.2 to the case when \( \rho = f\omega \) we obtain
\[ \text{div} (\omega (fH)) = \omega (f \text{div} H) + \text{tr}(H^\top \nabla (f\omega)). \quad (67.12) \]
Using \( \nabla (f\omega) = \omega \otimes \nabla f \) and using (25.9), (26.3), and (22.3), we obtain
\[ \text{tr}(H^\top \nabla (f\omega)) = \text{tr}(H^\top (\omega \otimes \nabla f)) = \nabla f (H^\top \omega) = \omega (H \nabla f). \]
Hence (67.12) and (67.7) yield \( \omega \text{div} (fH) = \omega (f \text{div} H + H(\nabla f)). \) Since \( \omega \in \mathcal{W}^* \) was arbitrary, (67.11) follows.

From now on we assume that \( \mathcal{E} \) has the structure of a Euclidean space, so that \( \mathcal{V} \) becomes an inner-product space and we can use the identification \( \mathcal{V}^* \cong \mathcal{V} \) (see Sect.41). Thus, if \( k : D \to \mathcal{W} \) is twice differentiable, we can consider the divergence of \( \nabla k : D \to \text{Lin}(\mathcal{V}, \mathcal{W}) \cong \text{Lin}(\mathcal{V}^*, \mathcal{W}) \).

**Definition 2:** Let \( k : D \to \mathcal{W} \) be twice differentiable. Then the Laplacian of \( k \) is defined to be
\[ \Delta k := \text{div} (\nabla k). \quad (67.13) \]

If \( \Delta k = 0 \), then \( k \) is called a harmonic function.

**Remark:** In the case when \( \mathcal{E} \) is not a genuine Euclidean space, but one whose translation space has index 1 (see Sect.47), the term D’Alembertian or Wave-Operator and the symbol \( \Box \) are often used instead of Laplacian and \( \Delta \).

The following result is a direct consequence of (66.5) and Props.3 and 4.

**Proposition 5:** Let \( k : D \to \mathcal{W} \) and \( f : D \to \mathbb{R} \) be twice differentiable. The Laplacian of \( f k : D \to \mathcal{W} \) is then given by
\[ \Delta (fk) = (\Delta f)k + f \Delta k + 2(\nabla k)(\nabla f). \quad (67.14) \]
CHAPTER 6. DIFFERENTIAL CALCULUS

The following result follows directly from the form (63.19) of the Chain Rule and Prop.3.

**Proposition 6:** Let \( I \) be an open interval and let \( f : D \to I \) and \( g : I \to W \) both be twice differentiable. The Laplacian of \( g \circ f : D \to W \) is then given by

\[
\Delta(g \circ f) = (\nabla f)^2(g^{**} \circ f) + (\Delta f)(g^* \circ f).
\] (67.15)

We now discuss an important application of Prop.6. We consider the case when \( f : D \to I \) is given by

\[
f(x) := (x - q)^2 \text{ for all } x \in D,
\] (67.16)

where \( q \in \mathcal{E} \) is given. We wish to solve the following problem: How can \( D, I \) and \( g : I \to W \) be chosen such that \( g \circ f \) is harmonic?

It follows directly from (67.16) and Prop.3 of Sect.66, applied to \( Q := sq \), that \( \nabla_x f = 2(x - q) \) for all \( x \in D \) and that hence \( (\nabla f)^* = 4f \). Also, \( \nabla(\nabla f) \) is the constant with value \( 21_V \) and hence, by (26.9),

\[
\Delta f = \text{div}(\nabla f) = 2tr_1_V = 2 \dim V = 2 \dim \mathcal{E}.
\]

Substitution of these results into (67.15) gives

\[
\Delta(g \circ f) = 4f(g^{**} \circ f) + 2(\dim \mathcal{E})g^* \circ f.
\] (67.17)

Hence, \( g \circ f \) is harmonic if and only if

\[
(2g^{**} + ng^*) \circ f = 0 \quad \text{with} \quad n := \dim \mathcal{E},
\] (67.18)

where \( \iota \) is the identity mapping of \( \mathbb{R} \), suitably adjusted (see Sect.08). Now, if we choose \( I := \text{Rng } f \), then (67.18) is satisfied if and only if \( g \) satisfies the ordinary differential equation \( 2\iota g^{**} + ng^* = 0 \). It follows that \( g \) must be of the form

\[
g = \begin{cases} 
(a \iota - (\frac{n}{2} - 1)) + b & \text{if } n \neq 2 \\
 a \log + b & \text{if } n = 2 
\end{cases},
\] (67.19)

where \( a, b \in W \), provided that \( I \) is an interval.

If \( \mathcal{E} \) is a genuine Euclidean space, we may take \( D := \mathcal{E} \setminus \{q\} \) and \( I := \text{Rng } f = \mathbb{P}^\times \). We then obtain the harmonic function

\[
h = \begin{cases} 
(a \iota - (n-2)) + b & \text{if } n \neq 2 \\
 a(\log \circ r) + b & \text{if } n = 2 
\end{cases},
\] (67.20)

where \( a, b \in W \) and where \( r : \mathcal{E} \setminus \{q\} \to \mathbb{P}^\times \) is defined by \( r(x) := |x - q| \).
If the Euclidean space $E$ is not genuine and $V$ is double-signed we have two possibilities. For all $n \in \mathbb{N}^\times$ we may take $D := \{x \in E \mid (x - q)^2 > 0\}$ and $I := \mathbb{P}^\times$. If $n$ is even and $n \neq 2$, we may instead take $D := \{x \in E \mid (x - q)^2 < 0\}$ and $I := -\mathbb{P}^\times$.

Notes 67

(1) In some of the literature on “Vector Analysis”, the notation $\nabla \cdot h$ instead of $\text{div } h$ is used for the divergence of a vector field $h$, and the notation $\nabla^2 f$ instead of $\Delta f$ for the Laplacian of a function $f$. These notations come from a formalistic and erroneous understanding of the meaning of the symbol $\nabla$ and should be avoided.

68. Local Inversion, Implicit Mappings

In this section, $D$ and $D'$ denote open subsets of flat spaces $E$ and $E'$, respectively.

To say that a mapping $\varphi : D \to D'$ is differentiable at $x \in D$ means, roughly, that $\varphi$ may be approximated, near $x$, by a flat mapping $\alpha$, the tangent of $\varphi$ at $x$. One might expect that if the tangent $\alpha$ is invertible, then $\varphi$ itself is, in some sense, “locally invertible near $x$”. To decide whether this expectation is justified, we must first give a precise meaning to “locally invertible near $x$”.

Definition: Let a mapping $\varphi : D \to D'$ be given. We say that $\psi$ is a local inverse of $\varphi$ if $\psi = (\varphi|_{\mathcal{N}'})^\leftarrow$ for suitable open subsets $\mathcal{N} = \text{Cod } \psi = \text{Rng } \psi$ and $\mathcal{N}' = \text{Dom } \psi$ of $D$ and $D'$, respectively. We say that $\psi$ is a local inverse of $\varphi$ near $x \in D$ if $x \in \text{Rng } \psi$. We say that $\varphi$ is locally invertible near $x \in D$ if it has some local inverse near $x$.

To say that $\psi$ is a local inverse of $\varphi$ means that

$$\begin{align*}
\psi \circ \varphi|_{\mathcal{N}'} &= 1_{\mathcal{N}'}, \\
\varphi|_{\mathcal{N}'} \circ \psi &= 1_{\mathcal{N}},
\end{align*}$$

for suitable open subsets $\mathcal{N} = \text{Cod } \psi$ and $\mathcal{N}' = \text{Dom } \psi$ of $D$ and $D'$, respectively.

If $\psi_1$ and $\psi_2$ are local inverses of $\varphi$ and $\mathcal{M} := (\text{Rng } \psi_1) \cap (\text{Rng } \psi_2)$, then $\psi_1$ and $\psi_2$ must agree on $\varphi_\leftarrow(\mathcal{M}) = \psi_1^\leftarrow(\mathcal{M}) = \psi_2^\leftarrow(\mathcal{M})$, i.e.

$$\psi_1|_{\varphi_\leftarrow(\mathcal{M})} = \psi_2|_{\varphi_\leftarrow(\mathcal{M})} \quad \text{if} \quad \mathcal{M} := (\text{Rng } \psi_1) \cap (\text{Rng } \psi_2).$$

Pitfall: A mapping $\varphi : D \to D'$ may be locally invertible near every point in $D$ without being invertible, even if it is surjective. In fact, if $\psi$ is a local inverse of $\varphi$, $\varphi^\leftarrow(\text{Dom } \psi)$ need not be included in $\text{Rng } \psi$. 

\begin{align*}
\psi \circ \varphi|_{\mathcal{N}'} &= 1_{\mathcal{N}'}, \\
\varphi|_{\mathcal{N}'} \circ \psi &= 1_{\mathcal{N}},
\end{align*}
For example, the function \( f : [-1, 1] \rightarrow [0, 1] \) defined by \( f(s) := |s| \) for all \( s \in [-1, 1] \) is easily seen to be locally invertible, surjective, and even continuous. The identity function \( 1_{[0,1]} \) of \([0,1]\) is a local inverse of \( f \), and we have

\[
f^{-1}(\text{Dom } 1_{[0,1]}) = f^{-1}([0,1]) = [-1, 1] = [0,1].
\]

The function \( h : [0,1] \rightarrow [-1,0] \) defined by \( h(s) = -s \) for all \( s \in [0,1] \) is another local inverse of \( f \).

If \( \dim \mathcal{E} \geq 2 \), one can even give counter-examples of mappings \( \varphi : D \rightarrow D' \) of the type described above for which \( D \) is convex (see Sect.74).

The following two results are recorded for later use.

**Proposition 1:** Assume that \( \varphi : D \rightarrow D' \) is differentiable at \( x \in D \) and locally invertible near \( x \). Then, if some local inverse near \( x \) is differentiable at \( \varphi(x) \), so are all others and all have the same gradient, namely \((\nabla_x \varphi)^{-1}\).

**Proof:** We choose a local inverse \( \psi_1 \) of \( \varphi \) near \( x \) such that \( \psi_1 \) is differentiable at \( \varphi(x) \). Applying the Chain Rule to (68.1) gives

\[
(\nabla_{\varphi(x)} \psi_1)(\nabla_x \varphi) = 1_{\psi_1} \quad \text{and} \quad (\nabla_{\varphi(x)} \psi_1)(\nabla_{\varphi(x)} \psi_1) = 1_{\psi_1},
\]

which shows that \( \nabla_x \varphi \) is invertible and \( \nabla_{\varphi(x)} \psi_1 = (\nabla_x \varphi)^{-1} \).

Let now a local inverse \( \psi_2 \) of \( \varphi \) near \( x \) be given. Let \( \mathcal{M} \) be defined as in (68.2). Since \( \mathcal{M} \) is open and since \( \psi_1 \), being differentiable at \( x \), is continuous at \( x \), it follows that \( \varphi_>( \mathcal{M} ) = \psi_1^{-1}(\mathcal{M}) \) is a neighborhood of \( \varphi(x) \) in \( \mathcal{E}' \). By (68.2), \( \psi_2 \) agrees with \( \psi_1 \) on the neighborhood \( \varphi_>( \mathcal{M} ) \) of \( \varphi(x) \). Hence \( \psi_2 \) must also be differentiable at \( \varphi(x) \), and \( \nabla_{\varphi(x)} \psi_2 = \nabla_{\varphi(x)} \psi_1 = (\nabla_x \varphi)^{-1} \).

**Proposition 2:** Assume that \( \varphi : D \rightarrow D' \) is continuous and that \( \psi \) is a local inverse of \( \varphi \) near \( x \).

(i) If \( \mathcal{M}' \) is an open neighborhood of \( \varphi(x) \) and \( \mathcal{M}' \subset \text{Dom } \psi \), then \( \psi_>( \mathcal{M}' ) = \varphi_<( \mathcal{M}' ) \cap \text{Rng } \psi \) is an open neighborhood of \( x \); hence the adjustment \( \psi|_{\mathcal{M}'(\mathcal{M}') \cap \text{Rng } \psi} \) is again a local inverse of \( \varphi \) near \( x \).

(ii) Let \( \mathcal{G} \) be an open subset of \( \mathcal{E} \) with \( x \in \mathcal{G} \). If \( \psi \) is continuous at \( \varphi(x) \) then there is an open neighborhood \( \mathcal{M} \) of \( x \) with \( \mathcal{M} \subset \text{Rng } \psi \cap \mathcal{G} \) such that \( \varphi_>( \mathcal{M} ) = \mathcal{M} \) is open; hence the adjustment \( \psi|_{\mathcal{M}'(\mathcal{M})} \) is again a local inverse of \( \varphi \) near \( x \).

**Proof:** Part (i) is an immediate consequence of Prop.3 of Sect.56. To prove (ii), we observe first that \( \psi_<( \text{Rng } \psi \cap \mathcal{G} ) \) must be a neighborhood of \( \varphi(x) = \psi^-(x) \). We can choose an open subset \( \mathcal{M}' \) of \( \psi_<( \text{Rng } \psi \cap \mathcal{G} ) \) with
\( \varphi(x) \in \mathcal{M}' \) (see Sect. 53). Application of (i) gives the desired result with \( \mathcal{M} := \psi_>(\mathcal{M}') \).

**Remark:** It is in fact true that if \( \varphi : D \rightarrow D' \) is continuous, then every local inverse of \( \varphi \) is also continuous, but the proof is highly non-trivial and goes beyond the scope of this presentation. Thus, in Part (ii) of Prop. 2, the requirement that \( \psi \) be continuous at \( \varphi(x) \) is, in fact, redundant.

**Pitfall:** The expectation mentioned in the beginning is not justified. A continuous mapping \( \varphi : D \rightarrow D' \) can have an invertible tangent at \( x \in D \) without being locally invertible near \( x \). An example is the function \( f : \mathbb{R} \rightarrow \mathbb{R} \) defined by

\[
f(t) := \begin{cases} 2t^2 \sin \left( \frac{1}{t} \right) + t & \text{if } t \in \mathbb{R}^\times \\ 0 & \text{if } t = 0 \end{cases}
\]

(68.3)

It is differentiable and hence continuous. Since \( f^\bullet(0) = 1 \), the tangent to \( f \) at \( 0 \) is invertible. However, \( f \) is not monotone, let alone locally invertible, near \( 0 \), because one can find numbers \( s \) arbitrarily close to \( 0 \) such that \( f^\bullet(s) < 0 \).

Let \( I \) be an open interval and let \( f : I \rightarrow \mathbb{R} \) be a function of class \( C^1 \). If the tangent to \( f \) at a given \( t \in I \) is invertible, i.e. if \( f^\bullet(t) \neq 0 \), one can easily prove, not only that \( f \) is locally invertible near \( t \), but also that it has a local inverse near \( t \) that is of class \( C^1 \). This result generalizes to mappings \( \varphi : D \rightarrow D' \), but the proof is far from easy.

**Local Inversion Theorem:** Let \( D \) and \( D' \) be open subsets of flat spaces, let \( \varphi : D \rightarrow D' \) be of class \( C^1 \) and let \( x \in D \) be such that the tangent to \( \varphi \) at \( x \) is invertible. Then \( \varphi \) is locally invertible near \( x \) and every local inverse of \( \varphi \) near \( x \) is differentiable at \( \varphi(x) \).

Moreover, there exists a local inverse \( \psi \) of \( \varphi \) near \( x \) that is of class \( C^1 \) and satisfies

\[
\nabla_y \psi = (\nabla_{\psi(y)} \varphi)^{-1} \quad \text{for all } y \in \text{Dom } \psi.
\]

Before proceeding with the proof, we state two important results that are closely related to the Local Inversion Theorem.

**Implicit Mapping Theorem:** Let \( \mathcal{E}, \mathcal{E}', \mathcal{E}'' \) be flat spaces, let \( A \) be an open subset of \( \mathcal{E} \times \mathcal{E}' \) and let \( \omega : A \rightarrow \mathcal{E}'' \) be a mapping of class \( C^1 \). Let \((x_o, y_o) \in A, \ z_o := \omega(x_o, y_o), \) and assume that \( \nabla_{y_o} \omega(x_o, \bullet) \) is invertible. Then there exist an open neighborhood \( D \) of \( x_o \) and a mapping \( \varphi : D \rightarrow \mathcal{E}' \), differentiable at \( x_o \), such that \( \varphi(x_o) = y_o \) and \( \text{Gr}(\varphi) \subset A \), and

\[
\omega(x, \varphi(x)) = z_o \quad \text{for all } x \in D.
\]

Moreover, \( D \) and \( \varphi \) can be chosen such that \( \varphi \) is of class \( C^1 \) and

\[
\nabla \varphi(x) = - (\nabla_{(2)} \omega(x, \varphi(x)))^{-1} \nabla_{(1)} \omega(x, \varphi(x)) \quad \text{for all } x \in D.
\]

(68.5)
Remark: The mapping \( \varphi \) is determined implicitly by an equation in this sense: for any given \( x \in D \), \( \varphi(x) \) is a solution of the equation

\[
\exists y \in \mathcal{E}', \quad \omega(x, y) = z_0. \tag{68.7}
\]

In fact, one can find a neighborhood \( \mathcal{M} \) of \((x_0, y_0)\) in \( \mathcal{E} \times \mathcal{E}' \) such that \( \varphi(x) \) is the only solution of

\[
\exists y \in \mathcal{M}(x, \bullet), \quad \omega(x, y) = z_0, \tag{68.8}
\]

where \( \mathcal{M}(x, \bullet) := \{ y \in \mathcal{E}' \mid (x, y) \in \mathcal{M} \} \). \( \square \)

Differentiation Theorem for InversionMappings: Let \( \mathcal{V} \) and \( \mathcal{V}' \) be linear spaces of equal dimension. Then the set \( \text{Lis}(\mathcal{V}, \mathcal{V}') \) of all linear isomorphisms from \( \mathcal{V} \) onto \( \mathcal{V}' \) is a (non-empty) open subset of \( \text{Lin}(\mathcal{V}, \mathcal{V}') \), the inversion mapping \( \text{inv} : \text{Lis}(\mathcal{V}, \mathcal{V}') \to \text{Lin}(\mathcal{V}', \mathcal{V}) \) defined by \( \text{inv}(L) := L^{-1} \) is of class \( C^1 \), and its gradient is given by

\[
(\nabla_{\text{inv}})M = -L^{-1}ML^{-1} \quad \text{for all} \quad M \in \text{Lin}(\mathcal{V}, \mathcal{V}'). \tag{68.9}
\]

The proof of these three theorems will be given in stages, which will be designated as lemmas. After a basic preliminary lemma, we will prove a weak version of the first theorem and then derive from it weak versions of the other two. The weak version of the last theorem will then be used, like a bootstrap, to prove the final version of the first theorem; the final versions of the other two theorems will follow.

**Lemma 1:** Let \( \mathcal{V} \) be a linear space, let \( \nu \) be a norm on \( \mathcal{V} \), and put \( B := C_{\nu}(\nu) \). Let \( f : B \to \mathcal{V} \) be a mapping with \( f(0) = 0 \) such that \( f - 1_{B\subset\mathcal{V}} \) is constricted with

\[
\kappa := \text{str}(f - 1_{B\subset\mathcal{V}}; \nu, \nu) < 1. \tag{68.10}
\]

Then the adjustment \( f^{(1-\kappa)B} \) (see Sect.03) of \( f \) has an inverse and this inverse is confined near \( 0 \).

**Proof:** We will show that for every \( w \in (1 - \kappa)B \), the equation

\[
\exists z \in B, \quad f(z) = w \tag{68.11}
\]

has a unique solution.

To prove uniqueness, suppose that \( z_1, z_2 \in B \) are solutions of (68.11) for a given \( w \in \mathcal{V} \). We then have

\[
(f - 1_{B\subset\mathcal{V}})(z_1) - (f - 1_{B\subset\mathcal{V}})(z_2) = z_2 - z_1
\]
and hence, by (68.10), $\nu(z_2 - z_1) \leq \kappa \nu(z_2 - z_1)$, which amounts to $(1 - \kappa)\nu(z_2 - z_1) \leq 0$. This is compatible with $\kappa < 1$ only if $\nu(z_1 - z_2) = 0$ and hence $z_1 = z_2$.

To prove existence, we assume that $w \in (1 - \kappa)\mathcal{B}$ is given and we put $\alpha := \frac{\nu(w)}{1 - \kappa}$, so that $\alpha \in [0, 1]$. For every $v \in \alpha \mathcal{B}$, we have $\nu(v) \leq \alpha$ and hence, by (68.10)

$$\nu((f - 1_{\mathcal{B} \cap \mathcal{Y}})(v)) \leq \kappa \nu(v) \leq \kappa \alpha,$$

which implies that

$$\nu(w - (f(v) - v)) \leq \nu(w) + \nu((f - 1_{\mathcal{B} \cap \mathcal{Y}})(v)) \leq (1 - \kappa)\alpha + \kappa \alpha = \alpha.$$

Therefore, it makes sense to define $h_w : \alpha \mathcal{B} \to \alpha \mathcal{B}$ by

$$h_w(v) := w - (f(v) - v) \quad \text{for all} \quad v \in \alpha \mathcal{B}. \quad (68.12)$$

Since $h_w$ is the difference of a constant with value $w$ and a restriction of $f - 1_{\mathcal{B} \cap \mathcal{Y}}$, it is constricted and has an absolute striction no greater than $\kappa$. Therefore, it is a contraction, and since its domain $\alpha \mathcal{B}$ is closed, the Contraction Fixed Point Theorem states that $h_w$ has a fixed point $z \in \alpha \mathcal{B}$.

It is evident from (68.12) that a fixed point of $h_w$ is a solution of (68.11).

We proved that (68.11) has a unique solution $z \in \mathcal{B}$ and that this solution satisfies

$$\nu(z) \leq \alpha := \frac{\nu(w)}{1 - \kappa}. \quad (68.13)$$

If we define $g : (1 - \kappa)\mathcal{B} \to f^\circ((1 - \kappa)\mathcal{B})$ in such a way that $g(w)$ is this solution of (68.11), then $g$ is the inverse we are seeking. It follows from (68.13) that $\nu(g(w)) \leq \frac{1}{1 - \kappa} \nu(w)$ for all $w \in (1 - \kappa)\mathcal{B}$. In view of part (ii) of Prop.2 of Sect.62, this proves that $g$ is confined. 

Lemma 2: The assertion of the Local Inversion Theorem, except possibly the statement starting with “Moreover”, is valid.

Proof: Let $\alpha : \mathcal{E} \to \mathcal{E}'$ be the (invertible) tangent to $\varphi$ at $x$. Since $\mathcal{D}$ is open, we may choose a norming cell $\mathcal{B}$ such that $x + \mathcal{B} \subset \mathcal{D}$. We define $f : \mathcal{B} \to \mathcal{Y}$ by

$$f := (\alpha^- |_{\mathcal{D}'}) \circ \varphi |_{x + \mathcal{B}} \circ (x + 1_{\mathcal{Y}}) |_{\mathcal{B}^c} - x. \quad (68.14)$$

Since $\varphi$ is of class $\mathcal{C}^1$, so is $f$. Clearly, we have

$$f(0) = 0, \quad \nabla_0 f = 1_{\mathcal{Y}}. \quad (68.15)$$
Put \( \nu := \nu_0 \), so that \( B = C e(\nu) \). Choose \( \varepsilon \in ]0, 1[ \) \( (\varepsilon := \frac{1}{2} \) would do). If we apply Prop. 4 of Sect. 64 to the case when \( \phi \) there is replaced by \( f \) and \( k \) by \( \{0\} \), we see that there is \( \delta \in ]0, 1[ \) such that
\[
\kappa := \text{str}(f|_B - 1_{\delta B \subset V}; \nu, \nu) \leq \varepsilon. \tag{68.16}
\]

Now, it is clear from (64.2) that a striction relative to \( \nu \) and \( \nu' \) remains unchanged if \( \nu \) and \( \nu' \) are multiplied by the same strictly positive factor. Hence, (68.16) remains valid if \( \nu \) there is replaced by \( \frac{1}{\delta} \nu \). Since \( C e(\frac{1}{\delta} \nu) = \delta C e(\nu) = \delta B \) (see Prop. 6 of Sect. 51), it follows from (68.16) that Lemma 1 can be applied when \( \nu, B, \) and \( f \) there are replaced by \( \frac{1}{\delta} \nu, \delta B \) and \( f|_\delta B \), respectively. We infer that if we put
\[
M_0 := (1 - \kappa)\delta B, \quad N_0 := f^\prec(M_o),
\]
then \( f|_{M_o}^{N_o} \) has an inverse \( g : M_o \rightarrow N_o \) that is confined near zero. Note that \( M_o \) and \( N_o \) are both open neighborhoods of zero, the latter because it is the pre-image of an open set under a continuous mapping. We evidently have
\[
g|_V - 1_{M_o \subset V} = (1_{N_o \subset V} - f|_{N_o}^\prec) \circ g.
\]
In view of (68.15), \( 1_{N_o \subset V} - f|_{N_o}^\prec \) is small near \( 0 \in V \). Since \( g \) is confined near \( 0 \), it follows by Prop. 3 of Sect. 62 that \( g|_V - 1_{M_o \subset V} \) is small near zero. By the Characterization of Gradients of Sect. 63, we conclude that \( g \) is differentiable at \( 0 \) with \( D_0 g = 1_V \).

We now define
\[
N := x + N_0, \quad N' := \phi(x) + (\nabla \alpha)^> (M_o).
\]
These are open neighborhoods of \( x \) and \( \phi(x) \), respectively. A simple calculation, based on (68.14) and \( g = (f|_{M_o}^{N_o})^{-1} \), shows that
\[
\psi := (x + 1_V)|_{N_o} \circ g \circ (a^-- x)|_{N'_o}^\prec \tag{68.17}
\]
is the inverse of \( \phi|_{N'_o} \). Using the Chain Rule and the fact that \( g \) is differentiable at \( 0 \), we conclude from (68.17) that \( \psi \) is differentiable at \( \phi(x) \).

**Lemma 3:** The assertion of the Implicit Mapping Theorem, except possibly the statement starting with “Moreover”, is valid.

**Proof:** We define \( \tilde{\omega} : A \rightarrow \mathcal{E} \times \mathcal{E}'' \) by
\[
\tilde{\omega}(x, y) := (x, \omega(x, y)) \quad \text{for all} \quad (x, y) \in A. \tag{68.18}
\]
Of course, \( \tilde{\omega} \) is of class \( C^1 \). Using Prop. 2 of Sect. 63 and (65.4), we see that
\[
(\nabla_{(x_o,y_o)} \tilde{\omega})(u,v) = (u, (\nabla_{x_o} \omega(\cdot,y_o))u + (\nabla_{y_o} \omega(x_o,\cdot))v)
\]
for all \((u,v) \in \mathcal{V} \times \mathcal{V}'\). Since \( \nabla_{y_o} \omega(x_o,\cdot) \) is invertible, so is \( \nabla_{(x_o,y_o)} \tilde{\omega} \). Indeed, we have
\[
(\nabla_{(x_o,y_o)} \tilde{\omega})^{-1}(u,w) = (u,(\nabla_{y_o} \omega(x_o,\cdot))^{-1}(w - (\nabla_{x_o} \omega(\cdot,y_o))u))
\]
for all \((u,w) \in \mathcal{V} \times \mathcal{V}''\). Lemma 2 applies and we conclude that \( \tilde{\omega} \) has a local inverse \( \rho \) with \((x_o,y_o) \in \text{Rng} \rho \) that is differentiable at \( \tilde{\omega}(x_o,y_o) = (x_o,z_o) \). In view of Prop. 2, (i), we may assume that the domain of \( \rho \) is of class \( C^1 \). Let \( D \times D \) be linear spaces of equal dimension. Then, by (68.18),
\[
\tilde{\omega}(\rho(x,y)) = (\psi(x,z),\omega(\psi(x,z),\psi'(x,z))) = (x,z)
\]
and hence
\[
\psi(x,z) = x, \quad \omega(x,\psi'(x,z)) = z \quad \text{for all} \quad x \in \mathcal{D}, \quad z \in \mathcal{D}''. \tag{68.19}
\]
Since \( \rho(x_o,z_o) = (\psi(x_o,z_o),\psi'(x_o,z_o)) = (x_o,y_o) \), we have \( \psi'(x_o,z_o) = y_o \). Therefore, if we define \( \varphi : \mathcal{D} \to \mathcal{E}'' \) by \( \varphi(x) := \psi'(x,z_o) \), we see that \( \varphi(x_o) = y_o \) and (68.5) are satisfied. The differentiability of \( \varphi \) at \( x_o \) follows from the differentiability of \( \rho \) at \((x_o,z_o)\).

If we define \( \mathcal{M} := \text{Rng} \rho \), it is immediate that \( \varphi(x) \) is the only solution of (68.8). \( \blacksquare \)

**Lemma 4:** Let \( \mathcal{V} \) and \( \mathcal{V}' \) be linear spaces of equal dimension. Then \( \operatorname{Lis}(\mathcal{V}, \mathcal{V}') \) is an open subset of \( \operatorname{Lin}(\mathcal{V}, \mathcal{V}') \) and the inversion mapping \( \operatorname{inv} : \operatorname{Lis}(\mathcal{V}, \mathcal{V}') \to \operatorname{Lin}(\mathcal{V}', \mathcal{V}) \) defined by \( \operatorname{inv}(\mathcal{L}) := \mathcal{L}^{-1} \) is differentiable.

**Proof:** We apply Lemma 3 with \( \mathcal{E} \) replaced by \( \operatorname{Lin}(\mathcal{V}, \mathcal{V}') \), \( \mathcal{E}' \) by \( \operatorname{Lin}(\mathcal{V}', \mathcal{V}) \), and \( \mathcal{E}'' \) by \( \operatorname{Lin}\mathcal{V} \). For \( \omega \) of Lemma 3 we take the mapping \((\mathcal{L}, \mathcal{M}) \mapsto \mathcal{M}\mathcal{L}\) from \( \operatorname{Lin}(\mathcal{V}, \mathcal{V}') \times \operatorname{Lin}(\mathcal{V}', \mathcal{V}) \) to \( \operatorname{Lin}\mathcal{V} \). Being bilinear, this mapping is of class \( C^1 \) (see Prop. 1 of Sect. 66). Its partial 2-gradient at \((\mathcal{L}, \mathcal{M})\) does not depend on \( \mathcal{M} \). It is the right-multiplication \( \operatorname{Ri}_\mathcal{L} \in \operatorname{Lin}(\operatorname{Lin}(\mathcal{V}', \mathcal{V}), \operatorname{Lin}\mathcal{V}) \) defined by \( \operatorname{Ri}_\mathcal{L}(\mathcal{K}) := \mathcal{K}\mathcal{L} \) for all \( \mathcal{K} \in \operatorname{Lin}(\mathcal{V}, \mathcal{V}') \). It is clear that \( \operatorname{Ri}_\mathcal{L} \) is invertible if and only if \( \mathcal{L} \) is invertible. (We have \( (\operatorname{Ri}_\mathcal{L})^{-1} = \operatorname{Ri}_{\mathcal{L}^{-1}} \) if this is the case.) Thus, given \( \mathcal{L}_o \in \operatorname{Lis}(\mathcal{V}, \mathcal{V}') \), we can
apply Lemma 3 with $L_0, L_0^{-1}$, and $1_V$ playing the roles of $x_0, y_0,$ and $z_0$, respectively. We conclude that there is an open neighborhood $D$ of $L_0$ in $\text{Lin}(\mathcal{V}, \mathcal{V}')$ such that the equation

$$?M \in \text{Lin}(\mathcal{V}', \mathcal{V}), \quad LM = 1_V$$

has a solution for every $L \in D$. Since this solution can only be $L^{-1} = \text{inv}(L)$, it follows that the role of the mapping $\varphi$ of Lemma 3 is played by $\text{inv}|_{D}$. Therefore, we must have $D \subset \text{Lis}(\mathcal{V}, \mathcal{V}')$ and $\text{inv}$ must be differentiable at $L_0$. Since $L_0 \in \text{Lis}(\mathcal{V}, \mathcal{V}')$ was arbitrary, it follows that $\text{Lis}(\mathcal{V}, \mathcal{V}')$ is open and that $\text{inv}$ is differentiable.

Completion of Proofs: We assume that hypotheses of the Local Inversion Theorem are satisfied. The assumption that $\varphi$ has an invertible tangent at $x \in D$ is equivalent to $\nabla_x \varphi \in \text{Lis}(\mathcal{V}, \mathcal{V}')$. Since $\nabla \varphi$ is continuous and since $\text{Lis}(\mathcal{V}, \mathcal{V}')$ is an open subset of $\text{Lin}(\mathcal{V}, \mathcal{V}')$ by Lemma 4, it follows that $\mathcal{G} := (\nabla \varphi)^{-1}(\text{Lis}(\mathcal{V}, \mathcal{V}'))$ is an open subset of $\mathcal{E}$. By Lemma 2, we can choose a local inverse of $\varphi$ near $x$ which is differentiable and hence continuous at $\varphi(x)$. By Prop. 2, (ii), we can adjust this local inverse such that its range is included in $\mathcal{G}$. Let $\psi$ be the local inverse so adjusted. Then $\varphi$ has an invertible tangent at every $z \in \text{Rng} \psi$. Let $z \in \text{Rng} \psi$ be given. Applying Lemma 2 to $z$, we see that $\varphi$ must have a local inverse near $z$ that is differentiable at $\varphi(z)$. By Prop. 1, it follows that $\psi$ must also be differentiable at $\varphi(z)$. Since $z \in \text{Rng} \psi$ was arbitrary, it follows that $\psi$ is differentiable. The formula (68.4) follows from Prop. 1. Since $\text{inv} : \text{Lis}(\mathcal{V}, \mathcal{V}') \to \text{Lin}(\mathcal{V}', \mathcal{V})$ is differentiable and hence continuous by Lemma 4, since $\nabla \varphi$ is continuous by assumption, and since $\psi$ is continuous because it is differentiable, it follows that the composite $\text{inv} \circ \nabla \varphi^{-1}_{\text{Rng} \psi} \circ \psi$ is continuous. By (68.4), this composite is none other than $\nabla \psi$, and hence the proof of the Local Inversion Theorem is complete.

Assume, now, that the hypotheses of the Implicit Mapping Theorem are satisfied. By the Local Inversion Theorem, we can then choose a local inverse $\rho$ of $\varphi$, as defined by (68.18), that is of class $C^1$. It then follows that the function $\varphi : D \to \mathcal{E}'$ defined in the proof of Lemma 3 is also of class $C^1$. The formula (68.6) is obtained easily by differentiating the constant $x \mapsto \omega(x, \varphi(x))$, using Prop. 1 of Sect. 65 and the Chain Rule.

To prove the Differentiation Theorem for Inversion Mappings, one applies the Implicit Mapping Theorem in the same way as Lemma 3 was applied to obtain Lemma 4. ■

Pitfall: Let $I$ and $I'$ be intervals and let $f : I \to I'$ be of class $C^1$ and surjective. If $f$ has an invertible tangent at every $t \in I$, then $f$ is not only
locally invertible near every \( t \in I \) but (globally) invertible, as is easily seen. This conclusion does not generalize to the case when \( I \) and \( I' \) are replaced by connected open subsets of flat spaces of higher dimension. The curvilinear coordinate systems discussed in Sect.74 give rise to counterexamples. One can even give counterexamples in which \( I \) and \( I' \) are replaced by convex open subsets of flat spaces.

Notes 68

(1) The Local Inversion Theorem is often called the “Inverse Function Theorem”. Our term is more descriptive.

(2) The Implicit Mapping Theorem is most often called the “Implicit Function Theorem”.

(3) If the sets \( D \) and \( D' \) of the Local Inversion Theorem are both subsets of \( \mathbb{R}^n \) for some \( n \in \mathbb{N} \), then \( \nabla_x \varphi \) can be identified with an \( n \)-by-\( n \) matrix (see (65.12)). This matrix is often called the “Jacobian matrix” and its determinant the “Jacobian” of \( \varphi \) at \( x \). Some textbook authors replace the condition that \( \nabla_x \varphi \) be invertible by the condition that the Jacobian be non-zero. I believe it is a red herring to drag in determinants here. The Local Inversion Theorem has an extension to infinite-dimensional spaces, where it makes no sense to talk about determinants.

69 Extreme Values, Constraints

In this section \( E \) and \( F \) denote flat spaces with translation spaces \( V \) and \( W \), respectively, and \( D \) denotes a subset of \( E \).

The following definition is “local” variant of the definition of an extremum given in Sect.08.

**Definition 1:** We say that a function \( f : D \to \mathbb{R} \) attains a local maximum [local minimum] at \( x \in D \) if there is \( N \in \text{Nhd}_x(D) \) (see (56.1)) such that \( f|_N \) attains a maximum [minimum] at \( x \). We say that \( f \) attains a local extremum at \( x \) if it attains a local maximum or local minimum at \( x \).

The following result is a direct generalization of the Extremum Theorem of elementary calculus (see Sect.08).

**Extremum Theorem:** If \( f : D \to \mathbb{R} \) attains a local extremum at \( x \in \text{Int}D \) and if \( f \) is differentiable at \( x \), then \( \nabla_x f = 0 \).

**Proof:** Let \( v \in V \) be given. It is clear that \( S_{x,v} := \{ s \in \mathbb{R} \mid x + sv \in \text{Int}D \} \) is an open subset of \( \mathbb{R} \) and the function \( (s \mapsto f(x + sv)) : S_{x,v} \to \mathbb{R} \) attains a local extremum at \( 0 \in \mathbb{R} \) and is differentiable at \( 0 \in \mathbb{R} \). Hence, by
CHAPTER 6. DIFFERENTIAL CALCULUS

the Extremum Theorem of elementary calculus, and by (65.13) and (65.14), we have

\[ 0 = \partial_0 (s \mapsto f(x + sv)) = (ddf)(x) = (\nabla_x f)v. \]

Since \( v \in V \) was arbitrary, the desired result \( \nabla_x f = 0 \) follows. □

From now on we assume that \( D \) is an open subset of \( E \).

The next result deals with the case when a restriction of \( f : D \to \mathbb{R} \) to a suitable subset of \( D \), but not necessarily \( f \) itself, has an extremum at a given point \( x \in D \). This subset of \( D \) is assumed by the set on which a given mapping \( \varphi : D \to F \) has a constant value, i.e. the set \( \varphi^{-1}(\{\varphi(x)\}) \). The assertion that \( f|_{\varphi^{-1}(\{\varphi(x)\})} \) has an extremum at \( x \) is often expressed by saying that \( f \) attains an extremum at \( x \) subject to the constraint that \( \varphi \) be constant.

**Constrained-Extremum Theorem:** Assume that \( f : D \to \mathbb{R} \) and \( \varphi : D \to F \) are both of class \( C^1 \) and that \( \nabla_x \varphi \in \text{Lin}(V, W) \) is surjective for a given \( x \in D \). If \( f|_{\varphi^{-1}(\{\varphi(x)\})} \) attains a local extremum at \( x \), then \( \nabla_x f \in \text{Rng}(\nabla_x \varphi)^\top \). (69.1)

**Proof:** We put \( U := \text{Null} \nabla_x \varphi \) and choose a supplement \( Z \) of \( U \) in \( V \). Every neighborhood of \( 0 \) in \( V \) includes a set of the form \( N + M \), where \( N \) is an open neighborhood of \( 0 \) in \( U \) and \( M \) is an open neighborhood of \( 0 \) in \( Z \). Since \( D \) is open, we may select \( N \) and \( M \) such that \( x + N + M \subset D \).

We now define \( \omega : N \times M \to F \) by

\[ \omega(u, z) := \varphi(x + u + z) \quad \text{for all} \quad u \in N, z \in M. \]  

It is clear that \( \omega \) is of class \( C^1 \) and we have

\[ \nabla_0 \omega(0, \cdot) = \nabla_x \varphi|_Z. \]  

(69.3)

Since \( \nabla_x \varphi \) is surjective, it follows from Prop. 5 of Sect. 13 that \( \nabla_x \varphi|_Z \) is invertible. Hence we can apply the Implicit Mapping Theorem of Sect. 68 to the case when \( A \) there is replaced by \( N \times M \), the points \( x_0, y_0 \) by \( 0 \), and \( z_0 \) by \( \varphi(x) \). We obtain an open neighborhood \( G \) of \( 0 \) in \( U \) with \( G \subset N \) and a mapping \( h : G \to Z \), of class \( C^1 \), such that \( h(0) = 0 \) and

\[ \omega(u, h(u)) = \varphi(x) \quad \text{for all} \quad u \in G. \]  

(69.4)

Since \( \nabla(1) \omega(0, 0) = \nabla_0 \omega(\cdot, 0) = \nabla_x \varphi|_U = 0 \) by the definition \( U := \text{Null} \nabla_x \varphi \), the formula (68.6) yields in our case that \( \nabla_0 h = 0 \), i.e. we have

\[ h(0) = 0, \quad \nabla_0 h = 0. \]  

(69.5)
It follows from (69.2) and (69.4) that $x + u + h(u) \in \varphi^<\{(\varphi(x))\}$ for all $u \in \mathcal{G}$ and hence that

$$(u \mapsto f(x + u + h(u))) : \mathcal{G} \to \mathbb{R}$$

has a local extremum at 0. By the Extremum Theorem and the Chain Rule, it follows that

$$0 = \nabla_0 (u \mapsto f(x + u + h(u))) = \nabla_x f(1_{U \subset V} + \nabla_0 h|_{V})$$

and therefore, by (69.5), that $\nabla_x f|_U = (\nabla_x f)1_{U \subset V} = 0$, which means that $\nabla_x f \in U^\perp = \text{Null } \nabla_x \varphi^\perp$. The desired result (69.1) is obtained by applying (22.9).

In the case when $F := \mathbb{R}$, the Constrained-Extremum Theorem reduces to

**Corollary 1:** Assume that $f, g : D \to \mathbb{R}$ are both of class $C^1$ and that $\nabla_x g \neq 0$ for a given $x \in D$. If $f|_{g^<(\{g(x)\})}$ attains a local extremum at $x$, then there is $\lambda \in \mathbb{R}$ such that

$$\nabla_x f = \lambda \nabla_x g. \quad (69.6)$$

In the case when $F := \mathbb{R}^I$, we can use (23.19) and obtain

**Corollary 2:** Assume that $f : D \to \mathbb{R}$ and all terms $g_i : D \to \mathbb{R}$ in a finite family $g := (g_i \mid i \in I) : D \to \mathbb{R}^I$ are of class $C^1$ and that $(\nabla_x g_i \mid i \in I)$ is linearly independent for a given $x \in D$. If $f|_{g^<(\{g(x)\})}$ attains a local extremum at $x$, then there is $\lambda \in \mathbb{R}^I$ such that

$$\nabla_x f = \sum_{i \in I} \lambda_i \nabla_x g_i. \quad (69.7)$$

**Remark 1:** If $E$ is two-dimensional then Cor.1 can be given a geometrical interpretation as follows: The sets $\mathcal{L}_g := g^<\{(g(x))\}$ and $\mathcal{L}_f := f^<\{(f(x))\}$ are the “level-lines” through $x$ of $f$ and $g$, respectively (see Figure).
If these lines cross at \( x \), then the value \( f(y) \) of \( f \) at \( y \in L_g \) strictly increases or decreases as \( y \) moves along \( L_g \) from one side of the line \( L_f \) to the other and hence \( f|_{L_g} \) cannot have an extremum at \( x \). Hence, if \( f|_{L_g} \) has an extremum at \( x \), the level-lines must be tangent. The assertion (69.6) expresses this tangency. The condition \( \nabla x g \neq 0 \) insures that the level-line \( L_g \) does not degenerate to a point. 

Notes 69

(1) The number \( \lambda \) of (69.6) or the terms \( \lambda_i \) occurring in (69.7) are often called “Lagrange multipliers”.

610 Integral Representations

Let \( I \in \text{Sub} \mathbb{R} \) be some genuine interval and let \( h : I \rightarrow V \) be a continuous process with values in a given linear space \( V \). For every \( \lambda \in V^* \) the composite \( \lambda h : I \rightarrow \mathbb{R} \) is then continuous. Given \( a, b \in I \) one can therefore form the integral \( \int_a^b \lambda h \) in the sense of elementary integral calculus (see Sect.08). It is clear that the mapping

\[
(\lambda \mapsto \int_a^b \lambda h) : V^* \rightarrow \mathbb{R}
\]

is linear and hence an element of \( V^{**} \cong V \).

**Definition 1:** Let \( h : I \rightarrow V \) be a continuous process with values in the linear space \( V \). Given any \( a, b \in I \) the integral of \( h \) from \( a \) to \( b \) is defined to be the unique element \( \int_a^b h \) of \( V \) which satisfies

\[
\lambda \int_a^b h = \int_a^b \lambda h \quad \text{for all} \quad \lambda \in V^*.
\]

(610.1)

**Proposition 1:** Let \( h : I \rightarrow V \) be a continuous process and let \( L \) be a linear mapping from \( V \) to a given linear space \( W \). Then \( Lh : I \rightarrow W \) is a continuous process and

\[
L \int_a^b h = \int_a^b Lh \quad \text{for all} \quad a, b \in I.
\]

(610.2)

**Proof:** Let \( \omega \in W^* \) and \( a, b \in I \) be given. Using Def.1 twice, we obtain

\[
\omega(\int_a^b h) = (\omega L) \int_a^b h = \int_a^b (\omega L)h = \int_a^b \omega(Lh) = \omega \int_a^b Lh.
\]
610. INTEGRAL REPRESENTATIONS

Since \( \omega \in W^* \) was arbitrary, (610.2) follows.

**Proposition 2:** Let \( \nu \) be a norm on the linear space \( V \) and let \( h : I \to V \) be a continuous process. Then \( \nu \circ h : I \to \mathbb{R} \) is continuous and, for every \( a, b \in I \), we have

\[
\nu \left( \int_a^b h \right) \leq \left| \int_a^b \nu \circ h \right|.
\]  

(610.3)

**Proof:** The continuity of \( \nu \circ h \) follows from the continuity of \( \nu \) (Prop. 7 of Sect. 56) and the Composition Theorem for Continuity of Sect. 56.

It follows from (52.14) that

\[
|\lambda h(t)| = |\lambda h(t)| \leq \nu^* (\lambda) \nu(h(t)) = \nu^*(\lambda)(\nu \circ h)(t)
\]

holds for all \( \lambda \in V^* \) and all \( t \in I \), and hence that

\[
|\lambda h| \leq \nu \circ h \quad \text{for all} \quad \lambda \in \text{Ce}(\nu^*).
\]

Using this result and (610.1), (08.40), and (08.42), we obtain

\[
\left| \lambda \int_a^b h \right| = \left| \int_a^b \lambda h \right| \leq \left| \int_a^b |\lambda h| \right| \leq \left| \int_a^b \nu \circ h \right|
\]

for all \( \lambda \in \text{Ce}(\nu^*) \) and all \( a, b \in I \). The desired result now follows from the Norm-Duality Theorem, i.e. (52.15).

Most of the rules of elementary integral calculus extend directly to integrals of processes with values in a linear space. For example, if \( h : I \to V \) is continuous and if \( a, b, c \in I \), then

\[
\int_a^b h = \int_a^c h + \int_c^b h.
\]  

(610.4)

The following result is another example.

**Fundamental Theorem of Calculus:** Let \( \mathcal{E} \) be a flat space with translation space \( V \). If \( h : I \to \mathcal{V} \) is a continuous process and if \( x \in \mathcal{E} \) and \( a \in I \) are given, then the process \( p : I \to \mathcal{E} \) defined by

\[
p(t) := x + \int_a^t h \quad \text{for all} \quad t \in I
\]

(610.5)

is of class \( C^1 \) and \( p^* = h \).

**Proof:** Let \( \lambda \in V^* \) be given. By (610.5) and (610.1) we have

\[
\lambda(p(t) - x) = \lambda \int_a^t h = \int_a^t \lambda h \quad \text{for all} \quad t \in I.
\]
256  CHAPTER 6. DIFFERENTIAL CALCULUS

By Prop.1 of Sect.61 \( p \) is differentiable and we have \( (\lambda(p - x))^* = \lambda^p^* \). Hence, the elementary Fundamental Theorem of Calculus (see Sect.08) gives \( \lambda^p^* = \lambda h \). Since \( \lambda \in \mathcal{V}^* \) was arbitrary, we conclude that \( p^* = h \), which is continuous by assumption. 

In the remainder of this section, we assume that the following items are given: (i) a flat space \( \mathcal{E} \) with translation space \( \mathcal{V} \), (ii) An open subset \( \mathcal{D} \) of \( \mathcal{E} \), (iii) An open subset \( \mathcal{D} \times \mathcal{E} \) such that, for each \( x \in \mathcal{D} \), \( \mathcal{D}(x) := \{s \in \mathbb{R} | (s, x) \in \mathcal{D}\} \) is a non-empty open interval, (iv) a linear space \( \mathcal{W} \).

Consider now a mapping \( h : \mathcal{D} \to \mathcal{W} \) such that \( h(\cdot, x) : \mathcal{D}(x) \to \mathcal{W} \) is continuous for all \( x \in \mathcal{D} \). Given \( a, b \in \mathbb{R} \) such that \( a, b \in \mathcal{D}(x) \) for all \( x \in \mathcal{D} \), we can then define \( k : \mathcal{D} \to \mathcal{W} \) by

\[
k(x) := \int_a^b h(\cdot, x) \text{ for all } x \in \mathcal{D}.
\]

We say that \( k \) is defined by an integral representation.

**Proposition 3:** If \( h : \mathcal{D} \to \mathcal{W} \) is continuous, so is the mapping \( k : \mathcal{D} \to \mathcal{W} \) defined by the integral representation \((610.6)\).

**Proof:** Let \( x \in \mathcal{D} \) be given. Since \( \mathcal{D} \) is open, we can choose a norm \( \nu \) on \( \mathcal{V} \) such that \( x + \mathcal{C}e(\nu) \subset \mathcal{D} \). We may assume, without loss of generality, that \( a < b \). Then \( [a, b] \) is a compact interval and hence, in view of Prop.4 of Sect.58, \( [a, b] \times \mathcal{C}e(\nu) \) is a compact subset of \( \mathcal{D} \). By the Uniform Continuity Theorem of Sect. 58, it follows that the restriction of \( h \) to \( [a, b] \times \mathcal{C}e(\nu) \) is uniformly continuous. Let \( \mu \) be a norm on \( \mathcal{W} \) and let \( \varepsilon \in \mathbb{P}^x \) be given. By Prop.4 of Sect.56, we can determine \( \delta \in [0, 1] \) such that

\[
\mu(h(s, y) - h(s, x)) < \frac{\varepsilon}{b - a}
\]

for all \( s \in [a, b] \) and all \( y \in x + \delta \mathcal{C}e(\nu) \). Hence, by \((610.6)\) and Prop.2, we have

\[
\mu(k(y) - k(x)) = \mu(\int_a^b (h(\cdot, y) - h(\cdot, x)) \, \nu) \leq \int_a^b \mu(h(\cdot, y) - h(\cdot, x)) \leq (b - a) \frac{\varepsilon}{b - a} = \varepsilon
\]

whenever \( \nu(y - x) < \delta \). Since \( \varepsilon \in \mathbb{P}^x \) was arbitrary, the continuity of \( k \) at \( x \) follows by Prop.1 of Sect.56. Since \( x \in \mathcal{D} \) was arbitrary, the assertion follows. 

The following is a stronger version of Prop. 3.

**Proposition 4:** Let \( \overline{D} \subset \mathbb{R} \times \mathbb{R} \times E \) be defined by
\[
\overline{D} := \{(a, b, x) \mid x \in D, (a, x) \in \overline{D}, (b, x) \in \overline{D}\}. \tag{610.7}
\]

Then, if \( h : \overline{D} \rightarrow W \) is continuous, so is the mapping \( (a, b, x) \mapsto \int_a^b h(\cdot, x) : \overline{D} \rightarrow W \).

**Proof:** Choose a norm \( \mu \) on \( W \). Let \( (a, b, x) \in \overline{D} \) be given. Since \( \mu \circ h \) is continuous at \( (a, x) \) and at \( (b, x) \) and since \( \overline{D} \) is open, we can find \( N \in \text{Nhd}_x(\overline{D}) \) and \( \sigma \in \mathbb{P}^x \) such that \( N := ((a+\sigma, \sigma] \cup (b+\sigma, \sigma]) \times N \) is a subset of \( \overline{D} \) and such that the restriction of \( \mu \circ h \) to \( N \) is bounded by some \( \beta \in \mathbb{P}^x \). By Prop. 2, it follows that
\[
\mu \left( \int_s^a h(\cdot, y) + \int_t^b h(\cdot, y) \right) \leq (|s - a| + |t - b|) \beta \tag{610.8}
\]

for all \( y \in N \), all \( s \in a + \sigma, \sigma \), and all \( t \in b + \sigma, \sigma \).

Now let \( \varepsilon \in \mathbb{P}^x \) be given. If we put \( \delta := \min\{\sigma, \frac{\varepsilon}{4|\beta|} \} \), it follows from (610.8) that
\[
\mu \left( \int_s^a h(\cdot, y) + \int_t^b h(\cdot, y) \right) < \frac{\varepsilon}{2} \tag{610.9}
\]

for all \( y \in N \), all \( s \in a + \delta \), and all \( t \in b + \delta \). On the other hand, by Prop. 3, we can determine \( M \in \text{Nhd}_x(\overline{D}) \) such that
\[
\mu \left( \int_a^b h(\cdot, y) - \int_a^b h(\cdot, x) \right) < \frac{\varepsilon}{2} \tag{610.10}
\]

for all \( y \in M \). By (610.4) we have
\[
\int_s^t h(\cdot, y) - \int_a^b h(\cdot, x) = \int_a^b h(\cdot, y) - \int_a^b h(\cdot, x) + \int_s^a h(\cdot, y) + \int_b^t h(\cdot, y).
\]

Hence it follows from (610.9) and (610.10) that
\[
\mu \left( \int_s^t h(\cdot, y) - \int_a^b h(\cdot, x) \right) < \varepsilon
\]
for all \( y \in \mathcal{M} \cap \mathcal{N} \in \text{Nhd}_x(D) \), all \( s \in a + [-\delta, \delta] \), and all \( t \in b + [-\delta, \delta] \). Since \( \varepsilon \in \mathbb{P}^\times \) was arbitrary, it follows from Prop.1 of Sect.56 that the mapping under consideration is continuous at \((a, b, x)\).

**Differentiation Theorem for Integral Representations:** Assume that \( h : D \to W \) satisfies the following conditions:

1. \( h(s, \bullet, x) \) is continuous for every \( x \in D \).

2. \( h(s, \bullet) \) is differentiable at \( x \) for all \((s, x) \in D\) and the partial 2-gradient \( \nabla_{(2)} h : D \to \text{Lin}(\mathcal{V}, \mathcal{W}) \) is continuous.

Then the mapping \( k : D \to W \) defined by the integral representation (610.6) is of class \( C^1 \) and its gradient is given by the integral representation

\[
\nabla_x k = \int_a^b \nabla_{(2)} h(s, \bullet, x) \text{ for all } x \in D. \tag{610.11}
\]

Roughly, this theorem states that if \( h \) satisfies the conditions (i) and (ii), one can differentiate (610.6) with respect to \( x \) by “differentiating under the integral sign”.

**Proof:** Let \( x \in D \) be given. As in the proof of Prop.3, we can choose a norm \( \nu \) on \( \mathcal{V} \) such that \( x + \delta \text{Ce}(\nu) \subset D \), and we may assume that \( a < b \). Then \([a, b] \times (x + \delta \text{Ce}(\nu))\) is a compact subset of \( \overline{D} \). By the Uniform Continuity Theorem of Sect.58, the restriction of \( \nabla_{(2)} h \) to \([a, b] \times (x + \delta \text{Ce}(\nu))\) is uniformly continuous. Hence, if a norm \( \mu \) on \( \mathcal{W} \) and \( \varepsilon \in \mathbb{P}^\times \) are given, we can determine \( \delta \in [0, 1] \) such that

\[
\|\nabla_{(2)} h(s, x + u) - \nabla_{(2)} h(s, x)\|_{\nu, \mu} < \frac{\varepsilon}{b-a} \tag{610.12}
\]

holds for all \( s \in [a, b] \) and \( u \in \delta \text{Ce}(\nu) \).

We now define \( n : (\overline{D} - (0, x)) \to W \) by

\[
n(s, u) := h(s, x + u) - h(s, x) - (\nabla_{(2)} h(s, x)) u. \tag{610.13}
\]

It is clear that \( \nabla_{(2)} n \) exists and is given by

\[
\nabla_{(2)} n(s, u) = \nabla_{(2)} h(s, x + u) - \nabla_{(2)} h(s, x).
\]

By (610.12), we hence have

\[
\|\nabla_{(2)} n(s, u)\|_{\nu, \mu} < \frac{\varepsilon}{b-a}
\]
for all $s \in [a, b]$ and $u \in \delta \text{Ce}(\nu)$. By the Striction Estimate for Differentiable Mapping of Sect.64, it follows that $n(s, \bullet)_{|\delta \text{Ce}(\nu)}$ is constricted for all $s \in [a, b]$ and that

$$\text{str}(n(s, \bullet)_{|\delta \text{Ce}(\nu)} : \nu, \mu) \leq \frac{\varepsilon}{b - a} \text{ for all } s \in [a, b].$$

Since $n(s, 0) = 0$ for all $s \in [a, b]$, the definition (64.1) shows that

$$\mu(n(s, u)) \leq \frac{\varepsilon}{b - a} \nu(u) \text{ for all } s \in [a, b]$$

and all $u \in \delta \text{Ce}(\nu)$. Using Prop.2, we conclude that

$$\mu \left( \int_a^b n(\bullet, u) \right) \leq \int_a^b \mu \circ (n(\bullet, u)) \leq \varepsilon \nu(u)$$

whenever $u \in \delta \text{Ce}(\nu)$. Since $\varepsilon \in \mathbb{P}^+$ was arbitrary, it follows that the mapping $(u \mapsto \int_a^b n(\bullet, u))$ is small near $0 \in \mathcal{V}$ (see Sect.62). Now, integrating (610.13) with respect to $s \in [a, b]$ and observing the representation (610.6) of $k$, we obtain

$$\int_a^b n(\bullet, u) = k(x + u) - k(u) - \left( \int_a^b \nabla^2 h(\bullet, x) \right) u$$

for all $u \in D - x$. Therefore, by the Characterization of Gradients of Sect.63, $k$ is differentiable at $x$ and its gradient is given by (610.11). The continuity of $\nabla k$ follows from Prop.3. $lacksquare$

The following corollary deals with generalizations of integral representations of the type (610.6).

**Corollary:** Assume that $h : D \to \mathcal{W}$ satisfies the conditions (i) and (ii) of the Theorem. Assume, further, that $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ are differentiable and satisfy

$$f(x), g(x) \in \overline{D}(\bullet, x) \text{ for all } x \in D.$$  

Then $k : D \to \mathcal{W}$, defined by

$$k(x) := \int_{f(x)}^{g(x)} h(\bullet, x) \text{ for all } x \in D,$$  

is of class $C^1$ and its gradient is given by

$$\nabla_x k = h(g(x), x) \otimes \nabla_x g - h(f(x), x) \otimes \nabla_x f + \int_{f(x)}^{g(x)} \nabla^2 h(\bullet, x) \text{ for all } x \in D.$$  

(610.14)
**Proof:** Consider the set $\overline{D} \subset \mathbb{R} \times \mathbb{R} \times \mathcal{E}$ defined by (610.7), so that $a, b \in \overline{D}_{(\star x)}$ for all $(a, b, x) \in \overline{D}$. Since $\overline{D}_{(\star x)}$ is assumed to be an interval for each $x \in D$, we can define $m : \overline{D} \rightarrow W$ by

$$m(a, b, x) := \int_{a}^{b} h(\bullet, x).$$

By the Theorem, $\nabla(3)m : \overline{D} \rightarrow \text{Lin}(\mathcal{V}, \mathcal{W})$ exists and is given by

$$\nabla(3)m(a, b, x) = \int_{a}^{b} \nabla(2)h(\bullet, x).$$

(610.17)

It follows from Prop.4 that $\nabla(3)m$ is continuous. By the Fundamental Theorem of Calculus and (610.16), the partial derivatives $m_1$ and $m_2$ exist, are continuous, and are given by

$$m_1(a, b, x) = -h(a, x), \quad m_2(a, b, x) = h(b, x).$$

(610.18)

By the Partial Gradient Theorem of Sect.65, it follows that $m$ is of class $C^1$. Since $\nabla(1)m = m_1 \otimes$ and $\nabla(2)m = m_2 \otimes$, it follows from (65.9), (610.17), and (610.18) that

$$\left(\nabla_{(a,b,x)m}\right)(\alpha, \beta, u) = h(b, x) \otimes \beta - h(a, x) \otimes \alpha + \left(\int_{a}^{b} \nabla(2)h(\bullet, x)\right)u$$

(610.19)

for all $(a, b, x) \in \overline{D}$ and all $(\alpha, \beta, u) \in \mathbb{R} \times \mathbb{R} \times \mathcal{V}$.

Now, since $f$ and $g$ are differentiable, so is $(f, g, 1_{D \subseteq \mathcal{E}}) : D \rightarrow \mathbb{R} \times \mathbb{R} \times \mathcal{E}$ and we have

$$\nabla_{x}(f, g, 1_{D \subseteq \mathcal{E}}) = (\nabla_{x}f, \nabla_{x}g, 1_{\mathcal{V}})$$

(610.20)

for all $x \in D$. (See Prop.2 of Sect.63). By (610.14) and (610.16) we have $k = m \circ (f, g, 1_{D \subseteq \mathcal{E}})|\overline{D}$. Therefore, by the General Chain Rule of Sect.63, $k$ is of class $C^1$ and we have

$$\nabla_{x}k = (\nabla_{(f(x), g(x), x)}m)\nabla_{x}(f, g, 1_{D \subseteq \mathcal{E}})$$

for all $x \in D$. Using (610.19) and (610.20), we obtain (610.15). □
611 Curl, Symmetry of Second Gradients

In this section, we assume that flat spaces $\mathcal{E}$ and $\mathcal{E}'$, with translation spaces $\mathcal{V}$ and $\mathcal{V}'$, and an open subset $\mathcal{D}$ of $\mathcal{E}$ are given.

If $H : \mathcal{D} \to \text{Lin}(\mathcal{V}, \mathcal{V}')$ is differentiable, we can apply the identification (24.1) to the codomain of the gradient

$$\nabla H : \mathcal{D} \to \text{Lin}(\mathcal{V} \otimes \mathcal{V}^2, \mathcal{V}')$$

and hence we can apply the switching defined by (24.4) to the values of $\nabla H$.

**Definition 1:** The **curl** of a differentiable mapping $H : \mathcal{D} \to \text{Lin}(\mathcal{V}, \mathcal{V}')$ is the mapping $\text{Curl} H : \mathcal{D} \to \text{Lin}^2(\mathcal{V}^2, \mathcal{V}')$ defined by

$$\text{Curl} H(x) := \nabla x H - (\nabla x H) \sim$$

for all $x \in \mathcal{D}$. (611.1)

The values of $\text{Curl} H$ are skew, i.e. $\text{Rng} \text{Curl} H \subset \text{Skew}^2(\mathcal{V}^2, \mathcal{V}')$.

The following result deals with conditions that are sufficient or necessary for the curl to be zero.

**Curl-Gradient Theorem:** Assume that $H : \mathcal{D} \to \text{Lin}(\mathcal{V}, \mathcal{V}')$ is of class $C^1$. If $H = \nabla \varphi$ for some $\varphi : \mathcal{D} \to \mathcal{E}'$ then $\text{Curl} H = 0$. Conversely, if $\mathcal{D}$ is convex and $\text{Curl} H = 0$ then $H = \nabla \varphi$ for some $\varphi : \mathcal{D} \to \mathcal{E}'$.

The proof will be based on the following

**Lemma:** Assume that $\mathcal{D}$ is convex. Let $q \in \mathcal{D}$ and $q' \in \mathcal{E}'$ be given and let $\varphi : \mathcal{D} \to \mathcal{E}'$ be defined by

$$\varphi(q + v) := q' + \left(\int_0^1 H(q + sv)ds\right) v$$

for all $v \in \mathcal{D} - q$. (611.2)

Then $\varphi$ is of class $C^1$ and we have

$$\nabla \varphi(q + v) = H(q + v) - \left(\int_0^1 s(C\text{Curl} H)(q + sv)ds\right) v$$

(611.3)

for all $v \in \mathcal{D} - q$.

**Proof:** We define $\overline{\mathcal{D}} \subset \mathbb{R} \times \mathcal{V}$ by

$$\overline{\mathcal{D}} := \{(s, u) \mid u \in \mathcal{D} - q, \ s + qu \in \mathcal{D}\}.$$

Since $\mathcal{D}$ is open and convex, it follows that, for each $u \in \mathcal{D} - q$, $\overline{\mathcal{D}}(s, u) := \{s \in \mathbb{R} \mid q + su \in \mathcal{D}\}$ is an open interval that includes $[0, 1]$. Since $H$ is of class $C^1$, so is

$$((s, u) \mapsto H(q + su)) : \overline{\mathcal{D}} \to \text{Lin}(\mathcal{V}, \mathcal{V}').$$
hence we can apply the Differentiation Theorem for Integral Representations of Sect.610 to conclude that \( u \mapsto \int_0^1 H(q + su)ds \) is of class \( C^1 \) and that its gradient is given by

\[
\nabla_v (u \mapsto \int_0^1 H(q + su)ds) = \int_0^1 sH(q + sv)ds
\]

for all \( v \in D - q \). By the Product Rule (66.7), it follows that the mapping \( \varphi \) defined by (611.2) is of class \( C^1 \) and that its gradient is given by

\[
(\nabla_{q + \varphi} \varphi)u = \left( \int_0^1 s\nabla H(q + sv)uds \right) v + \left( \int_0^1 H(q + sv)ds \right) u
\]

for all \( v \in D - q \) and all \( u \in V \). Applying the definitions (24.2), (24.4), and (611.1), and using the linearity of the switching and Prop.1 of Sect.610, we obtain

\[
(\nabla_{q + \varphi} \varphi)u = \begin{pmatrix} \int_0^1 s\text{Curl} \nabla H(q + sv)ds \end{pmatrix} (u, v) + \left( \int_0^1 H(q + sv)ds \right) u
\]

for all \( v \in D - q \) and all \( u \in V \). By the Product Rule (66.10) and the Chain Rule we have, for all \( s \in [0, 1] \),

\[
\partial_s (t \mapsto tH(q + tv)) = H(q + sv) + s(\nabla H(q + sv)v).
\]

Hence, by the Fundamental Theorem of Calculus, the second integral on the right side of (611.4) reduces to \( H(q + v) \). Therefore, since Curl \( H \) has skew values and since \( u \in V \) was arbitrary, (611.4) reduces to (611.3).

**Proof of the Theorem:** Assume, first, that \( D \) is convex and that \( \text{Curl} \, H = 0 \). Choose \( q \in D \), \( q \in E' \) and define \( \varphi : D \to E' \) by (611.2). Then \( H = \nabla \varphi \) by (611.3).

Conversely, assume that \( H = \nabla \varphi \) for some \( \varphi : D \to E' \). Let \( q \in D \) be given. Since \( D \) is open, we can choose a convex open neighborhood \( N \) of \( q \) with \( N \subset D \). Let \( v \in N - q \) by given. Since \( N \) is convex, we have \( q + sv \in N \) for all \( s \in [0, 1] \), and hence we can apply the Fundamental Theorem of Calculus to the derivative of \( s \mapsto \varphi(q + sv) \), with the result

\[
\varphi(q + v) = \varphi(q) + \left( \int_0^1 H(q + sv)ds \right) v.
\]
Since \( v \in \mathcal{N} - q \) was arbitrary, we see that the hypotheses of the Lemma are satisfied for \( \mathcal{N} \) instead of \( D \), \( q' := \varphi(q) \), \( \varphi|_{\mathcal{N}} \) instead of \( \varphi \), and \( H|_{\mathcal{N}} = \nabla \varphi|_{\mathcal{N}} \) instead of \( H \). Hence, by (611.3), we have

\[
\left( \int_0^1 s(\text{Curl } H)(q + sv)ds \right) v = 0 \tag{611.5}
\]

for all \( v \in \mathcal{N} - q \). Now let \( u \in \mathcal{V} \) be given. Since \( \mathcal{N} - q \) is a neighborhood of \( 0 \in \mathcal{V} \), there is a \( \delta \in \mathbb{P}^\times \) such that \( tu \in \mathcal{N} - q \) for all \( t \in [0, \delta] \). Hence, substituting \( tu \) for \( v \) in (611.5) and dividing by \( t \) gives

\[
\left( \int_0^1 s(\text{Curl } H)(q + stu)ds \right) u = 0 \quad \text{for all} \quad t \in [0, \delta].
\]

since \( \text{Curl } H \) is continuous, we can apply Prop.3 of Sect.610 and, in the limit \( t \to 0 \), obtain \((\text{Curl } H)(q)x = 0\). Since \( u \in \mathcal{V} \) and \( q \in D \) were arbitrary, we conclude that \( \text{Curl } H = 0 \).

**Remark 1:** In the second part of the Theorem, the condition that \( D \) be convex can be replaced by the weaker one that \( D \) be “simply connected”. This means, intuitively, that every closed curve in \( D \) can be continuously shrunk entirely within \( D \) to a point.

Since \( \text{Curl } \nabla \varphi = \nabla^{(2)} \varphi - (\nabla^{(2)} \varphi)^\sim \) by (611.1), we can restate the first part of the Curl-Gradient Theorem as follows:

**Theorem on Symmetry of Second Gradients:** If \( \varphi : D \to \mathcal{E}' \) is of class \( C^2 \), then its second gradient \( \nabla^{(2)} \varphi : D \to \text{Lin}_2(\mathcal{V}'^2, \mathcal{V}') \) has symmetric values, i.e. \( \text{Rng } \nabla^{(2)} \varphi \subset \text{Sym}_2(\mathcal{V}'^2, \mathcal{V}') \).

**Remark 2:** The assertion that \( \nabla_x(\nabla \varphi) \) is symmetric for a given \( x \in D \) remains valid if one merely assumes that \( \varphi \) is differentiable and that \( \nabla \varphi \) is differentiable at \( x \). A direct proof of this fact, based on the results of Sect.64, is straightforward although somewhat tedious.

We assume now that \( \mathcal{E} := \mathcal{E}_1 \times \mathcal{E}_2 \) is the set-product of flat spaces \( \mathcal{E}_1, \mathcal{E}_2 \) with translation spaces \( \mathcal{V}_1, \mathcal{V}_2 \), respectively. Assume that \( \varphi : D \to \mathcal{E}' \) is twice differentiable. Using Prop.1 of Sect.65 repeatedly, it is easily seen that

\[
\nabla^{(2)} \varphi(x) = \\
(\nabla^{(1)}_{(1)} \nabla^{(1)} \varphi)(x)(ev_1, ev_1) + (\nabla^{(1)}_{(1)} \nabla^{(2)} \varphi)(x)(ev_1, ev_2) + (611.6) \\
(\nabla^{(2)}_{(1)} \nabla^{(1)} \varphi)(x)(ev_2, ev_1) + (\nabla^{(2)}_{(2)} \nabla^{(2)} \varphi)(x)(ev_2, ev_2)
\]

for all \( x \in D \), where \( ev_1 \) and \( ev_2 \) are the evaluation mappings from \( \mathcal{V} := \mathcal{V}_1 \times \mathcal{V}_2 \) to \( \mathcal{V}_1 \) and \( \mathcal{V}_2 \), respectively (see Sect.04). Evaluation of (611.6) at
264 CHAPTER 6. DIFFERENTIAL CALCULUS

\[(u, v) = ((u_1, u_2), (v_1, v_2)) \in V^2 \text{ gives } \]
\[
\nabla^{(2)} \varphi(x)(u, v) = \]
\[
(\nabla^{(1)} \nabla^{(1)} \varphi)(x)(u_1, v_1) + (\nabla^{(1)} \nabla^{(2)} \varphi)(x)(u_1, v_2) + \]
\[
(\nabla^{(2)} \nabla^{(1)} \varphi)(x)(u_2, v_1) + (\nabla^{(2)} \nabla^{(2)} \varphi)(x)(u_2, v_2). \quad (611.7)
\]

The following is a corollary to the preceding theorem.

**Theorem on the Interchange of Partial Gradients:** Assume that \( D \) is an open subset of a product space \( \mathcal{E} := E_1 \times E_2 \) and that \( \varphi : D \to \mathcal{E}' \) is of class \( C^2 \). Then
\[
\nabla^{(1)} \nabla^{(2)} \varphi = (\nabla^{(2)} \nabla^{(1)} \varphi)^\sim, \quad (611.8)
\]
where the operation \( \sim \) is to be understood as value-wise switching.

**Proof:** Let \( x \in D \) and \( w \in V_1 \times V_2 \) be given. If we apply (611.7) to the case when \( u := (w_1, 0), \ v := (0, w_2) \) and then with \( u \) and \( v \) interchanged we obtain
\[
\nabla^{(2)} \varphi(x)(u, v) = (\nabla^{(1)} \nabla^{(2)} \varphi)(x)(w_1, w_2),
\]
\[
\nabla^{(2)} \varphi(x)(v, u) = (\nabla^{(2)} \nabla^{(1)} \varphi)(x)(w_2, w_1).
\]
Since \( w \in V_1 \times V_2 \) was arbitrary, the symmetry of \( \nabla^{(2)} \varphi(x) \) gives (611.8). \(

**Corollary:** Let \( I \) be a finite index set and let \( D \) be an open subset of \( \mathbb{R}^I \). If \( \varphi : D \to \mathcal{E}' \) is of class \( C^2 \), then the second partial derivatives \( \varphi_{,ik} : D \to V' \) satisfy
\[
\varphi_{,ik} = \varphi_{,ki} \quad \text{for all } i, k \in I. \quad (611.9)
\]

**Remark 3:** Assume \( D \) is an open subset of a Euclidean space \( \mathcal{E} \) with translation space \( \mathcal{V} \). A mapping \( h : D \to V \cong V^* \) is then called a vector-field (see Sect.71). If \( h \) is differentiable, then the range of \( \text{Curl} \, h \) is included in \( \text{Skew}(V^2, \mathbb{R}) \cong \text{Skew} \, V \). If \( V \) is 3-dimensional, there is a natural doubleton of orthogonal isomorphisms from \( \text{Skew} \, V \) to \( V \), as will be explained in Vol.II. If one of these two isomorphisms, say \( V \in \text{Orth}(\text{Skew} \, V, V) \), is singled out, we may consider the vector-curl \( \text{curl} \, h := V(\text{Curl} \, h)|_{\text{Skew} \, V} \) of the vector field \( h \) (Note the lower-case “c”). This vector-curl rather than the curl of Def.1 is used in much of the literature. \(

Notes 611

(1) In some of the literature on “Vector Analysis”, the notation \( \nabla \times h \) instead of \( \text{curl} \, h \) is used for the vector-curl of \( h \), which is explained in the Remark above. This notation should be avoided for the reason mentioned in Note (1) to Sect.67.
612. Lineonic Exponentials

We note the following generalization of a familiar result of elementary calculus.

**Proposition 1:** Let $E$ be a flat space with translation space $V$ and let $I$ be an interval. Let $g$ be a sequence of continuous processes in $\text{Map}(I, V)$ that converges locally uniformly to $h \in \text{Map}(I, V)$. Let $a \in I$ and $q \in E$ be given and define the sequence $z$ in $\text{Map}(I, E)$ by

$$z_n(t) := q + \int_a^t g_n \quad \text{for all } n \in \mathbb{N}^\times \text{ and } t \in I.$$  \hspace{1cm} (612.1)

Then $h$ is continuous and $z$ converges locally uniformly to the process $p : I \to E$ defined by

$$p(t) := q + \int_a^t h \quad \text{for all } t \in I.$$  \hspace{1cm} (612.2)

**Proof:** The continuity of $h$ follows from the Theorem on Continuity of Uniform Limits of Sect.56. Hence (612.2) is meaningful. Let $\nu$ be a norm on $V$. By Prop.2 of Sect.610 and (612.1) and (612.2) we have

$$\nu(z_n(t) - p(t)) = \nu\left(\int_a^t (g_n - h)\right) \leq \int_a^t \nu \circ (g_n - h) \quad \text{for all } n \in \mathbb{N}^\times \text{ and } t \in I.$$  \hspace{1cm} (612.3)

for all $n \in \mathbb{N}^\times$ and all $t \in I$. Now let $s \in I$ be given. We can choose a compact interval $[b, c] \subset I$ such that $a \in [b, c]$ and such that $[b, c]$ is a neighborhood of $s$ relative to $I$ (see Sect.56). By Prop.7 of Sect.58, $g|_{[b, c]}$ converges uniformly to $h|_{[b, c]}$. Now let $\epsilon \in \mathbb{P}^\times$ be given. By Prop.7 of Sect.55 we can determine $m \in \mathbb{N}^\times$ such that

$$\nu \circ (g_n - h)|_{[b, c]} < \frac{\epsilon}{c - b} \quad \text{for all } n \in m + \mathbb{N}.$$  

Hence, by (612.3), we have

$$\nu(z_n(t) - p(t)) < |t - a| \frac{\epsilon}{c - b} < \epsilon \quad \text{for all } n \in m + \mathbb{N}$$

and all $t \in [b, c]$. Since $\epsilon \in \mathbb{P}^\times$ was arbitrary, we can use Prop.7 of Sect.55 again to conclude that $z|_{[b, c]}$ converges uniformly to $p|_{[b, c]}$. Since $s \in I$ was arbitrary and since $[b, c] \in \text{Nhd}_s(I)$, the conclusion follows. 

From now on we assume that a linear space $V$ is given and we consider the algebra $\text{Lin}V$ of lineons on $V$ (see Sect.18).
Proposition 2: The sequence \( S \) in \( \text{Map} \left( \text{Lin} V, \text{Lin} V \right) \) defined by
\[
S_n(L) := \sum_{k \in \mathbb{N}} \frac{1}{k!} L^k
\] (612.4)
for all \( L \in \text{Lin} V \) and all \( n \in \mathbb{N}^\times \) converges locally uniformly. Its limit is called the (lineonic) exponential for \( V \) and is denoted by \( \exp V : \text{Lin} V \rightarrow \text{Lin} V \). The exponential \( \exp V \) is continuous.

Proof: We choose a norm \( \nu \) on \( V \). Using Prop.6 of Sect.52 and induction, we see that \( \| L^k \|_\nu \leq \| L \|_\nu ^k \) for all \( k \in \mathbb{N} \) and all \( L \in \text{Lin} V \) (see (52.9)). Hence, given \( \sigma \in \mathbb{P}^\times \), we have
\[
\left\| \frac{1}{k!} L^k \right\|_\nu \leq \frac{\sigma^k}{k!}
\]
for all \( k \in \mathbb{N} \) and all \( L \in \sigma \text{Ce} \left( \| \bullet \|_\nu \right) \). Since the sum-sequence of \( \left( \frac{\sigma^k}{k!} \right)_{k \in \mathbb{N}} \) converges (to \( e^\sigma \)), we can use Prop.9 of Sect.55 to conclude that the restriction of the sum-sequence \( S \) of \( \left( \frac{1}{k!} L^k \right)_{k \in \mathbb{N}} \) to \( \sigma \text{Ce} \left( \| \bullet \|_\nu \right) \) converges uniformly. Now, given \( L \in \text{Lin} V \), \( \sigma \text{Ce} \left( \| \bullet \|_\nu \right) \) is a neighborhood of \( L \) if \( \sigma > \| L \|_\nu \). Therefore \( S \) converges locally uniformly. The continuity of its limit \( \exp V \) follows from the Continuity Theorem for Uniform Limits. 

For later use we need the following:

Proposition 3: Let \( L \in \text{Lin} V \) be given. Then the constant zero is the only differentiable process \( D : \mathbb{R} \rightarrow \text{Lin} V \) that satisfies
\[
D^* = LD \quad \text{and} \quad D(0) = 0.
\] (612.5)

Proof: By the Fundamental Theorem of Calculus, (612.5) is equivalent to
\[
D(t) = \int_0^t (LD) \quad \text{for all} \quad t \in \mathbb{R}.
\] (612.6)
Choose a norm \( \nu \) on \( V \). Using Prop.2 of Sect.610 and Prop.6 of Sect.52, we conclude from (612.6) that
\[
\| D(t) \|_\nu \leq \int_0^t \| LD(s) \|_\nu ds \leq \| L \|_\nu \int_0^t \| D(s) \|_\nu ds,
\]
i.e., with the abbreviations
\[
\sigma(t) := \| D(t) \|_\nu \quad \text{for all} \quad t \in \mathbb{P}, \quad \kappa := \| L \|_\nu,
\] (612.7)
that
\[
0 \leq \sigma(t) \leq \kappa \int_0^t \sigma \quad \text{for all} \quad t \in \mathbb{P}.
\] (612.8)
We define \( \varphi : \mathbb{P} \rightarrow \mathbb{R} \) by
\[
\varphi(t) := e^{-\kappa t} \int_0^t \sigma \quad \text{for all } t \in \mathbb{P}.
\] (612.9)

Using elementary calculus, we infer from (612.9) and (612.8) that
\[
\varphi^*(t) = -\kappa e^{-\kappa t} \int_0^t \sigma + e^{-\kappa t} \sigma(t) \leq 0
\] for all \( t \in \mathbb{P} \) and hence that \( \varphi \) is antitone. On the other hand, it is evident from (612.9) that \( \varphi(0) = 0 \) and \( \varphi \geq 0 \). This can happen only if \( \varphi = 0 \), which, in turn, can happen only if \( \sigma(t) = 0 \) for all \( t \in \mathbb{P} \) and hence, by (612.7), if \( D(t) = 0 \) for all \( t \in \mathbb{P} \). To show that \( D(t) = 0 \) for all \( t \in -\mathbb{P} \), one need only replace \( D \) by \( D \circ (-\iota) \) and \( L \) by \( -L \) in (612.5) and then use the result just proved.

**Proposition 4:** Let \( L \in \text{Lin}\mathcal{V} \) be given. Then the only differentiable process \( E : \mathbb{R} \rightarrow \text{Lin}\mathcal{V} \) that satisfies
\[
E^* = LE \quad \text{and} \quad E(0) = 1_{\mathcal{V}}
\] (612.10)
is the one given by
\[
E := \exp \circ (\iota L).
\] (612.11)

**Proof:** We consider the sequence \( G \) in \( \text{Map} (\mathbb{R}, \text{Lin}\mathcal{V}) \) defined by
\[
G_n(t) := \sum_{k \in \mathbb{N}} \frac{t^k}{k!} L^{k+1}
\] (612.12)
for all \( n \in \mathbb{N}^\times \) and all \( t \in \mathbb{R} \). Since, by (612.4), \( G_n(t) = \text{LS}_n(tL) \) for all \( n \in \mathbb{N}^\times \) and all \( t \in \mathbb{R} \), it follows from Prop.2 that \( G \) converges locally uniformly to \( \text{Lexp} \circ (\iota L) : \mathbb{R} \rightarrow \text{Lin}\mathcal{V} \). On the other hand, it follows from (612.12) that
\[
1_{\mathcal{V}} + \int_0^t G_n = 1_{\mathcal{V}} + \sum_{k \in \mathbb{N}^\times} \frac{t^{k+1}}{(k+1)!} L^{k+1} = S_{n+1}(tL)
\]
for all \( n \in \mathbb{N}^\times \) and all \( t \in \mathbb{R} \). Applying Prop.1 to the case when \( \mathcal{E} \) and \( \mathcal{V} \) are replaced by \( \text{Lin}\mathcal{V} \), \( I \) by \( \mathbb{R} \), \( g \) by \( G \), \( a \) by 0, and \( q \) by \( 1_{\mathcal{V}} \), we conclude that
\[
\exp_\mathcal{V}(tL) = 1_{\mathcal{V}} + \int_0^t (L \exp_\mathcal{V} \circ (\iota L)) \quad \text{for all } t \in \mathbb{R},
\] (612.13)
which, by the Fundamental Theorem of Calculus, is equivalent to the assertion that (612.10) holds when \( E \) is defined by (612.11).

The uniqueness of \( E \) is an immediate consequence of Prop.3. □

**Proposition 5:** Let \( L \in \text{Lin} V \) and a continuous process \( F : \mathbb{R} \to \text{Lin} V \) be given. Then the only differentiable process \( D : \mathbb{R} \to \text{Lin} V \) that satisfies

\[
D^* = LD + F \quad \text{and} \quad D(0) = 0 \tag{612.14}
\]

is the one given by

\[
D(t) = \int_0^t E(t-s)F(s)ds \quad \text{for all} \quad t \in \mathbb{R}, \tag{612.15}
\]

where \( E : \mathbb{R} \to \text{Lin} V \) is given by (612.11).

**Proof:** Let \( D \) be defined by (612.15). Using the Corollary to the Differentiation Theorem for Integral Representations of Sect.610, we see that \( D \) is of class \( C^1 \) and that \( D^* \) is given by

\[
D^*(t) = E(0)F(t) + \int_0^t E^*(t-s)F(s)ds \quad \text{for all} \quad t \in \mathbb{R}.
\]

Using (612.10) and Prop.1 of Sect.610, and then (612.15), we find that (612.14) is satisfied.

The uniqueness of \( D \) is again an immediate consequence of Prop.3. □

**Differentiation Theorem for Lineonic Exponentials:** Let \( V \) be a linear space. The exponential \( \text{exp}_V : \text{Lin} V \to \text{Lin} V \) is of class \( C^1 \) and its gradient is given by

\[
(\nabla_L \text{exp}_V)M = \int_0^1 \text{exp}_V(sL)M\text{exp}_V((1-s)L)ds \tag{612.16}
\]

for all \( L, M \in \text{Lin} V \).

**Proof:** Let \( L, M \in \text{Lin} V \) be given. Also, let \( r \in \mathbb{R}^\times \) be given and put

\[
K := L + rM, \quad E_1 := \text{exp}_V \circ (tL), \quad E_2 := \text{exp}_V \circ (tK). \tag{612.17}
\]

By Prop.4 we have \( E_1^* = LE_1, \ E_2^* = KE_2, \) and \( E_1(0) = E_2(0) = 1_V \). Taking the difference and observing (612.17) we see that \( D := E_2 - E_1 \) satisfies

\[
D^* = KE_2 - LE_1 = LD + rM E_2, \quad D(0) = 0.
\]

Hence, if we apply Prop.5 with the choice \( F := rM E_2 \) we obtain

\[
(E_2 - E_1)(t) = r \int_0^t E_1(t-s)M E_2(s)ds \quad \text{for all} \quad t \in \mathbb{R}.
\]
For $t := 1$ we obtain, in view of (612.17),

$$
\frac{1}{r}(\exp_V(L + rM) - \exp_V(L)) = \int_0^1 (\exp_V((1 - s)L))M\exp_V(s(L + rM))ds.
$$

Since $\exp_V$ is continuous by Prop.2, and since bilinear mappings are continuous (see Sect.66), we see that the integrand on the right depends continuously on $(s, r) \in \mathbb{R}^2$. Hence, by Prop.3 of Sect.610 and by (65.13), in the limit $r \to 0$ we find

$$
(dd_M \exp_V)(L) = \int_0^1 (\exp_V((1 - s)L))M\exp_V(sL)ds. \tag{612.18}
$$

Again, we see that the integrand on the right depends continuously on $(s, L) \in \mathbb{R} \times \text{Lin}V$. Hence, using Prop.3 of Sect.610 again, we conclude that the mapping $dd_M \exp_V : \text{Lin}V \to \text{Lin}V$ is continuous. Since $M \in \text{Lin}V$ was arbitrary, it follows from Prop.7 of Sect.65 that $\exp_V$ is of class $C^1$. The formula (612.16) is the result of combining (65.14) with (612.18).

**Proposition 6:** Assume that $L, M \in \text{Lin}V$ commute, i.e. that $LM = ML$. Then:

(i) $M$ and $\exp_V(L)$ commute.

(ii) We have

$$
\exp_V(L + M) = (\exp_V(L))(\exp_V(M)). \tag{612.19}
$$

(iii) We have

$$
(\nabla_L \exp_V)M = (\exp_V(L))M. \tag{612.20}
$$

**Proof:** Put $E := \exp_V \circ (\iota_L)$ and $D := EM - ME$. By Prop.4, we find $D^* = E^*M - ME^* = LEM - MLE$. Hence, since $LM = ML$, we obtain $D^* = LD$. Since $D(0) = 1_VM - M1_V = 0$, it follows from Prop.3 that $D = 0$ and hence $D(1) = 0$, which proves (i).

Now put $F := \exp_V \circ (\iota_M)$. By the Product Rule (66.13) and by Prop.4, we find

$$
(EF)^* = E^*F + EF^* = LEF + EMF.
$$

Since $EM = ME$ by part (i), we obtain $(EF)^* = (L + M)EF$. Since $(EF)(0) = E(0)F(0) = 1_V$, by Prop.4, $EF = \exp_V \circ (\iota(L + M))$. Evaluation at 1 gives (612.19), which proves (ii).

Part (iii) is an immediate consequence of (612.16) and of (i) and (ii).
Pitfalls: Of course, when $\mathcal{V} = \mathbb{R}$, then $\text{Lin}\mathcal{V} = \text{Lin}\mathbb{R} \cong \mathbb{R}$ and the lineonic exponential reduces to the ordinary exponential $\exp$. Prop.6 shows how the rule $\exp(s+t) = \exp(s) \exp(t)$ for all $s, t \in \mathbb{R}$ and the rule $\exp^* = \exp$ generalize to the case when $\mathcal{V} \neq \mathbb{R}$. The assumption, in Prop.6, that $\mathbf{L}$ and $\mathbf{M}$ commute cannot be omitted. The formulas (612.19) and (612.20) need not be valid when $\mathbf{LM} \neq \mathbf{ML}$.

If the codomain of the ordinary exponential is restricted to $\mathbb{P}^\times$ it becomes invertible and its inverse $\log : \mathbb{P}^\times \to \mathbb{R}$ is of class $C^1$. If $\dim \mathcal{V} > 1$, then $\exp_\mathcal{V}$ does not have differentiable local inverses near certain values of $L \in \text{Lin}\mathcal{V}$ because for these values of $L$, the gradient $\nabla_L \exp_\mathcal{V}$ fails to be injective (see Problem 10). In fact, one can prove that $\exp_\mathcal{V}$ is not locally invertible near the values of $L$ in question. Therefore, there is no general lineonic analogue of the logarithm. See, however, Sect.85. 

613 Problems for Chapter 6

(1) Let $I$ be a genuine interval, let $\mathcal{E}$ be a flat space with translation space $\mathcal{V}$, and let $p : I \to \mathcal{E}$ be a differentiable process. Define $h : I^2 \to \mathcal{V}$ by

$$h(s, t) := \begin{cases} \frac{p(s) - p(t)}{s - t} & \text{if } s \neq t \\ p^*(s) & \text{if } s = t \end{cases}.$$  

(P6.1)

(a) Prove: If $p^*$ is continuous at $t \in I$, then $h$ is continuous at $(t, t) \in I^2$.

(b) Find a counterexample which shows that $h$ need not be continuous if $p$ is merely differentiable and not of class $C^1$.

(2) Let $f : \mathcal{D} \to \mathbb{R}$ be a function whose domain $\mathcal{D}$ is an open convex subset of a flat space $\mathcal{E}$ with translation space $\mathcal{V}$.

(a) Show: If $f$ is differentiable and $\nabla f$ is constant, then $f = a|_\mathcal{D}$ for some flat function $a \in \text{Flf}(\mathcal{E})$ (see Sect.36).

(b) Show: If $f$ is twice differentiable and $\nabla^{(2)} f$ is constant, then $f$ has the form

$$f = a|_\mathcal{D} + Q \circ (1_{\mathcal{D} \subset \mathcal{E}} - q_{\mathcal{D} \to \mathcal{E}}),$$  

(P6.2)

where $a \in \text{Flf}(\mathcal{E})$, $q \in \mathcal{E}$, and $Q \in \text{Qu}(\mathcal{V})$ (see Sect. 27).
(3) Let $\mathcal{D}$ be an open subset of a flat space $\mathcal{E}$ and let $\mathcal{W}$ be a genuine inner-product space. Let $k : \mathcal{D} \rightarrow \mathcal{W}^\times$ be a mapping that is differentiable at a given $x \in \mathcal{D}$.

(a) Show that the value-wise magnitude $|k| : \mathcal{D} \rightarrow \mathbb{P}^\times$, defined by $|k|(y) := |k(y)|$ for all $y \in \mathcal{D}$, is differentiable at $x$ and that

$$\nabla_x |k| = \frac{1}{|k(x)|} (\nabla_x k)\top k(x). \quad \text{(P6.3)}$$

(b) Show that $k/|k| : \mathcal{D} \rightarrow \mathcal{W}^\times$ is differentiable at $x$ and that

$$\nabla_x \left( \frac{k}{|k|} \right) = \frac{1}{|k(x)|^3} \left( |k(x)|^2 \nabla_x k - k(x) \otimes (\nabla_x k)\top k(x) \right). \quad \text{(P6.4)}$$

(4) Let $\mathcal{E}$ be a genuine inner-product space with translation space $\mathcal{V}$, let $q \in \mathcal{E}$ be given, and define $r : \mathcal{E} \setminus \{q\} \rightarrow \mathcal{V}^\times$ by

$$r(x) := x - q \quad \text{for all} \quad x \in \mathcal{E} \setminus \{q\} \quad \text{(P6.5)}$$

and put $r := |r|$ (see Part (a) of Problem 3).

(a) Show that $r/|r| : \mathcal{E} \setminus \{q\} \rightarrow \mathcal{V}^\times$ is of class $C^1$ and that

$$\nabla \left( \frac{r}{|r|} \right) = \frac{1}{r^2} (r^2 1_\mathcal{V} - r \otimes r). \quad \text{(P6.6)}$$

(Hint: Use Part (b) of Problem 3.)

(b) Let the function $h : \mathbb{P}^\times \rightarrow \mathbb{R}$ be twice differentiable. Show that $h \circ r : \mathcal{E} \setminus \{q\} \rightarrow \mathbb{R}$ is twice differentiable and that

$$\nabla^{(2)} (h \circ r) = \frac{1}{r^2} \left( r (h^{**} \circ r) - (h^\bullet \circ r) \right) r \otimes r + (h^\bullet \circ r) 1_\mathcal{V}. \quad \text{(P6.7)}$$

(c) Evaluate the Laplacian $\Delta (h \circ r)$ and reconcile your result with (67.17).

(5) A Euclidean space $\mathcal{E}$ of dimension $n$ with $n \geq 2$, a point $q \in \mathcal{E}$, and a linear space $\mathcal{W}$ are assumed given.

(a) Let $I$ be an open interval, let $\mathcal{D}$ be an open subset of $\mathcal{E}$, let $f : \mathcal{D} \rightarrow I$ be given by (67.16), let $a$ be a flat function on $\mathcal{E}$, let $g : I \rightarrow \mathcal{W}$ be twice differentiable, and put $h := a|_{\mathcal{D}}(g \circ f) : \mathcal{D} \rightarrow \mathcal{W}$. Show that

$$\Delta h = 2a|_{\mathcal{D}} \left( \left(2g^{**} + (n+2)g^\bullet \right) \circ f \right) - 4a(q)(g^\bullet \circ f). \quad \text{(P6.8)}$$
(b) Assuming that the Euclidean space $E$ is genuine, show that the function $h : E \setminus \{q\} \to W$ given by
\[
h(x) := ((e \cdot (x - q))|x - q|^{-n})a + b
\]
(P6.9)
is harmonic for all $e \in V := E - E$ and all $a, b \in W$. (Hint: Use Part (a) with $a$ determined by $\nabla a = e$, $a(q) = 0$.)

(6) Let $E$ be a flat space, let $q \in E$, let $Q$ be a non-degenerate quadratic form (see Sect.27) on $V := E - E$, and define $f : E \to \mathbb{R}$ by
\[
f(y) := Q(y - q) \quad \text{for all} \quad y \in E.
\]
(P6.10)
Let $a$ be a non-constant flat function on $E$ (see Sect.36).

(a) Prove: If the restriction of $f$ to the hyperplane $F := a < \{0\}$ attains an extremum at $x \in F$, then $x$ must be given by
\[
x = q - \lambda \underset{Q}{\nabla a}^{-1} (\nabla a), \quad \text{where} \quad \lambda := \frac{a(q)}{Q_{\nabla a, \nabla a}}.
\]
(P6.11)
(Hint: Use the Constrained-Extremum Theorem.)

(b) Under what condition does $f|_F$ actually attain a maximum or a minimum at the point $x$ given by (P6.11)?

(7) Let $E$ be a 2-dimensional Euclidean space with translation-space $V$ and let $J \in \text{Orth}V \cap \text{Skew}V$ be given (see Problem 2 of Chapt.4). Let $D$ be an open subset of $E$ and let $h : D \to V$ be a vector-field of class $C^1$.

(a) Prove that
\[
\text{Curl} h = -(\text{div}(Jh))J.
\]
(P6.12)

(b) Assuming that $D$ is convex and that $\text{div} h = 0$, prove that $h = J\nabla f$ for some function $f : D \to \mathbb{R}$ of class $C^2$.

Note: If $h$ is interpreted as the velocity field of a volume-preserving flow, then $\text{div} h = 0$ is valid and a function $f$ as described in Part (b) is called a “stream-function” of the flow.
(8) Let a linear space \( V \), a lineon \( L \in \text{Lin} V \), and \( u \in V \) be given. Prove: The only differentiable process \( h : \mathbb{R} \to V \) that satisfies

\[
\begin{align*}
    h^* &= Lh \quad \text{and} \quad h(0) = u \\
\end{align*}
\]

is the one given by

\[
    h := (\exp V \circ (\iota L))u. 
\]

(Hint: To prove existence, use Prop. 4 of Sect. 612. To prove uniqueness, use Prop.3 of Sect.612 with the choice \( D := h \otimes \lambda \) (value-wise), where \( \lambda \in V^* \) is arbitrary.)

(9) Let a linear space \( V \) and a lineon \( J \) on \( V \) that satisfies \( J^2 = -1_V \) be given. (There are such \( J \) if \( \dim V \) is even; see Sect.89.)

(a) Show that there are functions \( c : \mathbb{R} \to \mathbb{R} \) and \( d : \mathbb{R} \to \mathbb{R} \), of class \( C^1 \), such that

\[
\exp V \circ (iJ) = c1_V + dJ. 
\]

(Hint: Apply the result of Problem 8 to the case when \( V \) is replaced by \( C := Lsp(1_V, J) \subset \text{Lin} V \) and when \( L \) is replaced by \( L_J \in \text{Lin} C \), defined by \( L_J U = JU \) for all \( U \in C \).)

(b) Show that the functions \( c \) and \( d \) of Part (a) satisfy

\[
\begin{align*}
    d^* &= c, \quad c^* = -d, \quad c(0) = 1, \quad d(0) = 0, \\
\end{align*}
\]

and

\[
\begin{align*}
    c(t + s) &= c(t)c(s) - d(t)d(s) \\
    d(t + s) &= c(t)d(s) + d(t)c(s) \\
\end{align*}
\]

for all \( s, t \in \mathbb{R} \). (P6.17)

(c) Show that \( c = \cos \) and \( d = \sin \).

**Remark:** One could, in fact, use Part (a) to define \( \sin \) and \( \cos \).

(10) Let a linear space \( V \) and \( J \in \text{Lin} V \) satisfying \( J^2 = -1_V \) be given and put

\[
\mathcal{A} := \{L \in \text{Lin} V \mid LJ = -JL\}. 
\]
(a) Show that \( \mathcal{A} = \text{Null } (\nabla \pi J \exp_\mathcal{V}) \) and conclude that \( \nabla \pi J \exp_\mathcal{V} \) fails to be invertible when \( \dim \mathcal{V} > 0 \). (Hint: Use the Differentiation Theorem for Linearic Exponentials, Part (a) of Problem 9, and Part (d) of Problem 12 of Chap. 1).

(b) Prove that \( -1_\mathcal{V} \in \text{Bdy Rng } \exp_\mathcal{V} \) and hence that \( \exp_\mathcal{V} \) fails to be locally invertible near \( \pi J \).

(11) Let \( \mathcal{D} \) be an open subset of a flat space \( \mathcal{E} \) with translation space \( \mathcal{V} \) and let \( \mathcal{V}' \) be a linear space.

(a) Let \( H : \mathcal{D} \to \text{Lin}(\mathcal{V}, \mathcal{V}') \) be of class \( C^2 \). Let \( x \in \mathcal{D} \) and \( v \in \mathcal{V} \) be given. Note that

\[
\nabla_x \text{Curl} H \in \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \mathcal{V}'))) \cong \text{Lin}_2(\mathcal{V}^2, \text{Lin}(\mathcal{V}, \mathcal{V}'))
\]

and hence

\[
(\nabla_x \text{Curl} H) \sim \in \text{Lin}_2(\mathcal{V}^2, \text{Lin}(\mathcal{V}, \mathcal{V}')) \cong \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \mathcal{V}'))).
\]

Therefore we have

\[
G := (\nabla_x \text{Curl} H) \sim v \in \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \mathcal{V}')) \cong \text{Lin}_2(\mathcal{V}^2, \mathcal{V}').
\]

Show that

\[
G - G^\sim = (\nabla_x \text{Curl} H)v. \quad (P6.19)
\]

(Hint: Use the Symmetry Theorem for Second Gradients.)

(b) Let \( \eta : \mathcal{D} \to \mathcal{V}^\ast \) be of class \( C^2 \) and put

\[
W := \text{Curl } \eta : \mathcal{D} \to \text{Lin}(\mathcal{V}, \mathcal{V}^\ast) \cong \text{Lin}_2(\mathcal{V}^2, \mathbb{R}).
\]

Prove: If \( h : \mathcal{D} \to \mathcal{V} \) is of class \( C^1 \), then

\[
\text{Curl } (\mathcal{W}h) = (\nabla \mathcal{W})h + \mathcal{W} \nabla h + (\nabla h)^\top \mathcal{W}, \quad (P6.20)
\]

where value-wise evaluation and composition are understood.

(12) Let \( \mathcal{E} \) and \( \mathcal{E}' \) be flat spaces with \( \dim \mathcal{E} = \dim \mathcal{E}' \geq 2 \), let \( \varphi : \mathcal{E} \to \mathcal{E}' \) be a mapping of class \( C^1 \) and put

\[
\mathcal{C} := \{ x \in \mathcal{E} \mid \nabla_x \varphi \text{ is not invertible} \}.
\]
(a) Show that \( \varphi_>(C) \supset \text{Rng} \varphi \cap \text{Bdy} (\text{Rng} \varphi) \).

(b) Prove: If the pre-image under \( \varphi \) of every bounded subset of \( \mathcal{E}' \) is bounded and if \( \text{Acc} \varphi_>(C) = \emptyset \) (see Def. 1 of Sect. 57), then \( \varphi \) is surjective. (Hint: Use Problem 13 of Chap. 5.)

(13) By a **complex polynomial** \( p \) we mean an element of \( \mathbb{C}^{(\mathbb{N})} \), i.e. a sequence in \( \mathbb{C} \) indexed on \( \mathbb{N} \) and with finite support (see (07.10)). If \( p \neq 0 \), we define the **degree** of \( p \) by

\[
\deg p := \max \text{Supt } p = \max \{ k \in \mathbb{N} \mid p_k \neq 0 \}.
\]

By the **derivative** \( p' \) of \( p \) we mean the complex polynomial \( p' \in \mathbb{C}^{(\mathbb{N})} \) defined by

\[
(p')_k := (k + 1)p_{k+1} \quad \text{for all} \quad k \in \mathbb{N}.
\]  

(P6.21)

The **polynomial function** \( \tilde{p} : \mathbb{C} \to \mathbb{C} \) of \( p \) is defined by

\[
\tilde{p}(z) := \sum_{k \in \mathbb{N}} p_k z^k \quad \text{for all} \quad z \in \mathbb{C}.
\]  

(P6.22)

Let \( p \in (\mathbb{C}^{(\mathbb{N})})^\times \) be given.

(a) Show that \( \tilde{p} < (\{0\}) \) is finite and \( \sharp \tilde{p} < (\{0\}) \leq \deg p \).

(b) Regarding \( \mathbb{C} \) as a two-dimensional linear space over \( \mathbb{R} \), show that \( \tilde{p} \) is of class \( \mathcal{C}^2 \) and that

\[
(\nabla_z \tilde{p})w = \tilde{p}'(z)w \quad \text{for all} \quad z, w \in \mathbb{C}.
\]  

(P6.23)

(c) Prove that \( \tilde{p} \) is surjective if \( \deg p > 0 \). (Hint: Use Part (b) of Problem 12.)

(d) Show: If \( \deg p > 0 \), then the equation \( \tilde{p}(z) = 0 \) has at least one and no more than \( \deg p \) solutions.

**Note:** The assertion of Part (d) is usually called the “Fundamental Theorem of Algebra”. It is really a theorem of analysis, not algebra, and it is not all that fundamental.
Chapter 7

Coordinate Systems

In this chapter we assume again that all spaces (linear, flat, or Euclidean) under consideration are finite-dimensional.

71 Coordinates in Flat Spaces

We assume that a flat space \( \mathcal{E} \) with translation space \( \mathcal{V} \) is given. Roughly speaking, a coordinate system is a method for specifying points in \( \mathcal{E} \) by means of families of numbers. The coordinates are the functions that assign the specifying numbers to the points. It is useful to use the coordinate functions themselves as indices when describing the family of numbers that specifies a given point. For most “curvilinear” coordinate systems, the coordinate functions can be defined unambiguously only on a subset \( \mathcal{D} \) of \( \mathcal{E} \) obtained from \( \mathcal{E} \) by removing a suitable set of “exceptional” points. Examples will be considered in Sect.74.

Definition: A coordinate system on a given open subset \( \mathcal{D} \) of \( \mathcal{E} \) is a finite set \( \Gamma \subset \text{Map}(\mathcal{D},\mathbb{R}) \) of functions \( c: \mathcal{D} \to \mathbb{R} \), which are called coordinates, subject to the following conditions:

\( (C1) \) Every coordinate \( c \in \Gamma \) is of class \( C^2 \).

\( (C2) \) For every \( x \in \mathcal{D} \), the family \( (\nabla_x c \mid c \in \Gamma) \) of the gradients of the coordinates at \( x \) is a basis of \( \mathcal{V}^* \).

\( (C3) \) The mapping \( \Gamma: \mathcal{D} \to \mathbb{R}^\Gamma \), identified with the set \( \Gamma \) by self-indexing (see Sect.02) and term-wise evaluation (see Sect.04), i.e. \( \Gamma(x) := (c(x) \mid c \in \Gamma) \) for all \( x \in \mathcal{D} \), is injective.

277
For the remainder of this section we assume that an open subset $\mathcal{D}$ of $\mathcal{E}$ and a coordinate system $\Gamma$ on $\mathcal{D}$ are given. Using the identification $\mathcal{V}^* \Gamma = \text{Lin}(\mathcal{V}, \mathbb{R})^\Gamma \cong \text{Lin}(\mathcal{V}, \mathbb{R}^\Gamma)$ (see Sect.14), we obtain

$$\nabla_x \Gamma = (\nabla_x^c \mid c \in \Gamma). \quad (71.1)$$

The condition (C2) expresses the requirement that the gradient of $\Gamma : \mathcal{D} \rightarrow \mathbb{R}^\Gamma$ have invertible values. Using the Local Inversion Theorem of Sect.68 and (C3) we obtain the following result.

**Proposition 1:** The range $\overline{\mathcal{D}} := \text{Rng } \Gamma$ of $\Gamma$ is an open subset of $\mathbb{R}^\Gamma$, the mapping $\gamma := \Gamma|_{\overline{\mathcal{D}}} : \mathcal{D} \rightarrow \overline{\mathcal{D}}$ is invertible and has an inverse $\psi := \gamma^{-1} : \overline{\mathcal{D}} \rightarrow \mathcal{D}$ that is of class $C^2$.

By (65.11), the family $$(\psi_c \mid c \in \Gamma) \in \text{Map}(\overline{\mathcal{D}}, \mathcal{V}) \cong \text{Map}(\overline{\mathcal{D}}, \mathcal{V}^\Gamma)$$
of partial derivatives of $\psi$ with respect to the coordinates is related to the gradient $\nabla \psi : \overline{\mathcal{D}} \rightarrow \text{Lin}(\mathbb{R}^\Gamma, \mathcal{V})$ of $\psi$ by

$$\nabla \psi = \text{Inc}(\psi_c \mid c \in \Gamma), \quad (71.2)$$

where the value at $\xi \in \overline{\mathcal{D}}$ of the right side is understood to be the linear combination mapping of the family $(\psi_c(\xi) \mid c \in \Gamma)$. Since $\psi = \gamma^{-1}$, the Chain Rule shows that $\nabla \xi \psi$ is invertible with inverse $(\nabla \xi \psi)^{-1} = (\nabla \psi(\xi) \Gamma) = \nabla \psi(\xi) \Gamma$ for all $\xi \in \overline{\mathcal{D}}$. In view of (71.2), it follows that $(\psi_c(\xi) \mid c \in \Gamma)$ is a basis of $\mathcal{V}$ and, in view of (71.1), that

$$(\nabla \psi(\xi) \Gamma)^c_d(\xi) = \delta^c_d := \begin{cases} 0 & \text{if } c \neq d \\ 1 & \text{if } c = d \end{cases} \quad (71.3)$$

for all $c,d \in \Gamma$ and all $\xi \in \overline{\mathcal{D}}$. We use the notation

$$b_c := \psi_c \circ \gamma, \quad \beta^c := \nabla c \quad \text{for all } c \in \Gamma. \quad (71.4)$$

By (C1) and Prop.1, $b_c : \mathcal{D} \rightarrow \mathcal{V}$ and $\beta^c : \mathcal{D} \rightarrow \mathcal{V}^*$ are of class $C^1$ for all $c \in \Gamma$. It follows from (71.3) that

$$\beta^c b_d = \delta^c_d \quad \text{for all } c,d \in \Gamma. \quad (71.5)$$

By Prop.4 of Sect.23, we conclude that for each $x \in \mathcal{D}$, the family $\beta(x) := (\beta^c(x) \mid c \in \Gamma)$ in $\mathcal{V}^*$ is the dual basis of the basis $b(x) := (b_c(x) \mid c \in \Gamma)$ of $\mathcal{V}$. We call $b : \mathcal{D} \rightarrow \mathcal{V}^*$ the **basis field** and $\beta : \mathcal{D} \rightarrow \mathcal{V}^*\Gamma$ the **dual basis field** of the given coordinate system $\Gamma$. 
Remark: The reason for using superscripts rather than subscripts as indices to denote the terms of certain families will be explained in Sect.73. The placing of the indices is designed in such a way that in most of the summations that occur, the summation dummy is used exactly twice, once as a superscript and once as a subscript.

We often use the word “field” for a mapping whose domain is $D$ or an open subset of $D$. If the codomain is $\mathbb{R}$, we call it a scalar field; if the codomain is $\mathcal{V}$, we call it a vector field; if the codomain is $\mathcal{V}^*$, we call it a covector field; and if the codomain is Lin$\mathcal{V}$, we call it a lineon field. If $h$ is a vector field, we define the component-family

$$[h] := ([h]^c | c \in \Gamma) \in (\text{Map}(\text{Dom } h, \mathbb{R}))^\Gamma$$

of $h$ relative to the given coordinate system by $[h](x) := \ln c_{b(x)}^{-1} h(x)$ for all $x \in \text{Dom } h$, so that

$$h = \sum_{d \in \Gamma} [h]^d b_d, \quad [h]^c = \beta^c h \quad \text{for all } c \in \Gamma. \quad (71.6)$$

If $\eta$ is a covector field, we define the component family $[\eta] := ([\eta]^c | c \in \Gamma)$ of $\eta$ by $[\eta](x) := \ln c_{\beta(x)}^{-1} \eta(x)$ for all $x \in \text{Dom } \eta$, so that

$$\eta = \sum_{d \in \Gamma} [\eta]^d \beta^d, \quad [\eta]^c = \eta b_c \quad \text{for all } c \in \Gamma. \quad (71.7)$$

If $T$ is a lineon field, we define the component-matrix $[T] := ([T]^c_{d} | (c, d) \in \Gamma^2)$ of $T$ by

$$[T](x) := (\ln c_{(b_c \otimes \beta^d)(c, d) \in \Gamma^2})^{-1} T(x)$$

for all $x \in \text{Dom } T$, so that

$$T = \sum_{(c, d) \in \Gamma^2} [T]^c_{d} (b_c \otimes \beta^d). \quad (71.8)$$

The matrix $[T]$ is given by

$$[T]^c_{d} = \beta^d T b_d \quad \text{for all } c, d \in \Gamma \quad (71.9)$$

and characterized by

$$T b_c = \sum_{d \in \Gamma} [T]^d_{c} b_d \quad \text{for all } c \in \Gamma. \quad (71.10)$$
CHAPTER 7. COORDINATE SYSTEMS

In general, if \( F \) is a field whose codomain is a linear space \( W \) constructed from \( V \) by some natural construction, we define the component family \([F]\) of \( F \) as follows: For each \( x \in \text{Dom} F \), \([F](x)\) is the family of components of \( F \) relative to the basis of \( W \) induced by the basis \( b(x) \) of \( V \) by the construction of \( W \). For example, if \( F \) is a field with codomain \( \text{Lin}(V^*, V) \) then the component-matrix \([F] := ([F]_{cd} \mid (c, d) \in \Gamma^2)\) of \( F \) is determined by

\[
F = \sum_{(c,d) \in \Gamma^2} [F]_{cd}(b_c \otimes b_d). \quad (71.11)
\]

It is given by

\[
[F]_{cd} = \beta_c F \beta_d \quad \text{for all } c, d \in \Gamma \quad (71.12)
\]

and characterized by

\[
F \beta_c = \sum_{d \in \Gamma} [F]_{dc} b_d \quad \text{for all } c \in \Gamma. \quad (71.13)
\]

A field of any type is continuous, differentiable, or of class \( C^1 \) if and only if all of its components have the corresponding property.

We say that \( \Gamma \) is a flat coordinate system if the members of \( \Gamma \) are all flat functions or restrictions thereof. If also \( D = E \), then \( \gamma = \Gamma : E \to \mathbb{R}^\Gamma \) is a flat isomorphism. The point \( q := \psi(0) = \gamma^{-1}(0) \in \mathcal{E} \) is then called the origin of the given flat coordinate system. A flat coordinate system can be specified by prescribing the origin \( q \in \mathcal{E} \) and a set basis \( b \) of \( V \). Then each member \( \lambda \) of the dual basis \( b^* \) can be used to specify a coordinate \( c : \mathcal{E} \to \mathbb{R} \) by \( c(x) := \lambda(x - q) \) for all \( x \in \mathcal{E} \).

The basis field \( b \) (or its dual \( \beta \)) is constant if and only if the coordinate system is flat.

Let \( \xi \in \mathcal{D} \subset \mathbb{R}^\Gamma \) and \( c \in \Gamma \) be given and let \( \psi(\xi, c) \) be defined according to (04.25). It is easily seen that \( \text{Dom} \psi(\xi, c) \) is an open subset of \( \mathbb{R} \). We call \( \text{Rng} \psi(\xi, c) \) the coordinate curve through the point \( \psi(\xi) \) corresponding to the coordinate \( c \). If \( \Gamma \) is a flat coordinate system, then the coordinate curves are straight lines.

Pitfall: It is easy to give examples of non-flat coordinate systems which are “rectilinear” in the sense that all coordinate curves are straight lines. Thus, there is a distinction between non-flat coordinate systems and “curvilinear coordinate systems”, i.e. coordinate systems having some coordinate curves that are not straight.
(1) In all discussions of the theory of coordinate systems I have seen in the literature, the coordinates are assumed to be enumerated so as to form a list \( \{ c^i \mid i \in \mathbb{N} \} \), \( n := \dim E \). The numbers in \( \mathbb{N} \) are then used as indices in all formulas involving coordinates and components. However, when dealing with a specific coordinate system, most authors do not follow their theory but instead use the coordinates themselves as indices. I believe that one should do in theory what one does in practice and always use the set of coordinates itself as the index set. Enumeration of the coordinates is artificial and serves no useful purpose.

(2) Many textbooks contain a lengthy discussion about the transformation from one coordinate system to another. I believe that such “coordinate transformations” serve no useful purpose and may lead to enormous unnecessary calculations. In a specific situation, one should choose from the start a coordinate system that fits the situation and then stick with it. Even if a second coordinate system is considered, as is useful on rare occasions, one does not need “coordinate transformations”.

## 72 Connection Components

### Components of Gradients

Let a coordinate system \( \Gamma \) on an open subset \( D \) of a flat space \( E \) with translation space \( V \) be given. As we have seen in the previous section, the basis field \( b \) and the dual basis field \( \beta \) of the system \( \Gamma \) are of class \( C^1 \).

We use the notations

\[
C_{c \to e}^d := \beta^d(\nabla b_c)b_e \quad \text{for all} \quad c, d, e \in \Gamma \tag{72.1}
\]

and

\[
D_c := \text{div } b_c \quad \text{for all} \quad c \in \Gamma. \tag{72.2}
\]

The functions \( C_{c \to e}^d : D \to \mathbb{R} \) and \( D_c : D \to \mathbb{R} \) are called the \textbf{connection components} and the \textbf{deviation components}, respectively, of the system \( \Gamma \).

The connection components give the component matrices of the gradients not only of the terms of the basis field \( b \) but also of its dual \( \beta \):

**Proposition 1:** For each coordinate \( c \in \Gamma \), the component matrices of \( \nabla b_c \) and \( \nabla \beta^e \) are given by

\[
[\nabla b_c] = (C_{c \to e}^d \mid (d, e) \in \Gamma^2) \tag{72.3}
\]

and

\[
[\nabla \beta^e] = (-C_{d \to e}^c \mid (d, e) \in \Gamma^2). \tag{72.4}
\]
CHAPTER 7. COORDINATE SYSTEMS

**Proof:** We obtain (72.3) simply by comparing (72.1) with (71.9). To derive (72.4), we use the product rule (66.6) when taking the gradient of (71.5) and obtain

\[(\nabla \beta^c)^\top b_d + (\nabla b_d)^\top \beta^c = 0 \quad \text{for all } d \in \Gamma.\]

Value-wise operation on \(b_e\) gives

\[b_d(\nabla \beta^c)b_e + \beta^c(\nabla b_d)b_e = 0 \quad \text{for all } d, e \in \Gamma.\]

The result (72.4) follows now from (72.1) and the fact that the matrix \([\nabla \beta^c]\) is given by

\[\begin{bmatrix} \nabla \beta^c \end{bmatrix} = (b_d(\nabla \beta^c)b_e \mid (d, e) \in \Gamma^2).\]

**Proposition 2:** The connection components satisfy the symmetry relations

\[C^{d e c} = C^{ed c} \quad \text{for all } c, d, e \in \Gamma. \quad (72.5)\]

**Proof:** In view of the definition \(\beta^c := \nabla c\) we have \(\nabla \beta^c = \nabla \nabla c\). The assertion (72.5) follows from (72.4) and the Theorem on Symmetry of Second Gradients of Sect.611, applied to \(c\).

**Proposition 3:** The deviation components are obtained from the connection components by

\[D^c = \sum_{d \in \Gamma} C^{d e c} \quad \text{for all } c \in \Gamma. \quad (72.6)\]

**Proof:** By (72.2), (67.1), and (67.2), we have \(D^c(x) = \text{tr}(\nabla_x b_c)\) for all \(x \in \mathcal{D}\). The desired result (72.6) then follows from (72.3) and (26.8).

Let \(\mathcal{W}\) be a linear space and let \(\mathbf{F}\) be a differentiable field with domain \(\mathcal{D}\) and with codomain \(\mathcal{W}\). We use the notation

\[\mathbf{F}_{:e} := (\mathbf{F} \circ \psi)_c \circ \gamma \quad \text{for all } c \in \Gamma, \quad (72.7)\]

which says that \(\mathbf{F}_{:e}\) is obtained from \(\mathbf{F}\) by first looking at the dependence of the values of \(\mathbf{F}\) on the coordinates, then taking the partial \(c\)-derivative, and then looking at the result as a function of the point. It is easily seen that

\[(\nabla \mathbf{F})b_c = \mathbf{F}_{:c} \quad \text{for all } c \in \Gamma. \quad (72.8)\]

If \(\mathbf{F} := f\) is a differentiable scalar field, then (72.8) states that the component family \([\nabla f]\) of the covector field \(\nabla f : \mathcal{D} \to \mathcal{V}^*\) is given by

\[\begin{bmatrix} \nabla f \end{bmatrix} = (f_{:c} \mid c \in \Gamma). \quad (72.9)\]
The connection components are all zero if the coordinate system is flat. If it is not flat, one must know what the connection components are in order to calculate the components of gradients of vector and covector fields.

**Proposition 4:** Let \( h \) be a differentiable vector field and \( \eta \) a differentiable covector field with component-families \([h]\) and \([\eta]\), respectively. Then the component-matrices of \( \nabla h \) and \( \nabla \eta \) have the terms

\[
[\nabla h]_{cd}^e = [h]_{c:d}^e + \sum_{e \in \Gamma} [h]^e_c C_{e:d}^c, \tag{72.10}
\]

and

\[
[\nabla \eta]_{cd} = [\eta]_{c:d} - \sum_{e \in \Gamma} [\eta]^e_c C_{e:d}^c \tag{72.11}
\]

for all \( c, d \in \Gamma \).

**Proof:** To obtain (72.10), one takes the gradient of \((71.6)_1\), uses the product rule \((66.5)\), the formula \((71.9)\) with \( T := \nabla h \), the formula \((72.9)\) with \( f := [h]^e_c \) and \((72.3)\). The same procedure, starting with \((71.7)_1\) and using \((72.4)\), yields \((72.11)\).

It is easy to derive formulas analogous to \((72.10)\) and \((72.11)\) for the components of gradients of other kinds of fields.

**Proposition 5:** Let \( \eta \) be a differentiable covector field. Then the component-matrix of \( \text{Curl} \eta := \nabla \eta - (\nabla \eta)^\top : \text{Dom} \eta \rightarrow \text{Lin}(\mathcal{V}, \mathcal{V}^*) \) (see Sect.611) is given by

\[
[\text{Curl} \eta]_{cd} = [\eta]_{c:d} - [\eta]_{d:c} \quad \text{for all } c, d \in \Gamma. \tag{72.12}
\]

**Proof:** By Def.1 of Sect.611, the components of \( \text{Curl} \eta \) are obtained from the components of \( \nabla \eta \) by

\[
[\text{Curl} \eta]_{cd} = [\nabla \eta]_{cd} - [\nabla \eta]_{dc} \quad \text{for all } c, d \in \Gamma.
\]

After substituting \((72.11)\), we see from \((72.5)\) that the terms involving the connection components cancel and hence that \((72.12)\) holds.

Note that the connection components do not appear in \((72.12)\).

**Proposition 6:** Let \( h \) be a differentiable vector field. The divergence of \( h \) is given, in terms of the component family \([h]\) of \( h \) and the deviation components, by

\[
\text{div} h = \sum_{c \in \Gamma} ([h]_{c:e}^c + [h]_{c:D_c}). \tag{72.13}
\]

**Proof:** Using \((71.6)_1\) and Prop.1 of Sect.67, we obtain

\[
\text{div} h = \sum_{c \in \Gamma} ([\nabla ([h]^e_c)b_c + [h]^e_c \text{div} b_c).}
\]
Using (72.8) with $F_\Gamma := [h]^c$ and the definition (72.2) of $D_c$, we obtain the desired result (72.13).

**Proposition 7:** Let $F$ be a differentiable field with codomain $\text{Lin}(V^*, V)$. Then the component family of $\text{div} F : \text{Dom} F \to V$ (see Def. 1 of Sect. 67) is given by

$$[\text{div} F]^c = \sum_{d \in \Gamma} ([F]^{cd};_d + [F]^{cd}D_d) + \sum_{(e,d) \in \Gamma^2} [F]^{ed}C^c_{ed}$$

(72.14)

for all $c \in \Gamma$.

**Proof:** Let $c \in \Gamma$ be given. Then $\beta^c F$ is a differentiable vector field. If we apply Prop. 2 of Sect. 67 with $W := V$, $H := F$, and $\rho := \beta^c$, we obtain

$$\text{div}(\beta^c F) = \beta^c \text{div} F + \text{tr}(F^T \nabla \beta^c).$$

(72.15)

It follows from (72.4), Prop. 5 of Sect. 16, (23.9), and (26.8) that

$$\text{tr}(F^T \nabla \beta^c) = -\sum_{(e,d) \in \Gamma^2} [F]^{ed}C^c_{ed},$$

and from Prop. 6 that

$$\text{div}(\beta^c F) = \sum_{d \in \Gamma} ([\beta^c F]^d;_d + [\beta^c F]^dD_d).$$

Substituting these two results into (72.15) and observing that $[\text{div} F]^c = \beta^c \text{div} F$ and $[\beta^c F]^d = [F]^{cd}$ for all $d \in \Gamma$, we obtain the desired result (72.14).

Assume now that the coordinate system $\Gamma$ is flat and hence that its basis field $b$ and the dual basis field $\beta$ are constant. By (72.1) and (72.2), the connection components and deviation components are zero. The formulas (72.10) and (72.11) for the components of the gradients of a differentiable vector field $h$ and a differentiable covector field $\eta$ reduce to

$$[\nabla h]^c_{;d} = [h]^c_{;d}, \quad [\nabla \eta]_{cd} = [\eta]_{cd},$$

(72.16)

valid for all $c, d \in \Gamma$. The formula (72.13) for the divergence of a differentiable vector field $h$ reduces to

$$\text{div} h = \sum_{c \in \Gamma} [h]^c_{;c},$$

(72.17)

and the formula (72.14) for the divergence of a differentiable field $F$ with values in $\text{Lin}(V^*, V)$ becomes

$$[\text{div} F]^c = \sum_{d \in \Gamma} [F]^{cd;}_d \quad \text{for all} \quad c \in \Gamma.$$
Notes 72

(1) The term “connection components” is used here for the first time. More traditional terms are “Christoffel components”, “Christoffel symbols”, “three-index symbols”, or “Gamma symbols”. The last three terms are absurd because it is not the symbols that matter but what they stand for. The notations \( \Gamma_{\alpha \beta}^{\gamma} \) instead of our \( C_{c}^{\epsilon} \) are often found in the literature.

(2) The term “deviation components” is used here for the first time. I am not aware of any other term used in the literature.

(3) The components of gradients as described, for example, in Prop.4 are often called “covariant derivatives”.

73 Coordinates in Euclidean Spaces

We assume that a coordinate system \( \Gamma \) on an open subset of a (not necessarily genuine) Euclidean space \( \mathcal{E} \) with translation space \( \mathcal{V} \) is given. Since \( \mathcal{V} \) is an inner-product space, we identify \( \mathcal{V}^* \cong \mathcal{V} \), so that the basis field \( b \) as well as the dual basis field \( \beta \) of \( \Gamma \) have terms that are vector fields of class \( C^1 \). We use the notations

\[
G_{cd} := b_c \cdot b_d \quad \text{for all } c, d \in \Gamma \tag{73.1}
\]

and

\[
\overline{G}_{cd} := \beta_c \cdot \beta^d \quad \text{for all } c, d \in \Gamma, \tag{73.2}
\]

so that \( G := (G_{cd} | (c, d) \in \Gamma^2) \) and \( \overline{G} := (\overline{G}_{cd} | (c, d) \in \Gamma^2) \) are symmetric matrices whose terms are scalar fields of class \( C^1 \). These scalar fields are called the inner-product components of the system \( \Gamma \). In terms of the notation (41.13), \( G \) and \( \overline{G} \) are given by

\[
G(x) = G_{b(x)}, \quad \overline{G}(x) = G_{\beta(x)} \tag{73.3}
\]

for all \( x \in \mathcal{D} \). Since \( \beta(x) \) is the dual basis of \( b(x) \) for each \( x \in \mathcal{D} \), we obtain the following result from (41.16) and (41.15).

**Proposition 1:** For each \( x \in \mathcal{D} \), the matrix \( \overline{G}(x) \) is the inverse of \( G(x) \) and hence we have

\[
\sum_{\epsilon \in \Gamma} G_{\epsilon c} \overline{G}_{\epsilon d} = \delta_d^c \quad \text{for all } c, d \in \Gamma. \tag{73.4}
\]
The basis fields $\mathbf{b}$ and $\beta$ are related by

$$
\mathbf{b}_c = \sum_{d \in \Gamma} G_{cd} \beta^d, \quad \beta^c = \sum_{d \in \Gamma} \overline{G}^d_{cd} \mathbf{b}_d
$$

(73.5)

for all $c \in \Gamma$.

The following result shows that the connection components can be obtained directly from $G$ and $\overline{G}$.

**Theorem on Connection Components:** The connection components of a coordinate system on an open subset of a Euclidean space can be obtained from the inner-product components of the system by

$$
C_{cde} = \frac{1}{2} \sum_{f \in \Gamma} \overline{G}^{df}_{cd} (G_{fce} + G_{fde} - G_{efe})
$$

(73.6)

for all $c, d, e \in \Gamma$.

**Proof:** We use the abbreviation

$$
B_{cde} := \mathbf{b}_d \cdot (\nabla \mathbf{b}_c) \mathbf{b}_e \quad \text{for all} \quad c, d, e \in \Gamma.
$$

(73.7)

By (72.1) and Prop.1 we have

$$
C_{cde} = \frac{1}{2} \sum_{f \in \Gamma} \overline{G}^{df}_{cd} B_{cfe}, \quad B_{cde} = \sum_{f \in \Gamma} G_{df} C_{cfe}
$$

(73.8)

for all $c, d, e \in \Gamma$. We note that, in view of (73.8) and (72.5),

$$
B_{cde} = B_{edc} \quad \text{for all} \quad c, d, e \in \Gamma.
$$

(73.9)

Let $c, d, e \in \Gamma$ be given. Applying the Product Rule (66.9) to (73.1), we find that

$$
\nabla G_{cd} = (\nabla \mathbf{b}_c) ^\top \mathbf{b}_d + (\nabla \mathbf{b}_d) ^\top \mathbf{b}_c.
$$

Hence, by (72.8), (41.10), and (73.7) we obtain

$$
G_{cde} = (\nabla G_{cd}) \cdot \mathbf{b}_e = B_{cde} + B_{dce}.
$$

We can view this as a system of equations which can be solved for $B_{cde}$ as follows: Since $c, d, e \in \Gamma$ were arbitrary, we can rewrite the system with $c, d, e$ cyclically permuted and find that

$$
G_{cde} = B_{cde} + B_{dce},
$$

$$
G_{dec} = B_{dec} + B_{edc},
$$
are valid for all $c, d, e \in \Gamma$. If we subtract the last of these equations from the sum of the first two and observe (73.9), we obtain

$$G_{ced} + G_{dce} - G_{ecd} = 2B_{cde},$$

valid for all $c, d, e \in \Gamma$. Using (73.8), we conclude that (73.6) holds. 

For the following results, we need the concept of determinant, which will be explained only in Vol.II of this treatise. However, most readers undoubtedly know how to compute determinants of square matrices of small size, and this is all that is needed for the application of the results to special coordinate systems. We are concerned here with the determinant for $\mathbb{R}^\Gamma$. In Vol.II we will see that it is a mapping $\det: \text{Lin} \mathbb{R}^\Gamma \to \mathbb{R}$ of class $C^1$ whose gradient satisfies

$$(\nabla M \det)N = \det(M)\text{tr}(M^{-1}N) \quad (73.10)$$

for all $M \in \text{Lin} \mathbb{R}^\Gamma$ and all $N \in \text{Lin} \mathbb{R}^\Gamma$. If $M \in \text{Lin} \mathbb{R}^\Gamma \cong \mathbb{R}^{\Gamma \times \Gamma}$ is a diagonal matrix (see Sect.02), then $\det(M)$ is the product of the diagonal of $M$, i.e. we have

$$\det(M) = \prod_{c \in \Gamma} M_{cc}. \quad (73.11)$$

We have $\det(M) \neq 0$ if and only if $M \in \text{Lin} \mathbb{R}^\Gamma$ is invertible. Hence, since $G$ has invertible values by Prop.1, $\det \circ G : \mathcal{D} \to \mathbb{R}$ is nowhere zero.

**Theorem on Deviation Components:** The deviation components of a coordinate system on an open subset of a Euclidean space can be obtained from the determinant of the inner-product matrix of the system by

$$D_c = 2 \left( \frac{\det \circ G}{\det \circ G} \right)_c c \quad \text{for all} \quad c \in \Gamma. \quad (73.12)$$

**Proof:** By (73.10), the Chain Rule, and Prop.1 we obtain

$$(\det \circ G)_c = (\det \circ G)\text{tr}(\nabla (G_{c})). \quad (73.13)$$

On the other hand, by (73.6) and (72.6), we have

$$D_c = \frac{1}{2} \sum_{(d,f) \in \Gamma^2} \overline{G}^{df}(G_{fc;d} + G_{fd;c} - G_{cd;f}). \quad (73.14)$$

Since the matrices $G$ and $\overline{G}$ are symmetric we have

$$\sum_{(d,f) \in \Gamma^2} \overline{G}^{df} G_{fc;d} = \sum_{(f,d) \in \Gamma^2} \overline{G}^{fd} G_{dc;f} = \sum_{(d,f) \in \Gamma^2} \overline{G}^{df} G_{ed;f},$$
and hence, by (73.14), Prop. 5 of Sect. 16, and (26.8),

\[ D_c = \frac{1}{2} \sum_{(d,f) \in \Gamma^2} \overline{G}^{df} G_{f,c} = \frac{1}{2} \text{tr}(\overline{G}G_{:c}). \]

Comparing this result with (73.13), we obtain (73.12). \[\Box\]

In practice, it is useful to introduce the function

\[ g := \sqrt{|\det \circ G|}, \tag{73.15} \]

which is of class \( C^1 \) because \( \det \circ G \) is nowhere zero. Since \( \det \circ G = \pm g^2 \), (73.12) is equivalent to

\[ D_c = \frac{g^c}{g} \text{ for all } c \in \Gamma. \tag{73.16} \]

**Remark:** One can show that the sign of \( \det \circ G \) depends only on the signature of the inner product space \( \mathcal{V} \) (see Sect. 47). In fact \( \det \circ G \) is strictly positive if \( \text{sig}^{-\mathcal{V}} \) is even, strictly negative if \( \text{sig}^{-\mathcal{V}} \) is odd. In particular, if \( \mathcal{E} \) is a genuine Euclidean space, \( \det \circ G \) is strictly positive and the absolute value symbols can be omitted in (73.15). \[\Box\]

The identification \( \mathcal{V} \cong \mathcal{V}^* \) makes the distinction between vector fields and covector fields disappear. This creates ambiguities because a vector field \( h \) now has two component families, one with respect to the basis field \( b \) and the other with respect to the dual basis field \( \beta \). Thus, the symbol \( [h] \) becomes ambiguous. For the terms of \( [h] \), the ambiguity is avoided by careful attention to the placing of indices as superscripts or subscripts. Thus, we use \( ([h]^c | c \in \Gamma) := \beta \cdot h \) for the component family of \( h \) relative to \( b \) and \( ([h]_c | c \in \Gamma) := b \cdot h \) for the component family of \( h \) relative to \( \beta \) (see also (41.17)). The symbol \( [h] \) by itself can no longer be used. It follows from (73.5) that the two types of components of \( h \) are related by the formulas

\[ [h]^c = \sum_{d \in \Gamma} \overline{G}^{cd} [h]_d, \quad [h]_c = \sum_{d \in \Gamma} G_{cd}[h]^d, \tag{73.17} \]

valid for all \( c \in \Gamma \). To avoid clutter, we often omit the brackets and write \( h^c \) for \([h]^c \) and \( h_c \) for \([h]_c \) if no confusion can arise.

A lineon field \( T \) has four component matrices because of the identifications \( \text{Lin}\mathcal{V} \cong \text{Lin}\mathcal{V}^* \cong \text{Lin}(\mathcal{V}, \mathcal{V}^*) \cong \text{Lin}(\mathcal{V}^*, \mathcal{V}) \). The resulting ambiguity is avoided again by careful attention to the placing of indices as superscripts and subscripts. The four types of components are given by the formulas
valid for all \( c, d \in \Gamma \). The various types of components are related to each other by formulas such as

\[
[T]^{cd} = \sum_{e \in \Gamma} G^{de} [T]^e_c = \sum_{(e,f) \in \Gamma^2} \frac{G^{cf}}{G^{ef}} [T]^{ef},
\]

valid for all \( c, d \in \Gamma \). Again, we often omit the brackets to avoid clutter.

Using (73.16) and the product rule, we see that Props. 6 and 7 of Sect. 72 have the following corollaries.

**Proposition 2:** The divergence of a differentiable vector field \( \mathbf{h} \) is given by

\[
\text{div} \, \mathbf{h} = \frac{1}{g} \sum_{c \in \Gamma} (g[h]^c)_c.
\]  

**Proposition 3:** The components of the divergence of a differentiable lineon field \( \mathbf{T} \) are given by

\[
[\text{div} \, \mathbf{T}]^c = \frac{1}{g} \sum_{d \in \Gamma} (g[T]^{cd})_d + \sum_{(e,d) \in \Gamma^2} \frac{G^{cd}}{G^e} C^e_c d
\]

for all \( c \in \Gamma \).

Using Def. 2 of Sect. 67 and observing (72.9) and (73.17)1, we obtain the following immediate consequence of Prop. 2.

**Proposition 4:** The Laplacian of a twice differentiable scalar field \( f \) is given by

\[
\Delta f = \frac{1}{g} \sum_{(c,d) \in \Gamma^2} (gG^{cd} f)_c.
\]

Notes 73

1. If \( \mathbf{h} \) is a vector field, the components \([h]^c\) of \( \mathbf{h} \) relative to the basis field \( \mathbf{b} \) are often called the “contravariant components” of \( \mathbf{h} \), and the components \([h]_c\) of \( \mathbf{h} \) relative to the dual basis field \( \mathbf{\beta} \) are then called the “covariant components” of \( \mathbf{h} \). (See also Note (5) to Sect. 41.)

2. If \( \mathbf{T} \) is a lineon field, the components \([\mathbf{T}]_{cd}\) and \( T^{cd} \) are often called the “covariant components” and “contravariant components”, respectively, while \([\mathbf{T}]^{cd}\) and \([\mathbf{T}]^c_d\) are called “mixed components” of \( \mathbf{T} \).
74 Special Coordinate Systems

In this section a genuine Euclidean space $\mathcal{E}$ with translation space $\mathcal{V}$ is assumed given.

(A) Cartesian Coordinates: A flat coordinate system $\Gamma$ is called a Cartesian coordinate system if $(\nabla c \mid c \in \Gamma)$ is a (genuine) orthonormal basis of $\mathcal{V}$. To specify a Cartesian coordinate system, we may prescribe a point $q \in \mathcal{E}$ and an orthonormal basis set $\mathcal{e}$ of $\mathcal{V}$. For each $e \in \mathcal{e}$, we define a function $c : \mathcal{E} \to \mathbb{R}$ by

$$c(x) := e \cdot (x - q) \quad \text{for all } x \in \mathcal{E}.$$  \hfill (74.1)

The set $\Gamma$ of all functions $c$ defined in this way is then a Cartesian coordinate system on $\mathcal{E}$ with origin $q$. Since $\mathcal{e}$ is orthonormal, we have $b^c = b_c$ for all $c \in \Gamma$. The mapping $\Gamma : \mathcal{E} \to \mathbb{R}^\Gamma$ defined by $\Gamma(x) := (c(x) \mid c \in \Gamma)$ for all $x \in \mathcal{E}$ is invertible and its inverse $\psi : \mathbb{R}^\Gamma \to \mathcal{E}$ is given by

$$\psi(\xi) = q + \sum_{c \in \Gamma} \xi_c b_c \quad \text{for all } \xi \in \mathbb{R}^\Gamma.$$  \hfill (74.2)

The matrices of the inner-product components of $\Gamma$ are constant and given by $G = \overline{G} = 1_{\mathbb{R}^\Gamma}$. If $\mathbf{h}$ is a vector field, then $[\mathbf{h}]^c = [\mathbf{h}]_c$ for all $c \in \Gamma$ and if $\mathbf{T}$ is a lineon field, then $[\mathbf{T}]^c_d = [\mathbf{T}]^{cd} = [\mathbf{T}]_c^d = [\mathbf{T}]_{cd}$ for all $c, d \in \Gamma$. Thus, all indices can be written as subscripts without creating ambiguity. Since $\det \circ G = 1$ and hence $g = 1$, the formulas (73.20), (73.22), and (73.21) reduce to

$$\text{div } \mathbf{h} = \sum_{c \in \Gamma} [\mathbf{h}]_{cc},$$

$$\Delta f = \sum_{c \in \Gamma} f_{cc},$$

$$[\text{div } \mathbf{T}]_c = \sum_{d \in \Gamma} [\mathbf{T}]_{cd},$$  \hfill (74.3)

respectively.

(B) Polar Coordinates: We assume that $\dim \mathcal{E} = 2$. We prescribe a point $q \in \mathcal{E}$, a lineon $\mathbf{J} \in \text{Skew}\mathcal{V} \cap \text{Orth}\mathcal{V}$ (see Problem 2 of Chap.4) and a unit vector $e \in \mathcal{V}$. ($\mathbf{J}$ is a perpendicular turn as defined in Sect.87.) We define $\mathbf{h} : \mathbb{R} \to \mathcal{V}$ by

$$\mathbf{h} := \exp_{\mathcal{V}} \circ (\iota \mathbf{J}) e = \cos e + \sin \mathbf{Je}$$  \hfill (74.4)
(see Problem 9 of Chap. 6). Roughly, \( h(t) \) is obtained from \( e \), by a rotation with angle \( t \). We have

\[
h^\bullet(t) = Jh(t) \quad \text{for all} \quad t \in \mathbb{R},
\]

and hence

\[
h \cdot h^\bullet = 0, \quad |h| = |h^\bullet| = 1.
\]

We now consider the mapping \( \Psi : \mathbb{P}^x \times \mathbb{R} \rightarrow \mathcal{E} \) defined by

\[
\Psi(s, t) := q + sh(t) \quad \text{for all} \quad (s, t) \in \mathbb{P}^x \times \mathbb{R}.
\]

It is clear that \( \Psi \) is of class \( C^1 \), and, in view of (65.11), its gradient is given by

\[
\nabla_{(s, t)} \Psi = \ln_c(h(t), sh^\bullet(t)) \quad \text{for all} \quad (s, t) \in \mathbb{P}^x \times \mathbb{R}.
\]

It is clear from (74.6) that \( (h(t), sh^\bullet(t)) \) is a basis of \( V \) for every \( (s, t) \in \mathbb{P}^x \times \mathbb{R} \) and hence that \( \nabla \Psi \) has only invertible values. It follows from the Local Inversion Theorem of Sect. 68 that \( \Psi \) is locally invertible. It is not injective because \( \Psi(s, t) = \Psi(s, t + 2\pi) \) for all \( (s, t) \in \mathbb{P}^x \times \mathbb{R} \). An invertible adjustment of \( \Psi \) with open domain is \( \psi := \Psi|_{\overline{D}} \), where

\[
\overline{D} := \mathbb{P}^x \times ]0, 2\pi[, \quad D := \mathcal{E} \setminus (q + \mathbb{P}e).
\]

We put \( \gamma := \psi^- \) and define \( r : D \rightarrow \mathbb{R} \) and \( \theta : D \rightarrow \mathbb{R} \) by \( (r, \theta) := \gamma|^{\mathbb{R}^2} \). Then \( \Gamma := \{r, \theta\} \) is called a **polar coordinate system**. The function \( r \) is given by

\[
r(x) = |x - q| \quad \text{for all} \quad x \in D.
\]

and the function \( \theta \) is characterized by \( \text{Rng} \theta = ]0, 2\pi[ \) and

\[
h(\theta(x)) = \frac{x - q}{r(x)} \quad \text{for all} \quad x \in D.
\]

The values of the coordinates of a point \( x \) are indicated in Fig. 1; the argument \( x \) is omitted to avoid clutter.
If we interpret the first term in a pair as the r-term and the second as the θ-term of a family indexed on Γ = \{r, θ\}, then the mappings γ and ψ above coincide with the mappings denoted by the same symbols in Sect. 71. Since ψ is an adjustment of the mapping Ψ defined by (74.7), it follows from (71.4) that
\[ b_r = h \circ θ, \quad b_θ = r(h^* \circ θ). \] (74.11)

By (74.6) we have \( b_r \cdot b_θ = 0 \), \( |b_r| = 1 \), and \( |b_θ| = r \) (see Fig. 1). Hence the inner-product components (73.1) are
\[ G_{rr} = 0, \quad G_{θθ} = r^2. \] (74.12)

The value-wise inverse \( G \) of the matrix \( G \) is given by
\[ G^{rr} = 0, \quad G^{θθ} = \frac{1}{r^2}. \] (74.13)

By (73.11) we have \( \det G = r^2 \), and hence \( g := \sqrt{|\det G|} \) becomes
\[ g = r. \] (74.14)

Using the Theorem on Connection Components of Sect. 73, we find that the only non-zero connection components are
\[ C_{θr} = C_{rθ} = \frac{1}{r}, \quad C_{θθ} = -r. \] (74.15)

The formula (73.20) for the divergence of a differentiable vector field \( h \) becomes
\[ \text{div } h = \frac{1}{r}(r[h]^r)_{,r} + [h]^{θ,θ}. \] (74.16)
The formula (73.22) for Laplacian of a twice differentiable scalar field $f$ becomes
\[ \Delta f = \frac{1}{r} (r f_r)_r + \frac{1}{r^2} f_{\theta\theta}. \] (74.17)
The formula (73.21) for the components of the divergence of a differentiable lineon field $T$ yields
\[ [\text{div } T]^r = \frac{1}{r} (r T^{rr})_r + T^{r\theta} \cdot \theta - r T^{\theta\theta}, \]
\[ [\text{div } T]^\theta = \frac{1}{r} (r T^{r\theta})_r + T^{\theta\theta} \cdot \theta + \frac{1}{r} (T^{r\theta} + T^{\theta r}). \] (74.18)
(On the right sides, brackets are omitted to avoid clutter.)

\textbf{Remark:} The adjustment of $\Psi$ described above is only one of several that yield a suitable definition of a polar coordinate system. Another would be obtained by replacing (74.8) by
\[ \mathcal{D} := \{0\} \times [0, \pi) \times \mathbb{R}, \quad \mathcal{D} := \mathcal{E} \setminus (q - \mathbb{F}e). \] (74.19)
In this case $\theta$ would be characterized by (74.10) and $\text{Rng } \theta = [\pi, \pi[.$

\textbf{(C) Cylindrical Coordinates:} We assume that $\dim \mathcal{E} = 3.$ We first prescribe a point $q \in \mathcal{E}$ and a unit vector $f \in \mathcal{V}.$ We then put $\mathcal{U} := \{f\}^\perp$ and prescribe a lineon $J \in \text{Skew} \mathcal{U} \cap \text{Orth} \mathcal{U}$ and a unit vector $e \in \mathcal{U}.$ We define $h : \mathbb{R} \to \mathcal{V}$ by (74.4) and consider the mapping $\Psi : \mathbb{P}^\times \times \mathbb{R} \times \mathbb{R} \to \mathcal{E}$ defined by
\[ \Psi(s, t, u) := q + sh(t) + uf. \] (74.20)
It is easily seen that $\Psi$ is of class $C^1$ and locally invertible but not injective. An invertible adjustment of $\Psi$ with open domain is $\psi := \Psi|_{\mathcal{D}^P}$ where
\[ \mathcal{D} := \mathbb{P}^\times \times [0, 2\pi) \times \mathbb{R}, \quad \mathcal{D} := \mathcal{E} \setminus (q + \mathbb{F}e + \mathbb{R}f). \] (74.21)
We put $\gamma := \psi^{-1}$ and define the functions $r, \theta, z$, all with domain $\mathcal{D}$, by
\[ (r, \theta, z) := \gamma|_{\mathbb{R}^3}. \] Then $\Gamma := \{r, \theta, z\}$ is called a \textbf{cylindrical coordinate system}. let $E$ be the symmetric idempotent for which $\text{Rng } E = \mathcal{U}$ (see Prop.4 of Sect.41). Then the functions $r$ and $z$ are given by
\[ r(x) = |E(x - q)|, \quad z(x) = f \cdot (x - q) \quad \text{for all } x \in \mathcal{D}, \] (74.22)
and $\theta$ is characterized by $\text{Rng } \theta = [0, 2\pi[$ and
\[ h(\theta(x)) = \frac{1}{r(x)} E(x - q) \quad \text{for all } x \in \mathcal{D}. \] (74.23)
The values of the coordinates of a point $x$ are indicated in Fig.2; the argument $x$ is omitted to avoid clutter.
If we interpret the first term in a triple as the \( r \)-term, the second as the \( \theta \)-term, and the third as the \( z \)-term of a family indexed on \( \Gamma = \{ r, \theta, z \} \), then the notation \( \gamma \) and \( \psi \) above is in accord with the one used in Sect. 1. By the same reasoning as used in Example (B), we infer from (74.20) that the basis field \( b \) of \( \Gamma \) is given by

\[
b_r = h \circ \theta, \quad b_\theta = r (h^* \circ \theta), \quad b_z = f. \tag{74.24}
\]

The matrices \( G \) and \( \overline{G} \) of the inner-product components of \( \Gamma \) are diagonal matrices and their diagonals are given by

\[
G_{rr} = 1, \quad G_{\theta \theta} = r^2, \quad G_{zz} = 1, \tag{74.25}
\]

\[
\overline{G}^{rr} = 1, \quad \overline{G}^{\theta \theta} = \frac{1}{r^2}, \quad \overline{G}^{zz} = 1. \tag{74.26}
\]

The relation (74.14) remains valid and the only non-zero connection components of \( \Gamma \) are again given by (74.15). The formulas (74.16), (74.17), and (74.18) must be replaced by

\[
\text{div } h = \frac{1}{r} (rh^r)_r + h^\theta_{,\theta} + h^z_{,z}, \tag{74.27}
\]

\[
\Delta f = \frac{1}{r} (rf)_r + \frac{1}{r^2} f_{,\theta \theta} + f_{,z z}. \tag{74.28}
\]
and

\[
\begin{align*}
[\text{div } \mathbf{T}]^r &= \frac{1}{r}(r\mathbf{T}_{rr})_r + \mathbf{T}_{r\theta}^\theta + \mathbf{T}_{rz}^z - r\mathbf{T}_{\theta\theta}, \\
[\text{div } \mathbf{T}]^\theta &= \frac{1}{r}(r\mathbf{T}_{r\theta})_r + \mathbf{T}_{\theta\theta}^\theta + \mathbf{T}_{\theta z}^z + \frac{1}{r}(\mathbf{T}_{r\theta} + \mathbf{T}_{\theta r}), \\
[\text{div } \mathbf{T}]^z &= \frac{1}{r}(r\mathbf{T}_{xr})_r + \mathbf{T}_{x\theta}^\theta + \mathbf{T}_{zx}^z.
\end{align*}
\]

(74.29)

(Brackets are omitted on the right sides.)

**(D) Spherical Coordinates:** We assume \( \dim \mathcal{E} = 3 \) and prescribe \( q, f, \) and \( J \) as in Example (C). We replace the formula (74.20) by

\[
\Psi(s, t, u) := q + s(\cos(t)f + \sin(t)h(u)).
\]

(74.30)

The definition of \( \mathcal{D} \) in (74.21) must be replaced by

\[
\mathcal{D} := \mathbb{P}^\times \times [0, \pi] \times [0, 2\pi],
\]

(74.31)

but \( \mathcal{D} \) remains the same. Then \( \psi := \Psi|_{\mathcal{D}} \) is invertible, and the functions \( r, \theta, \varphi, \) all with domain \( \mathcal{D} \), are defined by \( \gamma := \psi^{-1} \) and \( (r, \theta, \varphi) := \gamma|^{\mathbb{R}^3} \). Then \( \Gamma := \{ r, \theta, \varphi \} \) is called a spherical coordinate system. The functions \( r \) and \( \theta \) are given by

\[
r(x) = |x - q|, \quad \theta(x) = \arccos\left( \frac{f \cdot (x - q)}{r(x)} \right) \quad \text{for all } x \in \mathcal{D}
\]

(74.32)

and \( \varphi \) is characterized by \( \text{Rng } \varphi = ]0, 2\pi[ \) and

\[
h(\varphi(x)) = \frac{1}{r(x)\sin(\theta(x))}E(x - q) \quad \text{for all } x \in \mathcal{D}
\]

(74.33)

where \( E \) is defined as in Example (C). The values of the coordinates of a point \( x \) are indicated in Fig.3, again with arguments omitted.
By the same procedure as used in the previous examples, we find that the basis field $b$ of $\Gamma$ is given by
\begin{align*}
b_r &= (\cos \circ \theta)f + (\sin \circ \theta)(h \circ \varphi), \\
b_\theta &= r(-\sin \circ \theta)f + (\cos \circ \theta)(h \circ \varphi)), \\
b_\varphi &= r(\sin \circ \theta)(h \circ \varphi).
\end{align*}
(74.34)

The matrices $G$ and $\overline{G}$ of the inner-product components are again diagonal matrices and their diagonals are given by
\begin{align*}
G_{rr} &= 1, \\
G_{\theta\theta} &= r^2, \\
G_{\varphi\varphi} &= r^2(\sin \circ \theta)^2,
\end{align*}
(74.35)
\begin{align*}
\overline{G}_{rr} &= 1, \\
\overline{G}^{\theta\theta} &= \frac{1}{r^2}, \\
\overline{G}^{\varphi\varphi} &= \frac{1}{r^2(\sin \circ \theta)^2}.
\end{align*}
(74.36)

By (73.11) we have $\det \circ G = r^4(\sin \circ \theta)^2$ and hence $g := \sqrt{|\det \circ G|}$ becomes
\begin{equation}
g = r^2(\sin \circ \theta).
\end{equation}
(74.37)

Using (73.6), we find that the only non-zero connection components are
\begin{align*}
C^\theta_r &= C^\theta_r = C^\varphi_r = C^\varphi_r = \frac{1}{r}, \\
C^\theta_\theta &= -r, \\
C^\varphi_\varphi &= \frac{1}{\tan \circ \theta}, \\
C^\varphi_r &= -r(\sin \circ \theta)^2, \\
C^\varphi_\varphi &= -(\sin \circ \cos \circ \theta).
\end{align*}
(74.38)
The formulas (73.20) and (73.22) for the divergence and for the Laplacian specialize to

\[
\text{div } h = \frac{1}{r^2}(r^2h^r)_r + \frac{1}{\sin \theta}((\sin \theta)h^\theta)_{,\theta} + h^\varphi_{,\varphi}
\]

and

\[
\Delta f = \frac{1}{r^2}(r^2f)_r + \frac{1}{r^2(\sin \theta)}((\sin \theta)f)_{,\theta} + \frac{1}{r^2(\sin \theta)^2}f_{,\varphi,\varphi}.\]

the formula (73.21) for the divergence of a lineon field gives

\[
[\text{div } T]^r = \frac{1}{r^2}(r^2T^{r\gamma})_{,r} + \frac{1}{\sin \theta}((\sin \theta)T^{r\theta})_{,\theta} + T^{r\varphi}_{,\varphi}
\]

\[
- rT^{\theta\theta} - r(\sin \theta)^2T^{\varphi\varphi},
\]

\[
[\text{div } T]^\theta = \frac{1}{r^2}(r^2T^{\theta\gamma})_{,r} + \frac{1}{\sin \theta}((\sin \theta)T^{\theta\theta})_{,\theta} + T^{\theta\varphi}_{,\varphi}
\]

\[
+ \frac{1}{r}(T^{r\theta} + T^{\theta r}) - ((\sin \cos \theta)T^{\varphi\varphi},
\]

\[
[\text{div } T]^\varphi = \frac{1}{r^2}(r^2T^{\varphi\gamma})_{,r} + \frac{1}{\sin \theta}((\sin \theta)T^{\varphi\theta})_{,\theta} + T^{\varphi\varphi}_{,\varphi}
\]

\[
+ \frac{1}{r}(T^{r\varphi} + T^{\varphi r}) + \frac{1}{\tan \theta}(T^{\varphi\theta} + T^{\theta\varphi}.
\]

(Brackets are omitted on the right sides to avoid clutter.)

Notes

(1) Unfortunately, there is no complete agreement in the literature on what letters to use for which coordinates. Often, the letter \( \varphi \) is used for our \( \theta \) and vice versa. In cylindrical coordinates, the letter \( \rho \) is often used for our \( r \). In spherical coordinates, one sometimes finds \( \omega \) for our \( \varphi \).

(2) Most of the literature is very vague about how one should choose the domain \( D \) for each of the curvilinear coordinate systems discussed in this section.

75 Problems for Chapter 7

(1) Let \( \Gamma \) be a coordinate system as in Def.1 of Sect.71 and let \( \beta \) be the dual basis field of \( \Gamma \). Let \( p : I \rightarrow \mathcal{D} \) be a twice differentiable process on some interval \( I \in \text{Sub } \mathbb{R} \). Define the component-functions of \( p^* \) and \( p^{**} \) by
CHAPTER 7. COORDINATE SYSTEMS

\[ [p^*]^c := (\beta^c \circ p)^*, \quad [p^{**}]^c := (\beta^c \circ p)^{**} \]  
\( (P7.1) \)

for all \( c \in \Gamma \).

(a) Show that

\[ [p^*]^c = (c \circ p)^* \]  
\( (P7.2) \)

and

\[ [p^{**}]^c = (c \circ p)^{**} + \sum_{(d,e) \in \Gamma^2} (C_d^e \circ p)(d \circ p)^*(e \circ p)^* \]  
\( (P7.3) \)

for all \( c \in \Gamma \), where \( C \) denotes the family of connection components.

(b) Write out the formula \((P7.3)\) for cylindrical and spherical coordinates.

(2) Let \( \Gamma \) be a coordinate system as in Def.1 of Sect.71 and let \( b \) be the basis field and \( \beta \) the dual basis field of \( \Gamma \). If \( F \) is a field whose codomain is \( \text{Lin}(\mathcal{V}, \text{Lin}\mathcal{V}) \cong \text{Lin}_2(\mathcal{V}^2, \mathcal{V}) \), then the component family \( \{F\} \in (\text{Map}(\text{Dom}F, \mathbb{R}))^\Gamma \) of \( F \) is given by

\[ [F]_{cde} := \beta^c F(b_e, b_d) \quad \text{for all} \quad c, d, e \in \Gamma. \]  
\( (P7.4) \)

(a) Show: The components of the gradient \( \nabla T \) of a differentiable lineon field \( T \) are given by

\[ [\nabla T]_{cde} = [T]_{cde} + \sum_{f \in \Gamma} (\{T\}_f^c C_f^e - \{T\}_c^f C_f^e) \]  
\( (P7.5) \)

for all \( c, d, e \in \Gamma \).

(b) Show that if the connection components are of class \( C^1 \), they satisfy

\[ C_{c d e}^c - C_{c f e}^d + \sum_{g \in \Gamma} (C_{c g}^e C_g^d - C_{c g}^f C_g^d) = 0 \]  
\( (P7.6) \)

for all \( c, d, e, f \in \Gamma \). (Hint: Apply the Theorem on Symmetry of Second Gradients to \( b_c \) and use Part (a).)

(3) Let \( \Gamma \) be a coordinate system on an open subset \( \mathcal{D} \) of a genuine Euclidean space \( \mathcal{E} \) with translation space \( \mathcal{V} \) and let \( b \) be the basis field of
Assume that the system is orthogonal in the sense that \( b_c \cdot b_d = 0 \) for all \( c, d \in \Gamma \) with \( c \neq d \). Define 

\[
b_c := |b_c| \quad \text{for all} \quad c \in \Gamma \quad \text{(P7.7)}
\]

and

\[
e_c := \frac{b_c}{b_c} \quad \text{for all} \quad c \in \Gamma, \quad \text{(P7.8)}
\]

so that, for each \( x \in D \), \( e(x) := (e_c(x) \mid c \in \Gamma) \) is an orthonormal basis of \( V \). If \( h \) is a vector field with \( \text{Dom} \ h \subset D \), we define the family \( \langle h \rangle := ((h)_c \mid c \in \Gamma) \) of **physical components** of \( h \) by \( \langle h \rangle(x) := \ln e(x) h(x) \) for all \( x \in D \), so that

\[
\langle h \rangle_c = h \cdot e_c \quad \text{for all} \quad c \in \Gamma. \quad \text{(P7.9)}
\]

Physical components of fields of other types are defined analogously.

(a) Derive a set of formulas that express the connection components in terms of the functions \( b_c \) and \( b_{cd} \), \( c, d \in \Gamma \). (Hint: Use the Theorem on Connection Components of Sect.73.)

(b) Show that the physical components of a vector field \( h \) are related to the components \( [h]^c \) and \( [h]_c, c \in \Gamma \), by

\[
[h]_c = b_c(h)_c, \quad [h]^c = \frac{1}{b_c}(h)_c \quad \text{for all} \quad c \in \Gamma \quad \text{(P7.10)}
\]

(c) Show that the physical components of the gradient \( \nabla h \) of a differentiable vector field \( h \) are given by

\[
\langle \nabla h \rangle_{c,d} = \frac{1}{b_d} \langle h \rangle_{c,d} - \frac{b_{cd}}{b_d b_c} \langle h \rangle_d \quad \text{if} \quad c \neq d,
\]

\[
\langle \nabla h \rangle_{c,c} = \frac{1}{b_c} \langle h \rangle_{c,c} + \sum_{e \in \Gamma \setminus \{c\}} \frac{b_{ce}}{b_c b_e} \langle h \rangle_e \quad \text{(P7.11)}
\]

for all \( c, d \in \Gamma \).

(4) Let a 2-dimensional genuine inner-product space \( \mathcal{E} \) be given. Assume that a point \( q \in \mathcal{E} \), an orthonormal basis \( (e, f) \) and a number \( \varepsilon \in \mathbb{P}^\times \) have been prescribed. Consider the mapping \( \Psi : (\varepsilon + \mathbb{P}^\times) \times ]0, \varepsilon[ \to \mathcal{E} \) defined by
\[ \Psi(s, t) := q + \frac{1}{2}(st \mathbf{e} + \sqrt{(s^2 - \varepsilon^2)(\varepsilon^2 - t^2)} \mathbf{f}). \]  

(P7.12)

(a) Compute the partial derivatives \( \Psi_{,1} \) and \( \Psi_{,2} \) and show that \( \Psi \) is locally invertible.

(b) Show that \( \Psi \) is injective and hence that \( \mathcal{D} := \text{Rng} \, \Psi \) is open and that \( \psi := \Psi_{|\text{Rng}} \) is invertible.

(c) Put \( \gamma := \psi^{-1} \) and define the functions \( \lambda, \mu \) from \( \mathcal{D} \) to \( \mathbb{R} \) by \( \lambda, \mu := \gamma_{|\mathcal{D}} \). Show that \( \Gamma := \{\lambda, \mu\} \) is a coordinate-system on \( \mathcal{D} \); it is called an elliptical coordinate system.

(d) Show that the coordinate curves corresponding to the coordinates \( \lambda \) and \( \mu \) are parts of ellipses and hyperbolas, respectively, whose foci are \( q - \varepsilon \mathbf{e} \) and \( q + \varepsilon \mathbf{e} \). Show that \( \mathcal{D} = q + \mathbb{P}^\times \mathbf{e} + \mathbb{P}^\times \mathbf{f} \).

(e) Using Part (a), write down the basis field \( \{\mathbf{b}_i \mid i \in \{\lambda, \mu\}\} \) of the system \( \{\lambda, \mu\} \), and show that the inner-product components are given by

\[ G_{\lambda,\mu} = 0, \quad G_{\lambda,\lambda} = \frac{\lambda^2 - \mu^2}{\lambda^2 - \varepsilon^2}, \quad G_{\mu,\mu} = \frac{\lambda^2 - \mu^2}{\varepsilon^2 - \mu^2}. \]  

(P7.13)

(f) Show that the Laplacian of a twice differentiable scalar field \( f \) with \( \text{Dom} \, f \subset \mathcal{D} \) is given by

\[ \Delta f = \frac{\sqrt{(\lambda^2 - \varepsilon^2)(\varepsilon^2 - \mu^2)}}{\lambda^2 - \mu^2} \left( \left( \sqrt{\frac{\lambda^2 - \varepsilon^2}{\lambda^2 - \mu^2}} f_{,\lambda} \right)_{,\lambda} + \left( \sqrt{\frac{\varepsilon^2 - \mu^2}{\lambda^2 - \mu^2}} f_{,\mu} \right)_{,\mu} \right). \]  

(P7.14)

(g) Compute the connection components of the system \( \{\lambda, \mu\} \).

(5) Let a 3-dimensional genuine inner-product space \( \mathcal{E} \) with translation space \( \mathcal{V} \) be given. Assume that \( q, \mathbf{f} \) and \( \mathbf{J} \) are prescribed as in Example (C) of Sect.74 and that \( \mathbf{h} : \mathbb{R} \to \mathcal{V} \) is defined by (74.4). Define \( \Psi : \mathbb{P}^\times \times \mathbb{P}^\times \times \mathbb{R} \to \mathcal{E} \) by

\[ \Psi(s, t, u) := q + \frac{1}{2}(s^2 - t^2)\mathbf{f} + st \mathbf{h}(u). \]  

(P7.15)

(a) Compute the partial derivatives of \( \Psi \) and show that \( \Psi \) is locally invertible.

(b) Specify an open subset \( \overline{\mathcal{D}} \) of \( \mathbb{P}^\times \times \mathbb{P}^\times \times \mathbb{R} \) and an open subset \( \mathcal{D} \) of \( \mathcal{E} \) such that \( \psi := \Psi_{|\mathcal{D}} \) is invertible.
(c) Show that $\gamma := \psi^{-1}$ and $(\alpha, \beta, \theta) := \gamma|^{\mathbb{R}^3}$ define a coordinate system $\Gamma := \{\alpha, \beta, \theta\}$ on $\mathcal{D}$; it is called a **paraboloidal coordinate system**.

(d) Show that the coordinate curves corresponding to the coordinates $\alpha$ and $\beta$ are parabolas with focus $q$.

(e) Using Part (a), write down the basis field $(b_i \mid i \in \{\alpha, \beta, \theta\})$ of the system and compute the inner-product components.

(f) Find the formula for the Laplacian of a twice differentiable scalar field in paraboloidal coordinates.

(g) Compute the connection components of the system $\{\alpha, \beta, \theta\}$.

(6) Let $\Gamma$ be a coordinate system on an open subset $\mathcal{D}$ of a Euclidean space $\mathcal{E}$ and let $\beta$ be the dual basis field of $\Gamma$. Let $\widetilde{G}^{cd}, c, d \in \Gamma,$ be defined by (73.2) and $g$ by (73.15).

(a) Show that the Laplacian of the coordinate $c \in \Gamma$ is given by

$$\Delta c = \frac{1}{g} \sum_{d \in \Gamma} (gG^{dc})_{;d}.$$  \hspace{1cm} (P7.16)

(b) Given $c \in \Gamma$, show that the components of the Laplacian of $\beta^c$ are given by

$$[\Delta \beta^c]_d = \left( \frac{1}{g} \sum_{e \in \Gamma} (gG^{ec})_{;e} \right)_{;d}. \hspace{1cm} (P7.17)$$

(Hint: Use $\beta^c = \nabla c$ and Part (a)).

(c) Let $\mathbf{h}$ be a twice differentiable vector field with $\text{Dom} \mathbf{h} \subset \mathcal{D}$. Show that the components of the Laplacian of $\mathbf{h}$ are given by

$$[\Delta \mathbf{h}]_c = \triangle_c [\mathbf{h}] +$$

$$\sum_{(d,e) \in \Gamma^2} \left( \frac{1}{g} (g\widetilde{G}^{cd}{}_{;e})_{;e} [\mathbf{h}]_d - 2 \sum_{f \in \Gamma} C^d_{cf} \bar{G}^{bf} \triangle_c [\mathbf{h}]_{df} \right) \hspace{1cm} (P7.18)$$

for all $c \in \Gamma$, where $C^d_{cf}, c, d, f \in \Gamma,$ are the connection components of $\Gamma$. (Hint: Use Prop.5 of Sect.67.)

(d) Write out the formula $(P7.18)$ for cylindrical coordinates.
Chapter 8

Spectral Theory

In this chapter, it is assumed that all linear spaces under consideration are over the real field $\mathbb{R}$ and that all linear and inner-product spaces are finite-dimensional. However, in Sects. 89 and 810 we deal with linear spaces over $\mathbb{R}$ that are at the same time linear spaces over the complex field $\mathbb{C}$. Most of this chapter actually deals with a given finite-dimensional genuine inner-product space. Some of the definitions remain meaningful and some of the results remain valid if the space is infinite-dimensional or if $\mathbb{R}$ is replaced by an arbitrary field.

81 Disjunct Families, Decompositions

We assume that a linear space $\mathcal{V}$ and a finite family $(\mathcal{U}_i \mid i \in I)$ of subspaces of $\mathcal{V}$ are given. There is a natural summing mapping from the product space $\prod (\mathcal{U}_i \mid i \in I)$ (see Sect. 14) to $\mathcal{V}$, defined by

$$ (u \mapsto \Sigma_i u) : \prod_{i \in I} \mathcal{U}_i \to \mathcal{V}. \quad (81.1) $$

This summing mapping is evidently linear.

**Definition 1:** We say that the finite family $(\mathcal{U}_i \mid i \in I)$ of subspaces of $\mathcal{V}$ is **disjunct** if the summing mapping (81.1) of the family is injective; we say that the family is a **decomposition** of $\mathcal{V}$ if the summing mapping is invertible.

It is clear that every restriction of a disjunct family is again disjunct. In other words, if $(\mathcal{U}_i \mid i \in I)$ is disjunct and $J \in \text{Sub} I$, then $(\mathcal{U}_j \mid j \in J)$ is again disjunct. The union of a disjunct family cannot be the same as its sum unless all but at most one of the terms are zero-spaces. A disjunct family
and only if, for every family \( u \) zero, we have the following result:

\[
(304)
\]

**CHAPTER 8. SPECTRAL THEORY**

Let \( (\mathcal{U}_i \mid i \in I) \) be a finite family of subspaces of \( \mathcal{V} \). Then the following are equivalent:

(i) The family \( (\mathcal{U}_i \mid i \in I) \) is disjunct [a decomposition of \( \mathcal{V} \)].

(ii) For every \( j \in I \), \( \mathcal{U}_j \) and \( \sum(\mathcal{U}_i \mid i \in I \setminus \{ j \}) \) are disjunct [supplementary in \( \mathcal{V} \)].

(iii) If \( I \neq \emptyset \) then for some \( j \in I \), \( \mathcal{U}_j \) and \( \sum(\mathcal{U}_i \mid i \in I \setminus \{ j \}) \) are disjunct [supplementary in \( \mathcal{V} \)] and the family \( (\mathcal{U}_i \mid i \in I \setminus \{ j \}) \) is disjunct.

**Proof:** We prove only the assertions concerning disjunctness. The rest then follows from

\[
\sum(\mathcal{U}_i \mid i \in I) = \mathcal{U}_j + \sum(\mathcal{U}_i \mid i \in I \setminus \{ j \}),
\]

valid for all \( j \in I \).

(i) \( \Rightarrow \) (ii): Assume that \( (\mathcal{U}_i \mid i \in I) \) is disjunct and that \( j \in I \) and \( w \in \mathcal{U}_j \cap \sum(\mathcal{U}_i \mid i \in I \setminus \{ j \}) \) are given. Then we may choose \( (u_i \mid i \in I \setminus \{ j \}) \) such that \( w = \sum(u_i \mid i \in I \setminus \{ j \}) \) and hence \(-w + \sum(u_i \mid i \in I \setminus \{ j \}) = 0\). By Prop.1, it follows that \( w = 0 \). We conclude that \( \mathcal{U}_j \cap \sum(\mathcal{U}_i \mid i \in I \setminus \{ j \}) = \{0\} \).

(ii) \( \Rightarrow \) (i): Assume that \( (\mathcal{U}_i \mid i \in I) \) fails to be disjunct. By Prop.1 we can then find \( j \in I \) and \( w \in \mathcal{U}_j \) such that \( w + \sum(u_i \mid i \in I \setminus \{ j \}) = 0 \) for a suitable choice of \( u \in \bigtimes(\mathcal{U}_i \mid i \in I \setminus \{ j \}) \). Then \( w \in \mathcal{U}_j \cap \sum(\mathcal{U}_i \mid i \in I \setminus \{ j \}) \) and hence, since \( w \neq 0 \), \( \mathcal{U}_j \) and \( \sum(\mathcal{U}_i \mid i \in I \setminus \{ j \}) \) are not disjunct.

(i) \( \Rightarrow \) (iii): This follows from (i) \( \Rightarrow \) (ii) and the fact that a restriction of a disjunct family is disjunct.

(iii) \( \Rightarrow \) (i): Choose \( j \in I \) according to (iii) and let \( u \in \bigtimes(\mathcal{U}_i \mid i \in I) \) such that \( \sum(u_i \mid i \in I) = 0 \) be given. Then \( u_i = -\sum(u_i \mid i \in I \setminus \{ j \}) \) and hence \( u_j = 0 \) because \( \mathcal{U}_j \) and \( \sum(\mathcal{U}_i \mid i \in I \setminus \{ j \}) \) are disjunct. It follows
that $\sum (u_i \mid i \in I \setminus \{j\}) = 0$ and hence $u|_{I \setminus \{j\}} = 0$ because $(U_i \mid i \in I \setminus \{j\})$ is disjunct. We conclude that $u = 0$ and hence, by Prop.1, that $(U_i \mid i \in I)$ is disjunct. 

It follows from Prop.2 that a pair $(U_1, U_2)$ of subspaces of $V$ is disjunct if and only if $U_1$ and $U_2$ are disjunct [supplementary in $V$] in the sense of Def.2 of Sect.12.

The following result is an easy consequence of Prop.2.

**Proposition 3:** Let $(U_i \mid i \in I)$ be a disjunct family of subspaces of $V$, let $J$ be a subset of $I$, and let $W_j$ be a subspace of $U_j$ for each $j \in J$. Then $(W_j \mid j \in J)$ is again a disjunct family of subspaces of $V$. Moreover, if $(W_j \mid j \in J)$ is a decomposition of $V$, so is $(U_i \mid i \in I)$ and we have $W_j = U_j$ for all $j \in J$ and $U_i = \{0\}$ for all $i \in I \setminus J$.

Using Prop.2 above and Prop.4 of Sect.17 one easily obtains the following result by induction.

**Proposition 4:** A finite family $(U_i \mid i \in I)$ of subspaces of $V$ is disjunct if and only if

$$\dim(\sum_{i \in I} U_i) = \sum_{i \in I} \dim U_i. \quad (81.2)$$

In particular, if $(U_i \mid i \in I)$ is a decomposition of $V$, then

$$\dim V = \sum (\dim U_i \mid i \in I).$$

Let $(f_i \mid i \in I)$ be a finite family of non-zero elements in $V$. This family is linearly independent, or a basis of $V$, depending on whether the family $(\text{Lsp}\{f_i\} \mid i \in I)$ of one-dimensional subspaces of $V$ is disjunct, or a decomposition of $V$, respectively.

The following result generalizes Prop.4 of Sect.19 and is easily derived from that proposition and Prop.2.

**Proposition 5:** Let $(U_i \mid i \in I)$ be a decomposition of $V$. Then there is a unique family $(P_i \mid i \in I)$ of projections $P_i : V \to U_i$ such that

$$v = \sum_{i \in I} P_i v \quad \text{for all} \quad v \in V, \quad (81.3)$$

and we have

$$U_i \subset \text{Null } P_j \quad \text{for all} \quad i, j \in I \quad \text{with} \quad i \neq j. \quad (81.4)$$

Also, there is a unique family $(E_i \mid i \in I)$ of idempotent lineons on $V$ such that $U_i = \text{Rng } E_i$,

$$\sum_{i \in I} E_i = 1_V, \quad (81.5)$$
and
\[ E_i E_j = 0 \quad \text{for all} \quad i, j \in I \quad \text{with} \quad i \neq j. \quad (81.6) \]

We have
\[ \text{Null} \ E_j = \sum_{i \in I \setminus \{j\}} \mathcal{U}_i \quad \text{for all} \quad j \in I \quad (81.7) \]

and
\[ E_i = P_i |^\mathcal{V} \quad \text{and} \quad P_i = E_i |^\mathcal{U}_i \quad \text{for all} \quad i \in I. \quad (81.8) \]

The family \((P_i \mid i \in I)\) is called the family of projections and the family \((E_i \mid i \in I)\) the family of idempotents associated with the given decomposition.

The following result, a generalization of Prop.5 of Sect.19 and Prop.2 of Sect.16, shows how linear mappings with domain \(\mathcal{V}\) are determined by their restrictions to each term of a decomposition of \(\mathcal{V}\).

**Proposition 6:** Let \((\mathcal{U}_i \mid i \in I)\) be a decomposition of \(\mathcal{V}\). For every linear space \(\mathcal{V}'\) and every family \((L_i \mid i \in I)\) with \(L_i \in \text{Lin}(\mathcal{U}_i, \mathcal{V}')\) for all \(i \in I\), there is exactly one \(L \in \text{Lin}(\mathcal{V}, \mathcal{V}')\) such that
\[ L_i = L |^\mathcal{U}_i \quad \text{for all} \quad i \in I. \quad (81.9) \]

It is given by
\[ L := \sum_{i \in I} L_i P_i, \quad (81.10) \]

where \((P_i \mid i \in I)\) is the family of projections associated with the decomposition.

**Notes 81**

(1) I introduced the term “disjunct” in the sense of Def.1 for pairs of subspaces (see Note (2) to Sect.12); J. J. Schäffer suggested that the term be used also for arbitrary families of subspaces.

(2) In most of the literature, the phrase “\(\mathcal{V}\) is the direct sum of the family \((\mathcal{U}_i \mid i \in I)\)” is used instead of “\((\mathcal{U}_i \mid i \in I)\) is a decomposition of \(\mathcal{V}\)” . The former phrase is actually absurd because it confuses a property of the family \((\mathcal{U}_i \mid i \in I)\) with a property of the sum of the family (see also Note (3) to Sect.12). Even Bourbaki falls into this trap.
82 Spectral Values and Spectral Spaces

We assume that a linear space \( V \) and a lineon \( L \in \text{Lin}V \) are given. Recall that a subspace \( U \) of \( V \) is called an \( L \)-space if it is \( L \)-invariant, i.e. if \( L(U) \subset U \) (Def.1 of Sect.18). Assume that \( U \) is an \( L \)-space. Then we can consider the adjustment \( L\lVert U \in \text{Lin}U \). It is clear that a subspace \( W \) of \( U \) is an \( L \)-space if and only if it is a \( L\lVert U \)-space.

**Proposition 1:** A subspace \( U \) of \( V \) is \( L \)-invariant if and only if its annihilator \( U^\perp \) is \( L^\top \)-invariant.

**Proof:** Assume that \( U \) is \( L \)-invariant, i.e. that \( L(U) \subset U \). By the Theorem on Annihilators and Transposes of Sect.21 we then have

\[
U^\perp \subset (L>U))^\perp = (L^\top)^<(U^\perp),
\]

which means that \( U^\perp \) is \( L^\top \)-invariant. Interchanging the roles of \( L = (L^\top)^\top \) and \( L^\top \) and of \( U^\perp \) and \( U = (U^\perp)^\perp \) we see that the \( L^\top \)-invariance of \( U^\perp \) implies the \( L \)-invariance of \( U \).

**Definition 1:** For every \( \sigma \in \mathbb{R} \), we write

\[
\text{Sps}_L(\sigma) := \text{Null } (L - \sigma 1_Y).
\]

This is an \( L \)-space; if it is non-zero, we call it the spectral space of \( L \) for \( \sigma \). The spectrum of \( L \) is defined to be

\[
\text{Spec}L := \{\sigma \in \mathbb{R} \mid \text{Sps}_L(\sigma) \neq \{0\}\},
\]

and its elements are called the spectral values of \( L \). If \( \sigma \in \text{Spec}L \), then the non-zero members of \( \text{Sps}_L(\sigma) \) are called spectral vectors of \( L \) for \( \sigma \). The family \( \{ \text{Sps}_L(\sigma) \mid \sigma \in \text{Spec}L \} \) of subspaces of \( L \) is called the family of spectral spaces of \( L \). The multiplicity function \( \text{mult}_L : \mathbb{R} \to \mathbb{N} \) of \( L \) is defined by

\[
\text{mult}_L(\sigma) := \dim(\text{Sps}_L(\sigma));
\]

its value \( \text{mult}_L(\sigma) \in \mathbb{N}^\times \) at \( \sigma \in \text{Spec}L \) is called the multiplicity of the spectral value \( \sigma \).

We note that

\[
\text{Spec}L = \{\sigma \in \mathbb{R} \mid \text{mult}_L(\sigma) \neq 0\},
\]

i.e. that \( \text{Spec}L \) is the support of \( \text{mult}_L \) (see Sect.07).

It is evident that \( \sigma \in \mathbb{R} \) is a spectral value of \( L \) if and only if

\[
Lu = \sigma u
\]
for some \( u \in V^\times \). Every \( u \in V^\times \) for which \((82.5)\) holds is a spectral vector of \( L \) for \( \sigma \). In fact, \( \text{Sp}_{L}(\sigma) \) is the largest among the subspaces \( \mathcal{U} \) of \( V \) for which

\[ L|_{\mathcal{U}} = \sigma 1_{\mathcal{U}} \subset V. \tag{82.6} \]

**Proposition 2:** If the lineons \( L \) and \( M \) commute, then every spectral space of \( M \) is \( L \)-invariant.

**Proof:** Note that \( L \) and \( M \) commute if and only if \( L \) and \( M - \sigma 1_{V} \) commute, no matter what \( \sigma \in \mathbb{R} \) is. Hence, by \((82.1)\), it is sufficient to show that \( \text{Null} \ M \) is \( L \)-invariant. But for every \( u \in \text{Null} \ M \), we have

\[
0 = L(Mu) = (LM)u = (ML)u = M(Lu),
\]

i.e. \( Lu \in \text{Null} \ M \), which proves what was needed. \[\square\]

Let \( E \in \text{Lin} V \) be an idempotent. By Prop.3, (iv) of Sect.19, we then have

\[
\text{Rng} \ E = \text{Sp}_{E}(1). \tag{82.7}
\]

It is easily seen that \( \text{Sp}_{E}(1) \) and \( \text{Sp}_{E}(0) = \text{Null} \ E \) are the only spaces of the form \( \text{Sp}_{E}(\sigma) \), \( \sigma \in \mathbb{R} \), that can be different from zero and hence that \( \text{Spec} \ E \subset \{0, 1\} \). In fact, we have \( \text{Spec} \ E = \{0, 1\} \) unless \( E = 0 \) or \( E = 1_{V} \). The multiplicities of the spectral values 0 and 1 of \( E \) are given, in view of \((26.11)\), by

\[
\text{mult}_{E}(1) = \text{tr} \ E, \quad \text{mult}_{E}(0) = \text{dim} \ V - \text{tr} \ E. \tag{82.8}
\]

The following result, which is easily proved, shows how linear isomorphisms affect invariance, spectra, spectral spaces, and multiplicity:

**Proposition 3:** Let \( V, V' \) be linear spaces, let \( A : V \to V' \) be a linear isomorphism, and let \( L \in \text{Lin} V \) be a lineon, so that \( ALA^{-1} \in \text{Lin} V' \).

(a) If \( \mathcal{U} \) is an \( L \)-space then \( A_{\geq} (\mathcal{U}) \) is an \((ALA^{-1})\)-space.

(b) For every \( \sigma \in \mathbb{R} \), we have

\[
A_{\geq}(\text{Sp}_{L}(\sigma)) = \text{Sp}_{ALA^{-1}}(\sigma).
\]

(c) \( \text{Spec} \ L = \text{Spec} (ALA^{-1}) \) and \( \text{mult}_{L} = \text{mult}_{ALA^{-1}} \).

**Theorem on Spectral Spaces:** The spectrum of a lineon on \( V \) has at most \( \text{dim} V \) members and the family of its spectral spaces is disjunct.

The proof will be based on the following Lemma:

**Lemma:** Let \( L \in \text{Lin} V \), let \( S \) be a finite subset of \( \text{Spec} L \) and let \( f := (f_{\sigma} \mid \sigma \in S) \) be such that \( f_{\sigma} \in (\text{Sp}_{L}(\sigma))^{\times} \) for all \( \sigma \in S \). Then \( f \) is linearly independent.
Proof: We proceed by induction over $\sharp S$. The assertion is trivial when $\sharp S = 0$, because $f$ must then be the empty family. Assume, then, that $\sharp S \geq 1$ and that the assertion is valid when $S$ is replaced by $S'$ with $\sharp S' = \sharp S - 1$. Let $\lambda \in \text{Null} (\text{lnc}_L) \subseteq \mathbb{R}^S$, so that $\sum (\lambda_\sigma f_\sigma \mid \sigma \in S) = 0$. Since $L f_\sigma = \sigma f_\sigma$, it follows that

$$0 = L \left( \sum_{\sigma \in S} \lambda_\sigma f_\sigma \right) = \sum_{\sigma \in S} \lambda_\sigma \sigma f_\sigma.$$  

Now choose $\tau \in S$ and put $S' := S \setminus \{\tau\}$. We then obtain

$$0 = \tau \sum_{\sigma \in S} \lambda_\sigma f_\sigma - \sum_{\sigma \in S} \lambda_\sigma \sigma f_\sigma = \sum_{\sigma \in S'} \lambda_\sigma (\tau - \sigma) f_\sigma.$$  

Since $\tau - \sigma \neq 0$ for all $\sigma \in S'$ and since $\sharp S' = \sharp S - 1$, we can apply the induction hypothesis to $((\tau - \sigma) f_\sigma \mid \sigma \in S')$ and conclude that $\lambda_\sigma = 0$ for all $\sigma \in S'$, so that $0 = \sum_{\sigma \in S'} \lambda_\sigma f_\sigma$ follows and hence that $\lambda = 0$. Since $\lambda \in \text{Null} (\text{lnc}_L)$ was arbitrary, we conclude that $\text{Null} (\text{lnc}_L) = \{0\}$. }

Proof of the Theorem: The fact that $\sharp \text{Spec} L \leq \dim \mathcal{V}$ follows from the Lemma and the Characterization of Dimension of Sect.17. To show that $(\text{Sps}_L(\sigma) \mid \sigma \in \text{Spec} L)$ is disjunct, we need only observe that $\sum (u_\sigma \mid \sigma \in \text{Spec} L) = 0$ with $u \in \bigtimes (\text{Sps}_L(\sigma) \mid \sigma \in \text{Spec} L)$ implies that $\sum (u_\sigma \mid \sigma \in S) = 0$, where $S := \text{Supt} u$. The Lemma states that this is possible only when $S = \emptyset$, i.e. when $u_\sigma = 0$ for all $\sigma \in \text{Spec} L$. By Prop.1, this shows that the family $(\text{Sps}_L(\sigma) \mid \sigma \in \text{Spec} L)$ is disjunct.

Since a disjunct family with more than one non-zero term cannot have $\mathcal{V}$ as its union, the following is an immediate consequence of the Theorem just proved.

Proposition 4: If every $v \in \mathcal{V}^\times$ is a spectral vector of $L \in \text{Lin} \mathcal{V}$ then $L = \lambda 1_\mathcal{V}$ for some $\lambda \in \mathbb{R}$.

Remark: If $\dim \mathcal{V} = 0$, i.e. if $\mathcal{V}$ is a zero space, then the spectrum of the only member $1_\mathcal{V} = 0$ of $\text{Lin} \mathcal{V}$ is empty.

Proposition 5: If the sum of the family $(\text{Sps}_L(\sigma) \mid \sigma \in \text{Spec} L)$ of spectral spaces of a lineon $L \in \text{Lin} \mathcal{V}$ is $\mathcal{V}$, and hence the family is a decomposition of $\mathcal{V}$, and if $(E_\sigma \mid \sigma \in \text{Spec} L)$ is the associated family of idempotents, then

$$L = \sum_{\sigma \in \text{Spec} L} \sigma E_\sigma$$  

and

$$\text{tr} L = \sum_{\sigma \in \text{Spec} L} (\text{mult}_L(\sigma)) \sigma.$$  

(82.9)
CHAPTER 8. SPECTRAL THEORY

Proof: In view of (82.6), we have \( L|_{\text{Spec}_L(\sigma)} = \sigma 1_{\text{Spec}_L(\sigma)} \subset \mathcal{V} \) for all \( \sigma \in \text{Spec } L \). It follows from Prop.6 of Sect.81 and (81.8) that (82.9) holds. The equation (82.10) follows from (82.9) and (82.3) using Prop.5 of Sect.26.

Proposition 6: Let \( L \in \text{Lin } \mathcal{V} \) be given. Assume that \( Z \) is a finite subset of \( \mathbb{R} \) and that \( (\mathcal{W}_\sigma \mid \sigma \in Z) \) is a decomposition of \( \mathcal{V} \) whose terms \( \mathcal{W}_\sigma \) are non-zero \( L \)-spaces and satisfy

\[
L|_{\mathcal{W}_\sigma} = \sigma 1_{\mathcal{W}_\sigma} \quad \text{for all } \sigma \in Z.
\]

Then \( Z = \text{Spec } L \) and \( \mathcal{W}_\sigma = \text{Spec}_L(\sigma) \) for all \( \sigma \in Z \).

Proof: Since \( \mathcal{W}_\sigma \) is non-zero for all \( \sigma \in Z \), it follows from (82.11) that \( \sigma \in \text{Spec } L \) and \( \mathcal{W}_\sigma \subset \text{Spec}_L(\sigma) \) for all \( \sigma \in Z \). The assertion is now an immediate consequence of Prop.3 of Sect.81.

Notes 82

1. A large number of terms for our “spectral value” can be found in the literature. The most common is “eigenvalue”. Others combine the adjectives “proper”, “characteristic”, “latent”, or “secular” in various combinations with the nouns “value”, “number”, or “root”. I believe it is economical to use the adjective “spectral”, which fits with the commonly accepted term “spectrum”.

2. In infinite-dimensional situations, one must make a distinction between the sets \( \{ \sigma \in \mathbb{R} \mid (L - \sigma 1_\mathcal{V}) \text{ is not invertible} \} \) and \( \{ \sigma \in \mathbb{R} \mid \text{Null } (L - \sigma 1_\mathcal{V}) \neq \{0\} \} \). It is the former that is usually called the spectrum. The latter is usually called the “point-spectrum”.

3. When considering spectral values or spectral spaces, most people replace “spectral” with “eigen”, or another of the adjectives mentioned in Note (1).

4. The notation \( \text{Spec}_L(\sigma) \) for a spectral space is introduced here for the first time. I could find only ad hoc notations in the literature.

83 Orthogonal Families of Subspaces

We assume that an inner-product space \( \mathcal{V} \) is given. We say that two subsets \( S \) and \( T \) of \( \mathcal{V} \) are orthogonal, and we write \( S \perp T \), if every element of \( S \) is orthogonal to every element of \( T \), i.e. if \( u \cdot v = 0 \) for all \( u \in S, \ v \in T \). This is the case if and only if \( S \subset T^\perp \). We say that a family \( (\mathcal{U}_i \mid i \in I) \) of subspaces of \( \mathcal{V} \) is orthogonal if its terms are pairwise orthogonal, i.e. if \( \mathcal{U}_i \perp \mathcal{U}_j \) for all \( i, j \in I \) with \( i \neq j \).
**Proposition 1:** A family \((U_i \mid i \in I)\) of subspaces of \(V\) is orthogonal if and only if for all \(j \in I\)
\[
\sum_{i \in I \setminus \{j\}} U_i \subset U_j^\perp.
\] (83.1)

**Proof:** If (83.1) holds for all \(j \in I\), then
\[
U_k \subset \sum_{i \in I \setminus \{j\}} U_i \subset U_j^\perp
\]
and hence \(U_k \perp U_j\) for all \(j, k \in I\) with \(j \neq k\). If the family is orthogonal, then \(U_i \perp U_j\) and hence \(U_i \subset U_j^\perp\) for all \(j \in I\) and \(i \in I \setminus \{j\}\). Since \(U_j^\perp\) is a subspace of \(V\), (83.1) follows.

Using Prop.2 of Sect.1 and the Characterization of Regular Subspaces of Sect.41 one immediately obtains the next two results.

**Proposition 2:** If the terms of an orthogonal family are regular subspaces then it is a disjunct family.

**Proposition 3:** A family \((U_i \mid i \in I)\) of non-zero subspaces of \(V\) is an orthogonal decomposition of \(V\) if and only if all the terms of the family are regular subspaces of \(V\) and
\[
\sum_{i \in I \setminus \{j\}} U_i = U_j^\perp \quad \text{for all} \quad j \in I.
\] (83.2)

**Proposition 4:** A decomposition of \(V\) is orthogonal if and only if all the terms of the family of idempotents associated with the decomposition are symmetric.

**Proof:** Let \((U_i \mid i \in I)\) be a decomposition of \(V\) and let \((E_i \mid i \in I)\) be the family of idempotents associated with it (see Prop.5 of Sect.81). From Prop.3 and (81.7) it is clear that the decomposition is orthogonal if and only if \(\text{Null } E_j = U_j^\perp = (\text{Rng } E_j)^\perp\) for all \(j \in I\); and this is equivalent, by Prop.3 of Sect.41, to the statement that all the terms of \((E_i \mid i \in I)\) are symmetric.

**Proposition 5:** The family of spectral spaces of a symmetric lineon is orthogonal.

**Proof:** Let \(S \in \text{Sym}V\) and \(\sigma, \tau \in \text{Spec}S\) with \(\sigma \neq \tau\) be given. For all \(u \in \text{Sps}_S(\sigma)\) and \(w \in \text{Sps}_S(\tau)\) we then have \(Su = \sigma u\) and \(Sw = \tau w\). It follows that
\[
\sigma(w \cdot u) = w \cdot (\sigma u) = w \cdot Su = (Sw) \cdot u = (\tau w) \cdot u = \tau(w \cdot u),
\]
and hence \((\sigma - \tau)(w \cdot u) = 0\) for all \(u \in \text{Sps}_S(\sigma)\) and \(w \in \text{Sps}_S(\tau)\). Since \(\sigma - \tau \neq 0\), we conclude that \(\text{Sps}_S(\sigma) \perp \text{Sps}_S(\tau)\).
84 The Structure of Symmetric Lineons

There are lineons with an empty spectrum, i.e. with no spectral values at all. The concepts of spectral value and spectral space are insufficient for the description of the structure of general lineons (see Chap.9). However, a symmetric lineon on a genuine inner-product space is completely determined by its family of spectral spaces. The following theorem, whose proof depends on the Theorem on Attainment of Extrema of Sect.58 and on the Constrained Extremum Theorem of Sect.69, is the key to the derivation of this result.

**Theorem on the Extreme Spectral Values of a Symmetric Lineon:** If $\mathcal{V}$ is a non-zero genuine inner-product space, then every symmetric lineon $S \in \text{Sym}\mathcal{V}$ has a non-empty spectrum. In fact, the least and greatest members of $\text{Spec} S$ are given by

$$
\min \text{Spec} S = \min \left| S \right|_{\text{Usph} \mathcal{V}}, \quad (84.1)
$$

$$
\max \text{Spec} S = \max \left| S \right|_{\text{Usph} \mathcal{V}}, \quad (84.2)
$$

where $S : \mathcal{V} \to \mathbb{R}$ is the quadratic form associated with $S$ (see (27.13)) and where $\text{Usph} \mathcal{V}$ is the unit sphere of $\mathcal{V}$ defined by (42.9).

**Proof:** $\text{Usph} \mathcal{V}$ is not empty because $\mathcal{V}$ is non-zero, and $\text{Usph} \mathcal{V}$ is closed and bounded because it is the boundary of the unit ball $\text{Ubl} \mathcal{V}$ (see Prop.3 of Sect.52 and Prop.12 of Sect.53). By Prop.3 of Sect.66, $S$ is of class $C^1$ and hence continuous. By the Theorem on Attainment of Extrema, it follows that $S |_{\text{Usph} \mathcal{V}}$ attains a maximum and a minimum. Assume that one of these extrema is attained at $u \in \text{Usph} \mathcal{V}$. By Cor.1 to the Constrained Extremum Theorem, we must have $\nabla_u S = \sigma \nabla_u \text{sq}$ for some $\sigma \in \mathbb{R}$. Since $(\nabla_u \text{sq})v = 2u \cdot v$ and $\nabla_u S v = \sigma (Su) \cdot v$ for all $v \in \mathcal{V}$ by (66.17), we conclude that $2(Su) \cdot v = \sigma (2u \cdot v)$ for all $v \in \mathcal{V}$, i.e. $Su = \sigma u$. Since $|u| = 1$, it follows that $\sigma \in \text{Spec} S$ and that $S(u) = u \cdot Su = \sigma |u|^2 = \sigma$. We conclude that the right sides of both (84.1) and (84.2) belong to $\text{Spec} S$.

Now let $\tau \in \text{Spec} S$ be given. We may then choose $u \in \text{Usph} \mathcal{V}$ such that $Su = \tau u$. Since $u \cdot u = 1$, we have $\tau = (\tau u) \cdot u = (Su) \cdot u = \overrightarrow{S}(u) \in \text{Rng} S |_{\text{Usph} \mathcal{V}}$. Since $\tau \in \text{Spec} S$ was arbitrary, the assertions (84.1) and (84.2) follow.

The next theorem, which completely describes the structure of symmetric lineons on genuine inner product spaces, is one of the most important theorems of all of mathematics. It has numerous applications not only in many branches of mathematics, but also in almost all branches of theoretical physics, both classical and modern. The reason is that symmetric lineons appear in many contexts, often unexpectedly.
The Structure of Symmetric Lineons

Spectral Theorem: Let $\mathcal{V}$ be a genuine inner-product space. A lineon on $\mathcal{V}$ is symmetric if and only if the family of its spectral spaces is an orthogonal decomposition of $\mathcal{V}$.

Proof: Let $S \in \text{Sym}\mathcal{V}$ be given. Since all spectral spaces of $S$ are $S$-invariant, so is their sum $U := \sum (\text{SpS}_\sigma | \sigma \in \text{Spec} S)$. By Prop.1 of Sect.82, the orthogonal supplement $U^\perp$ of $U$ is also $S$-invariant, and $S_{U^\perp} \in \text{Sym} U^\perp$ is meaningful. Now, by the preceding theorem, if $U^\perp$ is non-zero, $S_{U^\perp}$ must have a spectral vector $w \in (U^\perp)^\times$. Of course, $w$ is then also a spectral vector of $S$ and hence $w \in U$. Therefore, $U^\perp \neq \{0\}$ implies $U \cap U^\perp \neq \{0\}$, which contradicts the fact that, in a genuine inner-product space, all subspaces are regular. It follows that $U^\perp = \{0\}$ and hence $\mathcal{V} = U = \sum (\text{SpS}_\sigma | \sigma \in \text{Spec} S)$. By Props.2 and 5 of Sect.83 it follows that $(\text{SpS}_\sigma | \sigma \in \text{Spec} S)$ is an orthogonal decomposition of $\mathcal{V}$.

Assume, now, that $L \in \text{Lin}\mathcal{V}$ is such that $(\text{SpS}_L | \sigma \in \text{Spec} L)$ is an orthogonal decomposition of $\mathcal{V}$. We can then apply Prop.5 of Sect.82 and conclude that (82.9) holds. Since the decomposition was assumed to be orthogonal, each of the idempotents $E_\sigma$ in (82.9) is symmetric by Prop.4 of Sect.83.

We assume now that a genuine inner-product space $\mathcal{V}$ is given. We list two corollaries to the Spectral Theorem. The first follows by applying Prop.5 of Sect.81, Prop.5 of Sect.82, and Prop.4 of Sect.83.

Corollary 1: For every $S \in \text{Sym}\mathcal{V}$ there is exactly one family $(E_\sigma | \sigma \in \text{Spec} S)$ of non-zero symmetric idempotents such that

$$E_\sigma E_\tau = 0 \quad \text{for all } \sigma, \tau \in \text{Spec} S \quad \text{with } \sigma \neq \tau,$$

$$\Sigma_{\sigma \in \text{Spec} S} E_\sigma = 1_\mathcal{V},$$

and

$$\Sigma_{\sigma \in \text{Spec} S} \sigma E_\sigma = S.$$

The family $(E_\sigma | \sigma \in \text{Spec} S)$ is called the spectral resolution of $S$. Its terms are called the spectral idempotents of $S$.

The second corollary is obtained by choosing orthonormal bases in each of the spectral spaces of $S$.

Corollary 2: A lineon $S$ on $\mathcal{V}$ is symmetric if and only if there exists an orthonormal basis $e := (e_i | i \in I)$ of $\mathcal{V}$ whose terms are spectral vectors of $S$. If this is the case we have

$$Se_i = \lambda_i e_i \quad \text{for all } i \in I,$$
where \( \lambda := \{ \lambda_i \mid i \in I \} \) is a family of real numbers whose range is \( \text{Spec} \, S \). In fact each \( \sigma \in \text{Spec} \, S \) occurs exactly \( \text{mult}_S(\sigma) \) times in the family \( \lambda \).

The matrix of \( S \) relative to the basis \( e \) is diagonal.

Of course, the orthonormal basis \( e \) in Cor.2 is not uniquely determined by \( S \) if \( \dim V > 0 \). For example, if we replace any or all the terms of \( e \) by their opposites, we get another orthonormal basis whose terms are also spectral vectors of \( S \).

**Proposition 1:** Let \( V \) and \( V' \) be genuine inner product spaces. Then \( S \in \text{Sym} \, V \) and \( S' \in \text{Sym} \, V' \) have the same multiplicity function, i.e. \( \text{mult}_S = \text{mult}_{S'} \) if and only if there is an orthogonal isomorphism \( R : V \to V' \) such that

\[
S' = RSR^{-1} = RSR^\top. \tag{84.7}
\]

If (84.7) holds, then

\[
R^{-1}(\text{Sp}_{S}(\sigma)) = \text{Sp}_{S'}(\sigma) \quad \text{for all} \quad \sigma \in \text{Spec} \, S. \tag{84.8}
\]

**Proof:** Assume that \( \text{mult}_S = \text{mult}_{S'} \). By Cor.2 above, we may then choose orthonormal bases \( e := (e_i \mid i \in I) \) and \( e' := (e'_i \mid i \in I) \) of \( V \) and \( V' \), respectively, such that (84.6) and

\[
S'e'_i = \lambda_i e'_i \quad \text{for all} \quad i \in I \tag{84.9}
\]

hold with one and the same family \( \lambda := \{ \lambda_i \mid i \in I \} \). Let \( R : V \to V' \) be the linear isomorphism for which \( Re_i = e'_i \) for all \( i \in I \). By Prop.4 of Sect.43, \( R \) is an orthogonal isomorphism. By (84.9) and (84.6) we have

\[
S'R e_i = \lambda_i (R e_i) = R(\lambda_i e_i) = R S e_i
\]

for all \( i \in I \). Since \( e \) is a basis, it follows that \( S'R = R S \), which is equivalent to (84.7).

If (84.7) holds, then (84.8) and the equality \( \text{mult}_S = \text{mult}_{S'} \) follow from Prop.3 of Sect.82.

Using Prop.1 above and Prop.2 of Sect.82, we immediately obtain

**Proposition 2:** \( S \in \text{Sym} \, V \) commutes with \( R \in \text{Orth} \, V \) if and only if the spectral spaces of \( S \) are \( R \)-invariant.

In the special case when all spectral values of \( S \) have multiplicity 1, the condition of Prop.2 is equivalent to the following one: If \( e \) is an orthonormal basis such that the matrix of \( S \) relative to \( e \) is diagonal (see Cor.2 above), then the matrix of \( R \) relative to \( e \) is also diagonal and its diagonal terms are either 1 or -1.

We are now able to prove the result (52.20) already announced in Sect.52.
Proposition 3: If $V$ and $V'$ are genuine inner-product spaces and if $n := \dim V > 0$, then

$$|L| \geq \|L\| \geq \frac{1}{\sqrt{n}} |L| \quad \text{for all } L \in \text{Lin}(V, V').$$

(84.10)

Proof: Let $L \in \text{Lin}(V, V')$ be given. The inequality $|L| \geq \|L\|$ is an immediate consequence of (52.4), with both $\nu$ and $\nu'$ replaced by $|\cdot|$, and Prop.3 of Sect.44.

Since $S := L^\top L$ belongs to $\text{Sym}V$, we can consider the spectral resolution $(E_\sigma \mid \sigma \in \text{Spec} S)$ of $S$ mentioned in Cor.1 above. By (84.5) and (26.11), we obtain

$$\text{tr} S = \sum_{\sigma \in \text{Spec} S} \sigma \text{tr} E_\sigma = \sum_{\sigma \in \text{Spec} S} \sigma \dim \text{Rng} E_\sigma.$$  

(84.11)

Since $(\text{Rng} E_\sigma \mid \sigma \in \text{Spec} S)$ is a decomposition of $V$, it follows from (84.11) and Prop.4 of Sect.81 that

$$\text{tr} S \leq n \max(\text{Spec} S).$$

Using the Theorem on the Extreme Spectral Values of a Symmetric Lineon, we conclude that

$$\text{tr} S \leq n \max \{ |L|^2 \mid \text{v} \in \text{Usph} V \}.$$  

Since $S(v) = v \cdot Sv = v \cdot L^\top L v = |Lv|^2$ for all $v \in V$, it follows that

$$\text{tr} S \leq n \max \{ |Lv|^2 \mid v \in \text{Usph} V \}.$$  

Since $\text{tr} S = \text{tr}(L^\top L) = |L|^2$ by (44.13), it follows from (52.3), with both $\nu$ and $\nu'$ replaced by $|\cdot|$, that $|L|^2 \leq n \|L\|^2$. □

Remark: If $L$ is a tensor product, i.e. if $L = w \otimes v$ for some $v \in V \cong V^*$ and $w \in V'$, then $|L| = \|L\|$, as easily seen from (44.14) and Part (a) of Problem 7 in Chap.5. If $L$ is orthogonal, we have $\|L\| = \frac{1}{\sqrt{n}} |L|$ by Problem 5 in Chap.5. Hence either of the inequalities in (84.10) can become an equality. □

85 Lineonic Extensions

Lineonic Square Roots and Logarithms

We assume that a genuine inner-product space $V$ is given. In Sect.41, we noted the identification

$$\text{Sym} V \cong \text{Sym}_2(V^2, \mathbb{R}).$$
and in Sect. 27, we discussed the isomorphism 
\( \phi : \text{Sym}_2(V^2, \mathbb{R}) \to \text{Qu}_V \). We thus see that each symmetric lineon \( S \) on \( V \) corresponds to a quadratic form \( \overline{S} \) on \( V \) given by

\[
\overline{S}(u) := S(u, u) = (Su) \cdot u \quad \text{for all} \quad u \in V.
\]  

(85.1)

The identity lineon \( 1_V \in \text{Sym}_V \) is identified with the inner product \( ip \) on \( V \) and hence \( 1_V = ip = sq \). In view of Def. 1 of Sect. 27, it is meaningful to speak of \textit{positive} [\textit{strictly positive}] \textit{symmetric lineons}. We use the notations

\[
\text{Pos}_V := \{ S \in \text{Sym}_V \mid \overline{S}(u) \geq 0 \quad \text{for all} \quad u \in V \} \quad (85.2)
\]

and

\[
\text{Pos}^+_V := \{ S \in \text{Sym}_V \mid \overline{S}(u) > 0 \quad \text{for all} \quad u \in V^\times \} \quad (85.3)
\]

for the sets of all positive and strictly positive symmetric lineons, respectively. Since \( V \) is genuine, we have \( 1_V \in \text{Pos}^+_V \). By Prop. 3 of Sect. 27 we have

\[
L^\top SL = \overline{S} \circ L
\]

(85.4)

for all \( S \in \text{Sym}_V \) and all \( L \in \text{Lin}_V \).

The following result is an immediate consequence of (85.4):

\textbf{Proposition 1:} If \( S \in \text{Pos}_V \) and \( L \in \text{Lin}_V \), then \( L^\top SL \in \text{Pos}_V \). Moreover, \( L^\top SL \) is strictly positive if and only if \( S \) is strictly positive and \( L \) is invertible. In particular, we have \( L^\top L \in \text{Pos}_V \), and \( L^\top L \in \text{Pos}^+_V \) if and only if \( L \) is invertible.

The following result is a corollary to the Theorem on the Extreme Spectral Values of a Symmetric Lineon (see Sect. 84).

\textbf{Proposition 2:} A symmetric lineon is positive [strictly positive] if and only if all its spectral values are positive [strictly positive]. In other words,

\[
\text{Pos}_V = \{ S \in \text{Sym}_V \mid \text{Spec} S \subset \mathbb{P} \}, \quad (85.5)
\]

\[
\text{Pos}^+_V = \{ S \in \text{Sym}_V \mid \text{Spec} S \subset \mathbb{P}^\times \}. \quad (85.6)
\]

Let a function \( f : \mathbb{R} \to \mathbb{R} \) be given. We define the \textit{lineonic extension} \( f_{(V)} : \text{Sym}_V \to \text{Sym}_V \) of \( f \) by

\[
f_{(V)}(S) := \sum_{\sigma \in \text{Spec} S} f(\sigma)E_\sigma
\]

(85.7)
when \((E_\sigma \mid \sigma \in \Spec S)\) is the spectral resolution of \(S \in \Sym \mathcal{V}\). If the domain of \(f\) is \(\mathbb{P}\) or \(\mathbb{P}^\times\) instead of \(\mathbb{R}\), we still use (85.7), but, in view of Prop.2, the domain of \(f(\mathcal{V})\) must be taken to be \(\Pos \mathcal{V}\) or \(\Pos^+ \mathcal{V}\), respectively. A similar observation applies to the codomains of \(f\) and \(f(\mathcal{V})\).

Let \(f\) and \(g\) be functions whose domains and codomains are one of \(\mathbb{R}, \mathbb{P}, \mathbb{P}^\times\) each. The following rules are evident from the definition (85.7):

(I) For all \(S \in \Sym \mathcal{V}\), \(S \in \Pos \mathcal{V}\), or \(S \in \Pos^+ \mathcal{V}\), as appropriate, the spectrum of \(f(\mathcal{V})(S)\) is given by

\[
\Spec (f(\mathcal{V})(S)) = f(\Spec S).
\] (85.8)

Moreover, if \(f\mid_{\Spec S}\) is injective, then \(f(\mathcal{V})(S)\) and \(S\) have the same spectral spaces and the same spectral idempotents.

(II) If \(\Dom f = \Cod g\), then \(\Dom f(\mathcal{V}) = \Cod g(\mathcal{V})\) and

\[
(f \circ g)(\mathcal{V}) = f(\mathcal{V}) \circ g(\mathcal{V}).
\] (85.9)

(III) If \(f\) is invertible, so is \(f(\mathcal{V})\) and

\[
(f(\mathcal{V}))^\sim = (f^\sim)(\mathcal{V}).
\] (85.10)

As an example, we consider the lineonic extension \(\iota^n(\mathcal{V}) : \Sym \mathcal{V} \to \Sym \mathcal{V}\) of the real \(n\)th power function \(\iota^n : \mathbb{R} \to \mathbb{R}\).

By (85.7), we have

\[
\iota^n(\mathcal{V})(S) = \sum_{\sigma \in \Spec S} \sigma^n E_\sigma \quad \text{for all } \ n \in \mathbb{N}.
\] (85.11)

Hence, the lineonic extension of the real \(n\)th power is nothing but the adjustment \(\pow_n|_{\Sym \mathcal{V}}\) of the lineonic \(n\)th power defined by (66.18).

Of particular importance is the lineonic extension of the real square function \(\iota^2|_{\mathbb{P}}\), adjusted to \(\mathbb{P}\). This function is invertible and the inverse is the positive real square root function \(\sqrt{\mathcal{V}} := (\iota^2|_{\mathbb{P}})^\sim : \mathbb{P} \to \mathbb{P}\). We call the lineonic extension of \(\sqrt{\mathcal{V}}\) the **lineonic square root** and denote it by

\[
\sqrt{\mathcal{V}} := \sqrt{(\mathcal{V})} : \Pos \mathcal{V} \to \Pos \mathcal{V}.
\]
We also write $\sqrt{S} := \text{sqrt}(S)$ when $S \in \text{Pos}\mathcal{V}$. By (85.7) we have

$$\text{sqrt}(S) = \sqrt{S} = \sum_{\sigma \in \text{Spec}S} \sqrt{\sigma} E_{\sigma}$$

(85.13)

when $(E_{\sigma} | \sigma \in \text{Spec}S)$ is the spectral resolution of $S \in \text{Pos}\mathcal{V}$.

The following result follows directly from the rules (I) and (III) above and from (85.12).

**Lineonic Square Root Theorem:** For every $S \in \text{Pos}\mathcal{V}$, the lineonic square root $\sqrt{S}$ given by (85.13) is the only solution of the equation

$$\exists \textbf{D} \in \text{Pos}\mathcal{V}, \ \textbf{D}^2 = S.$$

(85.14)

The spectrum of $\sqrt{S}$ is given by

$$\text{Spec} \sqrt{S} = \{\sqrt{\sigma} | \sigma \in \text{Spec}S\},$$

(85.15)

and we have

$$\text{Sp}_{\sqrt{S}}(\sqrt{\sigma}) = \text{Sp}_{S}(\sigma) \text{ for all } \sigma \in \text{Spec}S.$$  

(85.16)

It turns out that the domain $\text{Pos}\mathcal{V}$ of the lineonic square root $\text{sqrt}$ is not an open subset of the linear space $\text{Sym}\mathcal{V}$ and hence it makes no sense to ask whether $\text{sqrt}$ is differentiable. However, as we will see, $\text{Pos}^{+}\mathcal{V}$ is an open subset of $\text{Sym}\mathcal{V}$ and the restriction of $\text{sqrt}$ to $\text{Pos}^{+}\mathcal{V}$ is of class $C^1$. Actually, $\text{Pos}^{+}\mathcal{V}$ is $\text{sqrt}$-invariant because the lineonic extension of the strictly positive real square function $(\iota^2 | \mathcal{X})^{-}$ must be the adjustment

$$\text{sqrt}^{+} := \text{sqrt}_{|\text{Pos}^{+}\mathcal{V}}$$

(85.17)

of the square root. By (85.12) and the rule (III) above, $\text{sqrt}^{+}$ is the inverse of the adjustment to $\text{Pos}^{+}\mathcal{V}$ of the lineonic square function $\text{pow}_2$:

$$\text{sqrt}^{+} = (\text{pow}_2^{+})^{-}, \text{ where } \text{pow}_2^{+} := \text{pow}_{2|\text{Pos}^{+}\mathcal{V}}.$$  

(85.18)

We call $\text{sqrt}^{+}$ the **strict lineonic square root**.

**Theorem on the Smoothness of the Strict Lineonic Square Root:** The set $\text{Pos}^{+}\mathcal{V}$ of strictly positive lineons is an open subset of $\text{Sym}\mathcal{V}$ and the strict lineonic square root $\text{sqrt}^{+} : \text{Pos}^{+}\mathcal{V} \rightarrow \text{Pos}^{+}\mathcal{V}$ is of class $C^1$.

**Proof:** Let $S \in \text{Pos}^{+}\mathcal{V}$ be given and let $\sigma$ be the least spectral value of $S$. By Prop.2, $\sigma$ is strictly positive and by (84.1), we have

$$(Sv) \cdot v \geq \sigma |v|^2 \text{ for all } v \in \mathcal{V}.$$ 

(85.19)
In view of (52.19), we have \( |Tv| \leq ||T|| \|v\| \) for every \( T \in \text{Sym} V \) and all \( v \in V \), where \( ||T|| \) is the operator-norm of \( T \). Therefore, the Inner-Product Inequality of Sect.42 yields

\[
(Tv) \cdot v \geq -||Tv|| \|v\| \geq -||T|| \|v\|^2 \quad \text{for all } v \in V.
\]

(85.20)

Combining (85.19) and (85.20), we see that

\[
((S + T)v) \cdot v \geq (\sigma - ||T||) \|v\|^2 \quad \text{for all } v \in V.
\]

It follows that \( S + T \) is strictly positive if \( ||T|| < \sigma \), i.e. if \( T \in \sigma \text{Ce}(|| \cdot ||) \). Hence \( \text{Pos}^+ V \) includes the neighborhood \( S + \sigma \text{Ce}(|| \cdot ||) \cap \text{Sym} V \) of \( S \) in \( \text{Sym} V \). Since \( S \in \text{Pos}^+ V \) was arbitrary, it follows that \( \text{Pos}^+ V \) is open in \( \text{Sym} V \).

Since \( \text{Pos}^+ V \) is open in \( \text{Sym} V \), it makes sense to say, and it is true by Prop.4 of Sect.66, that \( \text{pow}^2 \) is of class \( C^1 \). By (66.19), the gradient of \( \text{pow}^2 \) is given by

\[
(\nabla S \text{pow}^2)U = US + SU \quad \text{for all } U \in \text{Sym} V.
\]

(85.21)

Let \( U \in \text{Null } (\nabla S \text{pow}^2) \) be given, so that \( US + SU = 0 \). We then have

\[
(S + \gamma U)^2 = S^2 + \gamma(SU + US) + \gamma^2 U^2 = S^2 + \gamma^2 U^2 = S^2 - \gamma(SU + US) + \gamma^2 U^2 = (S - \gamma U)^2
\]

for all \( \gamma \in \mathbb{P}^\times \). Since \( \text{Pos}^+ V \) is open in \( \text{Sym} V \), we may choose \( \gamma \in \mathbb{P}^\times \) such that both \( S + \gamma U \) and \( S - \gamma U \) belong to \( \text{Pos}^+ V \). We then have

\[
\text{pow}^2(S + \gamma U) = \text{pow}^2(S - \gamma U).
\]

Since \( \text{pow}^2 \) is injective, it follows that \( S + \gamma U = S - \gamma U \) and hence that \( U = 0 \). It follows that \( \text{Null } (\nabla S \text{pow}^2) = \{0\} \). By the Pigeonhole Principle for Linear Mappings, it follows that \( \nabla S \text{pow}^2 \) is invertible. The Local Inversion Theorem yields that \( \sqrt{\text{pow}^2} = (\text{pow}^2)^{-\frac{1}{2}} \) is of class \( C^1 \).

Since, by (68.4), we have

\[
\nabla \sqrt{\text{pow}^2} = (\nabla \sqrt{\text{pow}^2})^{-1} \quad \text{for all } S \in \text{Pos}^+ V,
\]

we can conclude from (85.21) that for every \( S \in \text{Pos}^+ V \) and every \( V \in \text{Sym} V \), the value \( \nabla \sqrt{\text{pow}^2} V \) is the unique solution of the equation

\[
\nabla \sqrt{\text{pow}^2} \in \text{Sym} V, \quad V = U\sqrt{S} + \sqrt{SU}.
\]

(85.22)
If $V$ commutes with $S$ then the solution of (85.22) is given by

\[(\nabla S \sqrt{+}) V = \frac{1}{2} \sqrt{S}^{-1} V, \tag{85.23}\]

which is consistent with the formula for the derivative of the square-root function of elementary calculus. If $V$ does not commute with $S$, then there is no simple explicit formula for $(\nabla S \sqrt{+}) V$.

As another example of a lineonic extension we consider $(\exp | P \times) (V)$, where $\exp$ is the real exponential function (see Sect.08).

**Proposition 3:** The lineonic extension of $\exp | P \times$ coincides with an adjustment of the lineonic exponential defined in Prop.2 of Sect.612. Specifically, we have

\[(\exp | P \times)(V) = \exp_{\text{Pos} V} V. \tag{85.24}\]

**Proof:** Let $S \in \text{Sym} V$ be given and let $(E_\sigma | \sigma \in \text{Spec} S)$ be its spectral resolution. We define $F : \mathbb{R} \to \text{Lin} V$ by

\[F := \sum_{\sigma \in \text{Spec} S} (\exp \circ (i \sigma)) E_\sigma. \tag{85.25}\]

Differentiation gives

\[F' = \sum_{\sigma \in \text{Spec} S} (\exp \circ (i \sigma)) \sigma E_\sigma. \tag{85.26}\]

Using the fact that $E_\sigma^2 = E_\sigma$ for all $\sigma \in \text{Spec} S$ and that $E_\sigma E_\tau = 0$ for all $\sigma, \tau \in \text{Spec} S$ with $\sigma \neq \tau$, we infer from (84.5), (85.25), and (85.26) that $F' = SF$. Since, by (85.25) and (84.4), $F(0) = \sum (E_\sigma | \sigma \in \text{Spec} S) = 1_V$ we can use Prop.4 of Sect.612 to conclude that $F = \exp_V \circ (i S)$. Evaluation of (85.25) at $1 \in \mathbb{R}$ hence gives

\[(\exp | P \times)(V)(S) = F(1) = \exp_V(S). \]

Since $S \in \text{Sym} V$ was arbitrary, (85.24) follows.

Since $\exp | P \times$ is invertible, and since its inverse is the logarithm $\log := (\exp | P \times)^{-1} : P \times \to \mathbb{R}$, it follows from Prop.3 and rule (III) above that $\exp_{\text{Pos} V} \circ (\text{Sym} V)$ is invertible and that $\log_V$ is its inverse. We call $\log_V$ the lineonic logarithm and also write $\log S := \log_V(S)$ when $S \in \text{Pos} V$. In sum, we have the following result.

**Lineonic Logarithm Theorem:** For every $S \in \text{Pos} V$, the lineonic logarithm $\log S$, given by

\[\log S = \sum_{\sigma \in \text{Spec} S} (\log \sigma) E_\sigma, \tag{85.27}\]
where \((E_\sigma | \sigma \in \text{Spec } S)\) is the spectral resolution of \(S\), is the only solution of the equation

\[
T \in \text{Sym } V, \quad \exp_V(T) = S.
\] (85.28)

The spectrum of \(\log S\) is given by

\[
\text{Spec } (\log S) = \log > (\text{Spec } S)
\] (85.29)

and we have

\[
\text{Sps}_{\log S}(\log \sigma) = \text{Sps}_S(\sigma) \quad \text{for all } \sigma \in \text{Spec } S.
\] (85.30)

Remark: Since \(\text{Dom } \log (V) = \text{Pos}^+ V\) is an open subset of \(\text{Sym } V\) by the Theorem on Smoothness of the Strict Lineonic Square Root, it is meaningful to ask whether \(\log (V)\) is of class \(C^1\). In fact it is, but the proof is far from trivial. (See Problem 6 at the end of this chapter.)

Notes 85

(1) The terms “positive semidefinite” instead of “positive” and “positive definite” instead of “strictly positive” are often used in the literature in connection with symmetric lineons. (See also Note (1) to Sect.27.)

(2) The Theorem on the Smoothness of the Lineonic Square Root, with a proof somewhat different from the one above, was contained in notes that led to this book. These notes were the basis of the corresponding theorem and proof in “An Introduction to Continuum Mechanics” by M.E. Gurtin (Academic Press, 1981). I am not aware of any other place in the literature where the Theorem is proved or even mentioned.

86 Polar Decomposition

Let \(V\) be a genuine inner-product space. In view of the identifications \(\text{Lin } V \cong \text{Lin}_2(V^2, \mathbb{R})\), \(\text{Sym } V \cong \text{Sym}_2(V^2, \mathbb{R})\), and \(\text{Skew } V \cong \text{Skew}_2(V^2, \mathbb{R})\) (see Sect.41), we can phrase Prop.6 of Sect.24 in the following way:

**Additive Decomposition Theorem:** To every lineon \(L \in \text{Lin } V\) corresponds a unique pair \((S, A)\) of lineons such that

\[
L = S + A, \quad S \in \text{Sym } V, \quad A \in \text{Skew } V.
\] (86.1)

In fact, \(S\) and \(A\) are given by

\[
S = \frac{1}{2}(L + L^\top), \quad A = \frac{1}{2}(L - L^\top).
\] (86.2)
The pair \( (S, A) \) is called the additive decomposition of \( L \).

The following theorem asserts the existence and uniqueness of certain multiplicative decompositions for invertible lineons.

**Polar Decomposition Theorem** To every invertible lineon \( L \in \text{Lis}\mathcal{V} \) corresponds a unique pair \( (R, S) \) such that

\[
L = RS, \quad R \in \text{Orth}\mathcal{V}, \quad S \in \text{Pos}^+\mathcal{V}.
\]  

(86.3)

In fact, \( S \) and \( R \) are given by

\[
S = \sqrt{L^\top L}, \quad R = LS^{-1}.
\]  

(86.4)

Also, there is a unique pair \( (R', S') \) such that

\[
L = S'R', \quad R' \in \text{Orth}\mathcal{V}, \quad S' \in \text{Pos}^+\mathcal{V}.
\]  

(86.5)

In fact, \( R' \) coincides with \( R \), and \( S' \) is given by

\[
S' = RSR^\top.
\]  

(86.6)

The pair \( (R, S) \) \([\!(R', S')\!)\] for which (86.3) \([\!(86.5)\!)\] holds is called the right \([\!\text{left}\!]\) polar decomposition of \( L \).

**Proof:** Assume that (86.3) holds for a given pair \( (R, S) \). Since \( S^\top = S \) and \( R^\top R = 1\mathcal{V} \), we then have

\[
L^\top L = (RS)^\top (RS) = S^\top R^\top RS = S^2.
\]

Since \( L^\top L \) as well as \( S \) belongs to \( \text{Pos}^+\mathcal{V} \) (see Prop.1 of Sect.85), it follows from the Lineonic Square Root Theorem that (86.4) must be valid and hence that \( S \) and \( R \), if they exist, are uniquely determined by \( L \). To prove existence, we define \( S \) and \( R \) by (86.4). We then have \( L = RS \) and \( S \in \text{Pos}^+\mathcal{V} \). To prove that \( R \in \text{Orth}\mathcal{V} \) we need only observe that

\[
R^\top R = (LS^{-1})^\top (LS^{-1}) = S^{-1}L^\top LS^{-1} = S^{-1}S^2S^{-1} = 1\mathcal{V}.
\]

Assume now that (86.5) holds for a given pair \( (R', S') \). We then have

\[
L = S'R' = (R'R'^\top)S'R' = R'(R'^\top S'R').
\]

In view of Prop.1 of Sect.85, \( R'^\top S'R' \) belongs to \( \text{Pos}^+\mathcal{V} \). It follows from the already proved uniqueness of the pair \( (R, S) \) that \( R' = R \) and that \( S = R'^\top S'R' \), and hence (86.6) holds. Therefore, \( R' \) and \( S' \), if they exist, are uniquely determined by \( S \) and \( R \), and hence by \( L \). To prove existence,
we need only define $R'$ and $S'$ by $R' := R$ and $S' := R S R^\top$. It is then evident that (86.5) holds.

If we apply the Theorem to the case when $L$ is replaced by $L^\top$, we obtain

$$S' = \sqrt{LL^\top}, \quad R' = R = S'^{-1}L^\top. \quad (86.7)$$

The polar decompositions can be illustrated geometrically. Assume that (86.3) holds, choose an orthonormal basis $e := (e_i | i \in I)$ of spectral vectors of $S$ (see Cor.2 to the Spectral Theorem), and consider the “unit cube”

$$C := \text{Box}(\frac{1}{2}e) := \{v \in V | |v \cdot e_i| \leq \frac{1}{2} \text{ for all } i \in I\}.$$ Since $S$ is strictly positive, we have $S e_i = \lambda e_i$ for all $i \in I$ where $\lambda := (\lambda_i | i \in I)$ is a family of strictly positive numbers. The lineon $S$ will “stretch and/or compress” the cube $C = \text{Box}(\frac{1}{2}e)$ into the “rectangular box”

$$S_>(C) = \text{Box}(\frac{\lambda_i}{2} e_i | i \in I),$$

whose sides have the lengths $\lambda_i$, $i \in I$. Finally, the orthogonal lineon $R$ will “rotate” or “rotate and reflect” (see end of Sect.88) the rectangular box $S_>(C)$, the result being the rectangular box

$$L_>(C) = R_>(S_>(C)) = \text{Box}(\frac{\lambda_i}{2} R e_i | i \in I)$$

congruent to $S_>(C)$. Thus, the right polar decomposition (86.3) shows that the transformation of the cube $C$ into the rectangular box $L_>(C)$ can be obtained by first “stretching and/or compressing” and then “rotating” or “rotating and reflecting”. The left polar decomposition (86.5) can be used to show that $L_>(C)$ can be obtained from $C$ also by first “rotating” or “rotating and reflecting” and then “stretching and/or compressing”. In the case when $\dim V = 2$ these processes are illustrated in Figure 1.
The Polar Decomposition Theorem shows that there are mappings

\[ \text{or} : \text{Lis} \mathcal{V} \to \text{Lin} \mathcal{V}, \]

\[ \text{rp} : \text{Lis} \mathcal{V} \to \text{Pos}^+ \mathcal{V}, \]

\[ \text{lp} : \text{Lis} \mathcal{V} \to \text{Pos}^+ \mathcal{V}, \]

such that \( \text{Rng or} \subset \text{Orth} \mathcal{V} \) and

\[ \mathbf{L} = \text{or}(\mathbf{L}) \text{rp}(\mathbf{L}) = \text{lp}(\mathbf{L}) \text{or}(\mathbf{L}) \quad (86.8) \]

for all \( \mathbf{L} \in \text{Lis} \mathcal{V} \). For \( \mathbf{L} := \mathbf{1}_\mathcal{V} \), we get

\[ \text{or}(\mathbf{1}_\mathcal{V}) = \text{rp}(\mathbf{1}_\mathcal{V}) = \text{lp}(\mathbf{1}_\mathcal{V}) = \mathbf{1}_\mathcal{V}. \quad (86.9) \]

In view of (86.4), we have

\[ \text{rp}(\mathbf{L}) = \sqrt{\text{tr}}(\mathbf{L}^\top \mathbf{L}), \quad \text{or}(\mathbf{L}) = (\text{rp}(\mathbf{L}))^{-1} \quad (86.10) \]

for all \( \mathbf{L} \in \text{Lis} \mathcal{V} \), where \( \sqrt{\text{tr}} \) is the strict lineonic square-root function defined in the previous section. It is easily seen, also, that

\[ \text{lp}(\mathbf{L}) = \text{rp}(\mathbf{L}^\top) = \sqrt{\text{tr}}(\mathbf{L} \mathbf{L}^\top) \quad \text{for all} \quad \mathbf{L} \in \text{Lis} \mathcal{V}. \quad (86.11) \]

**Proposition 1:** The mappings \( \text{or}, \text{rp}, \) and \( \text{lp} \) characterized by the Polar Decomposition Theorem are of class \( C^1 \). Their gradients at \( \mathbf{1}_\mathcal{V} \) are given by

\[ \nabla_{\mathbf{1}_\mathcal{V}} \text{or} \mathbf{M} = \frac{1}{2}(\mathbf{M} - \mathbf{M}^\top), \quad (86.12) \]

\[ \nabla_{\mathbf{1}_\mathcal{V}} \text{rp} \mathbf{M} = \nabla_{\mathbf{1}_\mathcal{V}} \text{lp} \mathbf{M} = \frac{1}{2}(\mathbf{M} + \mathbf{M}^\top) \quad (86.13) \]

for all \( \mathbf{M} \in \text{Lin} \mathcal{V} \).

**Proof:** Since \( (\mathbf{L} \mapsto \mathbf{L}^\top \mathbf{L}) : \text{Lis} \mathcal{V} \to \text{Pos}^+ \mathcal{V} \) is of class \( C^1 \) by the General Product Rule and since \( \sqrt{\text{tr}}^+ : \text{Pos}^+ \mathcal{V} \to \text{Pos}^+ \mathcal{V} \) is of class \( C^1 \) by the Theorem on the Smoothness of the Strict Lineonic Square Root, it follows from (86.10) and the Chain Rule that \( \text{rp} \) is of class \( C^1 \). It is an immediate consequence of (86.11) that \( \text{lp} \) is also of class \( C^1 \) and that

\[ \nabla_{\mathbf{L}} \text{lp} \mathbf{M} = (\nabla_{\mathbf{L}^\top} \text{rp}) \mathbf{M}^\top \quad (86.14) \]

for all \( \mathbf{L} \in \text{Lis} \mathcal{V}, \ \mathbf{M} \in \text{Lin} \mathcal{V} \).

Using the Differentiation Theorem for Inversion Mappings of Sect.68 we conclude from (86.10) that or is also of class \( C^1 \).
If we differentiate (86.8) \(1\) with respect to \(L\) using the General Product Rule, we obtain

\[
M = ((\nabla L \text{or})M)rp(L) + or(L)((\nabla L \text{rp})M)
\]

for all \(L \in \text{Lis}\mathcal{V}, M \in \text{Lin}\mathcal{V}\). In view of (86.9), for \(L := 1\mathcal{V}\) this yields

\[
M = (\nabla 1\mathcal{V} \text{or})M + (\nabla 1\mathcal{V} \text{rp})M \quad \text{for all } M \in \text{Lin}\mathcal{V}.
\] (86.15)

Since the codomain \(\text{Pos}^+\mathcal{V}\) of \(rp\) is an open subset of \(\text{Sym}\mathcal{V}\), the codomain of \(\nabla L \text{rp}\) is \(\text{Sym}\mathcal{V}\) for all \(L \in \text{Lis}\mathcal{V}\). In particular,

\[
(\nabla 1\mathcal{V} \text{rp})M \in \text{Sym}\mathcal{V} \quad \text{for all } M \in \text{Lin}\mathcal{V}.
\] (86.16)

Since \(\text{Rng or} \subset \text{Orth}\mathcal{V}\), it follows from Prop.2 of Sect.66 that

\[
(\nabla 1\mathcal{V} \text{or})M \in \text{Skew}\mathcal{V} \quad \text{for all } M \in \text{Lin}\mathcal{V}.
\] (86.17)

Comparing (86.15)–(86.17) with (86.1), we see that for every \(M \in \text{Lin}\mathcal{V}\),

\[
U := (\nabla 1\mathcal{V} \text{rp})M, \quad A := (\nabla 1\mathcal{V} \text{or})M
\]
gives the additive decomposition \((U, A)\) of \(M\). Hence (86.12) and (86.13) follow from the Additive Decomposition Theorem.

There are no simple explicit formulas for \(\nabla L \text{rp}, \nabla L \text{or},\) and \(\nabla L \text{lp}\) when \(L \neq 1\mathcal{V}\).

The results (86.12) and (86.13) show, roughly, that for “infinitesimal” \(M \in \text{Lin}\mathcal{V}\), the right and left polar decompositions of \(1\mathcal{V} + M\) coincide and are given by \((1\mathcal{V} + A, 1\mathcal{V} + U)\) when \((U, A)\) is the additive decomposition of \(M\).

Notes 86

(1) In the literature on continuum mechanics (including some of my own papers, I must admit) Prop.1 is taken tacitly for granted. I know of no other place in the literature where it is proved.

87 The Structure of Skew Lineons

A genuine inner-product space \(\mathcal{V}\) is assumed to be given. First, we deal with an important special type of skew lineons:
Definition 1: We say that a lineon on \( \mathcal{V} \) is a **perpendicular turn** if it is both skew and orthogonal.

If \( U \) is a perpendicular turn and \( u \in \mathcal{V} \), then \( |Uu| = |u| \) because \( U \) is orthogonal, and \((Uu) \cdot u = 0\), i.e. \( Uu \perp u \), because \( U \) is skew. Hence \( U \) “turns” every vector into one that is perpendicular and has the same magnitude.

The following two results are immediate consequences of the definition.

**Proposition 1:** A lineon \( U \) on \( \mathcal{V} \) is a perpendicular turn if and only if it satisfies any two of the following three conditions: (a) \( U \in \text{Skew} \mathcal{V} \), (b) \( U \in \text{Orth} \mathcal{V} \), (c) \( U^2 = -1 \).

**Proposition 2:** Let \( U \) be a perpendicular turn on \( \mathcal{V} \) and let \( u \in \mathcal{V} \) be a unit vector. Then \((u, Uu)\) is an orthonormal pair and its span \( \text{Lsp}\{u, Uu\} \) is a two-dimensional \( U \)-space.

The structure of perpendicular turns is described by the following result.

**Structure Theorem for Perpendicular Turns:** If \( U \) is a perpendicular turn, then there exists an orthonormal subset \( c \) of \( \mathcal{V} \) such that \( (\text{Lsp}\{e, Ue\} \mid e \in c) \) is an orthogonal decomposition of \( \mathcal{V} \) whose terms are two-dimensional \( U \)-spaces. The set \( c \) can be chosen so as to contain any single prescribed unit vector. We have \( 2|c| = \dim \mathcal{V} \).

**Proof:** Consider orthonormal subsets \( c \) of \( \mathcal{V} \) such that \((\text{Lsp}\{e, Ue\} \mid e \in c)\) is an orthogonal family of subspaces of \( \mathcal{V} \). It is a trivial consequence of Prop.2 that the empty subset of \( \mathcal{V} \) and also singletons with unit vectors have this property. Choose a maximal set \( c \) with this property. It is clear that \( c \) can be chosen so as to contain any prescribed unit vector. Since all the spaces \( \text{Lsp}\{u, Uu\} \) are \( U \)-spaces by Prop.2, it follows that the sum

\[
U := \sum_{e \in c} \text{Lsp}\{e, Ue\}
\]

of the family \((\text{Lsp}\{e, Ue\} \mid e \in c)\) is also a \( U \)-space. Since \( U \) is skew, it follows from Prop.1 of Sect.82 that the orthogonal supplement \( U^\perp \) of \( U \) is also a \( U \)-space. Now if \( U^\perp \neq \{0\} \), we may choose a unit vector \( f \in U^\perp \). Since \( U^\perp \) is a \( U \)-space, we have \( \text{Lsp}\{f, Uf\} \subset U^\perp \). It follows that \( c \cup \{f\} \) is a subset of \( \mathcal{V} \) such that \((\text{Lsp}\{e, Ue\} \mid e \in c \cup \{f\})\) is an orthogonal family, which contradicts the maximality of \( c \). We conclude that \( U^\perp = \{0\} \) and hence \( U = \mathcal{V} \), which means that \((\text{Lsp}\{e, Ue\} \mid e \in c)\) is an orthogonal decomposition of \( \mathcal{V} \). By Prop.2, the terms of the decomposition are two-dimensional and by Prop.4 of Sect.81, we have \( 2|c| = \dim \mathcal{V} \).

**Corollary:** Perpendicular turns on \( \mathcal{V} \) can exist only when \( \dim \mathcal{V} \) is even. Let \( U \) be such a turn and let \( m := \frac{1}{2} \dim \mathcal{V} \). Then there exists an orthonormal
basis $e := (e_i \mid i \in (2m)^1)$ such that

$$Ue_i = \begin{cases} e_{i+1} & \text{when } i \text{ is odd} \\ -e_{i-1} & \text{when } i \text{ is even} \end{cases}$$

(87.1)

for all $i \in (2m)^1$.

**Proof:** The basis $e$ is obtained by choosing an orthonormal set $c$ as in the Theorem, so that $\sharp c = m$, and enumerating it by odd numbers, so that $c = \{e_i \mid i \in 2(m^1)-1\}$. The $e_i$ for $i \in 2(m^1)$ are then defined by $e_i = Ue_{i-1}$. Since $U^2 = -1_V$ by (c) of Prop.1, we have $Ue_i = U^2e_{i-1} = -e_{i-1}$ when $i \in 2(m^1)$.

If $m$ is small, then the $(2m)$-by-$(2m)$ matrix $[U]_e$ of $U$ relative to any basis for which (87.1) holds can be recorded explicitly in the form

$$[U]_e = \begin{bmatrix} 0 & -1 & 0 & 1 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 1 & 0 \\ & & & & & & \cdots & & & & \end{bmatrix},$$

(87.2)

where zeros are omitted.

In order to deal with skew lineons, we need the following analogue of Def.1 of Sect.82.

**Definition 2:** Let $L$ be a lineon on $V$. For every $\kappa \in P \times$, we write

$$Q_{sps} L(\kappa) := \text{Null } (L^2 + \kappa^2 1_V).$$

(87.3)

This is an $L$-space; if it is non-zero, we call it the **quasi-spectral space** of $L$ for $\kappa$. The **quasi-spectrum** of $L$ is defined to be

$$Q_{spec} L := \{\kappa \in P^\times \mid Q_{sps} L(\kappa) \neq \{0\}\}.$$  

(87.4)

In view of Def.1 of Sect.82 we have

$$Q_{spec} L = \{\sqrt{\sigma} \mid \sigma \in \text{Spec } (-L^2) \cap P^\times\}$$

(87.5)

and

$$Q_{sps} L(\kappa) = Sps_{-L^2}(\kappa^2) \text{ for all } \kappa \in P^\times$$

(87.6)

for all $L \in \text{Lin} V$. 

The proof of the following result is based on the Spectral Theorem.

**Structure Theorem for Skew Lineons:** A lineon \( A \) on \( V \) is skew if and only if (i) \( \text{Rng} \, A = (\text{Null} \, A)^\perp \), (ii) the family of quasi-spectral spaces of \( A \) is an orthogonal decomposition of \( \text{Rng} \, A \), and (iii) for each \( \kappa \in \text{Qspec} \, A \) we have

\[
A|_{W_\kappa} = \kappa U_\kappa,
\]

(87.7)

where \( W_\kappa := \text{Qsps}_A(\kappa) \) and where \( U_\kappa \in \text{Lin} \, W_\kappa \) is a perpendicular turn.

**Proof:** Assume that \( A \) is skew. It follows from (22.9) that \( \text{Rng} \, A = \text{Rng} (-A) = \text{Rng} (A^T) = (\text{Null} \, A)^\perp \), i.e. that (i) is valid. Since \( A^T A = (-A)A = -A^2 \) is positive by Prop.1 of Sect.85, it follows from Prop.2 of Sect.85 that \( \text{Spec} (-A^2) \subset \mathbb{P} \). Hence it follows from (87.5) that

\[
\text{Qspec} \, A = \{ \sqrt{\sigma} | \sigma \in \text{Spec} (-A^2) \setminus \{0\} \}.
\]

In view of (87.6) we conclude that the spectral spaces of \( -A^2 \) are \( \text{Qsps}_A(\kappa) \) with \( \kappa \in \text{Qspec} \, A \), and also \( \text{Sps}_{-A^2}(0) \) if this subspace is not zero. Since it is easily seen that \( \text{Sps}_{-A^2}(0) = \text{Null} \, A^2 = \text{Null} \, A \), it follows from the Spectral Theorem, applied to \( -A^2 \), that \( (\text{Qsps}_A(\kappa) | \kappa \in \text{Qspec} \, A) \) is an orthogonal decomposition of \( (\text{Null} \, A)^\perp = \text{Rng} \, A \), i.e. that (ii) is valid. Now let \( \kappa \in \text{Qspec} \, A \) be given. By (87.3) we have \((A^2 + \kappa^2 1_V)|_{W_\kappa} = 0\) when \( W_\kappa := \text{Qsps}_A(\kappa) \) and hence, since \( W_\kappa \) is an \( A \)-space, \((\frac{1}{\kappa} A|_{W_\kappa})^2 = -1_{W_\kappa}\).

Since \( \frac{1}{\kappa} A|_{W_\kappa} \in \text{Skew} \, W_\kappa \), it follows from Prop.1 that (iii) is valid.

Assume now that the conditions (i), (ii), and (iii) hold and put \( W_\circ := \text{Null} \, A \) and \( K := (\text{Qspec} \, A) \cup \{0\} \). Then \( (W_\kappa | \kappa \in K) \) is an orthogonal decomposition of \( V \). Let \((E_\kappa | \kappa \in K)\) be the associated family of idempotents (see Prop.5 of Sect.81). Using Prop.6 of Sect.81, (81.8), and (87.7), we find that

\[
A = \sum_{\kappa \in \text{Qspec} \, A} \kappa U_\kappa|_{E_\kappa}|W_\kappa.
\]

(87.8)

Since \( U_\kappa \in \text{Skew} \, W_\kappa \) and, by Prop.4 of Sect.83, \( E_\kappa \in \text{Sym} \, V \) for each \( \kappa \in \text{Qspec} \, A \), it is easily seen that each term in the sum on the right side of (87.8) is skew and hence that \( A \) is skew.

Assume that \( A \in \text{Skew} \, V \). It follows from (iii) of the Theorem just proved and the Corollary to the Structure Theorem for Perpendicular Turns that \( \dim W_\kappa \) is even for all \( \kappa \in \text{Qspec} \, A \). Hence, by (ii) and Prop.3 of Sect.81, we obtain the following

**Proposition 3:** If \( A \) is a skew lineon on \( V \), then \( \dim(\text{Rng} \, A) \) is even. In particular, if the dimension of \( V \) is odd, then there exist no invertible skew lineons on \( V \).
The following corollary to the Structure Theorem above is obtained by choosing, in each of the quasi-spectral spaces, an orthonormal basis as described in the Corollary to the Structure Theorem for Perpendicular Turns.

**Corollary:** Let \( \mathbf{A} \) be a skew lineon on \( \mathcal{V} \) and put \( n := \dim \mathcal{V}, m := \frac{1}{2} \dim(\text{Rng} \ A) \). Then there is an orthonormal basis \( \mathbf{e} := (\mathbf{e}_i \mid i \in n^\downarrow) \) of \( \mathcal{V} \) and a list \( (\kappa_k \mid k \in m^\downarrow) \) of strictly positive real numbers such that

\[
\begin{align*}
\mathbf{Ae}_{2k-1} &= \kappa_k \mathbf{e}_{2k} \\
\mathbf{Ae}_{2k} &= -\kappa_k \mathbf{e}_{2k-1}
\end{align*}
\] (87.9)

\[
\mathbf{Ae}_i = 0 \quad \text{for all} \quad i \in n^\downarrow \setminus (2m)^\downarrow.
\]

The only non-zero terms in the matrix \([\mathbf{A}]_e\) of \( \mathbf{A} \) relative to a basis \( \mathbf{e} \) for which (87.9) holds are those in blocks of the form

\[
\begin{bmatrix}
0 & -\kappa_k \\
\kappa_k & 0
\end{bmatrix}
\]

along the diagonal. For example, if \( \dim \mathcal{V} = 3 \) and \( \mathbf{A} \neq 0 \), then \([\mathbf{A}]_e\) has the form

\[
[\mathbf{A}]_e = \begin{bmatrix}
0 & -\kappa & 0 \\
\kappa & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\text{ for some } \kappa \in \mathbb{R}^\times.
\] (87.10)

Notes 87

1. I am introducing the term “perpendicular turn” here for the first time; I believe it is very descriptive.

2. The concepts of a quasi-spectrum, quasi-spectral space, and the notations of Def.2 are introduced here for the first time.

## 88 Structure of Normal and Orthogonal Lineons

Again, a genuine inner-product space \( \mathcal{V} \) is assumed to be given.

**Definition 1:** We say that a lineon on \( \mathcal{V} \) is **normal** if it commutes with its transpose.

It is clear that symmetric, skew, and orthogonal lineons are all normal.

**Proposition 1:** A lineon \( \mathbf{N} \in \text{Lin} \mathcal{V} \) is normal if and only if the two terms \( S := \frac{1}{2}(\mathbf{N} + \mathbf{N}^\top) \) and \( A := \frac{1}{2}(\mathbf{N} - \mathbf{N}^\top) \) of the additive decomposition \((S, A)\) of \( \mathbf{N} \) commute. If this is the case, then any two of \( \mathbf{N}, \mathbf{N}^\top, S \) and \( A \) commute.
Proof: We have $N = S + A$ and, since $S \in \text{Sym}V$, $A \in \text{Skew}V$, $N^\top = S - A$. It follows that $NN^\top - N^\top N = 2(AS - SA)$. Therefore, $N$ and $N^\top$ commute if and only if $S$ and $A$ commute. 

Proposition 2: A lineon $N \in \text{Lin}V$ is normal if and only if
\[
|Nv| = |N^\top v| \quad \text{for all } v \in V. \tag{88.1}
\]

If this is the case, then $\text{Null } L = \text{Null } L^\top$.

Proof: We have $|Nv|^2 = Nv \cdot Nv = v \cdot N^\top Nv$ and $|N^\top v|^2 = v \cdot NN^\top v$ for all $v \in V$. Hence (88.1) holds if and only if the quadratic forms corresponding to $N^\top N$ and $NN^\top$ are the same (see (85.1)). This is the case if and only if $N^\top N$ and $NN^\top$ are the same (see Sect.27).

Proposition 3: An invertible lineon $N \in \text{Lin}V$ is normal if and only if the two terms of its right [left] polar decomposition commute (see Sect.86).

Proof: It is clear from the Polar Decomposition Theorem of Sect.86 that the terms of the right or of the left polar decomposition commute if and only if the right and the left polar decompositions coincide, i.e. if and only if $S = S'$ in the notation of the statement of this theorem. But by (86.4)_1 and (86.7)_1, we have $S = S'$ if and only if $NN^\top = N^\top N$.

In order to deal with the structure of normal lineons, we need the following analogue of Def.2 of Sect.87.

Definition 2: Let $L$ be a lineon on $V$. For every $\nu \in \mathbb{R} \times P^\times$ we write
\[
P_{\text{spec}}(\nu) := \text{Null } ((L - \nu_1 1_V)^2 + \nu_2^2 1_V). \tag{88.2}
\]
This is an $L$-space; if it is not zero, we call it the pair-spectral space of $L$ for $\nu$. The pair-spectrum of $L$ is defined to be
\[
P_{\text{spec}}(\nu) := \{ \nu \in \mathbb{R} \times P^\times \mid P_{\text{spec}}(\nu) \neq \{0\} \}. \tag{88.3}
\]

The proof of the following theorem is based on the Structure Theorem for Skew Lineons and the Spectral Theorem.

Structure Theorem for Normal Lineons: A lineon $N$ on $V$ is normal if and only if the family of its pair-spectral spaces is an orthogonal decomposition of $V$ and, for each $\nu \in P_{\text{spec}}N$, we have
\[
N|_{W_{\nu}} = \begin{cases} 
\nu_1 1_{W_{\nu}} & \text{if } \nu_2 = 0 \\
\nu_1 1_{W_{\nu}} + \nu_2 U_{\nu} & \text{if } \nu_2 \neq 0
\end{cases}
\tag{88.4}
\]
where $W_{\nu} := P_{\text{spec}}N(\nu)$, and where $U_{\nu} \in \text{Lin}W_{\nu}$ is a perpendicular turn.

First we prove an auxiliary result:
**Lemma:** If $N \in \text{Lin}V$ is normal and if $(S, A)$ is the additive decomposition of $N$, then

$$P_{S_A}(\sigma, \kappa) = \begin{cases} S_{S_S}(\sigma) \cap Q_{S_A}(\kappa) & \text{if } \kappa \neq 0 \\ S_{S_S}(\sigma) \cap \text{Null } A & \text{if } \kappa = 0 \end{cases}$$

for all $\sigma \in \mathbb{R}$ and all $\kappa \in \mathbb{P}$.

**Proof:** Let $\sigma \in \mathbb{R}$ and $\kappa \in \mathbb{P}$ be given and put $M := N - \sigma 1_V$ and $T := S - \sigma 1_V$. Then $M$ is normal and $(T, A)$ is the additive decomposition of $M$. Using the fact that any two of $M, M^\top, T$ and $A$ commute (see Prop. 1), we obtain

$$(M^2 + \kappa^2 1_V)^\top (M^2 + \kappa^2 1_V) = (M^\top M)^2 + \kappa^2 (M^\top + M^2) + \kappa^4 1_V$$

$$= ((T - A)(T + A))^2 + \kappa^2 ((T - A)^2 + (T + A)^2) + \kappa^4 1_V$$

$$= T^4 + 2(A T)^\top (A T) + 2\kappa^2 T^2 + (A^2 + \kappa^2 1_V)^2.$$ 

Since $T, T^2$ and $(A^2 + \kappa^2 1_V)$ are symmetric, it follows that

$$|(M^2 + \kappa^2 1_V)v|^2 = |T^2 v|^2 + 2|A Tv|^2 + 2\kappa^2 |Tv|^2 + |(A^2 + \kappa^2 1_V)v|^2$$

for all $v \in V$, from which we infer that $v \in \text{Null } (M^2 + \kappa^2 1_V) = P_{S_A}(\sigma, \kappa)$ if and only if $v \in \text{Null } T \cap \text{Null } (A^2 + \kappa^2 1_V)$. Since $\text{Null } T = S_{S_S}(\sigma)$, since $\text{Null } (A^2 + \kappa^2 1_V) = Q_{S_A}(\kappa)$ when $\kappa \neq 0$ by Def. 2 of Sect. 82, and since $\text{Null } A^2 = \text{Null } A$ because $A$ is skew, (88.5) follows.

**Proof of Theorem:** Assume that $N$ is normal and that $(S, A)$ is its additive decomposition. Let $\sigma \in \text{Spec } S$ be given and put $U := S_{S_S}(\sigma)$. By Prop. 2 of Sect. 82, $U$ is $A$-invariant and we may hence consider the adjustment $A_{U} \in \text{Skew } U$. It is clear that $\text{Null } A_{U} = U \cap \text{Null } A$ and

$$Q_{S_A}(\kappa) = U \cap Q_{S_A}(\kappa) \quad \text{for all } \kappa \in \mathbb{P}^\times.$$ 

Hence, by the Lemma, we conclude that

$$P_{S_A}(\sigma, \kappa) = \begin{cases} Q_{S_A}(\kappa) & \text{for all } \kappa \in \mathbb{P}^\times \\ \text{Null } A_{U} & \text{if } \kappa = 0 \end{cases}.$$ 

Since $S_{U} = \sigma 1_U$, $U$ is $N$-invariant and

$$N_{U} = \sigma 1_U + A_{U}.$$ 

We now apply the Structure Theorem for Skew Lineons to $A_{U} \in \text{Skew } U$. In view of (88.6), and since $Q_{\text{Spec } A} \subset Q_{\text{Spec } A}$, we have
(Psps_{N}(\nu) \mid \nu \in \text{Pspec} \, N, \, \nu_1 = \sigma) \text{ is an orthogonal decomposition of } \mathcal{U},
and for all \nu \in \text{Pspec} \, N \text{ with } \nu_1 = \sigma \text{ we have }

A_{|_{W_{\nu}}} = 0 \text{ if } \nu_2 = 0 \text{ and } A_{|_{W_{\nu}}} = \nu_2 \mathbf{U}_{\nu} \text{ if } \nu_2 \neq 0,

where \, W_{\nu} := \text{Psps}_{N}(\nu) \text{ and where } \mathbf{U}_{\nu} \in \text{Lin} \, W_{\nu} \text{ is a perpendicular turn.}

Using this result and (88.7), adjusted to the subspace \, W_{\nu} = W(\sigma, \nu_2) \text{ of } \mathcal{U},
we infer that (88.4) is valid.

By the Spectral Theorem \,(\text{Sps}_{S}(\sigma) \mid \sigma \in \text{Spec} \, S)\text{ is an orthogonal decom-
position of } \mathcal{V}. \text{ Since, as we just have seen, } (\text{Psps}_{N}(\nu) \mid \nu \in \text{Pspec} \, N, \, \nu_1 = \sigma) \text{ is an orthogonal decomposition of } \text{Sps}_{S}(\sigma) \text{ for each } \sigma \in \text{Spec} \, S, \text{ it follows}
that } (\text{Psps}_{N}(\nu) \mid \nu \in \text{Pspec} \, N) \text{ is an orthogonal decomposition of } \mathcal{V}. \text{ This}
completes the proof of the “only-if” -part of the Theorem. \text{ The proof of}
the “if”-part goes the same way as that of the “if”-part of the Structure
Theorem for Skew Lineons. We leave the details to the reader.

For a symmetric lineon we have \, S, \, \text{Pspec} \, S = \text{Spec} \, S \times \{0\} \text{ and }
\text{Psps}_{S}((\sigma,0)) = \text{Sps}_{S}(\sigma) \text{ for all } \sigma \in \text{Spec} \, S. \text{ For a skew lineon } A \text{ we have}
\text{Pspec} \, A = \{0\} \times (\text{Spec} \, A \cup \text{Qspec} \, A) \text{ and } \text{Psps}_{A}(0, \kappa) = \text{Qsps}_{A}(\kappa) \text{ for all}
\kappa \in \text{Qspec} \, A, \text{ while } \text{Psps}_{A}((0,0)) = \text{Null} \, A \text{ when } A \text{ is not invertible.}

As in the previous section, we obtain a corollary to the Structure Theo-
rem just given as follows.

Corollary: Let \, N \text{ be a normal lineon on } \mathcal{V}. \text{ Then there exists a number}
m \in \mathbb{N} \text{ with } 2m \leq n := \text{dim} \, \mathcal{V}, \text{ an orthonormal basis } \mathbf{e} := \{(e_i \mid i \in n)\}
of \, \mathcal{V}, \text{ a list } (\mu_k \mid k \in m) \text{ in } \mathbb{R}, \text{ a list } (\kappa_k \mid k \in m) \text{ in } \mathbb{P}^\times, \text{ and a family}
(\lambda_i \mid i \in n \setminus (2m)) \text{ in } \mathbb{R} \text{ such that}

\begin{align*}
\mathbf{N} \mathbf{e}_{2k-1} &= \mu_k \mathbf{e}_{2k-1} + \kappa_k \mathbf{e}_{2k} \\
\mathbf{N} \mathbf{e}_{2k} &= \mu_k \mathbf{e}_{2k} - \kappa_k \mathbf{e}_{2k-1}
\end{align*}

(88.8)

\begin{align*}
\mathbf{N} \mathbf{e}_i &= \lambda_i \mathbf{e}_i \quad \text{for all } i \in n \setminus (2m).
\end{align*}

The only non-zero terms in the matrix \, [\mathbf{N}]_e \text{ of } \mathbf{N} \text{ relative to a basis } \mathbf{e}
for which (88.8) holds are either on the diagonal or in blocks of the form

\begin{bmatrix}
\mu_k & -\kappa_k \\
\kappa_k & \mu_k
\end{bmatrix}

along the diagonal. For example, if \, \text{dim} \, \mathcal{V} = 3 \text{ and if } \mathbf{N} \text{ is not}
symmetric, then \, [\mathbf{N}]_e \text{ has the form}

\begin{bmatrix}
\mu & -\kappa & 0 \\
\kappa & \mu & 0 \\
0 & 0 & \lambda
\end{bmatrix}, \quad \kappa \in \mathbb{P}^\times, \, \lambda, \, \mu \in \mathbb{R}. \quad (88.9)

Orthogonal lineons are special normal lineons. To describe their struc-
ture, the following definition is useful.
88. STRUCTURE OF NORMAL AND ORTHOGONAL LINEONS

**Definition 3:** Let $L$ be a lineon on $V$. For every $\theta \in [0, \pi[$ we write

$$\text{Aspec}_L(\theta) := \text{Psps}_L((\cos \theta, \sin \theta))$$

(see (88.2)). This is an $L$-space; if it is not zero, we call it the angle-spectral space of $L$ for $\theta$. The angle-spectrum of $L$ is defined to be

$$\text{Aspec } L := \{ \theta \in [0, \pi[ \mid \text{Psps}_L(\theta) \neq \{0\} \}.$$  (88.11)

The following result is a corollary to the Structure Theorem for Normal Lineons.

**Structure Theorem for Orthogonal Lineons:** A lineon $R$ on $V$ is orthogonal if and only if (i) $\text{Null } (R + 1_V) \perp \text{Null } (R - 1_V)$, (ii) the family of angle-spectral spaces is an orthogonal decomposition of $(\text{Null } (R + 1_V) + \text{Null } (R - 1_V))^\perp$, and (iii) for each $\theta \in \text{Aspec } R$ we have

$$R_{\mathcal{U}_\theta} = \cos \theta 1_{\mathcal{U}_\theta} + \sin \theta V_\theta,$$

(88.12) where $\mathcal{U}_\theta := \text{Psps}_R(\theta)$ and where $V_\theta \in \text{Lin } \mathcal{U}_\theta$ is a perpendicular turn.

**Proof:** The conditions (i), (ii), (iii) are satisfied if and only if the conditions of the Structure Theorem for Normal Lineons are satisfied with $N := R$ and

$$\text{Psps } R \subset \{(1, 0), (-1, 0)\} \cup \{ (\cos \theta, \sin \theta) \mid \theta \in \text{Aspec } R \}.$$ (88.13)

Thus we may assume that $R$ is normal. Since $(\text{Psps}_R(\nu) \mid \nu \in \text{Psps } R)$ is an orthogonal decomposition of $V$ whose terms $W_\nu := \text{Psps}_R(\nu)$ are $R$-invariant, it is easily seen that $R$ is orthogonal if and only if $R_{W_\nu}$ is orthogonal for each $\nu \in \text{Psps } R$. It follows from (88.4), with $N := R$, that this is the case if and only if $\nu_2 + \nu_2^2 = 1$ for all $\nu \in \text{Psps } R$, which is equivalent to (88.13). If $\nu_2 = 0$ we have $\nu = (1, 0)$ or $\nu = (-1, 0)$ and $\text{Psps}_R((1, 0)) = \text{Null } (R - 1_V)$ and $\text{Psps}_R((-1, 0)) = \text{Null } (R + 1_V)$. If $\nu_2 \neq 0$ we have $\nu = (\cos \theta, \sin \theta)$ for some $\theta \in [0, \pi[$ and (88.4) reduces to (88.12) with $V_\theta := U_{(\cos \theta, \sin \theta)}$.

As before we obtain a corollary as follows:

**Corollary:** Let $R$ be an orthogonal lineon on $V$. Then there exist numbers $m \in \mathbb{N}$ and $p \in \mathbb{N}$ with $2m \leq p \leq n := \dim V$, an orthonormal basis $e := (e_i \mid i \in n^1)$ of $V$, and a list $(\theta_k \mid k \in m^1)$ in $[0, \pi[$ such that

$$R e_{2k-1} = \cos \theta_k e_{2k-1} + \sin \theta_k e_{2k}$$

$$R e_{2k} = \cos \theta_k e_{2k} - \sin \theta_k e_{2k-1}$$

for all $k \in m^1$.
\[ \Re e_i = -e_i \quad \text{for all} \quad i \in p \setminus (2m), \]
\[ \Re e_i = e_i \quad \text{for all} \quad i \in n \setminus p. \] (88.14)

The only non-zero terms in the matrix \([R]_e\) of \(R\) relative to a basis \(e\) for which (88.14) holds are either 1 or \(-1\) on the diagonal or in blocks of the form
\[
\begin{bmatrix}
\cos \theta_k & -\sin \theta_k \\
\sin \theta_k & \cos \theta_k
\end{bmatrix}
\]
along the diagonal. For example, if \(\dim V = 3\) and if \(R\) is not symmetric, then \([R]_e\) has the form
\[
[R]_e = \begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & \pm 1
\end{bmatrix}, \quad \theta \in ]0, \pi[. \tag{88.15}
\]

If the + sign is appropriate, then \(R\) is a rotation by an angle \(\theta\) about an axis in the direction of \(e_3\). If the –sign is appropriate, then \(R\) is obtained from such a rotation by composing it with a reflection in the plane \(\{e_3\}^\perp\).

Notes 88

(1) The concepts of a pair-spectrum, pair-spectral space, angle-spectrum, and angle-spectral space, and the notations of Defs.2 and 3 are introduced here for the first time.

89 Complex Spaces, Unitary Spaces

**Definition 1:** A complex space is a linear space \(V\) (over \(\mathbb{R}\)) endowed with additional structure by the prescription of a linear \(J\) on \(V\) that satisfies \(J^2 = -1_V\). We call \(J\) the complexor of \(V\).

A complex space \(V\) acquires the natural structure of a linear space over the field \(\mathbb{C}\) of complex numbers if we stipulate that the scalar multiplication \(\text{sm}^C : \mathbb{C} \times V \to V\) of \(V\) as a space over \(\mathbb{C}\) be given by
\[
\text{sm}^C(\zeta, u) = (\Re \zeta)u + (\Im \zeta)Ju \quad \text{for all} \quad \zeta \in \mathbb{C}, \tag{89.1}
\]
where \(\Re \zeta\) and \(\Im \zeta\) denote the real and imaginary parts of \(\zeta\). Indeed, it is easily verified that \(\text{sm}^C\) satisfies the axioms (S1)–(S4) of Sect.11 if we take \(F := \mathbb{C}\). The scalar multiplication \(\text{sm} : \mathbb{R} \times V \to V\) of \(V\) as a space over \(\mathbb{R}\) is simply the restriction of \(\text{sm}^C\) to \(\mathbb{R} \times V\):
\[
\text{sm} = \text{sm}^C|_{\mathbb{R} \times V}. \tag{89.2}
\]
We use the simplified notation
\[ \zeta \mathbf{u} := \text{sm}^C(\zeta, \mathbf{u}) \quad \text{for all} \quad \zeta \in \mathbb{C}, \; \mathbf{u} \in \mathcal{V} \] (89.3)
as described in Sect.11.

Conversely, every linear space \( \mathcal{V} \) over the field \( \mathbb{C} \) of complex numbers has the natural structure of a complex space in the sense of Def.1. The structure of \( \mathcal{V} \) as a linear space over \( \mathbb{R} \) is obtained simply by restricting the scalar multiplication of \( \mathcal{V} \) to \( \mathbb{R} \times \mathcal{V} \), and the complexor of \( \mathcal{V} \) is the mapping \( (\mathbf{u} \mapsto i\mathbf{u}) : \mathcal{V} \rightarrow \mathcal{V} \).

Let \( \mathcal{V} \) be a complex space. Most of the properties and constructions involving \( \mathcal{V} \) discussed in Chaps.1 and 2 depend on whether \( \mathcal{V} \) is regarded as a linear space over \( \mathbb{R} \) or over \( \mathbb{C} \). To remove the resulting ambiguities, we use the prefix “\( \mathbb{C} \)” or a superscript \( \mathbb{C} \) if we refer to the structure of \( \mathcal{V} \) as a linear space over \( \mathbb{C} \). For example, a non-empty subset \( c \) of \( \mathcal{V} \) that is a \( \mathbb{C} \)-basis is not a basis relative to the structure of \( \mathcal{V} \) as a space over \( \mathbb{R} \). However, it is easily seen that \( c \in \text{Sub} \mathcal{V} \) is a \( \mathbb{C} \)-basis set if and only if \( c \cup J_>(c) \) is a basis set of \( \mathcal{V} \). If this is the case, then \( c \cap J_>(c) = \emptyset \) and hence, since \( J \) is injective, \( \sharp(c \cup J_>(c)) = 2(\sharp c) \). Therefore, the dimension of \( \mathcal{V} \) is a linear space over \( \mathbb{R} \) is twice the dimension of \( \mathcal{V} \) as a linear space over \( \mathbb{C} \):
\[ \dim \mathcal{V} = 2 \dim^C \mathcal{V}. \] (89.4)

It follows that \( \dim \mathcal{V} \) must be an even number.

Let \( \mathcal{V} \) and \( \mathcal{V}' \) be complex spaces, with complexors \( J \) and \( J' \), respectively. The set of all \( \mathbb{C} \)-linear mappings from \( \mathcal{V} \) to \( \mathcal{V}' \) is easily seen to be given by
\[ \text{Lin}^C(\mathcal{V}, \mathcal{V}') = \{ L \in \text{Lin}(\mathcal{V}, \mathcal{V}') \mid LJ = J'L \}. \] (89.5)

In particular, we have \( \text{Lin}^C(\mathcal{V}, \mathcal{V}) =: \text{Comm}J \), the commutant-algebra of the complexor \( J \) (see 18.2). The set \( \mathbb{C}1\mathcal{V} := \{ \zeta 1\mathcal{V} \mid \zeta \in \mathbb{C} \} \) is a subalgebra of \( \text{Lin}^C \mathcal{V} \) and we have
\[ J = i1\mathcal{V}. \] (89.6)

Let \( J \) be the complexor of a complex space \( \mathcal{V} \). Since \( (-J)^2 = J^2 = -1\mathcal{V} \), we can also consider \( \mathcal{V} \) as a complex space with \( -J \) rather than \( J \) as the designated complexor. We call the resulting structure of \( \mathcal{V} \) the conjugate-complex structure of \( \mathcal{V} \). In view of (89.1), the corresponding scalar multiplication, denoted by \( \text{sm}^C : \mathbb{C} \times \mathcal{V} \rightarrow \mathcal{V} \), is given by
\[ \text{sm}^C(\zeta, \mathbf{u}) = (\text{Re}\zeta)\mathbf{u} - (\text{Im}\zeta)J\mathbf{u} \quad \text{for all} \quad \zeta \in \mathbb{C} \] (89.7)
and we have
\[ \text{sn}^C(\zeta, u) = \overline{\zeta} u \quad \text{for all} \quad \zeta \in \mathbb{C}, \ u \in V. \] (89.8)
where \( \overline{\zeta} \) denotes the complex-conjugate of \( \zeta \).

Let \( V \) and \( V' \) be complex spaces, with complexors \( J \) and \( J' \), respectively.
We say that a mapping from \( V \) to \( V' \) is \textbf{conjugate-linear} if it is \( \mathbb{C} \)-linear when one of \( V \) and \( V' \) is considered as a linear space over \( \mathbb{C} \) relative to its conjugate-complex structure. The set of all conjugate-linear mappings from \( V \) to \( V' \) will be denoted by \( \text{Lin}^C(V, V') \), and we have
\[ \text{Lin}^C(V, V') = \{ L \in \text{Lin}(V, V') \mid LJ = -J'L \}. \] (89.9)
A mapping \( L \in \text{Lin}(V, V') \) is conjugate-linear if and only if
\[ L(\overline{\zeta} u) = \overline{\zeta} Lu \quad \text{for all} \quad \zeta \in \mathbb{C}, \ u \in V. \] (89.10)
A mapping \( L : V \to V' \) is \( \mathbb{C} \)-linear when both \( V \) and \( V' \) are considered as linear spaces over \( \mathbb{C} \) relative to their conjugate-complex structures if and only if it is \( \mathbb{C} \)-linear in the ordinary sense.

Let \( V \) be a complex space with complexor \( J \). Its dual \( V^* \) then acquires the natural structure of a complex space if we stipulate that the complexor of \( V^* \) be \( J^\top \in \text{Lin}^C(V^*, V^*) \). The dual \( V^* := \text{Lin}(V, \mathbb{R}) \) must be carefully distinguished from the “complex dual” \( \text{Lin}^C(V, \mathbb{C}) \) of \( V \). However, the following result shows that the two are naturally \( \mathbb{C} \)-linearly isomorphic.

**Proposition 1:** The mapping
\[ (\omega \mapsto \text{Re} \omega) : \text{Lin}^C(V, \mathbb{C}) \to V^*. \] (89.11)
is a \( \mathbb{C} \)-linear isomorphism. Its inverse \( \Gamma \in \text{Lis}^C(V^*, \text{Lin}^C(V, \mathbb{C})) \) is given by
\[ (\Gamma \lambda)u = \lambda u - i \lambda(Ju) \quad \text{for all} \quad u \in V, \quad \lambda \in V^*. \] (89.12)

**Proof:** We put \( V^\dagger := \text{Lin}^C(V, \mathbb{C}) \) and denote the mapping (89.11) by \( \Phi : V^\dagger \to V^* \), so that
\[ (\Phi \omega)u = \text{Re}(\omega u) \quad \text{for all} \quad \omega \in V^\dagger, \ u \in V. \] (89.13)
It is clear that \( \Phi \) is linear. By (22.3) and (89.6) we have
\[ ((J^\dagger \Phi) \omega)u = (\Phi \omega)(Ju) = \Phi \omega(iu) \]
\[ = \Phi(i(\omega u)) = ((\Phi(i1_{V^\dagger}))\omega)u. \]
for all $\omega \in V^\dagger$ and all $u \in V$, and hence

$$J^\top \Phi = \Phi(i1_{V^\dagger}).$$

Since $J^\top$ is the complexor of $V^*$ and $i1_{V^\dagger}$ the complexor of $V^\dagger$, it follows from (89.5) that $\Phi$ is $\mathbb{C}$-linear.

We now define $\Gamma : V^* \to V^\dagger$ by (89.12). Using (89.13), we then have

$$((\Gamma\Phi)(\omega))u = (\Gamma(\Phi(\omega)))u = (\Phi(\omega)u - i(\Phi(\omega)(Ju)) = Re(\omega u) - iRe(\omega Ju)$$

for all $\omega \in V^\dagger$ and all $u \in V$. It follows that $\Gamma\Phi = 1_{V^\dagger}$. In a similar way, one proves that $\Phi\Gamma = 1_{V^*}$ and hence that $\Gamma = \Phi^{-1}$.

**Definition 2:** A unitary space is an inner-product space $V$ endowed with additional structure by the prescription of a perpendicular turn $J$ on $V$ (see Def.1 of Sect.87).

We say that $V$ is a genuine unitary space if its structure as an inner-product space is genuine.

Since $J^2 = -1_V$ by Prop.1 of Sect.87, a unitary space has the structure of a complex space with complexor $J$ in the sense of Def.1. The identification $V^* \cong V$ has to be treated with some care. The natural complex-space structure of $V^*$ is determined by the complexor $J^\top$, but, since $J^\top = -J$ for a perpendicular turn, the complex structure of $V^*$, when identified with $V$, is the conjugate-complex structure of $V$, not the original complex structure. The identification mapping $(v \mapsto \cdot v) : V \to V^*$ (see Sect.41) is not $\mathbb{C}$-linear but conjugate-linear.

We define the unitary product $u_p : V \times V \to \mathbb{C}$ of a unitary space $V$ by

$$u_p(u, v) := \Gamma(v \cdot u) \quad \text{for all} \quad u, v \in V, \quad (89.14)$$

where $\Gamma \in \text{Lis}^C(V^*, \text{Lin}^C(V, \mathbb{C}))$ is the natural isomorphism given by (89.12). In view of the remarks above, $u_p(\cdot, v) : V \to \mathbb{C}$ is $\mathbb{C}$-linear for all $v \in V$ and $u_p(u, \cdot) : V \to \mathbb{C}$ is conjugate-linear for all $u \in V$. We use the simplified notation

$$\langle u | v \rangle := u_p(u, v) \quad \text{when} \quad u, v \in V. \quad (89.15)$$

It follows from (89.15), (89.14), and (89.12) that

$$\langle u | v \rangle = v \cdot u - i(v \cdot Ju) = u \cdot v + i(u \cdot Jv),$$

$$\text{Re} \langle u | v \rangle = u \cdot v, \quad \langle u | u \rangle = u^2. \quad (89.16)$$
CHAPTER 8. SPECTRAL THEORY

for all \( u, v \in \mathcal{V} \). The properties of \( u \) are reflected in the following rules, valid for all \( u, v, w \in \mathcal{V} \) and all \( \zeta \in \mathbb{C} \):

\[
\langle v | u \rangle = \overline{\langle u | v \rangle}, \quad (89.17)
\]

\[
\langle u + v | w \rangle = \langle u | w \rangle + \langle v | w \rangle, \quad (89.18)
\]

\[
\langle \zeta u | v \rangle = \zeta \langle u | v \rangle = \langle u | \zeta v \rangle. \quad (89.19)
\]

Let \( \mathcal{V} \) be a unitary space. The following result, which is easily proved, describes the properties of transposition of \( \mathbb{C} \)-lineons.

**Proposition 2:** If \( L \in \text{Lin}^\mathbb{C} \mathcal{V} \) then \( L^\top \in \text{Lin}^\mathbb{C} \mathcal{V} \) and \( L^\top \) is characterized by

\[
\langle Lu | v \rangle = \overline{\langle u | L^\top v \rangle} \quad \text{for all} \quad u, v \in \mathcal{V}. \quad (89.20)
\]

The mapping \( (L \mapsto L^\top) : \text{Lin}^\mathbb{C} \mathcal{V} \rightarrow \text{Lin}^\mathbb{C} \mathcal{V} \) is conjugate-linear, so that

\[
(\zeta L)^\top = \overline{\zeta L}^\top \quad \text{for all} \quad \zeta \in \mathbb{C}, \; L \in \text{Lin}^\mathbb{C} \mathcal{V}. \quad (89.21)
\]

We use the notation

\[
\text{Sym}^\mathbb{C} \mathcal{V} = \text{Sym} \mathcal{V} \cap \text{Lin}^\mathbb{C} \mathcal{V} \quad (89.22)
\]

for the set of all symmetric \( \mathbb{C} \)-lineons. This set is a subspace, but not a \( \mathbb{C} \)-subspace, of \( \text{Lin}^\mathbb{C} \mathcal{V} \). In fact, the following result shows how \( \text{Lin}^\mathbb{C} \mathcal{V} \) can be recovered from \( \text{Sym}^\mathbb{C} \mathcal{V} \):

**Theorem on Real and Imaginary Parts:** Let \( \mathcal{V} \) be a unitary space. We have

\[
i \text{Sym}^\mathbb{C} \mathcal{V} = \text{Skew} \mathcal{V} \cap \text{Lin}^\mathbb{C} \mathcal{V} \quad (89.23)
\]

and \( i \text{Sym}^\mathbb{C} \mathcal{V} \) is a supplement of \( \text{Sym}^\mathbb{C} \mathcal{V} \) in \( \text{Lin}^\mathbb{C} \mathcal{V} \), i.e. to every \( \mathbb{C} \)-lineon \( L \) corresponds a unique pair \( (H, G) \) such that

\[
L = H + i G, \quad \text{and} \quad H, G \in \text{Sym}^\mathbb{C} \mathcal{V}. \quad (89.24)
\]

\( H \) is called the **real part** and \( G \) the **imaginary part** of \( L \).

**Proof:** Let \( L \in \text{Lin}^\mathbb{C} \mathcal{V} \) be given. By (89.21) we have \( (iL)^\top = -iL^\top \) and hence \( L = -L^\top \) if and only if \( (iL) = (iL)^\top \). Since \( L \in \text{Lin}^\mathbb{C} \mathcal{V} \) was arbitrary, (89.23) follows.

Again let \( L \in \text{Lin}^\mathbb{C} \mathcal{V} \) be given and assume that (89.24) holds. Since \( iG \in \text{Skew} \mathcal{V} \) by (89.23), it follows that \( (H, iG) \) must be the additive decomposition of \( L \) described in the Additive Decomposition Theorem of Sect.86 and hence that

\[
H = \frac{1}{2}(L + L^\top), \quad G = \frac{i}{2}(L^\top - L), \quad (89.25)
\]
which proves the uniqueness of \((H, G)\). To prove existence, we define \(H\) and \(G\) by (89.25) and observe, using (89.21), that they have the required properties. 

Since \(\text{Sym}^C\mathcal{V}\) and \(i\text{Sym}^C\mathcal{V}\) are linearly isomorphic subspaces of \(\text{Lin}^C\mathcal{V}\) we can use Props. 5 and 7 of Sect. 17 to obtain the following consequence of the Theorem.

**Corollary:** Let \(n := \dim^C\mathcal{V} = \frac{1}{2} \dim \mathcal{V}\). Then \(\dim \text{Lin}^C\mathcal{V} = 2 \dim^C \text{Lin}^C\mathcal{V} = 2n^2\) and \(\dim \text{Sym}^C\mathcal{V} = \dim(\text{Skew}\mathcal{V} \cap \text{Lin}^C\mathcal{V}) = n^2\).

Let \(\mathcal{V}, \mathcal{V}'\) be complex spaces. We say that a mapping \(U : \mathcal{V} \rightarrow \mathcal{V}'\) is unitary if it is orthogonal and \(\mathbb{C}\)-linear. It is easily seen that \(U : \mathcal{V} \rightarrow \mathcal{V}'\) is unitary if and only if it is linear and preserves the unitary product defined by (89.14). The set of all unitary mappings from \(\mathcal{V}\) to \(\mathcal{V}'\) will be denoted by \(\text{Unit}(\mathcal{V}, \mathcal{V}')\), so that

\[
\text{Unit}(\mathcal{V}, \mathcal{V}') := \text{Orth}(\mathcal{V}, \mathcal{V}') \cap \text{Lin}^C(\mathcal{V}, \mathcal{V}')
\]  

(89.26)

If \(U\) is unitary and invertible, then \(U^{-1}\) is again unitary (see Prop.1 of Sect.43). In this case, \(U\) is called a unitary isomorphism. We write \(\text{Unit}\mathcal{V} := \text{Unit}(\mathcal{V}, \mathcal{V})\). It is clear that \(\text{Unit}\mathcal{V}\) is a subgroup of \(\text{Orth}\mathcal{V}\) and hence of \(\text{Lis}\mathcal{V}\). The group \(\text{Unit}\mathcal{V}\) is called the unitary group of the unitary space \(\mathcal{V}\).

**Remark:** Let \(I\) be a finite index set. The linear space \(\mathbb{C}^I\) over \(\mathbb{C}\) (see Sect.14) has the natural structure of a unitary space whose inner square is given by

\[
\lambda^2 := \sum_{i \in I} |\lambda_i|^2 \quad \text{for all} \quad \lambda \in \mathbb{C}^I.
\]  

(89.27)

Of course, the complexor is termwise multiplication by \(i\). The unitary product is given by

\[
\langle \lambda | \mu \rangle = \sum_{i \in I} \lambda_i \overline{\mu_i} \quad \text{for all} \quad \lambda, \mu \in \mathbb{C}^I
\]  

(89.28)

Notes 89

(1) As far as I know, the treatment of complex spaces as real linear spaces with additional structure appears nowhere else in the literature. The textbooks just deal with linear spaces over \(\mathbb{C}\). I believe that the approach taken here provides better insight, particularly into conjugate-linear structures and mappings.

(2) The term “complexor” is introduced here for the first time.
CHAPTER 8. SPECTRAL THEORY

(3) Some people use the term “anti-linear” or the term “semi-linear” instead of “conjugate-linear”.

(4) Most textbooks deal only with what we call “genuine” unitary spaces and use the term “unitary space” only in this restricted sense. The term “complex inner-product space” is also often used in this sense.

(5) What we call “unitary product” is very often called “complex inner-product” or “Hermitian”.

(6) The notations \[ \langle u | v \rangle \] and \((u, v)\) are often used for the value \(\langle u | v \rangle\) of the unitary product. (See also Note (3) to Sect. 41.)

(7) In most of the literature on theoretical physics, the order of \(u, v\) in the notation \(\langle u | v \rangle\) for the unitary product is reversed. Therefore, in textbooks on theoretical physics, (89.19) is replaced by
\[
\langle u | \zeta v \rangle = \zeta \langle u | v \rangle = \langle \zeta u | v \rangle.
\]

(8) Symmetric \(C\)-lineons are very often called “self-adjoint” or “Hermitian”. I do not think a special term is needed.

(9) Skew \(C\)-lineons are often called “skew-Hermitian” or “anti-Hermitian”.

(10) The notations \(U_n\) or \(U(n)\) are often used for the unitary group \(\text{Unit} C^n\).

810 Complex Spectra

Let \(V\) be a complex space. As we have seen in Sect. 89, \(V\) can be regarded both as a linear space over \(R\) and as a linear space over \(C\). We modify Def. 2 of Sect. 82 as follows: Let \(L \in \text{Lin} C V\) be given. For every \(\zeta \in C\), we write
\[
\text{Sps}_C L(\zeta) := \text{Null}(L - \zeta 1_V),
\]
and, if it is non-zero, call it the spectral \(C\)-space of \(L\) for \(\zeta\). Of course, if \(\zeta\) is real, we have \(\text{Sps}_C L(\zeta) = \text{Sps}_L(\zeta)\).

The spectral \(C\)-spaces are \(C\)-subspaces of \(V\). The \(C\)-spectrum of \(L\) is defined by
\[
\text{Spec}^C L := \{ \zeta \in C \mid \text{Sps}_C L(\zeta) \neq \{0\}\}.
\]

It is clear that
\[
\text{Spec} L = \text{Spec}^C L \cap R.
\]

The \(C\)-multiplicity function \(\text{mult}^C_L : C \to \mathbb{N}\) of \(L\) is defined by
\[
\text{mult}^C_L(\zeta) := \dim_C(\text{Sps}_C L(\zeta)) \text{ for all } \zeta \in C.
\]
In view of (89.4) we have
\[ \text{mult}_L(\sigma) = 2\text{mult}_L^C(\sigma) \quad \text{for all} \quad \sigma \in \mathbb{R}. \] (810.5)

\[ \text{Spec}^C_L \] is the support of mult\(_L^C\).

The Theorem on Spectral Spaces of Sect.82 has the following analogue, which is proved in the same way.

**Proposition 1:** The \( \mathbb{C} \)-spectrum of a \( \mathbb{C} \)-lineon \( L \) on a complex space \( V \) has at most \( \dim^C V \) members and the family \( (\text{Sps}_L^C(\zeta) \mid \zeta \in \text{Spec}^C_L) \) of the spectral \( \mathbb{C} \)-spaces of \( L \) is disjunct.

The following are analogues of Prop.5 and Prop.6 of Sect.82 and are proved in a similar manner.

**Proposition 2:** Let \( L \in \text{Lin}^C V \) be given and assume that the family \( (\text{Sps}_L(\zeta) \mid \zeta \in \text{Spec}_L) \) of spectral \( \mathbb{C} \)-spaces is a decomposition of \( V \). If \( (E_\zeta \mid \zeta \in \text{Spec}_L) \) is the associated family of idempotents then
\[ L = \sum_{\zeta \in \text{Spec}_L} \zeta E_\zeta. \] (810.6)

**Proposition 3:** Let \( L \in \text{Lin}^C V \) be given. Assume that \( Z \) is a finite subset of \( \mathbb{C} \) and that \( (W_\zeta \mid \zeta \in Z) \) is a decomposition of \( V \) whose terms are non-zero \( L \)-invariant \( \mathbb{C} \)-subspaces of \( V \) and satisfy
\[ L|_{W_\zeta} = \zeta 1_{W_\zeta} \quad \text{for all} \quad \zeta \in Z. \] (810.7)

Then \( Z = \text{Spec}_L \) and \( W_\zeta = \text{Sps}_L^C(\zeta) \) for all \( \zeta \in Z \).

From now on we assume that \( V \) is a genuine unitary space.

**Proposition 4:** The \( \mathbb{C} \)-spectrum of a symmetric \( \mathbb{C} \)-lineon \( S \in \text{Sym}^C V \) contains only real numbers, i.e. \( \text{Spec}^C_S = \text{Spec} S \).

**Proof:** Let \( \zeta \in \text{Spec}^C S \), so that \( Su = \zeta u \) for some \( u \in V^\times \). Using \( S = S^\dagger \), (89.20), (89.19), and (89.16) we find
\[ 0 = \langle Su \mid u \rangle - \langle u \mid Su \rangle = \langle \zeta u \mid u \rangle - \langle u \mid \zeta u \rangle = (\zeta - \overline{\zeta})\langle u \mid u \rangle = (\zeta - \overline{\zeta})u^2. \]

Since \( V \) is genuine, we must have \( u^2 \neq 0 \) and hence \( \zeta = \overline{\zeta} \), i.e. \( \zeta \in \mathbb{R} \).

**Remark:** A still shorter proof of Prop.4 can be obtained by applying the Spectral Theorem and Prop.2, but the proof above is more elementary in that it makes no use of the Spectral Theorem.

Let \( S \in \text{Sym}^C V \) be given. By Prop.4, the family of spectral \( \mathbb{C} \)-spaces of \( S \) coincides with the family of spectral spaces in the sense of Def.1 of Sect.82.
This family consists of \( \mathbb{C} \)-subspaces of \( V \) and, by the Spectral Theorem, is an orthogonal decomposition of \( V \). All the terms \( E_\sigma \) in the spectral resolution of \( S \) (see Cor.1 to the Spectral Theorem) are \( \mathbb{C} \)-linear. There exists a \( \mathbb{C} \)-basis \( e := (e_i \mid i \in I) \) of \( V \) and a family \( \lambda = (\lambda_i \mid i \in I) \) in \( \mathbb{R} \) such that (84.6) holds. Each \( \sigma \in \text{Spec} S \) occurs exactly \( \text{mult}_S(\sigma) \) times in the family \( \lambda \).

For \( \mathbb{C} \)-linear normal lineons, the Structure Theorem of Sect.88 can be replaced by the following simpler version.

**Spectral Theorem for Normal \( \mathbb{C} \)-Lineons:** Let \( V \) be a genuine unitary space. A \( \mathbb{C} \)-lineon on \( V \) is normal if and only if the family of its spectral \( \mathbb{C} \)-spaces is an orthogonal decomposition of \( V \).

**Proof:** Let \( N \in \text{Lin} \mathbb{C} V \) be normal. Let \( H, G \in \text{Sym} \mathbb{C} V \) be the real and imaginary parts of \( N \) as described in the Theorem on Real and Imaginary Parts of Sect.89, so that \( N = H + i G \). From this Theorem and Prop.1 of Sect.88 it follows that \( H \) and \( G \) commute. Let \( Z := \text{Spec} H = \text{Spec} \mathbb{C} H \). By Prop.2 of Sect.82, \( U_\sigma := \text{Sps}_H(\sigma) \) is \( G \)-invariant for each \( \sigma \in Z \). Hence we may consider \( D_\sigma := G|_{U_\sigma} \) for all \( \sigma \in Z \).

It is clear that \( D_\sigma \in \text{Sym\mathbb{C}} U_\sigma \) for all \( \sigma \in Z \). Put \( Q_\sigma := \text{Spec} D_\sigma = \text{Spec} \mathbb{C} D_\sigma \) and \( W_{(\sigma, \tau)} := \text{Sps}_{D_\sigma}(\tau) \) for all \( \sigma \in Z \) and all \( \tau \in Q_\sigma \). It is easily seen that \( W_{(\sigma, \tau)} \) is \( N \)-invariant and that

\[
N|_{W_{(\sigma, \tau)}} = \sigma 1_{W_{(\sigma, \tau)}} + i \tau 1_{W_{(\sigma, \tau)}} = (\sigma + i \tau) 1_{W_{(\sigma, \tau)}}
\]

for all \( \sigma \in Z \) and all \( \tau \in Q_\sigma \). Also, the family \( \{W_{(\sigma, \tau)} \mid \sigma \in Z, \tau \in Q_\sigma\} \) is an orthogonal decomposition of \( V \) whose terms are \( \mathbb{C} \)-subspaces of \( V \). It follows from Prop.3 that \( \text{Spec} \mathbb{C} N = \{\sigma + i \tau \mid \sigma \in Z, \tau \in Q_\sigma\} \) and that \( \text{Sps}_{\mathbb{C} N}(\zeta) = W_{(\text{Re} \zeta, \text{Im} \zeta)} \) for all \( \zeta \in \text{Spec} \mathbb{C} N \).

Assume now that \( L \in \text{Lin} \mathbb{C} V \) is such that \( \{\text{Sps}_{\mathbb{C} L}(\zeta) \mid \zeta \in \text{Spec} \mathbb{C} L\} \) is an orthogonal decomposition of \( V \). We can apply Prop.2 and conclude that (810.6) holds. By Prop.4 of Sect.83, each of the idempotents \( E_\zeta, \zeta \in \text{Spec} \mathbb{C} L \), belongs to \( \text{Sym} \mathbb{C} V \). Using this fact and (89.21) we derive from (810.6) that

\[
L^\top = \sum_{\zeta \in \text{Spec} \mathbb{C} L} \zeta E_\zeta
\]

and hence, by (81.6), that

\[
LL^\top = \sum_{\zeta \in \text{Spec} \mathbb{C} L} (\zeta \bar{\zeta}) E_\zeta = L^\top L,
\]

i.e. that \( L \) is normal. \( \blacksquare \)
The $\mathbb{C}$-spectrum of a normal $\mathbb{C}$-lineon is related to its pair-spectrum (see Sect.8.8) by

$$
\text{Pspec } N = \{(\text{Re } \zeta, |\text{Im } \zeta|) \mid \zeta \in \text{Spec}^\mathbb{C} N\}. \quad (810.8)
$$

A normal $\mathbb{C}$-lineon is symmetric, skew, or unitary depending on whether its $\mathbb{C}$-spectrum consists only of real numbers, imaginary numbers, or numbers of absolute value 1, respectively. The angle-spectrum of a unitary lineon $U$ is related to its $\mathbb{C}$-spectrum by

$$
\text{Aspec } U = \{\varphi \in [0, \pi] \mid e^{i\varphi} \in \text{Spec}^\mathbb{C} U \text{ or } e^{-i\varphi} \in \text{Spec}^\mathbb{C} U\}. \quad (810.9)
$$

The Spectral Theorem for Normal $\mathbb{C}$-lineons has two corollaries that are analogues of Cors.1 and 2 to the Spectral Theorem in Sect.8.4. We leave the formulation of these corollaries to the reader.

### 811 Problems for Chapter 8

1. Let $\mathcal{V}$ be a linear space, let $(\mathcal{U}_i \mid i \in I)$ be a finite family of subspaces of $\mathcal{V}$, let $\Pi$ be a partition of $I$ (see Sect.0.1) and put

$$
\mathcal{W}_J := \sum_{j \in J} \mathcal{U}_j \quad \text{for all } J \in \Pi \quad \text{(P8.1)}
$$

(see (07.14)). Prove that the family $(\mathcal{U}_i \mid i \in I)$ is disjunct [a decomposition of $\mathcal{V}$] if and only if the family $(\mathcal{W}_J \mid J \in \Pi)$ is disjunct [a decomposition of $\mathcal{V}$] and, for each $J \in \Pi$, the family $(\mathcal{U}_j \mid j \in J)$ is disjunct.

2. Let $(E_i \mid i \in I)$ be a finite family of lineons on a given linear space $\mathcal{V}$.

   (a) Show: If (81.5) and (81.6) hold, then $(\text{Rng } E_i \mid i \in I)$ is a decomposition of $\mathcal{V}$ and $(E_i \mid i \in I)$ is the family of idempotents associated with this decomposition.

   (b) Prove: If (81.5) alone holds and if $E_i$ is idempotent for every $i \in I$, then $(\text{Rng } E_i \mid i \in I)$ is already a decomposition of $\mathcal{V}$ and hence (81.6) follows. (Hint: Apply Prop.4 of Sect.8.1.)

3. Let $\mathcal{V}$ be a genuine inner-product space and let $S$ and $T$ be symmetric lineons on $\mathcal{V}$.

   (a) Show that $ST$ is symmetric if and only if $S$ and $T$ commute.
(b) Prove: If $S$ and $T$ commute then
\[ \text{Spec}(ST) \subset \{ \sigma \tau \mid \sigma \in \text{Spec } S, \tau \in \text{Spec } T \}. \]  
(P8.2)

(4) Let $\mathcal{V}$ be a genuine inner-product space and let $S$ be a symmetric lineon on $\mathcal{V}$. Show that the operator norm of $S$, as defined by (52.19), is given by
\[ ||S|| = \max\{|\sigma| \mid \sigma \in \text{Spec } S\}. \]  
(P8.3)

(5) Let a genuine Euclidean space $\mathcal{E}$ with translation space $\mathcal{V}$, a point $q \in \mathcal{E}$, and a strictly positive symmetric lineon $S \in \text{Pos}^+ \mathcal{V}$ be given. Then the set
\[ S := \{ x \in \mathcal{E} \mid \bar{S}(x - q) = 1 \} \]
is called an ellipsoid centered at $q$, and the spectral spaces of $S$ are called the principal directions of the ellipsoid. Put $T := (\sqrt{S})^{-1}$ (see Sect.85). The spectral values of $T$ are called the semi-axes of $S$.

(a) Show that
\[ S = q + T_{\geq}(\text{Usph } \mathcal{V}). \]  
(P8.4)

(see (42.9)).

(b) Show: If $e$ is a spectral unit vector of $S$, then
\[ S \cap (q + \mathbb{R}e) = \{ q - ae, q + ae \}, \]  
(P8.5)

where $a$ is a semi-axis of $S$.

(c) Let $e$ be an orthonormal basis-set whose terms are spectral vectors of $S$ (see Cor.2 to the Spectral Theorem). Let $\Gamma$ be the Cartesian coordinate system with origin $q$ which satisfies $\{ \nabla c \mid c \in \Gamma \} = e$ (see Sect.74). Show that
\[ S = \left\{ x \in \mathcal{E} \mid \sum_{c \in \Gamma} \frac{e(x)^2}{a_c^2} = 1 \right\}, \]  
(P8.6)

where $(a_c \mid c \in \Gamma)$ is a family whose terms are the semi-axes of the ellipsoid $S$. 
(d) Let \( p \in \mathcal{E} \) and \( \rho \in \mathbb{P}^\times \) be given and let \( \alpha : \mathcal{E} \to \mathcal{E} \) be an invertible flat mapping. Show that the image \( \alpha_{\geq} \left( \text{Sph}_{p,\rho} \mathcal{E} \right) \) under \( \alpha \) of the sphere \( \text{Sph}_{p,\rho} \mathcal{E} \) (see (46.11)) is an ellipsoid centered at \( \alpha(p) \) whose semi-axes form the set
\[
\left\{ \rho \sqrt{\sigma} \mid \sigma \in \text{Spec} \left( (\nabla \alpha)^T \nabla \alpha \right) \right\}.
\]

(6) Let \( S \) be a symmetric lineon on a genuine inner-product space \( \mathcal{V} \) and let \( (E_\sigma \mid \sigma \in \text{Spec} S) \) be the spectral resolution of \( S \).

(a) Show that, for all \( \sigma \in \text{Spec} S \), we have
\[
\int_0^1 \exp_{\mathcal{V}} o(\iota(S - \sigma 1_\mathcal{V})) = E_\sigma + \sum_{\tau \in \text{Spec} S \setminus \{\sigma\}} \frac{e^{\tau} - e^\sigma - 1}{\tau - \sigma} E_\tau. \quad (P8.7)
\]

(b) Show that, for every \( M \in \text{Lin} \mathcal{V} \), every \( \sigma \in \text{Spec} S \), and every \( v \in \text{Sp}_S(\sigma) \),
\[
((\nabla_S \exp_{\mathcal{V}})M)v = \left( e^\sigma E_\sigma + \sum_{\tau \in \text{Spec} S \setminus \{\sigma\}} \frac{e^\tau - e^\sigma}{\tau - \sigma} E_\tau \right) Mv. \quad (P8.8)
\]

(c) Show that \( \exp_{\mathcal{V}} \) is locally invertible near \( S \).

(d) Prove that the lineonic logarithm \( \log_{\mathcal{V}} \) defined in Sect.85 is of class \( C^1 \).

(7) Consider \( L \in \text{Lin} \mathbb{R}^3 \) and \( T \in \text{Sym} \mathbb{R}^3 \) as given by
\[
L := \frac{1}{5} \begin{bmatrix} 10 & 0 & 5 \\ 3 & 8 & 6 \\ 4 & -6 & 8 \end{bmatrix}, \quad T := \begin{bmatrix} 5 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 5 \end{bmatrix}.
\]

(a) Determine the spectrum and the spectral idempotents of \( T \).

(b) Determine the right and left polar decompositions of \( L \) (see Sect.86).

(c) Find the spectrum and angle-spectrum of \( R := \text{or}(L) \in \text{Orth} \mathbb{R}^3 \) (see Sect.88).

(8) Let \( L \) be an invertible lineon on a genuine inner-product space \( \mathcal{V} \), and let \( (R, S) \) be the right polar decomposition and \( (R, T) \) the left polar decomposition of \( L \). Show that the following are equivalent:
(i) \( L \in \text{Skew} \mathcal{V} \),
(ii) \( R \) is a perpendicular turn that commutes with \( S \) or with \( T \),
(iii) \( S = T \) and \( R \in \text{Skew} \mathcal{V} \),
(iv) \( \text{Sp}_S(\kappa) = Q\text{sp}_L(\kappa) \) for all \( \kappa \in \text{Spec} S \).

(9) Let \( \mathcal{V} \) be a genuine inner-product space, let \( I \in \text{Sub} \mathbb{R} \) be a genuine interval, and let \( F : I \to \text{Lis} \mathcal{V} \) be a lineonic process of class \( C^1 \) with invertible values. Define \( L : I \to \text{Lin} \mathcal{V} \) by
\[
L(t) := F^*(t)F(t)^{-1} \quad \text{for all} \quad t \in I, \tag{P8.9}
\]
and define \( D : I \to \text{Sym} \mathcal{V} \) and \( W : I \to \text{Skew} \mathcal{V} \) by the requirements that \( (D(t), W(t)) \) be the additive decomposition of \( L(t) \) for all \( t \in I \) (see Sect.86). For each \( t \in I \), define \( F(t) : I \to \text{Lis} \mathcal{V} \) by
\[
F(t)(s) := F(s)F(t)^{-1} \quad \text{for all} \quad s \in I \tag{P8.10}
\]
and put \( R(t) := \circ F(t), \ U(t) := \text{rp} \circ F(t), \ V(t) := \text{lp} \circ F(t) \) (see Sect.86). Prove that
\[
R^*(t)(t) = W(t), \ U^*(t)(t) = V^*(t)(t) = D(t) \quad \text{for all} \quad t \in I \tag{P8.11}
\]
(Hint: Use Prop.1 of Sect.86).

(10) Let \( \mathcal{V} \) be a genuine inner-product space. Prove: If \( N \in \text{Lin} \mathcal{V} \) is normal and satisfies \( N^2 = -1_\mathcal{V} \), then \( N \) is a perpendicular turn (see Def.1 of Sect.87 and Def.1 of Sect.88).

(11) Let \( S \) be a lineon on a genuine inner product space \( \mathcal{V} \). Prove that the following are equivalent:
(i) \( S \in \text{Sym} \mathcal{V} \cap \text{Orth} \mathcal{V} \),
(ii) \( S \in \text{Sym} \mathcal{V} \) and \( \text{Spec} S \subset \{1, -1\} \),
(iii) \( S \in \text{Orth} \mathcal{V} \) and \( S^2 = 1_\mathcal{V} \),
(iv) \( S = E - F \) for some symmetric idempotents \( E \) and \( F \) that satisfy \( EF = 0 \) and \( E + F = 1_\mathcal{V} \),
(v) \( S \in \text{Orth} \mathcal{V} \) and \( \text{Aspec} S = \emptyset \).
(12) Let $U$ be a skew lineon on a genuine inner product space. Show that $U$ is a perpendicular turn if and only if $\text{Null } U = \{0\}$ and $\text{Aspec } U = \{ \frac{\pi}{2} \}$.

(13) Let $A$ be a skew lineon on a genuine inner-product space and let $R := \exp_{\mathcal{V}}(A)$.

(a) Show that $R$ is orthogonal. (Hint: Use the Corollary to Prop.2 of Sect.66.)

(b) Show that the angle-spectrum of $R$ is given, in terms of the quasi-spectrum of $A$, by

$$\text{Aspec } R = \left\{ \kappa - \pi \left\lfloor \frac{\kappa}{\pi} \right\rfloor \bigg| \kappa \in \text{Qspec } A \right\}, \tag{P8.12}$$

where the floor $\lfloor a \rfloor$ of $a \in \mathbb{R}$ is defined by $\lfloor a \rfloor := \max\{n \in \mathbb{Z} \mid n \leq a\}$. (Hint: Use the Structure Theorems of Sects.87 and 88 and the results of Problem 9 of Chap.6)

Remark: The assertion of Part (a) is valid even if the inner-product space $\mathcal{V}$ is not genuine.

(14) Let $\mathcal{V}$ be a linear space and define $J \in \text{Lin } \mathcal{V}^2$ by

$$J(v_1, v_2) = (-v_2, v_1) \quad \text{for all } v \in \mathcal{V}^2. \tag{P8.13}$$

(a) Show that $J^2 = -1_{\mathcal{V}^2}$, and hence that $\mathcal{V}^2$ has the natural structure of a complex space with complexor $J$ (see Def.1 of Sect.89). The space $\mathcal{V}^2$, with this complex-space structure, is called the complexification of $\mathcal{V}$.

(b) Show that

$$(\xi + i \eta)v = (\xi v_1 - \eta v_2, \eta v_1 + \xi v_2) \tag{P8.14}$$

for all $\xi, \eta \in \mathbb{R}$ and all $v \in \mathcal{V}^2$.

(c) Prove: If $L, M \in \text{Lin } \mathcal{V}$, then the cross-product $L \times M \in \text{Lin } \mathcal{V}^2$ (see (04.18)) is $\mathbb{C}$-linear if and only if $M = L$ and it is conjugate-linear if and only if $M = -L$ (see Sect.89).

(d) Let $L \in \text{Lin } \mathcal{V}$ and $\zeta \in \mathbb{C}$ be given. Show that $\zeta \in \text{Spec}^C(L \times L)$ if and only if either $\text{Im } \zeta = 0$ and $\zeta \in \text{Spec } L$ or else $\text{Im } \zeta \neq 0$ and $(\text{Re } \zeta, |\text{Im } \zeta|) \in \text{Pspec } L$ (see Def.2 of Sect.88). Conclude that
(c) Let \( L \in \text{Lin} V \) and \( \zeta \in \text{Spec}^C(L \times L) \) be given and put \( \xi := \text{Re} \zeta, \eta := \text{Im} \zeta \), so that \( \zeta = \xi + i \eta \). Show that
\[
\text{Sps}^C_{L \times L}(\zeta) = (\text{Sps}_L(\xi))^2 \quad \text{if} \quad \eta = 0
\]
while
\[
\text{Sps}^C_{L \times L}(\zeta) = \left\{ (v, \frac{1}{\eta}(\xi 1_V - L)v) \mid v \in \text{Psp}_L(\xi, |\eta|) \right\}
\]
if \( \eta \neq 0 \).

(15) Let \( V \) be an inner-product space and let \( V^2 \) be the complexification of \( V \) as described in Problem 14. Recall that \( V^2 \) carries the natural structure of an inner-product space (see Sect. 44).

(a) Show that the complexor \( J \in \text{Lin} V^2 \) defined by (P8.14) is a perpendicular turn and hence endows \( V^2 \) with the structure of a unitary space.

(b) Show that the unitary product of \( V^2 \) is given by
\[
(u \mid v) = u_1 \cdot v_1 + u_2 \cdot v_2 + i(u_2 \cdot v_1 - u_1 \cdot v_2)
\]
for all \( u, v \in V^2 \).

(c) Show that \( H \in \text{Lin} V^2 \) is a symmetric \( \mathbb{C} \)-lineon on \( V^2 \) if and only if \( H = T \times T + J \circ (A \times A) \) for some \( T \in \text{Sym} V \) and some \( A \in \text{Skew} V \).
Chapter 9

The Structure of General Lineons

In this chapter, it is assumed again that all linear spaces under consideration are finite-dimensional except when a statement to the contrary is made. However, they may be spaces over an arbitrary field \( F \). If \( F \) is to be the field \( \mathbb{R} \) of real numbers or the field \( \mathbb{C} \) of complex numbers, it will be explicitly stated. We will make frequent use of the definitions and results of Sects.81 and 82, which remain valid for finite-dimensional spaces over \( F \) even if \( F \) is not \( \mathbb{R} \).

91 Elementary Decompositions

Let \( L \) be a lineon on a given linear space \( V \).

**Definition 1:** We say that a subspace \( M \) of \( V \) is a minimal \( L \)-space [maximal \( L \)-space] if \( M \) is minimal [maximal] (with respect to inclusion) among all \( L \)-subspaces of \( V \) that are different from the zero space \( \{0\} \) [the whole space \( V \)].

If \( V \) is not itself a zero-space, then there always exist minimal and maximal \( L \)-spaces. In fact, if \( U \) is an \( L \)-space and \( U \neq \{0\} \) [\( U \neq V \)], then \( U \) includes [is included in] a minimal [maximal] \( L \)-space. The following result follows directly from Def.1.

**Proposition 1:** Let \( U \) be an \( L \)-subspace of \( V \). If \( M \) is a minimal \( L \)-space, then either \( M \cap U = \{0\} \) or else \( M \subset U \). If \( W \) is a maximal \( L \)-space, then either \( W + U = V \) or else \( U \subset W \).

The next result is an immediate consequence of Prop.2 of Sect.21 and Prop.1 of Sect.82.
Proposition 2: A subspace $M$ of $V$ is a minimal [maximal] $L$-space if and only if the annihilator $M^\perp$ of $M$ is a maximal [minimal] $L^\perp$-space.

Definition 2: We say that the lineon $L$ on $V$ is an elementary lineon if there is exactly one minimal $L$-space. If $L$ is an arbitrary lineon on $V$, an $L$-subspace $U$ of $V$ is called an elementary $L$-space if $L_{|U}$ is elementary. We say that a decomposition of $V$ (see Sect. 81) is an elementary $L$-decomposition of $V$ if all of its terms are elementary $L$-spaces.

The following result is an immediate consequence of Def. 2.

Proposition 3: Let $L \in \text{Lin} V$ be an elementary lineon and $M$ its (only) minimal $L$-space. If $U$ is a non-zero $L$-space, then $M \subset U$ and $L_{|U}$ is again an elementary lineon and the (only) minimal $L_{|U}$-space is again $M$.

We say that a lineon $L \in \text{Lin} V$ is simple if $V \neq \{0\}$ and if there are no $L$-subspaces other than $\{0\}$ and $V$. A simple lineon is elementary; its only minimal $L$-space is $V$ and its only maximal $L$-space is $\{0\}$. If $L$ is elementary but not simple and if $W$ is a maximal $L$-space, then $W$ includes the only minimal $L$-space of $V$. A lineon of the form $\lambda 1_V$, $\lambda \in \mathbb{F}$, is elementary (and then simple) if and only if $\dim V = 1$.

The significance of Def. 2 lies in the following theorem, which reduces the study of the structure of general lineons to that of elementary lineons.

**Elementary Decomposition Theorem:** For every lineon $L$ there exists an elementary $L$-decomposition.

The proof is based on three lemmas. The proof of the first of these will be deferred to Sect. 93; it is the same as Cor. 2 to the Structure Theorem for Elementary Lineons.

**Lemma 1:** A lineon $L$ is elementary if and only if there is exactly one maximal $L$-space.

**Proof:** By Prop. 1 we have $E \nsubseteq W$ and hence $E \cap W$ is an $L_{|E}$-subspace of $V$. Now let $U$ be an $L$-space that is properly included in $E$, i.e. an $L_{|E}$-subspace of $E$ other than $E$. In view of the assumed minimality of $E$, we have $W + U \neq W$. We conclude that $W \cap E$ is the only maximal $L_{|E}$-space and hence, by Lemma 1, that $E$ is elementary.

**Lemma 2:** Let $L \in \text{Lin} V$, let $W$ be a maximal $L$-space and let $E$ be an $L$-space that is minimal among all subspaces of $V$ with the property $W + E = V$. Then $E$ is an elementary $L$-space.

**Proof:** By Prop. 1 we have $E \nsubseteq W$ and hence $E \cap W$ is a maximal $L_{|E}$-space of $V$. Using Prop. 1 again, it follows that $U \subset W$ and hence $U \subset E \cap W$. We conclude that $W \cap E$ is the only maximal $L_{|E}$-space and hence, by Lemma 1, that $E$ is elementary.

**Lemma 3:** Let $L \in \text{Lin} V$ and let $E$ be an elementary $L$-space with greatest possible dimension. Then there is an $L$-space that is a supplement of $E$ in $V$.

**Proof:** Let $M$ be the (only) minimal $L$-space included in $E$. By Prop. 2, $M^\perp$ is then a maximal $L^\perp$-space. We choose a subspace $E'$ of $V^*$ that is
minimal among all subspaces of $V^*$ with the property $M^\perp + \tilde{E} = V^*$. By Lemma 2, $\tilde{E}$ is an elementary $L^\top$-space.

In view of (21.11) and (22.5), we have \( \{0\} = V^* \perp = (M^\perp + \tilde{E})^\perp = M^\perp \cap \tilde{E}^\perp = M \cap \tilde{E}^\perp \) and hence $M \not\subset \tilde{E}^\perp$ and also $M \not\subset \tilde{E}^\perp \cap E$. Since $\tilde{E}^\perp \cap E$ is an $L$-subspace of $V$, it follows from Prop.1 that

$$\tilde{E}^\perp \cap E = \{0\}. \quad (91.1)$$

Using Prop.4 of Sect.17 and the Formula (21.15) for Dimension of Annihilators, we easily conclude from (91.1) that

$$\dim \tilde{E} \geq \dim E. \quad (91.2)$$

Since $\tilde{E}$ is elementary, we can repeat the argument above with $L$ replaced by $L^\top$ and $E$ replaced by $\tilde{E}$. Since $L^{\top\top} = L$, we thus find an elementary $L$-subspace $E'$ of $V$ such that

$$\dim E' \geq \dim \tilde{E}. \quad (91.3)$$

Since $E$ was assumed to have the greatest possible dimension, we must have $\dim E' \leq \dim \tilde{E}$ and hence, by (91.2) and (91.3), $\dim \tilde{E} = \dim E$, which gives

$$\dim V = \dim \tilde{E}^\perp + \dim \tilde{E} = \dim \tilde{E}^\perp + \dim E.$$ 

In view of (91.1), it follows from Prop.5 of Sect.17 that $\tilde{E}^\perp$ is a supplement of $E$. \[ \Box \]

**Proof of the Theorem:** We will show that there exists a collection of elementary $L$-spaces which, when interpreted as a self-indexed family, is a decomposition of $V$.

We proceed by induction over the dimension of $\text{Dom} L$. If $\dim \text{Dom} L = 0$, then the empty collection is an elementary $L$-decomposition. Assume then, that $L \in \text{Lin} V$ with $\dim V > 0$ is given and that the assertion is valid for every lineon whose domain has a dimension strictly less than $\dim V$.

Since $V$ is not the zero-space, there exist minimal $L$-spaces and hence elementary $L$-spaces. We choose an elementary $L$-space $E$ with greatest possible dimension. By Lemma 3, we can choose an $L$-subspace $U$ that is a supplement of $E$ in $V$. Since $\dim U = \dim V - \dim E < \dim V$, we can apply the induction hypothesis to $L_U$ and determine a collection $\mathcal{F}$ of elementary $L_U$-spaces which, self-indexed, is a decomposition of $U$. The elements of $\mathcal{F}$ are elementary $L$-spaces. Since $E$ is a supplement of
\[ U = \sum (F \mid F \in \mathfrak{F}) \], it follows from Prop. 2, (iii) of Sect. 8.1 that \( \mathfrak{F} \cup \{E\} \) is an elementary \( \mathbf{L} \)-decomposition of \( \mathcal{V} \).

We say that a lineon \( \mathbf{L} \in \text{Lin} \mathcal{V} \) is **semi-simple** if there is an \( \mathbf{L} \)-decomposition \( (U_i \mid i \in I) \) of \( \mathcal{V} \) such that \( \mathbf{L}_{|U_i} \) is simple for each \( i \in I \). Of course, such a decomposition is an elementary decomposition.

If \( \mathcal{V} \) is a genuine inner product space, then every normal lineon \( \mathbf{N} \) (and hence every skew, symmetric, or orthogonal lineon) on \( \mathcal{V} \) is semi-simple. Indeed, if we choose a basis \( \mathbf{e} \) as in the Corollary to the Structure Theorem for Normal Lineons of Sect. 8.8 and then define

\[
U_k := \begin{cases} 
Lsp\{\mathbf{e}_{2k-1}, \mathbf{e}_{2k}\} & \text{for all } k \in m \setminus 1 \\
Lsp\{\mathbf{e}_{2m+k}\} & \text{for all } k \in (n-m) \setminus m 
\end{cases}
\]

then \( (U_k \mid k \in (n-m) \setminus 1) \) is a decomposition of \( \mathcal{V} \) such that \( \mathbf{L}_{|U_k} \) is simple for each \( k \in (n-m) \setminus 1 \).

**Pitfall:** The collection of subspaces that are the terms of an elementary \( \mathbf{L} \)-decomposition is, in general, not uniquely determined by \( \mathbf{L} \). For example, if \( \mathbf{L} = 1_{\mathcal{V}} \), and if \( \mathbf{b} := (\mathbf{b}_i \mid i \in I) \) is a basis of \( \mathcal{V} \), then \( \{Lsp\{\mathbf{b}_i\} \mid i \in I\} \) is an elementary \( 1_{\mathcal{V}} \)-decomposition of \( \mathcal{V} \). If \( \dim \mathcal{V} > 1 \), then the collection \( \{Lsp\{\mathbf{b}_i\} \mid i \in I\} \) depends on the choice of the basis \( \mathbf{b} \) and is not uniquely determined by \( \mathcal{V} \).

Notes 91

(1) The concept of an elementary lineon and the term "elementary" in the sense of Def. 2, were introduced by me in 1970 (see Part E of the Introduction). I chose this term to correspond to the commonly accepted term "elementary divisor". In fact, the elementary divisors of a lineon can be matched, if counted with appropriate multiplicity, with the terms of any elementary decomposition of the lineon (see Sect. 95).

(2) The approach to the structure of general lineons presented in this Chapter was developed by me before 1970 and is very different from any that I have seen in the literature. The conventional approaches usually arrive at an elementary decomposition only after an intermediate (often called "primary") decomposition, which I was able to bypass. Also, the conventional treatments make heavy use of the properties of the ring of polynomials over \( \mathbb{F} \) (unique factorization, principal ideal ring property, etc.). My approach uses only some of the most basic concepts and facts of linear algebra as presented in Chaps. 1 and 2. Not even determinants are needed. The proofs are all based, in essence, on dimensional considerations.
Let \( F \) be any field. The elements of \( F^\mathbb{N} \), i.e. the sequences in \( F \) indexed on \( \mathbb{N} \) and with finite support (see (07.10)), are called polynomials over \( F \). As we noted in Sect.14, \( F^\mathbb{N} \) is a linear space over \( F \). It is infinite-dimensional.

We use the abbreviation
\[
\iota := \delta_1^\mathbb{N},
\]
where \( \delta^\mathbb{N} \) is the standard basis of \( F^\mathbb{N} \) (see Sect.16). The space \( F^\mathbb{N} \) acquires the structure of a commutative ring (see Sect.06) if we define its multiplication \( (p, q) \mapsto pq \) by
\[
(pq)_n := \sum_{k \in \mathbb{N}} p_k q_{n-k} \quad \text{for all} \quad n \in \mathbb{N}.
\]
This multiplication is characterized by the condition that the \( n \)'th power \( \iota^n \) of \( \iota \) be given by
\[
\iota^n = \delta_n^\mathbb{N} \quad \text{for} \quad n \in \mathbb{N}.
\]
The unity of \( F^\mathbb{N} \) is \( \iota^0 = \delta_0^\mathbb{N} \). We identify \( F \) with the subring \( F\iota^0 \) of \( F^\mathbb{N} \) and hence write
\[
\xi = \xi \iota^0 = \xi \delta_0^\mathbb{N} \quad \text{for all} \quad \xi \in F,
\]
so that \( 1 = \iota^0 \) denotes the unity of both \( F \) and \( F^\mathbb{N} \). We have
\[
p = \sum_{k \in \mathbb{N}} p_k \iota^k \quad \text{for all} \quad p \in F^\mathbb{N}.
\]

The degree of a non-zero polynomial \( p \) is defined by
\[
\deg p := \max \{ k \in \mathbb{N} \mid p_k \neq 0 \}
\]
(see (07.9)). If \( p \in F^\mathbb{N} \) is zero or if \( \deg p \leq n \), then
\[
p = \sum_{k \in (n+1)^\mathbb{N}} p_k \iota^k.
\]
We have
\[
\deg (pq) = \deg p + \deg q
\]
for all \( p, q \in (F^\mathbb{N})^\times \). We say that \( p \in (F^\mathbb{N})^\times \) is a monic polynomial if \( p_{\deg p} = 1 \). If this is the case, then
\[
p = \iota^{\deg p} + \sum_{k \in (\deg p)^\mathbb{N}} p_k \iota^k.
\]
The polynomial function \((\xi \mapsto p(\xi)) : \mathbb{F} \to \mathbb{F}\) associated with a given polynomial \(p \in \mathbb{F}(\mathbb{N})\) is defined by
\[
p(\xi) := \sum_{k \in \mathbb{N}} p_k \xi^k \quad \text{for all } \xi \in \mathbb{F}.
\] (92.14)

We call \(p(\xi)\) the value of \(p\) at \(\xi\). For every given \(\xi \in \mathbb{F}\), the mapping \((p \mapsto p(\xi)) : \mathbb{F}(\mathbb{N}) \to \mathbb{F}\) is linear and preserves products, i.e. we have
\[
p(\xi)q(\xi) = (pq)(\xi) \quad \text{for all } p, q \in \mathbb{F}(\mathbb{N}).
\] (92.15)

Remark: If \(\mathbb{F} := \mathbb{R}\), then there is a one-to-one correspondence between polynomials and the associated polynomial functions. In fact, using only basic facts of elementary calculus, one can show that \(f : \mathbb{R} \to \mathbb{R}\) is a polynomial function if and only if, for some \(n \in \mathbb{N}\), \(f\) is \(n\) times differentiable and \(f^{(n)} = 0\) (see Problem 4 in Chap.1). If this is the case, then there is exactly one \(p \in \mathbb{R}(\mathbb{N})\) such that \(f = (\xi \mapsto p(\xi))\) and, if \(f \neq 0\), then \(\deg p < n\). In the case when \(\mathbb{F} := \mathbb{R}\), one often identifies the ring \(\mathbb{R}(\mathbb{N})\) of polynomials over \(\mathbb{R}\) with the ring of associated polynomial functions; then \(i\) as defined by (92.1) becomes identified with the identity mapping of \(\mathbb{R}\) as in the abbreviation (08.26).

If \(\mathbb{F}\) is any infinite field, one can easily show that there is still a one-to-one correspondence between polynomials and polynomial functions. However, if \(\mathbb{F}\) is a finite field, then every function from \(\mathbb{F}\) to \(\mathbb{F}\) is a polynomial function associated with infinitely many different polynomials.

Let \(\mathcal{V}\) be a linear space over \(\mathbb{F}\). The lineonic polynomial function \((\mathbf{L} \mapsto p(\mathbf{L})) : \text{Lin}\mathcal{V} \to \text{Lin}\mathcal{V}\) associated with a given \(p \in \mathbb{F}(\mathbb{N})\) is defined by
\[
p(\mathbf{L}) := \sum_{k \in \mathbb{N}} p_k \mathbf{L}^k \quad \text{for all } \mathbf{L} \in \text{Lin}\mathcal{V}.
\] (92.16)

We call \(p(\mathbf{L})\) the value of \(p\) at \(\mathbf{L}\). Again, for every given \(\mathbf{L} \in \text{Lin}\mathcal{V}\), the mapping \((p \mapsto p(\mathbf{L})) : \mathbb{F}(\mathbb{N}) \to \text{Lin}\mathcal{V}\) is linear and preserves products, i.e. we have
\[
p(\mathbf{L})q(\mathbf{L}) = (pq)(\mathbf{L}) \quad \text{for all } p, q \in \mathbb{F}(\mathbb{N}).
\] (92.17)

The following statements are easily seen to be valid for every lineon \(\mathbf{L} \in \text{Lin}\mathcal{V}\) and every polynomial \(p \in \mathbb{F}(\mathbb{N})\).

(I) \(\mathbf{L}\) commutes with \(p(\mathbf{L})\).

(II) Every \(\mathbf{L}\)-invariant subspace of \(\mathcal{V}\) is also \(p(\mathbf{L})\)-invariant.
92. LINEONIC POLYNOMIAL FUNCTIONS

(III) Null $p(L)$ and Rng $p(L)$ are $L$-invariant.

(IV) If $U$ is an $L$-subspace of $V$, then

\[ p(L|_U) = p(L)|_U. \]  \hspace{1cm} (92.18)

(V) We have

\[ p(L^\top) = (p(L))^\top. \]  \hspace{1cm} (92.19)

Pitfall: To call $p(\xi)$, as defined by (92.10), and $p(L)$, as defined by (92.12), the value of the polynomial $p$ at $\xi$ and $L$, respectively, is merely a figure of speech. One must often carefully distinguish between the polynomial $p$, the polynomial function associated with $p$, and the lineonic polynomial functions associated with $p$. For example, if $p$ is a polynomial over $\mathbb{C}$, the polynomial function associated with the termwise complex-conjugate $\overline{p}$ of $p$ is not the same as the value-wise complex-conjugate of the polynomial function associated with $p$. In some contexts it is useful to introduce explicit notations for the polynomial functions associated with $p$ (see, e.g. Problem 13 in Chapt.6).

Let a lineon $L \in \text{Lin}V$ be given. It is clear that the intersection of any collection of $L$-invariant subspaces of $V$ is again $L$-invariant. We denote the span-mapping corresponding to the collection of all $L$-subspaces of $V$ by $\text{Lsp}_L$ (see Sect.03). If $S \in \text{Sub}V$ we call $\text{Lsp}_L S$ the \textbf{linear $L$-span} of $S$: it is the smallest $L$-space that includes the set $S$. It is easily seen that

\[ \text{Lsp}_L S = \text{Lsp}\{L^k v \mid v \in S, \ k \in \mathbb{N}\}, \]  \hspace{1cm} (92.20)

and, for each $v \in V$,

\[ \text{Lsp}_L \{v\} = \{p(L)v \mid p \in \mathbb{P}(\mathbb{N})\}. \]  \hspace{1cm} (92.21)

**Proposition 1:** If $p \in (\mathbb{P}(\mathbb{N}))^\times$ and $v \in \text{Null} \ p(L)$, then

\[ \dim \text{Lsp}_L \{v\} \leq \deg p. \]  \hspace{1cm} (92.22)

**Proof:** We put $m := \deg p$. Since $v \in \text{Null} \ p(L)$, we have

\[ 0 = p(L)v = p_m(L^m v) + \sum_{k \in m^1} p_k(L^k v). \]

Since $p_m \neq 0$, it follows that $L^m v \in \text{Lsp}\{L^k v \mid k \in m^1\}$. Hence, for every $r \in \mathbb{N}$, we have

\[ L^{m+r} v \in (L^r)_{>}(\text{Lsp}\{L^k v \mid k \in m^1\}) = \text{Lsp}\{L^{k+r} v \mid k \in m^1\}. \]
Using induction over \( r \in \mathbb{N} \), we conclude that \( L^{m+r}v \in \text{Lsp}\{L^kv \mid k \in m\} \) for all \( r \in \mathbb{N} \) and hence, in view of (92.16), that

\[
\text{Lsp}_L\{v\} = \text{Lsp}\{L^kv \mid k \in m\}.
\]

Since the list \( \{L^kv \mid k \in m\} \) has \( m = \deg p \) terms, it follows from the Characterization of Dimension of Sect.17 that (92.18) holds.

**Proposition 2:** For every lineon \( L \in \text{Lin}\mathcal{V} \), there is a unique monic polynomial \( p \) of least degree whose value at \( L \) is zero. This polynomial \( p \) is called the minimal polynomial of \( L \).

**Proof:** Since \( \text{Lin}\mathcal{V} \) is finite-dimensional, the sequence \( \{L^k\} \) must be linearly dependent. In view of (92.12) this means that there is a \( p \in (\mathbb{F}(\mathbb{N}))^\times \) such that \( p(L) = 0 \). It follows that there is a monic polynomial \( q \) of least degree such that \( q(L) = 0 \). If \( q' \) were another such polynomial, then \( (q - q')(L) = 0 \). If \( q \neq q' \), then \( \deg q - \deg q' < \deg q = \deg q' \), and hence a suitable multiple of \( q - q' \) would be a monic polynomial with a degree strictly less than \( \deg q \) whose value at \( L \) is still zero.

If \( L \in \text{Lin}\mathcal{V} \) is given and if \( \mathcal{U} \) is an \( L \)-subspace of \( \mathcal{V} \) then the minimal polynomial of \( L|_{\mathcal{U}} \) is the monic polynomial of least degree such that \( \mathcal{U} \subseteq \text{Null} q(L) \). Using poetic license, we sometimes call this polynomial the minimal polynomial of \( \mathcal{U} \).

**Proposition 3:** Let \( L \in \text{Lin}\mathcal{V} \) and \( v \in \mathcal{V} \) be given, and let \( q \) be the minimal polynomial of \( L \). If \( \mathcal{V} = \text{Lsp}_L\{v\} \), then \( \dim \mathcal{V} = \deg q \).

**Proof:** Put \( n := \dim \mathcal{V} \) and consider the list \( \{L^k v \mid k \in (n+1)\} \). Since this list has \( n + 1 \) terms, it follows from the Characterization of Dimension that it must be a linearly dependent list. This means that

\[
0 = \sum_{k \in (n+1)} h_k(L^k v) = h(L)v
\]

for some non-zero polynomial \( h \) with \( \deg h \leq n \). By (92.13), we have

\[
h(L)(p(L)v) = (hp)(L)v = ((ph)(L))v = p(L)(h(L)v) = 0
\]

for all \( p \in \mathbb{F}(\mathbb{N}) \). Since \( \mathcal{V} = \{p(L)v \mid p \in \mathbb{F}(\mathbb{N})\} \) by (92.17), it follows that \( h \) has the value zero at \( L \). Therefore, since \( q \) is the minimal polynomial of \( L \), we have \( \deg q \leq \deg h \leq n \). On the other hand, application of Prop.1 to \( q \) yields \( n = \dim \mathcal{V} \leq \deg q \).

We say that a polynomial \( p \) is prime if (i) \( p \) is monic, (ii) \( \deg p > 0 \), (iii) \( p \) is not the product of two monic polynomials that are both different from \( p \). We will see in the next section that in order to understand the structure
of elementary lineons, one must know the structure of prime polynomials. Unfortunately, for certain fields $F$ (for example for $F := \mathbb{Q}$) it is very difficult to describe all possible prime polynomials over $F$. However in the case when $F$ is $C$ or $R$, the following theorem gives a simple description.

**Theorem on Prime Polynomials Over $C$ and $R$:** A polynomial $p$ over $C$ is prime if and only if it is of the form $p = \iota - \zeta$ for some $\zeta \in C$.

A polynomial $p$ over $R$ is prime if and only if it is either of the form $p = \iota - \lambda$ for some $\lambda \in R$ or else of the form $p = (\iota - \mu)^2 + \kappa^2$ for some $(\mu, \kappa) \in R \times \mathbb{P}^\times$.

The proof of this theorem depends on the following Lemma.

**Lemma 1:** If $p$ is a polynomial over $C$ with $\deg p \geq 1$, then the equation $\exists \ z \in C, \ p(z) = 0$ has at least one solution.

The assertion of this Lemma is included in Part (d) of Problem 13 in Chap.6. The proof is quite difficult, but Problem 13 in Chap.6 gives an outline from which the reader can construct a detailed proof.

**Lemma 2:** Let $p$ be a monic polynomial over a field $F$ and let $\xi \in F$ be such that $p(\xi) = 0$. Then there is a unique monic polynomial $q$ over $F$ such that

$$p = (\iota - \xi)q. \quad (92.23)$$

The proof of this lemma, which is easy, is based on what is usually called “long division” of polynomials. We also leave the details to the reader.

**Proof of the Theorem:** Let $p$ be a prime polynomial over $C$. By Lemma 1, we can find $\zeta \in C$ such that $p(\zeta) = 0$. By Lemma 2, we then have $p = (\iota - \zeta)q$ for some monic polynomial $q$ and hence, since $p$ is prime, $p = \iota - \zeta$ and $q = 1$.

Let now $p$ be a prime polynomial over $R$. If there exists a $\lambda \in R$ such that $p(\lambda) = 0$ we have $p = (\iota - \lambda)q$ for some monic polynomial $q$ and hence, since $p$ is prime, $p = (\iota - \lambda)$ and $q = 1$. Assume, then, that $p(\lambda) \neq 0$ for all $\lambda \in R$. Since $R \subset C$ and hence $R^{(N)} \subset C^{(N)}$, $p$ is also a monic polynomial over $C$. Hence, by Lemma 1, we can find $\zeta \in C \setminus R$ such that $p(\zeta) = 0$ and hence, by Lemma 2, a monic polynomial $q \in C^{(N)}$ such that $p = (\iota - \zeta)q$. Since $p \in R^{(N)}$, we have

$$0 = p(\overline{\zeta}) = p(\overline{\zeta}) = p(\overline{\zeta}) = 0,$$

where $\overline{p}$ denotes the termwise complex-conjugate of $p$. It follows that $(\overline{\zeta} - \zeta)q(\overline{\zeta}) = 0$. Since $\zeta \in C \setminus R$ we have $\overline{\zeta} - \zeta \neq 0$ and hence $q(\overline{\zeta}) = 0$. Using Lemma 2 again, we can find a monic polynomial $r \in C^{(N)}$ such that $q = (\iota - \overline{\zeta})r$ and hence

$$p = (\iota - \zeta)q = (\iota - \zeta)(\iota - \overline{\zeta})r.$$
If we put $\mu := \Re \zeta$ and $\kappa := |\Im \zeta|$ then $(\iota - \zeta)(\iota - \overline{\zeta}) = (\iota - \mu)^2 + \kappa^2$ and hence

$$p = ((\iota - \mu)^2 + \kappa^2)r.$$  \hfill (92.24)

Now, since $p = \overline{p}$, it follows from (92.20) that

$$0 = ((\iota - \mu)^2 + \kappa^2)(r - \overline{r}).$$

Since $r - \overline{r} \in \mathbb{R}^{(N)}$ and since the polynomial function associated with $((\iota - \mu)^2 + \kappa^2)$ has strictly positive values, the polynomial function associated with $r - \overline{r}$ must be the zero function. In view of the Remark above, it follows that $r = \overline{r}$ and hence $r \in \mathbb{R}^{(N)}$. Since $p$ is prime, it follows from (92.20) that $p = (\iota - \mu)^2 + \kappa^2$ and $r = 1$.

It is easy to see that polynomials of the forms described in the Theorem are in fact prime. \hfill \blacksquare

93 The Structure of Elementary Lineons

The following result is the basis for understanding the structure of elementary lineons.

**Structure Theorem for Elementary Lineons:** Let $L \in \text{Lin}V$ be an elementary lineon and let $M$ be the (only) minimal $L$-space. Then:

(a) There is a unique monic polynomial $q$ such that

$$M = \text{Null } q(L) \text{ and } \deg q = \dim M.$$  \hfill (93.1)

This polynomial $q$ is prime.

(b) $\dim V$ is a multiple of $\dim M$, i.e. there is a (unique) $d \in \mathbb{N}^\times$ such that

$$\dim V = d \dim M.$$  \hfill (93.2)

(c) There are exactly $d + 1$ $L$-spaces; they are given by

$$\mathcal{H}_k := \text{Null } q^k(L) = \text{Rng } q^{d-k}(L), \quad k \in (d+1)^\mathbb{N},$$  \hfill (93.3)

and their dimensions are

$$\dim \mathcal{H}_k = k \dim M, \quad k \in (d+1)^\mathbb{N}.$$  \hfill (93.4)
In particular, we have

$$H_0 = \{0\}, \quad H_1 = \mathcal{M}, \quad H_d = \mathcal{V}.$$  \hfill (93.5)

The polynomial $q$ will be called the **prime polynomial** of $L$ and the number $d$ the **depth** of $L$.

**Proof of Part (a):** Let $q$ be a polynomial that satisfies (93.1). We choose $v \in \mathcal{M} \setminus \{0\}$. Since $\{v\} \subset \mathcal{M}$ and since $\text{Lsp}_L\{v\}$ is the smallest $L$-space that includes $\{v\}$, it follows that $\text{Lsp}_L\{v\} \subset \mathcal{M}$. On the other hand, since $\text{Lsp}_L\{v\}$ is a non-zero $L$-subspace, it follows from Prop.3 of Sect.91 that $\mathcal{M} \subset \text{Lsp}_L\{v\}$. We conclude that $\mathcal{M} = \text{Lsp}_L\{v\} = \text{Lsp}_{L|M}\{v\}$. By Prop.3 of Sect.92, it follows that $\dim \mathcal{M}$ is the degree of the minimal polynomial of $L|M$. Since $q$ has the value zero at $L|M$ by (93.1) and since $\deg q = \dim \mathcal{M}$, it follows that $q$ must be the minimal polynomial of $L|M$, and hence is uniquely determined by $L$.

On the other hand, if $q$ is defined to be the minimal polynomial of $L|M$, it is not hard to verify, using Prop.3 of Sect.91 and Props.1 and 3 of Sect.92, that (93.1) holds.

To prove that $q$ is prime, assume that $q = q_1q_2$, where $q_1$ and $q_2$ are monic polynomials. It cannot happen that both $\text{Null } q_1(L) := \{0\}$ and $\text{Null } q_2(L) = \{0\}$, because this would imply $\mathcal{M} = \text{Null } p(L) = \{0\}$. Assume, for example, that $\text{Null } q_1(L) \neq \{0\}$ and choose $v \in (\text{Null } q_1(L))^\times$. By Prop.3 of Sect.91 it follows that $\mathcal{M} \subset \text{Lsp}_L\{v\}$ and hence, by Prop.1 of Sect.92, that $\deg q = \dim \mathcal{M} \leq \dim \text{Lsp}_L\{v\} \leq \deg q_1$. In view of (92.8), this is possible only when $\deg q = \deg q_1$ and $\deg q_2 = 0$ and hence $q_1 = q$ and $q_2 = 1$.

The proof of the remaining assertions will be based on the following

**Lemma:** There is a $d \in \mathbb{N}^\times$ such that

$$q^d(L) = \mathcal{0}, \quad \dim \mathcal{V} = d \dim \mathcal{M},$$  \hfill (93.6)

and

$$\dim \text{Rng } q^k(L) = (d - k) \dim \mathcal{M} \quad \text{for all } k \in d^\mathbb{N}.$$  \hfill (93.7)

**Proof:** Let $j \in \mathbb{N}$ be given and assume that $q^j(L) \neq \mathcal{0}$, i.e. that $\mathcal{U} := \text{Rng } q^j(L)$ is not the zero-space. By Prop.3 of Sect.91 we then have $\mathcal{M} \subset \mathcal{U}$. By Part (a), it follows that $\mathcal{M} = \text{Null } q(L) = \text{Null } (q(L)|_{\mathcal{U}})$. Noting that $\text{Rng } (q(L)|_{\mathcal{U}}) = \text{Rng } q^{j+1}(L)$, we can apply the Theorem on Dimensions of Range and Nullspace of Sect.17 to $q(L)|_{\mathcal{U}}$ and obtain

$$\dim \text{Rng } q^{j+1}(L) + \dim \mathcal{M} = \dim \text{Rng } q^j(L)$$  \hfill (93.8)
for all \( j \in \mathbb{N} \) such that \( q^j(L) \neq 0 \). For \( j = 0 \) (93.8) reduces to
\[
\dim \text{Rng} \, q(L) + \dim \mathcal{M} = \dim \mathcal{V}. \tag{93.9}
\]
Using induction, we conclude from (93.8) and (93.9) that
\[
\dim \text{Rng} \, q^{j+1}(L) = \dim \mathcal{V} - (j + 1) \dim \mathcal{M} \tag{93.10}
\]
for all \( j \in \mathbb{N} \) for which \( q^j(L) \neq 0 \). Since the right side of (93.10) becomes negative for large enough \( j \), there must be a \( d \in \mathbb{N} \) such that \( q^d(L) = 0 \) and \( 0 = \dim \mathcal{V} - d(\dim \mathcal{M}) \), which proves (93.6). We have \( q^{k-1}(L) \neq 0 \) for all \( k \in d \). Hence, if we substitute \( j := k - 1 \) and \( \dim \mathcal{V} = d(\dim \mathcal{M}) \) into (93.10), we obtain (93.7).

The part (93.6)_2 of the Lemma asserts the validity of Part (b) of the Theorem.

**Proof of Part (c):** Let \( k \in (d + 1)^i \) be given and put
\[
\mathcal{H}_k := \text{Null} \, q^k(L). \tag{93.11}
\]
We apply the Theorem on Dimensions of Range and Nullspace to \( q^k(L) \) and use the Lemma to obtain
\[
\dim \mathcal{H}_k = \dim \mathcal{V} - \dim \text{Rng} \, q^k(L) = k \dim \mathcal{M}, \tag{93.12}
\]
which proves (93.4). On the other hand, by (93.6)_1, we have
\[
\{0\} = \text{Rng} \, q^d(L) = q^k(L) \supset (\text{Rng} \, q^{d-k}(L))
\]
and hence \( \text{Rng} \, q^{d-k}(L) \subset \mathcal{H}_k \). But by (93.12) and (93.7) we have \( \dim \mathcal{H}_k = \dim \text{Rng} \, q^{d-k}(L) \) and hence \( \mathcal{H}_k = \text{Rng} \, q^{d-k}(L) \), which proves (93.3).

Now let \( \mathcal{U} \) be an \( L \)-subspace of \( \mathcal{V} \). If \( \mathcal{U} = \{0\} \), then \( \mathcal{U} = \mathcal{H}_0 \). If \( \mathcal{U} \neq \{0\} \), then Prop.3 of Sect.81 applies. Hence we can apply the Lemma to \( L_{|\mathcal{U}} \) instead of to \( L \) and conclude that there is a \( k \in \mathbb{N} \) such that \( \dim \mathcal{U} = k \dim \mathcal{M} \) and
\[
0 = q^k(L_{|\mathcal{U}}) = q^k(L)_{|\mathcal{U}}, \quad \text{i.e.} \quad \mathcal{U} \subset \text{Null} \, q^k(L) = \mathcal{H}_k.
\]
Since \( \dim \mathcal{U} = k \dim \mathcal{M} = \dim \mathcal{H}_k \) by (93.12), it follows that \( \mathcal{H}_k = \mathcal{U} \). Since \( \mathcal{U} \) was an arbitrary \( L \)-subspace of \( \mathcal{V} \), Part (c) follows.

We now list several corollaries. The first is an immediate consequence of Part (c) of the Theorem.

**Corollary 1:** A lineon \( L \in \text{Lin} \mathcal{V} \) is elementary if and only if the collection of \( L \)-spaces is totally ordered by inclusion, which means that, given any two \( L \)-spaces, one of them must be included in the other.
Corollary 2: A lineon \( L \) is elementary if and only if there is exactly one maximal \( L \)-space.

Corollary 3: If the lineon \( L \) is elementary, so is \( L^\top \), and \( L^\top \) has the same prime polynomial and the same depth as \( L \).

Proof: If \( L \) is elementary, with prime polynomial \( q \) and depth \( d \), then, by Part (c) of the Theorem, \( \mathcal{H}_{d-1} := \text{Null } q^{d-1}(L) \) is the only maximal \( L \)-space.

If there is only one maximal \( L \)-space \( W \), then \( W^\perp \) is the only minimal \( L^\top \)-space and hence \( L^\top \) is elementary. In particular, if \( L \) is elementary, so is \( L^\top \). Hence, if \( L \) has only one maximal \( L \)-space, we can apply this observation to \( L^\top \) and conclude that \( (L^\top)^\top = L \) must be elementary. Thus, Cor.2 and the first assertion of Cor.3 are proved.

If \( q \) is the prime polynomial of \( L \), then \( W := \mathcal{H}_{d-1} = \text{Rng } q(L) \) is the only maximal \( L \)-space and hence

\[
W^\perp = (\text{Rng } q(L))^\perp = \text{Null } q(L)^\top = \text{Null } q(L^\top)
\]

(see (21.13) and (92.15)) is the only minimal \( L^\top \)-space. We have

\[
\dim W^\perp = \dim V - \dim W = \dim V - \dim \text{Rng } q(L)
\]

\[
= \dim \text{Null } q(L) = \dim M = \deg q
\]

by the Formula for Dimension of Annihilators, the Theorem on Dimensions of Range and Nullspace and Part (a) of the Theorem. The uniqueness assertion of Part (a) of the Theorem, applied to \( L^\top \) instead of to \( L \), shows that \( q \) is also the prime polynomial of \( L^\top \). Since \( \dim V^* = \dim V \) and \( \dim W^\perp = \dim M \), it follows from part (b) of the Theorem that \( d \) is also the depth of \( L^\top \).

Corollary 4: If \( L \in \text{Lin} V \) is elementary, if \( W \) is the (only) maximal \( L \)-space and if \( v \in V \setminus W \), then \( V = \text{Lsp}_L \{v\} \).

Proof: We have \( \text{Lsp}_L \{v\} \nsubseteq W \). Since all \( L \)-spaces other than \( V \) must be included in \( W \), it follows that \( \text{Lsp}_L \{v\} \neq V \).

Corollary 5: Let \( L \in \text{Lin} V \) be an elementary lineon with prime polynomial \( q \) and depth \( d \). Then \( \dim V = \deg q^d \) and \( q^d \) is the minimal polynomial of \( L \).

Proof: Since \( V = \text{Null } q^d(L) \) by Part (c) of the Theorem, \( q^d \) has the value zero at \( L \). By Cor.4 and Prop.3 of Sect.92, the degree of the minimal polynomial of \( L \) must be \( \dim V \). On the other hand, by (92.8), (93.1)_2, and (93.2), we have \( \deg q^d = d \deg q = d(\dim M) = \dim V \). Hence \( q^d \) must be the minimal polynomial of \( L \).
Remark: Another proof of the Structure Theorem can be based on the fact, known to readers familiar with algebra, that the ring $F((N)$ of polynomials is a principal-ideal ring. The proof goes as follows: First, one shows that if $L \in \text{Lin} V$ is cyclic in the sense that $V = \text{Lsp}_L \{v\}$ for some $v \in V$, then a subspace $U$ of $V$ is an $L$-space if and only if $U = \text{Rng}_p(L)$ for some monic divisor $p$ of the minimal polynomial of $L$. One deduces from this that if there is only one maximal $L$-space, the minimal polynomial of $L$ must be of the form $q^d$, where $q$ is a prime polynomial. Using an argument like the one used in the proofs of Cor.2, and 3, one proves that Cor.2 and 3 are valid. The remainder of the proof is then very easy.

As in the case when $F := \mathbb{R}$ (see Sect.82) the spectrum of a lineon $L$ on a linear space $V$ over $F$ is defined by

$$\text{Spec } L := \{\sigma \in F | \text{ Null } (L - \sigma 1_V) \neq \{0\}\}.$$

If $L$ is elementary, then Spec $L$ is non-empty only in the exceptional case described as follows.

**Proposition 1:** If $L \in \text{Lin} V$ is elementary and has a non-empty spectrum, then this spectrum is a singleton, the only minimal $L$-space is one-dimensional, the prime polynomial of $L$ is $\iota - \sigma$ when $\sigma \in \text{Spec } L$, and the depth of $L$ is $n := \dim V$. Also, there are exactly $n+1$ $L$-subspaces and they are given by

$$\mathcal{H}_k := \text{Null } (L - \sigma 1_V)^k = \text{Rng } (L - \sigma 1_V)^{n-k}, \quad k \in (n+1)\mathbb{Z}.$$  

**Proof:** Let $L \in \text{Lin} V$ by given. If $\sigma \in \text{Spec } L$, then every one-dimensional subspace of $\text{Null } (L - \sigma 1_V)$ is evidently a minimal $L$-space. Therefore, if Spec $L$ has more than one element or if $\dim \text{Null } (L - \sigma 1_V) \geq 2$ for some $\sigma \in \text{Spec } L$, there are at least two distinct minimal $L$-spaces. Hence, if $L$ is elementary, Spec $L$ can have only one element. If $\sigma$ is this element, then $\text{Null } (L - \sigma 1_V)$ is one-dimensional and it is the only minimal $L$-space. The remaining statements follow immediately from the Structure Theorem for Elementary Lineons.

A lineon $L$ is said to be nilpotent if $L^m = 0$ for some $m \in \mathbb{N}^\times$. The least such $m$ is then called the nilpotency of $L$.

**Proposition 2:** A lineon $L \in \text{Lin} V$ is both elementary and non-invertible if and only if it is nilpotent with a nilpotency equal to $\dim V$.

**Proof:** By Prop.1 of Sect.18 and (93.13), $L$ is non-invertible if and only if $0 \in \text{Spec } L$. In view of this fact, the assertion follows immediately from Prop.1.
94. CANONICAL MATRICES

Notes 93

(1) The use of the terms “prime polynomial” and “depth” in the sense described in the Structure Theorem for Elementary Lineons was recently proposed by J. J. Schäffer.

(2) What we call simply the “nilpotency” of a nilpotent lineon is sometimes called the “index of nilpotence”.

94 Canonical Matrices

We assume that a linear space \( V \) over the field \( F \) is given. In this section, we suggest methods for finding bases of \( V \) relative to which a given lineon on \( L \) has a matrix of a simple and illuminating form. Such a matrix is called a canonical matrix. In view of the Elementary Decomposition Theorem of Sect.91, it is sufficient to consider only elementary lineons. Indeed, let \((E_i \mid i \in I)\) be an elementary \( L \)-decomposition for a given \( L \in \text{Lin} V \) and let, for each \( i \in I \), a basis \( b^{(i)} \) be determined such that the matrix \( M_i \) of \( L \mid E_i \) relative to \( b^{(i)} \) is canonical. Then one can “put together” (possibly with the help of reindexing) the bases \( b^{(i)} \), \( i \in I \), to obtain a basis \( b \) of \( V \) such that the only non-zero terms of the matrix \( M \) of \( L \) relative to \( b \) are those in blocks of the form \( M_i \) along the diagonal. An illustration will be given at the end of this section.

We first deal with elementary lineons having a non-empty spectrum.

**Proposition 1:** If \( L \in \text{Lin}^n V \) is elementary and if its spectrum is not empty, then there is a list-basis \( b := (b_i \mid i \in n) \), \( n := \dim V \), and \( \sigma \in F \) such that
\[
Lb_i = \begin{cases} 
\sigma b_i & \text{if } i = 1 \\
\sigma b_i + b_{i-1} & \text{if } i \in n \setminus \{1\} 
\end{cases}
\] (94.1)

**Proof:** By Prop.1 of Sect.93, there is \( \sigma \in F \) such that \( \text{Spec} L = \{\sigma\} \) and the only maximal \( L \)-space is \( H_{n-1} = \text{Rng} (L - \sigma 1_V) \). We choose \( v \in V \setminus H_{n-1} \) and define the list \( b \) by \( b_i := (L - \sigma 1_V)^{n-i} v \) for all \( i \in n \). We then have \((L - \sigma 1_V)b_1 = (L - \sigma 1_V)^n v = 0 \) and \((L - \sigma 1_V)b_i = b_{i-1} \) for all \( i \in n \setminus \{1\} \), which proves (94.1). \( \blacksquare \)

The matrix \([L]_b\) of \( L \) relative to a basis \( b \) for which (94.1) holds is a canonical matrix. It is given by
\[
([L]_b)_{i,j} = \begin{cases} 
\sigma & \text{if } j = i \\
1 & \text{if } j = i + 1 \\
0 & \text{otherwise}
\end{cases}
\] (94.2)
If \( n \) is small, it can be recorded explicitly in the form
\[
[L]_b = \begin{bmatrix}
\sigma & 1 \\
\sigma & 1 \\
& & \ddots & \ddots \\
& & & \sigma & 1 \\
& & & & \sigma
\end{bmatrix},
\tag{94.3}
\]
where zeros are omitted.

From now on we confine ourselves to the case when \( \mathbb{F} := \mathbb{C} \) or \( \mathbb{F} := \mathbb{R} \).

Let \( V \) be a linear space over \( \mathbb{C} \), let \( L \in \text{Lin}V \) be an elementary lineon, and let \( q \) be its prime polynomial. By the Theorem on Prime Polynomials over \( \mathbb{C} \) and \( \mathbb{R} \) of Sect. 92, \( q \) must be of the form \( q = \iota - \sigma \) for some \( \sigma \in \mathbb{C} \).

By (93.1) the minimal \( L \)-space is \( M = \text{Null}(L) = \text{Null}(L - \sigma 1_V) \). Since \( M \neq \{0\} \), it follows that \( \sigma \) belongs to the \( \mathbb{C} \)-spectrum of \( L \). Hence this spectrum is not empty. Therefore, Prop. 1 applies and one can find a basis \( b := (b_i \mid i \in n) \) such that the matrix \([L]_b\) is given by (94.2).

In sum, if \( \mathbb{F} := \mathbb{C} \), the situation covered by Prop. 1 of Sect. 93 and Prop. 1 above is general rather than exceptional.

We now assume that \( V \) is a linear space over \( \mathbb{R} \) and that \( L \in \text{Lin}V \) is an elementary lineon. Let \( q \) be its prime polynomial. By the Theorem on Prime Polynomials over \( \mathbb{C} \) and \( \mathbb{R} \), we must have either \( q = \iota - \lambda \) for some \( \lambda \in \mathbb{R} \) or else \( q = (\iota - \mu)^2 + \kappa^2 \) for some \( (\mu, \kappa) \in \mathbb{R} \times \mathbb{P} \). In the former case, Prop. 1 applies again. The latter case is covered by the following result.

**Proposition 2:** Let \( L \in \text{Lin}V \) be an elementary lineon whose prime polynomial is \( q = (\iota - \mu)^2 + \kappa^2 \), \( (\mu, \kappa) \in \mathbb{R} \times \mathbb{P} \), and whose depth is \( d \). Then there is a list-basis \( b := (b_i \mid i \in n) \), \( n := \dim V = 2d \), such that
\[
Lb_1 = \mu b_1 + \kappa b_2, \\
Lb_2 = -\kappa b_1 + \mu b_2, \\
Lb_{2k+1} = \mu b_{2k+1} + \kappa b_{2k+2} + b_{2k-1}, \\
Lb_{2k+2} = -\kappa b_{2k+1} + \mu b_{2k+2} + b_{2k}
\]
for all \( k \in (d - 1) \). \( 94.5 \)

For each \( k \in d \), the space \( \text{Lsp}\{b_{2k-1}, b_{2k}\} \) is a supplement of the \( L \)-space \( \mathcal{H}_k \) in the \( L \)-space \( \mathcal{H}_{k-1} \) (see (93.3)).

**Proof:** By the Structure Theorem for Elementary Lineons of Sect. 93, the \( L \)-subspaces are given by (93.3), and \( \mathcal{H}_{d-1} = \text{Null}(q^{d-1}(L)) \) is the only maximal \( L \)-space. We use the abbreviation \( D := L - \mu 1_V \), so that \( q(L) = \)
$D^2 + \kappa^2 I_V$. We choose $v \in V \setminus H_{d-1}$ and define the lists $e := (e_k \mid k \in \{d\})$ and $f := (f_k \mid k \in \{d\})$ recursively by

$$e_1 := v, \quad f_1 := \frac{1}{\kappa} De_1,$$

(94.6)

$$e_{k+1} := De_k + \kappa f_k, \quad f_{k+1} := Df_k - \kappa e_k \text{ for all } k \in \{d-1\}.$$

(94.7)

An easy calculation shows that

$$\kappa e_1 = \frac{1}{\kappa} q(L)e_1 - Df_1,$$

$$\kappa e_{k+1} = q(L)f_k - Df_{k+1} \text{ for all } k \in \{d-1\},$$

and

$$\kappa e_{k+1} = q(L)(Df_{k-1} - f_k) + 2\kappa^2 f_k \text{ for all } k \in \{d-1\}.$$

Since $q(L)^d = 0$, it follows that

$$q^{d-1}(L)e_k = -\frac{1}{\kappa} Dq^{d-1}(L)f_k \text{ for all } k \in \{d\},$$

$$q^{d-1}(L)f_k = \frac{\kappa}{2\kappa} q^{d-1}(L)e_{k+1} \text{ for all } k \in \{d-1\}.$$

(94.8)

Using the fact that $v = e_1 \notin H_{d-1} = \text{Null } q^{d-1}(L)$ and hence $q(L)^{d-1} e_1 \neq 0$, we conclude from (94.8), by induction, that

$$q(L)^{d-1} e_k \neq 0 \text{ and } q(L)^{d-1} f_k \neq 0 \text{ for all } k \in \{d\}.$$

(94.9)

We now define the list $b := (b_i \mid i \in \{n\})$ by

$$b_{2k-1} := q(L)^{d-k}e_k, \quad b_{2k} := q(L)^{d-k}f_k \text{ for all } k \in \{d\}.$$

(94.10)

It follows easily from (94.6) and (94.7) that (94.4) and (94.5) hold and hence that $\text{Lsp}\{b_i \mid i \in \{n\}\}$ is $L$-invariant. On the other hand, we have $b_{n-1} = b_{2d-1} = e_d$ by (94.10), and hence $b_{n-1} \notin \text{Null } q(L)^{d-1} = H_{d-1}$ by (94.9). Since $H_{d-1}$ is the only maximal $L$-space, it follows from Cor.4 of Sect.93 that $\text{Lsp}_{L}\{b_{n-1}\} = V$. We conclude that $\text{Lsp}\{b_i \mid i \in \{n\}\} = V$ and hence, by the Theorem on Characterization of Dimension, that $b$ is indeed a basis of $V$.

The fact that $\text{Lsp}\{b_{2k-1}, b_{2k}\}$ is a supplement of $H_k$ in $H_{k-1}$ is an immediate consequence of (94.9), (94.10), and (93.3)."
The matrix \( [L]_b \) of \( L \) relative to a basis \( b \) for which (94.4) and (94.5) hold is a canonical matrix. It is given by

\[
([L]_b)_{i,j} = \begin{cases} 
\mu & \text{if } j = i \\
\kappa & \text{if } j = i + 1 \text{ and } i \text{ is odd} \\
-k & \text{if } j = i - 1 \text{ and } i \text{ is even} \\
1 & \text{if } j = i + 2 \\
0 & \text{otherwise}
\end{cases}
\]  
(94.11)

If \( n \) is small, it can be recorded explicitly in the form

\[
[L]_b = \begin{bmatrix}
\mu & \kappa & 1 & 0 \\
-k & \mu & 0 & 1 \\
\mu & \kappa & 1 & 0 \\
-k & \mu & 0 & 1 \\
\mu & \kappa & 1 & 0 \\
-k & \mu & 0 & 1 \\
\mu & \kappa & 1 & 0 \\
-k & \mu & 0 & 1 \\
\mu & \kappa & 1 & 0 \\
-k & \mu & 0 & 1 \\
\end{bmatrix}
\]  
(94.12)

where zeros are omitted.

We now illustrate our results by considering a linear space \( \mathcal{V} \) over \( \mathbb{R} \) with \( \dim \mathcal{V} = 4 \). If \( L \) is a lineon on \( \mathcal{V} \), we can then find a basis \( b := (b_i \mid i \in 4) \) such that the matrix \( [L]_b \) has one of the following 9 forms (zeros are not written):

\[
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\lambda_3 \\
\lambda_4
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\mu & \kappa \\
-k & \mu
\end{bmatrix},
\begin{bmatrix}
\mu & \kappa \\
-k & \mu
\end{bmatrix},
\begin{bmatrix}
\mu & \kappa \\
-k & \mu
\end{bmatrix},
\begin{bmatrix}
\mu & \kappa \\
-k & \mu
\end{bmatrix},
\begin{bmatrix}
\mu & \kappa \\
-k & \mu
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\sigma & 1 \\
\sigma & 1
\end{bmatrix},
\begin{bmatrix}
\mu & \kappa \\
-k & \mu
\end{bmatrix},
\begin{bmatrix}
\mu & \kappa \\
-k & \mu
\end{bmatrix}.
\]
The number of terms in an elementary decomposition of $L$ is 4 in the first form, 3 in the second and third, 2 in the fourth up to the seventh, and 1 in the last two. The first, third, and seventh forms apply when $L$ is semi-simple.

**Pitfall:** If $V$ is not the zero space, the basis relative to which the matrix of an elementary lineon has the form (94.2) or (94.11) is never uniquely determined by $L$. Indeed, the construction of these bases involved the choice of an element outside the maximal $L$-space, and different choices give rise to different bases.

**Remark:** Props.1 and 2 can be used to prove the following result: If $V$ is a field over $\mathbb{C}$ or $\mathbb{R}$, then every $L \in \text{Lin}V$ has an additive decomposition $(N, S)$, $N, S \in \text{Lin}V$, such that $L = N + S$, $N$ is nilpotent, $S$ is semi-simple, and $N$ and $S$ commute. Indeed, if $L$ is elementary, one can define $S$ to be the lineon whose matrix, relative to a basis for which (94.2) or (94.11) hold, is obtained by replacing 1 on the right side of (94.2) or (94.11) by 0. Then one can define $N := L - S$ and prove that $(N, S)$ is a decomposition with the desired properties. The case when $L$ is not elementary can be reduced to the case when it is by applying the Elementary Decomposition Theorem.

Actually the result just described remains valid for a large class of fields (often called perfect fields), and the decomposition can be proved to be unique.

Notes 94

1. A matrix of the type described by (94.2) or (94.3) is often called an “elementary Jordan matrix”. A matrix whose only non-zero terms are in blocks of the form described by (94.2) or (94.3) along the diagonal is often called a “Jordan matrix”, “Jordan form”, or “Jordan canonical form”. Using this terminology, we may say, in consequence of the results of Sects.91-93 and of Prop.1, that for every lineon on a linear space over $\mathbb{C}$, we can find a basis such that the matrix of the lineon relative to that basis is a Jordan matrix.

95  **Similarity, Elementary Divisors**

**Definitions** We say that the lineon $L \in \text{Lin}V$ is similar to the lineon $L' \in \text{Lin}V'$ if there is a linear isomorphism $A : V \to V'$ such that

$$L' = ALA^{-1}, \quad \text{i.e.} \quad L'A = AL.$$  

(95.1)

Assume that $L' \in \text{Lin}V'$ is similar to $L \in \text{Lin}V$. By Cor. 2 to the Characterization of Dimension (Sect.17) we then must have $\dim V = \dim V'$. Also, the following results, all easily proved, are valid.
(I) A subspace $\mathcal{U}$ of $\mathcal{V}$ is $L$-invariant if and only if $A_>(\mathcal{U})$ is $L'$-invariant. (See Prop.3 of Sect.8).

(II) $L$ is elementary if and only if $L'$ is elementary. If this is the case, then $L$ and $L'$ have the same prime polynomial and the same depth.

(III) A decomposition $(E_i | i \in I)$ of $\mathcal{V}$ is an elementary $L$-decomposition if and only if $(A_>(E_i) | i \in I)$ is an elementary $L'$-decomposition of $\mathcal{V}'$.

Roughly, similar lineons have the same intrinsic structure.

The following result shows, roughly, that elementary decompositions of similar lineons can be made to correspond.

**Similarity Theorem for Lineons:** Let $(E_i | i \in I)$ and $(E'_i | i \in I')$ be elementary decompositions of the given lineons $L \in \text{Lin}\mathcal{V}$ and $L' \in \text{Lin}\mathcal{V}'$, respectively. If $L'$ is similar to $L$, then there is an invertible mapping $\varphi : I \to I'$ such that $L_{E_i}$ is similar to $L'_{E'_i}$ for all $i \in I$.

The proof will be based on the following:

**Lemma:** Assume that (95.1) holds for given $L \in \text{Lin}\mathcal{V}$, $L' \in \text{Lin}\mathcal{V}'$, and $A \in \text{Lis}(\mathcal{V}, \mathcal{V}')$. Let $E, \mathcal{U}$ be supplementary $L$-subspaces of $\mathcal{V}$ and let $E', \mathcal{U}'$ be supplementary $L'$-subspaces of $\mathcal{V}'$. Let $E \in \text{Lin}\mathcal{V}$ be the idempotent for which $\text{Null } E = E$ and $\text{Rng } E = \mathcal{U}$ and let $F \in \text{Lin}\mathcal{V}'$ be the idempotent for which $\text{Null } F = \mathcal{U}'$ and $\text{Rng } F = E'$ (see Prop.4 of Sect.19). Then:

(a) $\text{Null } (FA)$ is an $L$-subspace of $\mathcal{V}$ and $\text{Null } (EA^{-1})$ is an $L'$-subspace of $\mathcal{V}'$.

(b) $A_>(\text{Null } (FA)|_E)) = \text{Null } (EA^{-1}|_{E'})$.

(c) If $\dim E \geq \dim E'$ and $\text{Null } (FA|_E) = \{0\}$, then $L_{E_E}$ is similar to $L'_{E'_E}$, and $L_{E_E}$ is similar to $L'_{E'_E}$.

**Proof:** It follows immediately from the definitions of $E$ and $F$ that $E$ commutes with $L$ and that $F$ commutes with $L'$.

Let $v \in \text{Null } (FA)$ be given. Using (95.1) we obtain $0 = L'(FA)v = F(L'A)v = F(AL)v = FA(Lv)$ and hence $Lv \in \text{Null } (FA)$. Since $v \in \text{Null } (FA)$ was arbitrary, it follows that $\text{Null } (FA)$ is $L$-invariant. Interchanging the roles of $L$ and $L'$ we find that $\text{Null } (EA^{-1})$ is $L'$-invariant. Hence Part (a) is proved.

Let $v \in \mathcal{V}$ be given. Then

$$v \in \text{Null } (FA|_E) = \text{Null } (FA) \cap E$$

$$\iff v \in E = \text{Null } E \text{ and } F(Av) = 0$$

$$\iff 0 = Ev = (EA^{-1})(Av) \text{ and } Av \in \text{Null } F = \mathcal{U}'$$

$$\iff Av \in \text{Null } (EA^{-1}) \cap \mathcal{U}' = \text{Null } (EA^{-1}|_{E'})$$.
95. SIMILARITY, ELEMENTARY DIVISORS 369

Since \( v \in V \) was arbitrary, Part (b) follows.

If \( \dim \mathcal{E} \geq \dim \mathcal{E}' \) and \( \text{Null} (\mathbf{F}A|_{\mathcal{E}}) = \{0\} \) we can apply the Pigeonhole Principle for Linear Mappings of Sect.17 to \( (\mathbf{F}A)|_{\mathcal{E}'} \) and conclude that \( \dim \mathcal{E} = \dim \mathcal{E}' \) and that \( (\mathbf{F}A)|_{\mathcal{E}'} \) is invertible. Using (95.1) we obtain

\[
(\mathbf{F}A)|_{\mathcal{E}'} \mathbf{L}|_{\mathcal{E}} = (\mathbf{FAL})|_{\mathcal{E}'} = (\mathbf{FL'}\mathbf{A})|_{\mathcal{E}'} = (\mathbf{L'FA})|_{\mathcal{E}'} = \mathbf{L'}|_{\mathcal{E}'} (\mathbf{F}A)|_{\mathcal{E}'}
\]

and hence that \( \mathbf{L'}|_{\mathcal{E}'} \) is similar to \( \mathbf{L}|_{\mathcal{E}} \). By part (b) we also have \( \text{Null} (\mathbf{E}A^{-1}|_{\mathcal{E}'}) = \{0\} \). Since

\[
\dim \mathcal{U}' = \dim \mathcal{V} - \dim \mathcal{E}' = \dim \mathcal{V} - \dim \mathcal{E} = \dim \mathcal{U}
\]

we can apply the same argument as just given to the case when \( \mathbf{L} \) and \( \mathbf{L}' \) are interchanged and conclude that \( \mathbf{L}|_{\mathcal{U}} \) is similar to \( \mathbf{L}'|_{\mathcal{U}} \). [1]

**Proof of the Theorem:** We choose a space of greatest dimension from the collection \( \{\mathcal{E}_i \mid i \in I\} \cup \{\mathcal{E}'_i \mid i \in I'\} \). Without loss of generality we may assume that we have chosen \( \mathcal{E}_j, j \in I \). Let \( \mathcal{M} \) be the (only) minimal \( \mathbf{L} \)-space included in \( \mathcal{E}_j \). Let \( (\mathbf{E}'_i) \mid i \in I \) be the family of idempotents associated with the decomposition \( (\mathcal{E}'_i) \mid i \in I' \) (see Prop.5 of Sect.81). By (81.5), we have

\[
\sum_{i \in I'} (\mathbf{E}'_i \mathbf{A})|_{\mathcal{M}} = (\sum_{i \in I'} \mathbf{E}'_i) \mathbf{A}|_{\mathcal{M}} = \mathbf{A}|_{\mathcal{M}}.
\]

Since \( \mathbf{A} \) is invertible and \( \mathcal{M} \neq \{0\} \), we may choose \( j' \in I' \) such that \( (\mathbf{E}'_{j'} \mathbf{A})|_{\mathcal{M}} \neq \{0\} \). We now abbreviate \( \mathcal{E} := \mathcal{E}_j, \mathcal{E}' := \mathcal{E}'_{j'} \),

\( \mathcal{U} := \sum (\mathcal{E}_i \mid i \in I \setminus \{j\}) \), \( \mathcal{U}' := \sum (\mathcal{E}'_i \mid i \in I' \setminus \{j'\}) \) and \( \mathbf{F} := \mathbf{E}'_{j'} \). The Lemma then applies. By part (a), \( \text{Null} (\mathbf{F}A) \) is an \( \mathbf{L} \)-subspace of \( \mathcal{V} \) and so is \( \mathcal{E} \cap \text{Null} (\mathbf{F}A) = \text{Null} (\mathbf{F}A|_{\mathcal{E}}) \). Since \( \mathbf{F}A|_{\mathcal{M}} \neq \{0\} \) and \( \mathcal{M} \subset \mathcal{E} \), we cannot have \( \mathcal{M} \subset \text{Null} (\mathbf{F}A|_{\mathcal{E}}) \). Hence, since \( \mathcal{E} \) is elementary, it follows from Prop.3 of Sect.91 that \( \text{Null} (\mathbf{F}A|_{\mathcal{E}}) = \{0\} \). In view of the maximality assumption on the dimension of \( \mathcal{E} := \mathcal{E}_j \), we have

\[
\dim \mathcal{E} = \dim \mathcal{E}'.
\]

Hence we can apply Part (c) of the Lemma to conclude that \( \dim \mathcal{E}'_j = \dim \mathcal{E}'_{j'} \), that \( \mathbf{L}|_{\mathcal{E}_j} \) is similar to \( \mathbf{L}'|_{\mathcal{E}'_j} \), and that \( \mathbf{L}|_{\mathcal{U}_j} \) is similar to \( \mathbf{L}'|_{\mathcal{U}'_{j'}} \). The desired result now follows immediately by induction, using Prop.6 of Sect.81. [1]

If we apply the Similarity Theorem to the case when \( \mathbf{L}' = \mathbf{L} \) we obtain

**Corollary 1:** All elementary \( \mathbf{L} \)-decompositions for a given lineon \( \mathbf{L} \) have the same number of terms. Moreover, if two such decompositions are given, then the terms of one decomposition can be matched with the terms of the other such that the adjustments of \( \mathbf{L} \) to the matched \( \mathbf{L} \)-spaces are similar.
In view of the statement (II) above, this Corollary implies that the prime polynomials and the depths of the adjustments of a lineon \( L \) to the terms in an elementary \( L \)-decomposition, if each is counted with an appropriate “multiplicity”, do not depend on the decomposition but only on \( L \). More precisely, denoting the set of all powers of prime polynomials over \( \mathbb{F} \) by \( \mathfrak{P} \), one can associate with each lineon \( L \) a unique elementary multiplicity function

\[
\text{emult}_L : \mathfrak{P} \to \mathbb{N}
\]

with the following property: In every elementary decomposition of \( L \), there are exactly \( \text{emult}_L(q^d) \) terms whose minimal polynomial is \( q^d \). The support of \( \text{emult}_L \) is finite and the members of this support are called the elementary divisors of \( L \). They are all divisors of the characteristic polynomial \( \text{chp} \) of \( L \), which is defined by

\[
\text{chp}_L := \prod_{q^d \in \mathfrak{P}} (q^d)^{\text{emult}_L(q^d)}.
\]

It follows from Cor.5 of Sect.93 and Prop.4 of Sect.81 that the degree of the characteristic polynomial (95.2) is \( \dim V \). It is evident that \( \text{chp}_L(L) = 0 \). Therefore the degree of the minimal polynomial of \( L \) cannot exceed the dimension of the domain \( V \) of \( L \).

**Remark 1:** In Vol.II we will give another definition of the characteristic polynomial, a definition in terms of determinants. Then (95.2) and \( \text{chp}_L(L) = 0 \) will become theorems. The determinant of \( L \) will turn out to be given by \( \det L = (-1)^{\dim V} (\text{chp}_L)_0 \).

The following consequence of the Similarity Theorem is now evident.

**Corollary 2:** If the lineon \( L' \) is similar to the lineon \( L \), then \( L' \) and \( L \) have the same elementary multiplicity functions, the same elementary divisors, and the same characteristic polynomial, i.e. we have \( \text{emult}_L = \text{emult}_{L'} \) and \( \text{chp}_L = \text{chp}_{L'} \).

**Remark 2:** The converse of Cor.2 is also true: if \( \text{emult}_L = \text{emult}_{L'} \) then \( L' \) is similar to \( L \) (see Problem 2). However, if \( L \) and \( L' \) merely have the same set of elementary divisors (without counting multiplicity), or if \( L \) and \( L' \) merely have the same characteristic polynomial, then they need not be equivalent.
96 Problems for Chapter 9

(1) Consider

\[
L := \begin{bmatrix}
3 & 0 & 1 \\
1 & 2 & 1 \\
-1 & 0 & 1
\end{bmatrix} \in \text{Lin} \mathbb{R}^3.
\]

(a) Find an elementary \(L\)-decomposition of \(\mathbb{R}^3\).

(b) Determine the elementary divisors, the characteristic polynomial, and the minimal polynomial of \(L\).

(c) Find a basis \(b := (b_1, b_2, b_3)\) of \(\mathbb{R}^3\) such that the matrix \([L]_b\) is canonical in the sense described in Sect.94.

(2) Let \(V\) and \(V'\) be linear spaces (over any field \(F\)). Let \(L \in \text{Lin} V\), \(L' \in \text{Lin} V'\) be given and let \(q\) and \(q'\) be their respective minimal polynomials.

(a) Assume that \(q = q'\) and \(\deg q = \dim V = \dim V'\). Show that \(L'\) is then similar to \(L\).

(d) Assume that \(L\) and \(L'\) are both elementary and have the same prime polynomial and depth. Show that \(L'\) is then similar to \(L\). (Hint: Use Cor.5 to the Structure Theorem for Elementary Lineons.)

(c) Assume that \(L\) and \(L'\) have the same elementary multiplicity function, i.e. that \(\text{emult}_L = \text{emult}_{L'}\) (see Sect.95). Show that \(L'\) is then similar to \(L\). (Hint: Choose elementary decompositions for \(L\) and \(L'\), use Part (b) above, and then Prop.6 of Sect.81.)

(3) Let \((\mathcal{E}_i \mid i \in I)\) be a decomposition of a given linear space \(V\) and let \(L \in \text{Lin} V\).

(a) Put

\[
\mathcal{E}_j' := \bigcap_{i \in I \setminus \{j\}} \mathcal{E}_i^\perp \quad \text{for all } j \in I.
\]

(P9.1)

Show that \((\mathcal{E}_j' \mid j \in I)\) is a decomposition of \(V^*\).

(b) Prove: If \((\mathcal{E}_i \mid i \in I)\) is an elementary \(L\)-decomposition, then \((\mathcal{E}_j' \mid j \in I)\) as defined by \((P9.1)\), is an elementary \(L^\top\)-decomposition and \((L_{|\mathcal{E}_i})^\top\) is similar to \((L^\top)|_{\mathcal{E}_i'}\) for each \(i \in I\).
(c) Prove that $L^\top$ is similar to $L$. (Hint: Use Cor. 3 to the Structure Theorem for Elementary Lineons and the results of Problem 2.)

(4) Let $\mathcal{V}$ be a linear space over $\mathbb{R}$ and let $L$ be a lineon on $\mathcal{V}$.

(a) Show: If $L$ is elementary, then either $\text{Spec} L$ is a singleton, in which case $\text{Pspec} L$ (see Def. 2 of Sect. 88) is empty, or else $\text{Spec} L$ is empty, in which case $\text{Pspec} L$ is a singleton.

(b) Assume that $L$ is elementary and that $\text{Spec} L$ is a singleton. Put $\sigma \in \text{Spec} L$. Show that

$$L = \sigma 1_{\mathcal{V}} + N \quad (P9.2)$$

for some nilpotent lineon $N$ with nilpotency $n := \dim \mathcal{V}$ (see Sect. 93).

(c) Assume that $L$ is elementary and that $\text{Spec} L$ is empty. Put $(\mu, \kappa) \in \text{Pspec} L$. Show that $\dim \mathcal{V}$ must be even and that

$$L = \mu 1_{\mathcal{V}} + \kappa J + N \quad (P9.3)$$

for some $J \in \text{Lin} \mathcal{V}$ satisfying $J^2 = -1_{\mathcal{V}}$ and some nilpotent lineon $N$ that commutes with $L$ and has nilpotency $d := \frac{1}{2} \dim \mathcal{V}$.

(d) Prove: $\text{Spec} L$ cannot be empty when $\dim \mathcal{V}$ is odd and $\text{Pspec} L$ cannot be empty when $\text{Spec} L$ is empty.

(5) Let $\mathcal{V}$ be a linear space over $\mathbb{R}$ and let $L$ be a lineon on $\mathcal{V}$.

(a) Let $(\mathcal{E}_i \mid i \in I)$ be a decomposition of $\mathcal{V}$ all of whose terms are $L$-spaces. Prove that the terms are then also $(\exp \mathcal{V}(L))$-spaces, that

$$(\exp \mathcal{V}(L))|_{\mathcal{E}_i} = \exp \mathcal{E}_i (L|_{\mathcal{E}_i}) \quad \text{for all} \quad i \in I, \quad (P9.4)$$

and that

$$\exp \mathcal{V}(L) = \sum_{i \in I} \exp \mathcal{E}_i (L|_{\mathcal{E}_i})^\mathcal{V} P_i, \quad (P9.5)$$

where $(P_i \mid i \in I)$ is the family of projections associated with the decomposition (see Prop. 5 of Sect. 81). (Hint: Use a proof similar to the one of Prop. 3 of Sect. 85.)
Remark: Since this result applies, in particular, to elementary decompositions, we see that the problem of evaluating the exponential of an arbitrary lineon is reduced to the problem of evaluating the exponential of elementary lineons.

(b) Assume that $L$ is elementary. Show that

$$\exp_V(L) = e^\sigma \sum_{k \in n^1} \frac{1}{k!} N^k$$  \hspace{1cm} (P9.6)

if $L$ is of the form (P9.2) of Problem 4 and that

$$\exp_V(L) = e^\mu (\cos \kappa 1_V + \sin \kappa J) \sum_{k \in d^1} \frac{1}{k!} N^k$$  \hspace{1cm} (P9.7)

if $L$ is of the form (P9.3). (Hint: Use the results of Problem 9 in Chap.6.)
### Index of Theorem Titles

<table>
<thead>
<tr>
<th>Theorem Title</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additive Decomposition Theorem</td>
<td>86</td>
<td>321</td>
</tr>
<tr>
<td>Annihilators and Transposes, Th. on</td>
<td>21</td>
<td>73</td>
</tr>
<tr>
<td>Attainment of Extrema, Th. on</td>
<td>08, 58</td>
<td>35, 203</td>
</tr>
<tr>
<td>Basic Convergence Criterion</td>
<td>55</td>
<td>188</td>
</tr>
<tr>
<td>Cell-Inclusion Theorem</td>
<td>52</td>
<td>173</td>
</tr>
<tr>
<td>Chain Rule for Flat Mappings</td>
<td>33</td>
<td>111</td>
</tr>
<tr>
<td>Characterization of Bases</td>
<td>15</td>
<td>54</td>
</tr>
<tr>
<td>Characterization of Dimension</td>
<td>17</td>
<td>58</td>
</tr>
<tr>
<td>Characterization of Gradients</td>
<td>63</td>
<td>219</td>
</tr>
<tr>
<td>Characterization of Regular Subspaces</td>
<td>41</td>
<td>137</td>
</tr>
<tr>
<td>Characterization of the Trace</td>
<td>26</td>
<td>89</td>
</tr>
<tr>
<td>Cluster Point Theorem</td>
<td>08, 55</td>
<td>33, 187</td>
</tr>
<tr>
<td>Compact Image Theorem</td>
<td>58</td>
<td>200</td>
</tr>
<tr>
<td>Compactness Theorem</td>
<td>58</td>
<td>201</td>
</tr>
<tr>
<td>Composition Theorem for Continuity</td>
<td>56</td>
<td>194</td>
</tr>
<tr>
<td>Composition Theorem for Uniform Continuity</td>
<td>56</td>
<td>195</td>
</tr>
<tr>
<td>Connection Components, Th. on</td>
<td>73</td>
<td>286</td>
</tr>
<tr>
<td>Congruence Theorem</td>
<td>46</td>
<td>153</td>
</tr>
<tr>
<td>Constrained-Extremum Theorem</td>
<td>69</td>
<td>252</td>
</tr>
<tr>
<td>Continuity of Uniform Limits, Th. on</td>
<td>56</td>
<td>196</td>
</tr>
<tr>
<td>Contraction Fixed Point Theorem</td>
<td>64</td>
<td>227</td>
</tr>
<tr>
<td>Convex Hull Theorem</td>
<td>37</td>
<td>124</td>
</tr>
<tr>
<td>Curl-Gradient Theorem</td>
<td>611</td>
<td>261</td>
</tr>
<tr>
<td>Deviation Components, Th. on</td>
<td>73</td>
<td>287</td>
</tr>
<tr>
<td>Difference-Quotient Theorem</td>
<td>08, 61</td>
<td>35, 211</td>
</tr>
<tr>
<td>Differentiation Theorem for Lineonic Exponentials</td>
<td>612</td>
<td>268</td>
</tr>
<tr>
<td>Differentiation Theorem for Integral Representations</td>
<td>610</td>
<td>258</td>
</tr>
<tr>
<td>Differentiation Theorem for Inversion Mappings</td>
<td>68</td>
<td>246</td>
</tr>
<tr>
<td>Dimension of Annihilators, Formula for</td>
<td>21</td>
<td>73</td>
</tr>
<tr>
<td>Theorem Title</td>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>Dimensions of Range and Nullspace, Th. on.</td>
<td>17</td>
<td>60</td>
</tr>
<tr>
<td>Elementary Decomposition Theorem</td>
<td>91</td>
<td>350</td>
</tr>
<tr>
<td>Elimination of Unknowns, Th. on.</td>
<td>16</td>
<td>57</td>
</tr>
<tr>
<td>Extreme Spectral Values of a Symmetric Lineon, Th. on.</td>
<td>84</td>
<td>312</td>
</tr>
<tr>
<td>Extremum Theorem</td>
<td>08, 69</td>
<td>35, 251</td>
</tr>
<tr>
<td>Flat Span Theorem</td>
<td>35</td>
<td>117</td>
</tr>
<tr>
<td>Fundamental Theorem of Calculus</td>
<td>08, 610</td>
<td>36, 255</td>
</tr>
<tr>
<td>General Chain Rule</td>
<td>63</td>
<td>219</td>
</tr>
<tr>
<td>General Product Rule</td>
<td>66</td>
<td>235</td>
</tr>
<tr>
<td>Half-Space Inclusion Theorem</td>
<td>38</td>
<td>126</td>
</tr>
<tr>
<td>Half-Space Intersection Theorem</td>
<td>54</td>
<td>183</td>
</tr>
<tr>
<td>Induced Inner-Product, Th. on the</td>
<td>44</td>
<td>144</td>
</tr>
<tr>
<td>Implicit Mapping Theorem</td>
<td>68</td>
<td>245</td>
</tr>
<tr>
<td>Inner-Product Index Theorem</td>
<td>47</td>
<td>157</td>
</tr>
<tr>
<td>Inner-Product Inequality</td>
<td>42</td>
<td>139</td>
</tr>
<tr>
<td>Inner-Product Signature Theorem</td>
<td>47</td>
<td>155</td>
</tr>
<tr>
<td>Interchange of Partial Gradients, Th. on.</td>
<td>611</td>
<td>264</td>
</tr>
<tr>
<td>Inversion Rule for Flat Mappings</td>
<td>33</td>
<td>111</td>
</tr>
<tr>
<td>Lineonic Logarithm Theorem</td>
<td>85</td>
<td>320</td>
</tr>
<tr>
<td>Lineonic Square-Root Theorem</td>
<td>85</td>
<td>318</td>
</tr>
<tr>
<td>Local Inversion Theorem</td>
<td>68</td>
<td>245</td>
</tr>
<tr>
<td>Norm-Duality Theorem</td>
<td>52</td>
<td>175</td>
</tr>
<tr>
<td>Norm-Equivalence Theorem</td>
<td>51</td>
<td>170</td>
</tr>
<tr>
<td>Partial Gradient Theorem</td>
<td>65</td>
<td>232</td>
</tr>
<tr>
<td>Pigeonhole Principle</td>
<td>05</td>
<td>21</td>
</tr>
<tr>
<td>Pigeonhole Principle for Linear Mappings</td>
<td>17</td>
<td>60</td>
</tr>
<tr>
<td>Polar Decomposition Theorem</td>
<td>86</td>
<td>322</td>
</tr>
<tr>
<td>Prime Polynomials over $\mathbb{C}$ and $\mathbb{R}$, Th. on.</td>
<td>92</td>
<td>357</td>
</tr>
<tr>
<td>Real and Imaginary Parts, Th. on.</td>
<td>89</td>
<td>338</td>
</tr>
</tbody>
</table>
### Index of Theorem Titles

<p>| Representation Theorem for Linear Forms on a Space of Linear Mappings | 26 | 89 |
| Similarity Theorem for Lineons | 95 | 368 |
| Smoothness of the Strict Lineonic Square Root, Th. on. | 85 | 318 |
| Specification of Flat Mappings, Th. on. | 33 | 109 |
| Spectral spaces, Th. on. | 82 | 308 |
| Spectral Theorem | 84 | 313 |
| Spectral Theorem for Normal C-Lineons | 810 | 342 |
| Striction Estimate for Differentiable Mappings | 64 | 224 |
| Strong Convex Hull Theorem | 37 | 125 |
| Structure Theorem for Elementary Lineons | 93 | 358 |
| Structure Theorem for Normal Lineons | 88 | 330 |
| Structure Theorem for Orthogonal Lineons | 88 | 333 |
| Structure Theorem for Perpendicular Turns | 87 | 326 |
| Structure Theorem for Skew Lineons | 87 | 328 |
| Subadditivity of Magnitude | 42 | 139 |
| Symmetry of Second Gradients, Th. on. | 611 | 263 |
| Uniform Continuity Theorem | 58 | 200 |
| Unique Existence of Barycenters, Th. on. | 34 | 112 |</p>
<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S \subset T$</td>
<td>($S$ is a subset of $T$)</td>
<td>01</td>
<td>3</td>
</tr>
<tr>
<td>$S \not\subset T$</td>
<td>($S$ is a proper subset of $T$)</td>
<td>01</td>
<td>3</td>
</tr>
<tr>
<td>$S \setminus T$</td>
<td>(set-difference of $S$ and $T$)</td>
<td>01</td>
<td>4</td>
</tr>
<tr>
<td>$(a_i \mid i \in I)$</td>
<td>(family with index set $I$)</td>
<td>02</td>
<td>7</td>
</tr>
<tr>
<td>$S^I$</td>
<td>(set of all families in $S$ with index set $I$)</td>
<td>02</td>
<td>7</td>
</tr>
<tr>
<td>$M^{(I)}$</td>
<td>(set of all families in $M$ with index set $I$ and finite support)</td>
<td>07</td>
<td>28</td>
</tr>
<tr>
<td>$(\mathbb{R}^{(I)})_{\nu}$</td>
<td>(set of all families in $\mathbb{R}$ with index set $I$, finite support, and sum $\nu$)</td>
<td>35, 37</td>
<td>116, 124</td>
</tr>
<tr>
<td>$\times (A_i \mid i \in I)$</td>
<td>(set-product of the family $(A_i \mid i \in I)$ of sets)</td>
<td>04</td>
<td>15</td>
</tr>
<tr>
<td>$f \times g$</td>
<td>(cross-product of the mappings $f$ and $g$)</td>
<td>04</td>
<td>17</td>
</tr>
<tr>
<td>$\times (f_i \mid i \in I)$</td>
<td>(cross-product of the family $(f_i \mid i \in I)$ of mappings)</td>
<td>04</td>
<td>17</td>
</tr>
<tr>
<td>$g^{\times I}$</td>
<td>($I$-cross-power of the mapping $g$)</td>
<td>04</td>
<td>18</td>
</tr>
<tr>
<td>$f_&gt;$</td>
<td>(image mapping of $f$)</td>
<td>03</td>
<td>12</td>
</tr>
<tr>
<td>$f&lt;$</td>
<td>(pre-image mapping of $f$)</td>
<td>03</td>
<td>12</td>
</tr>
<tr>
<td>$f^-$</td>
<td>(inverse of the mapping $f$)</td>
<td>03</td>
<td>11</td>
</tr>
<tr>
<td>$f^{on}$</td>
<td>($n$th iterate of the mapping $f$)</td>
<td>03</td>
<td>11</td>
</tr>
<tr>
<td>$f</td>
<td>A$</td>
<td>(restriction of $f$ to $A$)</td>
<td>03</td>
</tr>
<tr>
<td>$f</td>
<td>A^B$</td>
<td>(adjustment of $f$ range)</td>
<td>03</td>
</tr>
<tr>
<td>$f</td>
<td>A_{\text{tag}}$</td>
<td>(adjustment of $f$ to range)</td>
<td>03</td>
</tr>
<tr>
<td>$f</td>
<td>A_{A}$</td>
<td>($A$-adjustment of $f$ when $A$ is $f$-invariant)</td>
<td>03</td>
</tr>
<tr>
<td>$f</td>
<td>A_{(V)}$</td>
<td>(lineonic extension of $f$)</td>
<td>85</td>
</tr>
<tr>
<td>$c_{D-C}$</td>
<td>(constant with domain $D$, codomain $C$, and range ${c}$)</td>
<td>03</td>
<td>11</td>
</tr>
<tr>
<td>1$S$</td>
<td>(identity mapping of $S$)</td>
<td>03</td>
<td>11</td>
</tr>
<tr>
<td>1$U\subset S$</td>
<td>(inclusion mapping of $U$ into $S$)</td>
<td>03</td>
<td>11</td>
</tr>
<tr>
<td>$(a, \cdot), (\cdot, b)$</td>
<td>(“insertion” into a product of two sets)</td>
<td>04</td>
<td>18</td>
</tr>
<tr>
<td>$(c, j)$</td>
<td>(“insertion” into a product of a family of sets)</td>
<td>04</td>
<td>19</td>
</tr>
<tr>
<td>$\sharp S$</td>
<td>(cardinal of $S$)</td>
<td>05</td>
<td>20</td>
</tr>
<tr>
<td>$[x, y]$</td>
<td>(closed interval; segment joining the points $x$ and $y$)</td>
<td>08, 37</td>
<td>32, 123</td>
</tr>
<tr>
<td>$]x, y[$</td>
<td>(open interval; open segment joining the points $x$ and $y$)</td>
<td>08, 51</td>
<td>32, 163</td>
</tr>
<tr>
<td>Notation</td>
<td>Description</td>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>([a, b[, a, b])</td>
<td>(half-open intervals)</td>
<td>08</td>
<td>31</td>
</tr>
<tr>
<td>(\mathbb{R})</td>
<td>(extended-real-number set)</td>
<td>08</td>
<td>32</td>
</tr>
<tr>
<td>(\mathbb{P})</td>
<td>(extended-positive-number set)</td>
<td>08</td>
<td>32</td>
</tr>
<tr>
<td>(\nu)</td>
<td>(identity-mapping of (\mathbb{R}); “indeterminate”)</td>
<td>08, 92</td>
<td>34, 353</td>
</tr>
<tr>
<td>(\partial_t f)</td>
<td>(derivative of (f) at (t))</td>
<td>08, 61</td>
<td>34, 209</td>
</tr>
<tr>
<td>(\partial f)</td>
<td>(derivative-function of (f))</td>
<td>08, 61</td>
<td>34, 209</td>
</tr>
<tr>
<td>(f^*)</td>
<td>(derivative-function of (f))</td>
<td>08, 61</td>
<td>35, 210</td>
</tr>
<tr>
<td>(\partial^n f, f^{(n)})</td>
<td>(derivative of order (n))</td>
<td>08, 61</td>
<td>35, 209</td>
</tr>
<tr>
<td>(\nabla)</td>
<td>(gradient)</td>
<td>33, 63</td>
<td>108, 218</td>
</tr>
<tr>
<td>(\nabla(1), \nabla(2))</td>
<td>(partial gradients)</td>
<td>65</td>
<td>228, 229</td>
</tr>
<tr>
<td>(\varphi_1, \varphi_2)</td>
<td>(partial derivatives)</td>
<td>65</td>
<td>228, 229</td>
</tr>
<tr>
<td>(\nabla(j), \varphi_j)</td>
<td>(partial gradients and derivatives)</td>
<td>65</td>
<td>231</td>
</tr>
<tr>
<td>(\Delta)</td>
<td>(Laplacian)</td>
<td>67</td>
<td>241</td>
</tr>
<tr>
<td>(L_1 \oplus L_2)</td>
<td>(“evaluation-sum” of (L_1) and (L_2))</td>
<td>14</td>
<td>49</td>
</tr>
<tr>
<td>(\bigoplus(L_i</td>
<td>i \in I)</td>
<td>(“evaluation-sum” of a family of linear mappings)</td>
<td>14</td>
</tr>
<tr>
<td>(\delta I)</td>
<td>(standard basis of (F(I)))</td>
<td>16</td>
<td>55</td>
</tr>
<tr>
<td>(V^*)</td>
<td>(dual of the linear space (V))</td>
<td>21</td>
<td>71</td>
</tr>
<tr>
<td>(b^*)</td>
<td>(dual of the basis (b) )</td>
<td>23</td>
<td>78</td>
</tr>
<tr>
<td>(L^T)</td>
<td>(transpose of the linear mapping (L))</td>
<td>21</td>
<td>71</td>
</tr>
<tr>
<td>(S^\perp)</td>
<td>(annihilator of the set (S); orthogonal supplement)</td>
<td>21, 41</td>
<td>72, 137</td>
</tr>
<tr>
<td>(B^\sim)</td>
<td>(switch of the bilinear mapping (B))</td>
<td>24</td>
<td>83</td>
</tr>
<tr>
<td>(w \otimes \lambda)</td>
<td>(tensor product of (w) and (\lambda))</td>
<td>25</td>
<td>86</td>
</tr>
<tr>
<td>(\mathcal{S})</td>
<td>(quadratic form corresponding to the bilinear form (S))</td>
<td>27</td>
<td>94</td>
</tr>
<tr>
<td>(\mathcal{Q})</td>
<td>(bilinear form corresponding to the quadratic form (Q))</td>
<td>28</td>
<td>94</td>
</tr>
<tr>
<td>(x\overline{y})</td>
<td>(line passing through the points (x) and (y))</td>
<td>32</td>
<td>107</td>
</tr>
<tr>
<td>(\mathbf{v}^2)</td>
<td>(inner square of (v))</td>
<td>41</td>
<td>133</td>
</tr>
<tr>
<td>(u \cdot v)</td>
<td>(inner product of (u) and (v))</td>
<td>41</td>
<td>133</td>
</tr>
<tr>
<td>(\langle u</td>
<td>v \rangle)</td>
<td>(unitary product of (u) and (v))</td>
<td>89</td>
</tr>
<tr>
<td>(</td>
<td>v</td>
<td>)</td>
<td>(magnitude of (v))</td>
</tr>
<tr>
<td>(</td>
<td>L</td>
<td>_{\nu, \nu'})</td>
<td>(operator norm of (L) relative to (\nu, \nu'))</td>
</tr>
<tr>
<td>(</td>
<td>L</td>
<td>_{\nu})</td>
<td>(operator norm of the lineon (L) relative to (\nu))</td>
</tr>
<tr>
<td>(</td>
<td>L</td>
<td>)</td>
<td>(operator norm of (L) relative to magnitude)</td>
</tr>
<tr>
<td>([L]_{\mathbf{b}})</td>
<td>(matrix of the lineon (L) relative to the basis (\mathbf{b}))</td>
<td>18</td>
<td>63</td>
</tr>
<tr>
<td>([h]^c, [h]_{c}, [T]^c)</td>
<td>(components relative to a coordinate system)</td>
<td>71, 73</td>
<td>279, 289</td>
</tr>
</tbody>
</table>
## Index of Multiple-Letter Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>(set of accumulation points, of a set)</td>
<td>57</td>
<td>197</td>
</tr>
<tr>
<td>add</td>
<td>(addition mapping)</td>
<td>11</td>
<td>39</td>
</tr>
<tr>
<td>Aspec</td>
<td>(angle-spectrum, of a lineon)</td>
<td>88</td>
<td>333</td>
</tr>
<tr>
<td>Asps</td>
<td>(angle-spectral space, of a lineon)</td>
<td>88</td>
<td>333</td>
</tr>
<tr>
<td>Ball</td>
<td>(ball, in a genuine Euclidean space)</td>
<td>46</td>
<td>153</td>
</tr>
<tr>
<td>Bdy</td>
<td>(boundary, of a set)</td>
<td>53</td>
<td>179</td>
</tr>
<tr>
<td>Box</td>
<td>(norming box, determined by a basis)</td>
<td>51</td>
<td>168</td>
</tr>
<tr>
<td>Ce</td>
<td>(Norming cell, of a norm)</td>
<td>51</td>
<td>164</td>
</tr>
<tr>
<td>ch</td>
<td>(characteristic family or function, of a set)</td>
<td>02, 03</td>
<td>8, 10</td>
</tr>
<tr>
<td>chp</td>
<td>(characteristic polynomial, of a lineon)</td>
<td>95</td>
<td>370</td>
</tr>
<tr>
<td>Clo</td>
<td>(closure, of a set)</td>
<td>53</td>
<td>178</td>
</tr>
<tr>
<td>Cod</td>
<td>(codomain, of a mapping)</td>
<td>03</td>
<td>9</td>
</tr>
<tr>
<td>Comm</td>
<td>(commutant algebra, of a lineon)</td>
<td>18</td>
<td>62</td>
</tr>
<tr>
<td>Conf</td>
<td>(set of confined mappings)</td>
<td>62</td>
<td>213, 216</td>
</tr>
<tr>
<td>Curl</td>
<td>(curl, of a mapping)</td>
<td>611</td>
<td>261</td>
</tr>
<tr>
<td>cxc</td>
<td>(convex-combination mapping, of a family in a flat space)</td>
<td>37</td>
<td>124</td>
</tr>
<tr>
<td>Cxh</td>
<td>(convex hull, of a subset of a flat space)</td>
<td>37</td>
<td>123</td>
</tr>
<tr>
<td>dd</td>
<td>(directional derivative)</td>
<td>65</td>
<td>233</td>
</tr>
<tr>
<td>deg</td>
<td>(degree, of a polynomial)</td>
<td>92</td>
<td>353</td>
</tr>
<tr>
<td>det</td>
<td>(determinant)</td>
<td>73</td>
<td>287</td>
</tr>
<tr>
<td>diam</td>
<td>(diameter)</td>
<td>52</td>
<td>173</td>
</tr>
<tr>
<td>diff</td>
<td>(point-difference mapping)</td>
<td>32</td>
<td>103</td>
</tr>
<tr>
<td>dim</td>
<td>(dimension, of a linear space or a flat space)</td>
<td>17, 32</td>
<td>58, 107</td>
</tr>
<tr>
<td>div</td>
<td>(divergence)</td>
<td>67</td>
<td>239</td>
</tr>
<tr>
<td>Dmd</td>
<td>(norming diamond, determined by a basis)</td>
<td>51</td>
<td>168</td>
</tr>
<tr>
<td>Dom</td>
<td>(domain, of a mapping)</td>
<td>03</td>
<td>9</td>
</tr>
<tr>
<td>dst</td>
<td>(distance function, of a genuine Euclidean space)</td>
<td>46</td>
<td>152</td>
</tr>
<tr>
<td>Eis</td>
<td>(group of Euclidean automorphisms)</td>
<td>45</td>
<td>149</td>
</tr>
<tr>
<td>emult</td>
<td>(elementary multiplicity function, of a lineon)</td>
<td>95</td>
<td>370</td>
</tr>
<tr>
<td>ev</td>
<td>(evaluation, on a set-product or a set of mappings)</td>
<td>04, 22</td>
<td>16, 17, 74</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>exp</td>
<td>(exponential, lineonic exponential)</td>
<td>08, 612</td>
<td>34, 266</td>
</tr>
<tr>
<td>Fin</td>
<td>(set of all finite subsets, of a set)</td>
<td>05</td>
<td>21</td>
</tr>
<tr>
<td>Fis</td>
<td>(group of flat automorphisms)</td>
<td>33</td>
<td>111</td>
</tr>
<tr>
<td>flc</td>
<td>(flat combination mapping, of a family in a flat space)</td>
<td>35</td>
<td>116</td>
</tr>
<tr>
<td>Flf</td>
<td>(space of flat functions)</td>
<td>36</td>
<td>120</td>
</tr>
<tr>
<td>Fsp</td>
<td>(flat span, of a subset of a flat space)</td>
<td>32</td>
<td>107</td>
</tr>
<tr>
<td>Gr</td>
<td>(graph, of a mapping)</td>
<td>03</td>
<td>10</td>
</tr>
<tr>
<td>ind</td>
<td>(index, of an inner-product space)</td>
<td>47</td>
<td>157</td>
</tr>
<tr>
<td>Inj</td>
<td>(set of all injective mappings from a given set to another)</td>
<td>04</td>
<td>16</td>
</tr>
<tr>
<td>inf</td>
<td>(infinum, of a set)</td>
<td>08</td>
<td>32</td>
</tr>
<tr>
<td>ins</td>
<td>(insertion mapping)</td>
<td>14, 15</td>
<td>48, 52</td>
</tr>
<tr>
<td>Int</td>
<td>(interior, of a set)</td>
<td>53</td>
<td>177</td>
</tr>
<tr>
<td>inv</td>
<td>(inversion mapping)</td>
<td>68</td>
<td>246</td>
</tr>
<tr>
<td>ip</td>
<td>(inner-product)</td>
<td>41</td>
<td>133</td>
</tr>
<tr>
<td>Ker</td>
<td>(kernel, of a homomorphism)</td>
<td>06</td>
<td>24</td>
</tr>
<tr>
<td>lim</td>
<td>(limit)</td>
<td>08, 55, 57 34, 186, 198</td>
<td></td>
</tr>
<tr>
<td>Lin</td>
<td>(space of linear mappings, from a given linear space to another; algebra of lineons)</td>
<td>14, 18</td>
<td>47, 61</td>
</tr>
<tr>
<td>Lin₂</td>
<td>(space of bilinear mappings)</td>
<td>24</td>
<td>81</td>
</tr>
<tr>
<td>Lis</td>
<td>(set of linear isomorphism, from a given linear space to another; linear group)</td>
<td>14, 18</td>
<td>48, 62</td>
</tr>
<tr>
<td>lnc</td>
<td>(linear combination mapping, of a family in a linear space)</td>
<td>15</td>
<td>51</td>
</tr>
<tr>
<td>log</td>
<td>(lineonic logarithm)</td>
<td>85</td>
<td>320</td>
</tr>
<tr>
<td>lp</td>
<td>(polar decomposition, left positive part)</td>
<td>86</td>
<td>324</td>
</tr>
<tr>
<td>Lsp</td>
<td>(linear span, of a subset of a linear space)</td>
<td>12, 92</td>
<td>42, 355</td>
</tr>
<tr>
<td>Map</td>
<td>(set of all mappings, from a given set to another)</td>
<td>04</td>
<td>16</td>
</tr>
<tr>
<td>max</td>
<td>(maximum, of a set)</td>
<td>08</td>
<td>32</td>
</tr>
<tr>
<td>min</td>
<td>(minimum, of a set)</td>
<td>08</td>
<td>32</td>
</tr>
<tr>
<td>mult</td>
<td>(multiplicity function, of a lineon)</td>
<td>82, 810</td>
<td>307,340</td>
</tr>
<tr>
<td>Nhd</td>
<td>(collection of neighborhoods, of a point)</td>
<td>53</td>
<td>177</td>
</tr>
</tbody>
</table>
### Index of Multiple-Letter Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>(norm, of a norming cell)</td>
<td>51</td>
<td>165</td>
</tr>
<tr>
<td>Null</td>
<td>(nullspace, of a linear mapping)</td>
<td>13</td>
<td>46</td>
</tr>
<tr>
<td>opp</td>
<td>(opposition mapping)</td>
<td>11</td>
<td>39</td>
</tr>
<tr>
<td>or</td>
<td>(polar decomposition, orthogonal part)</td>
<td>86</td>
<td>324</td>
</tr>
<tr>
<td>Orth</td>
<td>(set of orthogonal mappings, from a given inner-product space to another; orthogonal group)</td>
<td>43</td>
<td>141, 142</td>
</tr>
<tr>
<td>Perm</td>
<td>(set of all permutations, of a given set)</td>
<td>04</td>
<td>16</td>
</tr>
<tr>
<td>Pos</td>
<td>(set of positive symmetric lineons)</td>
<td>85</td>
<td>316</td>
</tr>
<tr>
<td>Pos⁺</td>
<td>(set of strictly positive symmetric lineons)</td>
<td>85</td>
<td>316</td>
</tr>
<tr>
<td>pow</td>
<td>(lineonic power)</td>
<td>66</td>
<td>237</td>
</tr>
<tr>
<td>Pspec</td>
<td>(pair-spectrum, of a lineon)</td>
<td>88</td>
<td>330</td>
</tr>
<tr>
<td>Psps</td>
<td>(pair-spectral space, of a lineon)</td>
<td>88</td>
<td>330</td>
</tr>
<tr>
<td>Qu</td>
<td>(space of quadratic forms)</td>
<td>27</td>
<td>94</td>
</tr>
<tr>
<td>Qspec</td>
<td>(quasi-spectrum, of a lineon)</td>
<td>87</td>
<td>327</td>
</tr>
<tr>
<td>Qsps</td>
<td>(quasi-spectral space, of a lineon)</td>
<td>87</td>
<td>327</td>
</tr>
<tr>
<td>Rng</td>
<td>(range, of a family or a mapping)</td>
<td>02, 03</td>
<td>7, 10</td>
</tr>
<tr>
<td>rp</td>
<td>(polar decomposition, right positive part)</td>
<td>86</td>
<td>324</td>
</tr>
<tr>
<td>sep</td>
<td>(separation function, of a Euclidean space)</td>
<td>45</td>
<td>148</td>
</tr>
<tr>
<td>sgn</td>
<td>(sign-function)</td>
<td>08</td>
<td>32</td>
</tr>
<tr>
<td>sig</td>
<td>(signature, of an inner-product space)</td>
<td>47</td>
<td>155</td>
</tr>
<tr>
<td>Skew</td>
<td>(space of skew linear mappings or lineons)</td>
<td>27, 41</td>
<td>92, 135</td>
</tr>
<tr>
<td>Skew₂</td>
<td>(space of skew bilinear mappings)</td>
<td>24</td>
<td>83</td>
</tr>
<tr>
<td>sm</td>
<td>(scalar-multiple mapping)</td>
<td>11, 89</td>
<td>39, 335</td>
</tr>
<tr>
<td>Small</td>
<td>(set of small mappings)</td>
<td>62</td>
<td>212, 216</td>
</tr>
<tr>
<td>Spec</td>
<td>(spectrum, of a lineon)</td>
<td>82, 810</td>
<td>307, 340</td>
</tr>
<tr>
<td>Sph</td>
<td>(sphere, in a genuine Euclidean space)</td>
<td>46</td>
<td>153</td>
</tr>
<tr>
<td>Sps</td>
<td>(spectral space, of a lineon)</td>
<td>82, 810</td>
<td>307, 340</td>
</tr>
<tr>
<td>sq</td>
<td>(inner square)</td>
<td>41</td>
<td>133</td>
</tr>
<tr>
<td>sqrt</td>
<td>(lineonic square root)</td>
<td>85</td>
<td>317</td>
</tr>
<tr>
<td>sqrt⁺</td>
<td>(strict lineonic square root)</td>
<td>85</td>
<td>317</td>
</tr>
<tr>
<td>ssq</td>
<td>(sum-sequence, of a sequence)</td>
<td>08, 55</td>
<td>33, 191</td>
</tr>
<tr>
<td>str</td>
<td>(striction, of a mapping relative to given norms; absolute striction)</td>
<td>64</td>
<td>223, 227</td>
</tr>
</tbody>
</table>
**Index of Multiple-Letter Symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub</td>
<td>(subsetset, of a set)</td>
<td>01</td>
<td>3</td>
</tr>
<tr>
<td>sum</td>
<td>(summation mapping)</td>
<td>15</td>
<td>51</td>
</tr>
<tr>
<td>sup</td>
<td>(supremum of a set)</td>
<td>08</td>
<td>32</td>
</tr>
<tr>
<td>Supt</td>
<td>(support, of a family)</td>
<td>07</td>
<td>28</td>
</tr>
<tr>
<td>Sym</td>
<td>(space of symmetric linear mappings or lineons)</td>
<td>27, 41</td>
<td>92, 135</td>
</tr>
<tr>
<td>Sym_2</td>
<td>(space of symmetric bilinear mappings)</td>
<td>24</td>
<td>83</td>
</tr>
<tr>
<td>tr</td>
<td>(trace, of a lineon)</td>
<td>26</td>
<td>89</td>
</tr>
<tr>
<td>Ubl</td>
<td>(unit ball, in a genuine inner-product space)</td>
<td>42</td>
<td>140</td>
</tr>
<tr>
<td>Unit</td>
<td>(set of unitary mappings, from a given unitary space to another; unitary group)</td>
<td>89</td>
<td>339</td>
</tr>
<tr>
<td>Usph</td>
<td>(unit sphere, in a genuine inner-product space)</td>
<td>42</td>
<td>140</td>
</tr>
</tbody>
</table>
Index of Terminology

<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l^1$-norm</td>
<td>172</td>
</tr>
<tr>
<td>$l^\infty$-norm</td>
<td>172</td>
</tr>
<tr>
<td>abelian</td>
<td>26</td>
</tr>
<tr>
<td>absolute striction</td>
<td>227</td>
</tr>
<tr>
<td>absolute value</td>
<td>31</td>
</tr>
<tr>
<td>accumulation point</td>
<td>34, 197</td>
</tr>
<tr>
<td>action, of a group</td>
<td>101</td>
</tr>
<tr>
<td>addition</td>
<td>25, 39</td>
</tr>
<tr>
<td>additive decomposition</td>
<td>322</td>
</tr>
<tr>
<td>adjoint</td>
<td>74</td>
</tr>
<tr>
<td>adjustment, of a mapping</td>
<td>13</td>
</tr>
<tr>
<td>affine function</td>
<td>122</td>
</tr>
<tr>
<td>affine hull</td>
<td>108</td>
</tr>
<tr>
<td>affine mapping</td>
<td>112</td>
</tr>
<tr>
<td>affine space</td>
<td>108</td>
</tr>
<tr>
<td>affine subset</td>
<td>108</td>
</tr>
<tr>
<td>affine transformation</td>
<td>111</td>
</tr>
<tr>
<td>algebra of lineons</td>
<td>62</td>
</tr>
<tr>
<td>alternating</td>
<td>85</td>
</tr>
<tr>
<td>angle-spectrum</td>
<td>333</td>
</tr>
<tr>
<td>angle-spectrum space</td>
<td>333</td>
</tr>
<tr>
<td>annihilator</td>
<td>72</td>
</tr>
<tr>
<td>anti-Hermitian</td>
<td>340</td>
</tr>
<tr>
<td>anti-linear</td>
<td>340</td>
</tr>
<tr>
<td>anticommutate</td>
<td>70</td>
</tr>
<tr>
<td>antiderivative</td>
<td>36</td>
</tr>
<tr>
<td>antisymmetric</td>
<td>85</td>
</tr>
<tr>
<td>antitome</td>
<td>32</td>
</tr>
<tr>
<td>average</td>
<td>120</td>
</tr>
<tr>
<td>ball</td>
<td>152, 169, 172</td>
</tr>
<tr>
<td>barycenter</td>
<td>112, 116</td>
</tr>
<tr>
<td>barycentric coordinates</td>
<td>119</td>
</tr>
<tr>
<td>basis</td>
<td>52</td>
</tr>
<tr>
<td>big oh</td>
<td>217</td>
</tr>
<tr>
<td>bijection</td>
<td>11</td>
</tr>
<tr>
<td>bilinear form</td>
<td>92</td>
</tr>
<tr>
<td>bilinear mapping</td>
<td>80</td>
</tr>
<tr>
<td>Borel-Lebesgue Theorem</td>
<td>204</td>
</tr>
<tr>
<td>boundary</td>
<td>126, 179</td>
</tr>
<tr>
<td>bounded</td>
<td>32</td>
</tr>
<tr>
<td>bounded above [bounded below]</td>
<td>32</td>
</tr>
<tr>
<td>bounded sequence</td>
<td>187</td>
</tr>
<tr>
<td>bounded set</td>
<td>172</td>
</tr>
<tr>
<td>Bourbaki</td>
<td>306</td>
</tr>
<tr>
<td>box</td>
<td>168</td>
</tr>
<tr>
<td>Bunyakovsky’s inequality</td>
<td>141</td>
</tr>
<tr>
<td>canonical matrix</td>
<td>363</td>
</tr>
<tr>
<td>Carathéodory’s Theorem</td>
<td>126</td>
</tr>
<tr>
<td>cardinal</td>
<td>20</td>
</tr>
<tr>
<td>Cartesian Coordinates</td>
<td>290</td>
</tr>
<tr>
<td>Cartesian decomposition</td>
<td>85</td>
</tr>
<tr>
<td>Cartesian product</td>
<td>9</td>
</tr>
<tr>
<td>Cauchy Convergence Criterion</td>
<td>192</td>
</tr>
<tr>
<td>Cauchy sequence</td>
<td>192</td>
</tr>
<tr>
<td>Cauchy’s inequality</td>
<td>141</td>
</tr>
<tr>
<td>cell</td>
<td>163</td>
</tr>
<tr>
<td>Term</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>cell modelled on a norming cell</td>
<td>167</td>
</tr>
<tr>
<td>center of mass</td>
<td>116</td>
</tr>
<tr>
<td>centralizer</td>
<td>64</td>
</tr>
<tr>
<td>centroid</td>
<td>113, 116</td>
</tr>
<tr>
<td>characteristic family</td>
<td>8</td>
</tr>
<tr>
<td>characteristic function</td>
<td>10</td>
</tr>
<tr>
<td>characteristic polynomial</td>
<td>370</td>
</tr>
<tr>
<td>characteristic value</td>
<td>310</td>
</tr>
<tr>
<td>charge distribution</td>
<td>111</td>
</tr>
<tr>
<td>Christoffel symbols</td>
<td>288</td>
</tr>
<tr>
<td>class $C^0$</td>
<td>35, 209</td>
</tr>
<tr>
<td>class $C^\infty$</td>
<td>35, 209</td>
</tr>
<tr>
<td>class $C^1$</td>
<td>218</td>
</tr>
<tr>
<td>class $C^2$</td>
<td>218</td>
</tr>
<tr>
<td>closed ball</td>
<td>152</td>
</tr>
<tr>
<td>closed interval</td>
<td>32</td>
</tr>
<tr>
<td>closed set</td>
<td>179</td>
</tr>
<tr>
<td>closure</td>
<td>177</td>
</tr>
<tr>
<td>closure, of a cell</td>
<td>167</td>
</tr>
<tr>
<td>closure, of a norming cell</td>
<td>164</td>
</tr>
<tr>
<td>cluster point</td>
<td>33, 186</td>
</tr>
<tr>
<td>codomain, of a mapping</td>
<td>9</td>
</tr>
<tr>
<td>collection</td>
<td>3</td>
</tr>
<tr>
<td>column, of a matrix</td>
<td>8</td>
</tr>
<tr>
<td>combination, in a</td>
<td>22</td>
</tr>
<tr>
<td>pre-monoid</td>
<td>62</td>
</tr>
<tr>
<td>commutant-algebra</td>
<td>25</td>
</tr>
<tr>
<td>commutative</td>
<td>26</td>
</tr>
<tr>
<td>commutative ring</td>
<td>11</td>
</tr>
<tr>
<td>commute</td>
<td>199</td>
</tr>
<tr>
<td>compact</td>
<td>5</td>
</tr>
<tr>
<td>complement</td>
<td>44</td>
</tr>
<tr>
<td>complementary</td>
<td>336</td>
</tr>
<tr>
<td>complex dual space</td>
<td>340</td>
</tr>
<tr>
<td>complex inner-product</td>
<td>340</td>
</tr>
<tr>
<td>complex space</td>
<td>334</td>
</tr>
<tr>
<td>complexification</td>
<td>347</td>
</tr>
<tr>
<td>complexor</td>
<td>334</td>
</tr>
<tr>
<td>component</td>
<td>52</td>
</tr>
<tr>
<td>component-family</td>
<td>279</td>
</tr>
<tr>
<td>composite</td>
<td>11</td>
</tr>
<tr>
<td>Composition Rule</td>
<td>222</td>
</tr>
<tr>
<td>confined</td>
<td>213, 216</td>
</tr>
<tr>
<td>congruence</td>
<td>154</td>
</tr>
<tr>
<td>congruent</td>
<td>153</td>
</tr>
<tr>
<td>conjugate-complex structure</td>
<td>335</td>
</tr>
<tr>
<td>conjugate-linear mapping</td>
<td>336</td>
</tr>
<tr>
<td>connected</td>
<td>226</td>
</tr>
<tr>
<td>connection components</td>
<td>281</td>
</tr>
<tr>
<td>constant</td>
<td>11</td>
</tr>
<tr>
<td>constraint</td>
<td>252</td>
</tr>
<tr>
<td>constricted mapping</td>
<td>222</td>
</tr>
<tr>
<td>contained, in a set</td>
<td>3</td>
</tr>
<tr>
<td>continuous</td>
<td>34, 192</td>
</tr>
<tr>
<td>contraction</td>
<td>230</td>
</tr>
<tr>
<td>contravariant components</td>
<td>138, 289</td>
</tr>
<tr>
<td>convergence, of a sequence</td>
<td>33, 186</td>
</tr>
<tr>
<td>convex hull</td>
<td>123</td>
</tr>
<tr>
<td>convex set</td>
<td>123</td>
</tr>
<tr>
<td>convex-combination</td>
<td>124</td>
</tr>
<tr>
<td>coordinate</td>
<td>54</td>
</tr>
<tr>
<td>coordinate curve</td>
<td>280</td>
</tr>
<tr>
<td>coordinate system</td>
<td>277</td>
</tr>
<tr>
<td>coordinate transformation</td>
<td>281</td>
</tr>
<tr>
<td>cos</td>
<td>273</td>
</tr>
<tr>
<td>covariant components</td>
<td>138, 289</td>
</tr>
<tr>
<td>covariant derivatives</td>
<td>285</td>
</tr>
<tr>
<td>covector</td>
<td>71</td>
</tr>
<tr>
<td>covector field</td>
<td>279</td>
</tr>
<tr>
<td>cover</td>
<td>5, 199</td>
</tr>
<tr>
<td>cross-polytope</td>
<td>172</td>
</tr>
<tr>
<td>cross-power</td>
<td>18</td>
</tr>
<tr>
<td>cross-product, of a family of mappings</td>
<td>18</td>
</tr>
<tr>
<td>Term</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>curl</td>
<td>261</td>
</tr>
<tr>
<td>curvilinear coordinate system</td>
<td>280</td>
</tr>
<tr>
<td>cyclic lineon</td>
<td>362</td>
</tr>
<tr>
<td>cylindrical coordinates</td>
<td>293</td>
</tr>
<tr>
<td>D'Alembertian</td>
<td>241</td>
</tr>
<tr>
<td>decomposition, of a</td>
<td></td>
</tr>
<tr>
<td>linear space</td>
<td>303</td>
</tr>
<tr>
<td>definite quadratic form</td>
<td>92</td>
</tr>
<tr>
<td>degree, of a polynomial</td>
<td>275, 353</td>
</tr>
<tr>
<td>depth, of an elementary lineon</td>
<td>359</td>
</tr>
<tr>
<td>derivative</td>
<td>34, 209</td>
</tr>
<tr>
<td>determinant</td>
<td>287, 370</td>
</tr>
<tr>
<td>deviation components</td>
<td>281</td>
</tr>
<tr>
<td>diagonal matrix</td>
<td>8</td>
</tr>
<tr>
<td>diagonal, of a matrix</td>
<td>8</td>
</tr>
<tr>
<td>diameter</td>
<td>173</td>
</tr>
<tr>
<td>diamond</td>
<td>168</td>
</tr>
<tr>
<td>differentiable</td>
<td>34, 209, 218</td>
</tr>
<tr>
<td>differential</td>
<td>222</td>
</tr>
<tr>
<td>dimension</td>
<td>58, 107</td>
</tr>
<tr>
<td>Dini's Theorem</td>
<td>208</td>
</tr>
<tr>
<td>direct product</td>
<td>9</td>
</tr>
<tr>
<td>direct sum</td>
<td>44, 51, 306</td>
</tr>
<tr>
<td>direction space</td>
<td>106</td>
</tr>
<tr>
<td>directional derivative</td>
<td>233</td>
</tr>
<tr>
<td>disjoint</td>
<td>4, 5</td>
</tr>
<tr>
<td>disjunct</td>
<td>43, 303</td>
</tr>
<tr>
<td>distance</td>
<td>196</td>
</tr>
<tr>
<td>distance function</td>
<td>152</td>
</tr>
<tr>
<td>distributive law</td>
<td>25</td>
</tr>
<tr>
<td>divergence</td>
<td>239</td>
</tr>
<tr>
<td>domain, of a mapping</td>
<td>9</td>
</tr>
<tr>
<td>dot product</td>
<td>138</td>
</tr>
<tr>
<td>double-signed</td>
<td>94</td>
</tr>
<tr>
<td>doubleton</td>
<td>3</td>
</tr>
<tr>
<td>dual</td>
<td>74</td>
</tr>
<tr>
<td>dual basis field</td>
<td>278</td>
</tr>
<tr>
<td>Term</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>flat isomorphism</td>
<td>111</td>
</tr>
<tr>
<td>flat mapping</td>
<td>108</td>
</tr>
<tr>
<td>flat space</td>
<td>101</td>
</tr>
<tr>
<td>flat span</td>
<td>106</td>
</tr>
<tr>
<td>flatly-combination</td>
<td>117</td>
</tr>
<tr>
<td>flatly independent family</td>
<td>118</td>
</tr>
<tr>
<td>flatly spanning family</td>
<td>118</td>
</tr>
<tr>
<td>frame</td>
<td>120</td>
</tr>
<tr>
<td>free action</td>
<td>102</td>
</tr>
<tr>
<td>frontier</td>
<td>182</td>
</tr>
<tr>
<td>function</td>
<td>9</td>
</tr>
<tr>
<td>function of two variables</td>
<td>18</td>
</tr>
<tr>
<td>functional</td>
<td>9</td>
</tr>
<tr>
<td>fundamental sequence</td>
<td>192</td>
</tr>
<tr>
<td>Fundamental Theorem of Algebra</td>
<td>275</td>
</tr>
<tr>
<td>future-directed</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma symbols</td>
<td>285</td>
</tr>
<tr>
<td>Gaussian elimination</td>
<td>58</td>
</tr>
<tr>
<td>general linear group</td>
<td>64</td>
</tr>
<tr>
<td>genuine Euclidean space</td>
<td>152</td>
</tr>
<tr>
<td>genuine inner product</td>
<td>133</td>
</tr>
<tr>
<td>genuine interval</td>
<td>31</td>
</tr>
<tr>
<td>genuine unitary space</td>
<td>337</td>
</tr>
<tr>
<td>genuinely orthonormal</td>
<td>136</td>
</tr>
<tr>
<td>gradient</td>
<td>218</td>
</tr>
<tr>
<td>Gram-Schmidt</td>
<td>159</td>
</tr>
<tr>
<td>orthogonization</td>
<td></td>
</tr>
<tr>
<td>graph, of a mapping</td>
<td>10</td>
</tr>
<tr>
<td>group</td>
<td>22</td>
</tr>
<tr>
<td>group-isomorphism</td>
<td>24</td>
</tr>
<tr>
<td>groupable</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>half-line</td>
<td>32</td>
</tr>
<tr>
<td>half-open interval</td>
<td>32</td>
</tr>
<tr>
<td>half-space</td>
<td>126</td>
</tr>
<tr>
<td>harmonic function</td>
<td>241</td>
</tr>
<tr>
<td>Heine-Borel Theorem</td>
<td>204</td>
</tr>
<tr>
<td>Hermitian</td>
<td>340</td>
</tr>
<tr>
<td>homomorphism</td>
<td>24</td>
</tr>
</tbody>
</table>
# Index of Terminology

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Page</th>
<th>Definition</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>isotone</td>
<td>32</td>
<td>linear span</td>
<td>42</td>
</tr>
<tr>
<td>isotropic</td>
<td>158</td>
<td>linear transformation</td>
<td>46, 64</td>
</tr>
<tr>
<td>iterate</td>
<td>11</td>
<td>linearly dependent family</td>
<td>52</td>
</tr>
<tr>
<td>Jacobian</td>
<td>251</td>
<td>lineon</td>
<td>61</td>
</tr>
<tr>
<td>Jordan matrix</td>
<td>367</td>
<td>lineon field</td>
<td>279</td>
</tr>
<tr>
<td>Jordan-algebra</td>
<td>98</td>
<td>lineon-group</td>
<td>62</td>
</tr>
<tr>
<td>kernel</td>
<td>24, 46</td>
<td>lineon nth power</td>
<td>237, 317</td>
</tr>
<tr>
<td>Kronecker delta</td>
<td>58</td>
<td>lineonic polynomial function</td>
<td>354</td>
</tr>
<tr>
<td>(\ell_\infty)-norm</td>
<td>172</td>
<td>Lipschitz number</td>
<td>228</td>
</tr>
<tr>
<td>(\ell_1)-norm</td>
<td>172</td>
<td>list</td>
<td>7</td>
</tr>
<tr>
<td>L-space</td>
<td>62, 307</td>
<td>local inverse</td>
<td>243</td>
</tr>
<tr>
<td>Lagrange multipliers</td>
<td>254</td>
<td>local maximum</td>
<td></td>
</tr>
<tr>
<td>Laplacian</td>
<td>241</td>
<td>[local minimum]</td>
<td>251</td>
</tr>
<tr>
<td>latent root</td>
<td>310</td>
<td>locally invertible</td>
<td>243</td>
</tr>
<tr>
<td>Lebesgue number</td>
<td>204</td>
<td>locally uniform convergence</td>
<td>190</td>
</tr>
<tr>
<td>Lebesgue’s Covering Lemma</td>
<td>204</td>
<td>logarithm, lineonic</td>
<td>320</td>
</tr>
<tr>
<td>left-inverse</td>
<td>11</td>
<td>Lorentz group</td>
<td>143</td>
</tr>
<tr>
<td>left-multiplication mapping</td>
<td>67</td>
<td>lower bound</td>
<td>32</td>
</tr>
<tr>
<td>length, of a list</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lie-algebra</td>
<td>98</td>
<td>magnitude, of a vector</td>
<td>139</td>
</tr>
<tr>
<td>limit point</td>
<td>37</td>
<td>map</td>
<td>10</td>
</tr>
<tr>
<td>limit, of a function</td>
<td>34</td>
<td>mapping</td>
<td>9</td>
</tr>
<tr>
<td>limit, of a mapping</td>
<td>198</td>
<td>mass-point</td>
<td>113</td>
</tr>
<tr>
<td>limit, of a sequence</td>
<td>186</td>
<td>matrix</td>
<td>8, 55, 63</td>
</tr>
<tr>
<td>limit, of a sequence</td>
<td>33</td>
<td>matrix, of a bilinear form</td>
<td>93</td>
</tr>
<tr>
<td>line</td>
<td>107</td>
<td>matrix-algebra</td>
<td>64</td>
</tr>
<tr>
<td>linear (L)-span</td>
<td>355</td>
<td>maximal (L)-space</td>
<td>349</td>
</tr>
<tr>
<td>linear combination</td>
<td>51</td>
<td>maximum [minimum]</td>
<td>32, 35</td>
</tr>
<tr>
<td>linear form</td>
<td>71</td>
<td>Mean-Value Theorem</td>
<td>37</td>
</tr>
<tr>
<td>linear function</td>
<td>122</td>
<td>median</td>
<td>114</td>
</tr>
<tr>
<td>linear functional</td>
<td>74</td>
<td>member-wise difference</td>
<td>25</td>
</tr>
<tr>
<td>linear isomorphism</td>
<td>45</td>
<td>member-wise opposite</td>
<td>25</td>
</tr>
<tr>
<td>linear manifold</td>
<td>108</td>
<td>member-wise product</td>
<td>24</td>
</tr>
<tr>
<td>linear mapping</td>
<td>44</td>
<td>member-wise reciprocal</td>
<td>25</td>
</tr>
<tr>
<td>linear operator</td>
<td>46</td>
<td>member-wise sum</td>
<td>25, 29</td>
</tr>
<tr>
<td>linear operator</td>
<td>64</td>
<td>midpoint</td>
<td>114, 163</td>
</tr>
<tr>
<td>linear space</td>
<td>39</td>
<td>minimal (L)-space</td>
<td>349</td>
</tr>
<tr>
<td>Term</td>
<td>Page</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>minimal polynomial</td>
<td>356</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mixed components</td>
<td>289</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monic polynomial</td>
<td>353</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monoid</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monoidable</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>multiple</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>multiplication</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>multiplication, of lineons</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>multiplicative group</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>multiplicity, of a spectral value</td>
<td>307</td>
<td></td>
<td></td>
</tr>
<tr>
<td>negative</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>negative quadratic or bilinear form</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>negative-regular neighborhood</td>
<td>155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>neutral, of a monoid</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>neutral-subgroup</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>neutrality law</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nilpotency</td>
<td>69, 362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nilpotent lineon</td>
<td>69, 362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-degenerate quadratic or bilinear form</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-isotropic subspace</td>
<td>138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-singular subspace</td>
<td>138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>norm</td>
<td>164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>normal lineon</td>
<td>334</td>
<td></td>
<td></td>
</tr>
<tr>
<td>normal box</td>
<td>167</td>
<td></td>
<td></td>
</tr>
<tr>
<td>norming cell</td>
<td>163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>norming diamond</td>
<td>168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>null set</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nullity</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nullspace</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>one-to-one</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>one-to-one correspondence</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>onto</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>open interval</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>open segment</td>
<td>163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>open set</td>
<td>179, 192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>open-half-space</td>
<td>126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>operator</td>
<td>9, 64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>operator norm</td>
<td>174</td>
<td></td>
<td></td>
</tr>
<tr>
<td>opposite</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>opposition</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orbit, under an action</td>
<td>102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>origin</td>
<td>41, 280</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthogonal</td>
<td>134, 310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthogonal family of subspaces</td>
<td>310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthogonal group</td>
<td>142</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthogonal isomorphism</td>
<td>141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthogonal lineon</td>
<td>142</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthogonal mapping</td>
<td>141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthogonal projection</td>
<td>138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthogonal supplement</td>
<td>137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthonormal</td>
<td>136</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pair</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pair-spectral space</td>
<td>330</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pair-spectrum</td>
<td>330</td>
<td></td>
<td></td>
</tr>
<tr>
<td>paraboloidal coordinates</td>
<td>301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>parallel flats</td>
<td>106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>parallelepiped</td>
<td>172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>parallelogram</td>
<td>130, 172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>parallelogram law</td>
<td>160</td>
<td></td>
<td></td>
</tr>
<tr>
<td>parallelope</td>
<td>172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>partial derivative</td>
<td>228, 231</td>
<td></td>
<td></td>
</tr>
<tr>
<td>partial-gradient</td>
<td>228, 231</td>
<td></td>
<td></td>
</tr>
<tr>
<td>partition</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>past-directed</td>
<td>161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>perfect field</td>
<td>374</td>
<td></td>
<td></td>
</tr>
<tr>
<td>permutation</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>permutation-group</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>perpendicular projection</td>
<td>137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>perpendicular turn</td>
<td>290, 326</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Page</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>physical components</td>
<td>299</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pieces, of a partition</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>plane</td>
<td>107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>point</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>point-difference</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>point-spectrum</td>
<td>310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polar Coordinates</td>
<td>290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>polar decomposition</td>
<td>322</td>
<td></td>
<td></td>
</tr>
<tr>
<td>polynomial</td>
<td>275, 353</td>
<td></td>
<td></td>
</tr>
<tr>
<td>polynomial function</td>
<td>275, 354</td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive definite</td>
<td>96, 321</td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive quadratic or bilinear form</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive semidefinite</td>
<td>96, 321</td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive symmetric lineons</td>
<td>316</td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive-regular</td>
<td>155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>power set</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>power-space</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-image mapping</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-monoid</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-ring</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>primary decomposition</td>
<td>352</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prime polynomial</td>
<td>356</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prime polynomial, of an elementary lineon</td>
<td>359</td>
<td></td>
<td></td>
</tr>
<tr>
<td>principal direction, of an ellipsoid</td>
<td>344</td>
<td></td>
<td></td>
</tr>
<tr>
<td>process</td>
<td>209</td>
<td></td>
<td></td>
</tr>
<tr>
<td>product</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>product, of a family</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>product, of lineons</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>product-space</td>
<td>48, 49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>projection</td>
<td>19, 64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>projections, family of</td>
<td>306</td>
<td></td>
<td></td>
</tr>
<tr>
<td>proper number</td>
<td>310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>proper subset</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pseudo-Euclidean space</td>
<td>151</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quadratic form</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quadruple</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quasi-spectral space</td>
<td>327</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quasi-spectrum</td>
<td>327</td>
<td></td>
<td></td>
</tr>
<tr>
<td>range, of a family</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>range, of a mapping</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rank</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>real part, of a $\mathbb{C}$-lineon</td>
<td>338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reciprocal</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reflection</td>
<td>334</td>
<td></td>
<td></td>
</tr>
<tr>
<td>regular subspace</td>
<td>136, 148, 155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relativity</td>
<td>158</td>
<td></td>
<td></td>
</tr>
<tr>
<td>resultant, of a mapping</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>resultant force</td>
<td>115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reversion law</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reversion, in a group</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>right-inverse</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ring</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rng</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rotation</td>
<td>334</td>
<td></td>
<td></td>
</tr>
<tr>
<td>row, of a matrix</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scalar field</td>
<td>279</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scalar product</td>
<td>138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scale</td>
<td>167</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schwarz’s inequality</td>
<td>141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>second dual</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>second gradient</td>
<td>218</td>
<td></td>
<td></td>
</tr>
<tr>
<td>secular value</td>
<td>310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>segment</td>
<td>123</td>
<td></td>
<td></td>
</tr>
<tr>
<td>self-adjoint</td>
<td>340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>self-indexed family</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>semi-axes, of an ellipsoid</td>
<td>344</td>
<td></td>
<td></td>
</tr>
<tr>
<td>semi-linear</td>
<td>340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>semi-simple lineon</td>
<td>352</td>
<td></td>
<td></td>
</tr>
<tr>
<td>semigroup</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>separation function</td>
<td>148</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separation Theorem</td>
<td>129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sequence</td>
<td>8, 186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>series</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>set-difference</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Page 1</td>
<td>Page 2</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>set-power</td>
<td>7, 8</td>
<td></td>
<td>standard basis</td>
</tr>
<tr>
<td>set-product</td>
<td>8, 15</td>
<td></td>
<td>stream-function</td>
</tr>
<tr>
<td>set-square</td>
<td>8</td>
<td></td>
<td>striction</td>
</tr>
<tr>
<td>sign</td>
<td>31</td>
<td></td>
<td>strictly antitone</td>
</tr>
<tr>
<td>signal-like</td>
<td>158</td>
<td></td>
<td>strictly isotone</td>
</tr>
<tr>
<td>signature</td>
<td>155</td>
<td></td>
<td>strictly negative</td>
</tr>
<tr>
<td>similar lineons</td>
<td>367</td>
<td></td>
<td>strictly negative quadratic or</td>
</tr>
<tr>
<td>simple lineon</td>
<td>350</td>
<td></td>
<td>bilinear form</td>
</tr>
<tr>
<td>simply connected</td>
<td>263</td>
<td></td>
<td>strictly positive</td>
</tr>
<tr>
<td>sin</td>
<td>273</td>
<td></td>
<td>strictly positive quadratic or</td>
</tr>
<tr>
<td>single-signed</td>
<td>94</td>
<td></td>
<td>bilinear form</td>
</tr>
<tr>
<td>singlet</td>
<td>7</td>
<td></td>
<td>strictly positive symmetric</td>
</tr>
<tr>
<td>singleton</td>
<td>3</td>
<td></td>
<td>lineons</td>
</tr>
<tr>
<td>singleton-partition</td>
<td>6</td>
<td></td>
<td>strip</td>
</tr>
<tr>
<td>singular matrix</td>
<td>58</td>
<td></td>
<td>sub-pre-monoid</td>
</tr>
<tr>
<td>singular subspace</td>
<td>155</td>
<td></td>
<td>subgroup</td>
</tr>
<tr>
<td>skew bilinear mapping</td>
<td>83</td>
<td></td>
<td>submonoid</td>
</tr>
<tr>
<td>skew lineon</td>
<td>135</td>
<td></td>
<td>subring</td>
</tr>
<tr>
<td>skew-Hermitian</td>
<td>340</td>
<td></td>
<td>subsequence</td>
</tr>
<tr>
<td>skewsymmetric</td>
<td>85</td>
<td></td>
<td>subspace</td>
</tr>
<tr>
<td>small mapping</td>
<td>212, 216</td>
<td></td>
<td>subspace generated by</td>
</tr>
<tr>
<td>small oh</td>
<td>217</td>
<td></td>
<td>sum</td>
</tr>
<tr>
<td>space of linear mappings</td>
<td>47</td>
<td></td>
<td>sum-sequence</td>
</tr>
<tr>
<td>space-like</td>
<td>158</td>
<td></td>
<td>summation mapping</td>
</tr>
<tr>
<td>span-mapping</td>
<td>13</td>
<td></td>
<td>supplement</td>
</tr>
<tr>
<td>spanned</td>
<td>42</td>
<td></td>
<td>supplementary</td>
</tr>
<tr>
<td>spanning family</td>
<td>52</td>
<td></td>
<td>supremum</td>
</tr>
<tr>
<td>spectral idempotents</td>
<td>313</td>
<td></td>
<td>surjective mapping</td>
</tr>
<tr>
<td>spectral radius</td>
<td>228</td>
<td></td>
<td>Sylvester’s Law</td>
</tr>
<tr>
<td>spectral resolution</td>
<td>313</td>
<td></td>
<td>symmetric bilinear mapping</td>
</tr>
<tr>
<td>spectral space</td>
<td>307, 340, 345</td>
<td></td>
<td>symmetric lineon</td>
</tr>
<tr>
<td>spectral value</td>
<td>307</td>
<td></td>
<td>symmetric matrix</td>
</tr>
<tr>
<td>spectral vector</td>
<td>307</td>
<td></td>
<td>symmetry-group</td>
</tr>
<tr>
<td>spectrum</td>
<td>307, 340, 362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sphere</td>
<td>152</td>
<td></td>
<td>tangent</td>
</tr>
<tr>
<td>Spherical Coordinates</td>
<td>295</td>
<td></td>
<td>tensor</td>
</tr>
<tr>
<td>square matrix</td>
<td>8</td>
<td></td>
<td>tensor product</td>
</tr>
<tr>
<td>square root, lineonic</td>
<td>317</td>
<td></td>
<td>tensor product space</td>
</tr>
<tr>
<td>square root, strict lineonic</td>
<td>318</td>
<td></td>
<td>term, of a family</td>
</tr>
</tbody>
</table>
## Index of Terminology

<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
<th>Related Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>term-wise evaluation</td>
<td>17</td>
<td>value-wise n’th power</td>
</tr>
<tr>
<td>tetrahedron</td>
<td>114</td>
<td>value-wise opposite</td>
</tr>
<tr>
<td>three-index symbols</td>
<td>285</td>
<td>value-wise pair formation</td>
</tr>
<tr>
<td>time-like</td>
<td>158</td>
<td>value-wise product</td>
</tr>
<tr>
<td>total charge</td>
<td>111</td>
<td>value-wise sum</td>
</tr>
<tr>
<td>totally isotropic</td>
<td>158</td>
<td>vector</td>
</tr>
<tr>
<td>totally singular</td>
<td>155</td>
<td>Vector Analysis</td>
</tr>
<tr>
<td>trace</td>
<td>89, 111</td>
<td>vector field</td>
</tr>
<tr>
<td>transformation</td>
<td>9</td>
<td>vector space</td>
</tr>
<tr>
<td>transformation-monoid</td>
<td>23</td>
<td>vector-curl</td>
</tr>
<tr>
<td>transitive action</td>
<td>102</td>
<td>void</td>
</tr>
<tr>
<td>translated subspace</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>translation</td>
<td>103</td>
<td>Wave-Operator</td>
</tr>
<tr>
<td>translation space</td>
<td>103</td>
<td>weak boundedness</td>
</tr>
<tr>
<td>transpose</td>
<td>8</td>
<td>Weierstrass Comparison Test</td>
</tr>
<tr>
<td>transpose, of a linear mapping</td>
<td>71</td>
<td>weight</td>
</tr>
<tr>
<td>transposition</td>
<td>72</td>
<td>world-vector</td>
</tr>
<tr>
<td>triple</td>
<td>8</td>
<td>zero</td>
</tr>
<tr>
<td>trivial partition</td>
<td>6</td>
<td>zero-mapping</td>
</tr>
<tr>
<td>uniformly continuous</td>
<td>194</td>
<td>zero-space</td>
</tr>
<tr>
<td>union</td>
<td>4, 5, 15</td>
<td></td>
</tr>
<tr>
<td>unit</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>unit ball</td>
<td>140, 172</td>
<td></td>
</tr>
<tr>
<td>unit cube</td>
<td>323</td>
<td></td>
</tr>
<tr>
<td>unit matrix</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>unit sphere</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>unit vector</td>
<td>139</td>
<td></td>
</tr>
<tr>
<td>unitary group</td>
<td>339</td>
<td></td>
</tr>
<tr>
<td>unitary isomorphism</td>
<td>339</td>
<td></td>
</tr>
<tr>
<td>unitary mapping</td>
<td>339</td>
<td></td>
</tr>
<tr>
<td>unitary product</td>
<td>337</td>
<td></td>
</tr>
<tr>
<td>unitary space</td>
<td>337</td>
<td></td>
</tr>
<tr>
<td>unity</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>upper bound</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>value, of a mapping</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>value-wise absolute value</td>
<td>33</td>
<td></td>
</tr>
</tbody>
</table>