Solid Mechanics and Its Applications

Uwe Mühlich

Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds

Springer

Solid Mechanics and Its Applications

Volume 230

Aims and Scope of the Series

The fundamental questions arising in mechanics are: *Why*?, *How*?, and *How much*? The aim of this series is to provide lucid accounts written by authoritative researchers giving vision and insight in answering these questions on the subject of mechanics as it relates to solids.

The scope of the series covers the entire spectrum of solid mechanics. Thus it includes the foundation of mechanics; variational formulations; computational mechanics; statics, kinematics and dynamics of rigid and elastic bodies: vibrations of solids and structures; dynamical systems and chaos; the theories of elasticity, plasticity and viscoelasticity; composite materials; rods, beams, shells and membranes; structural control and stability; soils, rocks and geomechanics; fracture; tribology; experimental mechanics; biomechanics and machine design.

The median level of presentation is to the first year graduate student. Some texts are monographs defining the current state of the field; others are accessible to final year undergraduates; but essentially the emphasis is on readability and clarity.

More information about this series at http://www.springer.com/series/6557

Uwe Mühlich

Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds

Springer

Uwe Mühlich
Faculty of Applied Engineering
University of Antwerp
Antwerp
Belgium

ISSN 0925-0042 ISSN 2214-7764 (electronic)
Solid Mechanics and Its Applications
ISBN 978-3-319-56263-6 ISBN 978-3-319-56264-3 (eBook)
DOI 10.1007/978-3-319-56264-3

Library of Congress Control Number: 2017935015

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Continuum mechanics and its respective subtopics such as strength of materials, theory of elasticity, and plasticity are of utmost importance for mechanical and civil engineers. Tensors of different types, such as vectors and forms, appear most naturally in this context. Ultimately, when it comes to large inelastic deformations, operations like push-forward, pull-back, covariant derivative, and Lie derivative become inevitable. The latter form a part of modern differential geometry, also known as tensor calculus on differentiable manifolds.

Unfortunately, in many academic institutions, an engineering education still relies on conventional vector calculus and concepts like dual vector space, and exterior algebra are successfully ignored. The expression "manifold" arises more or less as a fancy but rather diffuse technical term. Analysis on manifolds is only mastered and applied by a very limited number of engineers. However, the manifold concept has been established now for decades, not only in physics but, at least in parts, also in certain disciplines of structural mechanics like theory of shells. Over the years, this has caused a large gap between the knowledge provided to engineering students and the knowledge required to master the challenges, continuum mechanics faces today.

The objective of this book is to decrease this gap. But, as the title already indicates, it does not aim to give a comprehensive introduction to smooth manifolds. On the contrary, at most it opens the door by presenting fundamental concepts of analysis in Euclidean space in a way which makes the transition to smooth manifolds as natural as possible.

The book is based on the lecture notes of an elective course on tensor calculus taught at TU-Bergakademie Freiberg. The audience consisted of master students in Mechanical Engineering and Computational Materials Science, as well as doctoral students of the Faculty of Mechanical, Process and Energy Engineering. This introductory text has a special focus on those aspects of which a thorough understanding is crucial for applying tensor calculus safely in Euclidean space, particularly for understanding the very essence of the manifold concept. Mathematical proofs are omitted not only because they are beyond the scope of the book but also because the author is an engineer and not a mathematician. Only in some particular cases are proofs sketched in order to raise awareness of the effort made by mathematicians to work out the tools we are using today. In most cases, however, the interested reader is referred to corresponding literature. Furthermore, invariants, isotropic tensor functions, etc., are not discussed, since these subjects can be found in many standard textbooks on continuum mechanics or tensor calculus.

Prior knowledge in real analysis, i.e., analysis in $\mathbb{R}$, is assumed. Furthermore, students should have a prior education in undergraduate engineering mechanics, including statics and strength of materials. The latter is surely helpful for understanding the differences between the traditional and modern version of tensor calculus.

Antwerp, Belgium
February 2017

Uwe Mühlich

Acknowledgements

I would like to express my gratitude to my former colleagues of the solid mechanics group of the Institute of Mechanics and Fluid Dynamics at TU-Bergakademie Freiberg, and in particular to Michael Budnitzki, Thomas Linse, and Christoph Häusler. Without their open-minded interest in the subject and their constant readiness to discuss it, this book would not exist. I am especially grateful to Heinz Gründemann for his support and inspiring discussions about specific topics related to the subject. In this context, I would also like to thank Jarno Hein, who brought us into contact.

Additionally, this book could not have been accomplished without the education I received in engineering mechanics, and I would like to take this opportunity to express my sincere gratitude to my teachers, Reinhold Kienzler and Wolfgang Brocks.

I would like to thank TU-Bergakademie Freiberg and especially Meinhard Kuna for the opportunity to teach a course about tensor calculus and for providing the freedom to teach a continuum mechanics lecture during the summer term of 2014 which deviated significantly from the approach commonly applied in engineering education. In addition, I would like to express my appreciation to all students and colleagues who attended the courses mentioned above for their interest and academic curiosity, but especially for critical remarks which helped to uncover mistakes and to clarify doubts.

Besides, I would like to acknowledge the support provided by the Faculty of Applied Engineering of the University of Antwerp in terms of infrastructure and time necessary to finalize the process of converting the course notes into a book.

Furthermore, I would like to thank Nathalie Jacobs, Almitra Gosh, Naga Kumar Natti, Henry Pravin, and Cynthia Feenstra from Springer Scientific Publishing for their support throughout the process of publishing this book. My thanks are also due to Marc Beschler for language editing the book.

Contents

Selected Symbols

$\mathbb{N}$	Set of natural numbers
$\mathbb{Z}$	Set of integer numbers
$\mathbb{Q}$	Set of rational numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\cap$	Intersection of sets: $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$
$\cup$	Union of sets: $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$
$\subset$	Subset $A \subset B$
A^{c}	Complement of a set A
$A \to B$	Mapping from set A to set B
$\mathbf{u}$, $\underset{\sim}{\boldsymbol{\omega}}$	Vector, dual vector
$\oplus$, $\boxplus$, $+$	Symbols used to indicate addition of elements of a vector space
$\odot$, $\boxdot$	Symbols used to indicate multiplication of elements of a vector space by a real number
$\dotplus$	Hybrid addition between points and vectors
$\wedge$	Exterior product
$*\mathbf{u}$	Star operator applied to a vector
$g_\star, g^\star$	Push forward and pull back operation under a mapping g
$\top$	General tensor
δ_i^k	Kronecker symbol, see (3.9)
$\underset{\sim}{\boldsymbol{d}}\vert_p f(\mathbf{u})$	Directional derivative of the function f at point p in direction $\mathbf{u}$
$\boldsymbol{\partial}_i\vert_p$	Tangent vector at a point p on a smooth manifold
$\underset{\sim}{\boldsymbol{\partial}}^i\vert_p$	Cotangent vector at a point p on a smooth manifold

$\frac{\partial}{\partial \boldsymbol{x}^i}\Big|_{\mu(p)}$ Tangent vector at the image of point p in a chart generated by a mapping μ with coordinates x^i

$\underset{\sim}{\boldsymbol{d}}\big|_{\mu(p)} x^i$ Cotangent vector at the image of point p in a chart generated by a mapping μ with coordinates x^i

Chapter 1
Introduction

Abstract The introduction aims to remind the reader that engineering mechanics is derived from classical mechanics, which is a discipline of general physics. Therefore, engineering mechanics also relies on a proper model for space, and the relations between space and geometry are discussed briefly. The idea of expressing geometrical concepts by means of linear algebra is sketched together with the concept of vectors as geometrical objects. Although this book provides only the very first steps of the manifold concept, this chapter intends to make its importance for modern continuum mechanics clear by raising a number of questions which cannot be answered by the conventional approach. Furthermore, aspects regarding mathematical notation used in subsequent chapters are discussed briefly.

1.1 Space, Geometry, and Linear Algebra

One of the most important tools for mechanical and civil engineers is certainly mechanics, and in particular continuum mechanics, together with its subtopics, linear and nonlinear elasticity, plasticity, etc. Mechanics is about the motion of bodies in space. Therefore, a theory of mechanics first requires concepts of space and time, particularly space–time, and, in addition, a concept to make the notion of a material body more precise. Traditionally, the scientific discipline concerned with the formulation of models for space is geometry. Starting from a continuum perspective of space–time, the simplest model one can think of is obtained by assuming euclidean geometry for space and by treating time separately as a simple scalar parameter.

In two-dimensional euclidean geometry, points and straight lines are primitive objects and the theory of euclidean space is based on a series of axioms. The most crucial one is the fifth postulate, which distinguishes euclidean geometry from other possible geometries. It introduces a particular notion of parallelism, and one way to express this in two dimensions is as follows. Given a straight line L, there is only one straight line through a point P not on L which never meets L.

It took mankind centuries to understand that euclidean geometry is only one of many possible geometries, see e.g., Holme [2] and BBC [1]. But, once this fundamental fact had been discovered, it became apparent rather quickly that the question

U. Mühlich, *Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds*, Solid Mechanics and Its Applications 230,
DOI 10.1007/978-3-319-56264-3_1

as to which space or space–time model is appropriate for developing a particular theory of physics can only be answered through experiments and not just through logical reasoning.

Historically, mechanics has been developed based on the assumption that euclidean space is an appropriate model for our physical space, i.e., the space we live in performing observations and measurements. The idea of force as an abstract concept of what causes a change in the motion of a body was a cornerstone of this development. However, at the beginning of the twentieth century, it became apparent to theoretical physicists that physical space is actually curved, and gravitation is seen nowadays as the cause for space curvature. Models for space, other than euclidean geometry, had to be employed in order to develop the corresponding theories.

In the course of this development, it became clear as well that new mathematical tools are required in order to encode new physical ideas properly. Tensor analysis using Ricci calculus was elaborated, Grassmann algebra was rediscovered, the differentiable manifold concept gained more and more use in theoretical physics, etc. Nowadays, these tools are matured and rather well established. However, the process of using them for reformulating existing theories in order to gain deeper understanding, is still going on.

Differentiable manifolds are by definition objects which are at least locally euclidean. Therefore, affine and euclidean geometry are crucial for the treatment of more general differentiable manifolds. Although linear algebra is not the same as euclidean geometry, the former provides concepts which can be used to formulate euclidean geometry in an elegant way by employing the vector space concept. A straight line in the language of linear algebra is a point combined with a the scalar multiple of a vector. Euclidean parallelism is expressed by the notion of linear dependence of vectors, etc.

1.2 Vectors as Geometrical Objects

Most of us were confronted with vectors for the first time in the following way. Starting with the definition that a vector $\mathbf{a}$ is an object which possesses direction and magnitude, this object is visualized by taking, e.g., a sheet of paper, drawing a straight line of some length, related to the magnitude, and indicating the direction by an arrow. Afterwards, it is agreed upon that two vectors are equal if they have same length and direction. A further step consists in the definition of two operations, namely the addition of vectors and multiplication of a vector with a real number. At this stage, these operations are defined only graphically, which can be expressed by the following definitions.

Definition 1.1 (*Addition of two vectors (graphically)*) In order to determine the sum $\mathbf{c}$ of two vectors $\mathbf{a}$ and $\mathbf{b}$ graphically, written as $\mathbf{c} = \mathbf{a} \oplus \mathbf{b}$

1. displace $\mathbf{a}$ and $\mathbf{b}$ in parallel such that the head of $\mathbf{a}$ meets the tail of $\mathbf{b}$ and
2. $\mathbf{c}$ is the directed line from the tail of $\mathbf{a}$ to the head of $\mathbf{b}$,

where displace in parallel means to move the corresponding vector without changing its length or direction.

Definition 1.2 (*Multiplication of a vector with a real number (graphically)*) The multiplication of a vector **a** of length $||\mathbf{a}||$ with a real number α gives a vector **d** with length $|\alpha|\,||\mathbf{a}||$. If $\alpha > 0$, then the directions of **a** and **d** coincide, and if $\alpha < 0$ then **d** has the direction opposite to that of **a**.

Of course, it is rather tedious to perform this operations graphically. This already gives motivation for an algebraic definition, which allows for a computation with numbers instead of performing graphical operations by means of drawings and eventually extracting the results with rulers. This algebraic definition, which will be discussed with more rigor later on, is based on the idea of defining vectors by their behavior under the operations mentioned above, leading to the notion of the vector space.

However, there is more which motivates an algebraic treatment of vectors. A very common picture from structural mechanics is the following. Some structure, for example, a straight beam, is drawn on a sheet of paper together with two force vectors acting on it. The position of the beam in space, for which the paper is a kind of image or model, is indicated by position vectors after choosing an origin. The resulting force is now obtained by the method defined by Definition 1.1. Although, this picture is seemingly quite intuitive, and is widely used also by the author, there are some serious objections against it. The most important one is, that the space of observation $\mathcal{S}$, and vector spaces with different underlying meanings are freely mixed together. This may, and often does, cause severe confusions when approaching more complex problems.

1.3 Differentiable Manifolds: First Contact

A general idea about space is that it consists of points together with a topology. Only in the presence of the latter does it make sense to talk about neighborhoods, interior and boundary of objects, etc. For a large number of problems, a three dimensional euclidean space is an appropriate model for physical space. Objects and operations between them can be defined in accordance with the properties of euclidean space, but those operations can only be performed symbolically. In order to allow for computation with numbers, the points of this space are labeled by triples of real numbers after choosing an origin which gets the label $(0, 0, 0)$. These labels are called coordinates. Through such a labeling, we actually generate a chart of our space of observation and all computations are performed within this chart. For a three-dimensional euclidean space, a single chart is sufficient which coincides with the $\mathbb{R}^3$. Technically, we map points from one euclidean space into a particular representative of an euclidean space in which computations can be carried out. In order to quantify observations in space, like the motion of some body, these observations are first transferred, and hence mapped, into the chart.

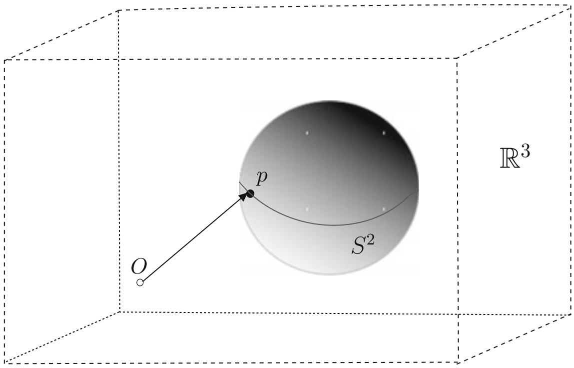

Fig. 1.1 $\mathbb{R}^3$ containing the surface of a sphere S^2, a chosen origin O and some curve on S^2 which passes through the point p

In the following, we consider a space given by the surface of an ordinary sphere $\mathcal{S}^2$. The surface itself is a non-euclidean object because the usual straight lines from euclidean space just do not exist on it. In order to describe $\mathcal{S}^2$ using charts there are two general options:

(I) We can make use of the fact that $\mathcal{S}^2$ can be mapped onto the surface S^2, embedded in $\mathbb{R}^3$. The embedding allows for describing the position of points on the surface either implicitly as the solution of a nonlinear equation or explicitly by position vectors, visualized as straight lines pointing from an origin to the location of the point on the surface, as sketched in Fig. 1.1.

(II) Using ideas from cartography, $\mathcal{S}^2$ can be described by a collection of two-dimensional charts covering $\mathcal{S}^2$ entirely. Since, in general, more than one chart is necessary, a smooth transition between charts has to be ensured. Such a collection is called an atlas. By means of an atlas, the ambient space can be bypassed.

While option (I) means to apply the traditional approach, option (II) eventually leads to the differentiable manifold concept. Although the first option seemingly requires less intellectual effort, there are several objections regarding it. First of all, it is rather common that not S^2 but rather functions defined exclusively on it and not in the remaining ambient space are of primary interest. Such problems require calculus on S^2 rather than calculus in $\mathbb{R}^3$ and a precursor of the differentiable manifold concept has to be applied.

Furthermore, the limitations of option (I) can be illustrated by the widely known "flatlander" gedankenexperiment. Assume that $\mathcal{S}^2$ is the space where the so-called flatlanders live. These flatlanders are two-dimensional beings with a two-dimensional perception, and their embedding in a three-dimensional euclidean space has no meaning to them. How can these flatlanders explore space, a question which leads to the further question, how, e.g., can calculus be performed on a surface without making use of a natural or artificial ambient space? Last but not least, non-euclidean n-dimensional objects can by no means always be embedded into an euclidean space with dimension $n + 1$, but the required dimension of the euclidean embedding can be far higher.

Nevertheless, there are usually objections with regard to the second option too. The most common of these is:

> For the problems I am working on, the concept of euclidean space is sufficient.

However, just by drawing the sketch and thinking about it before the addition of any technical details, important questions arise from the argument that physical phenomena are just there and do not depend on our way of describing them, e.g., by using a natural or artificial embedding. Some of these questions are:

- Global vector balances, like the global balance of momentum in continuum mechanics, rely on the parallel transport of vectors in euclidean space. Therefore, those balances only have a meaning in euclidean space. Can such statements actually be general laws of physics at all?
- Another concept widely used in continuum mechanics is the concept of invariance under finite rigid body motion. Imagine continuum mechanics on a surface with changing curvature. A finite rigid body motion is just not possible in this case.
- If something is moving in a non-euclidean space, there is still a velocity for which it makes sense to talk about a direction and a magnitude. Hence, the vector concept will be needed in non-euclidean spaces as well. But how to work with vectors in this case without any embedding?

At the end of this short introduction, another argument employed frequently to avoid the manifold concept should be mentioned. It goes as follows:

> Even if the manifold approach could give me more insight, it seems far too complicated, and traditional vector calculus works fine for me.

It is true that the mere act of obtaining even a basic understanding of the calculus on differentiable manifolds requires considerable effort. However, some tedious matters, for example integration theorems, become much easier. Furthermore, traditional vector calculus also relies on an underlying manifold approach. One just does not see it because it remains hidden all the time. A main objective of this book is to uncover this underlying approach.

Last but not least, most engineers design and develop constructions, machinery, and devices in order to get things done. Smart engineering solutions usually also possess certain aesthetics easily discerned by other engineers. In the end, linear algebra, analysis, etc., as well as calculus on manifolds are mental constructions designed to get things done safely and as efficiently as possible. And most likely, greater familiarity with these constructions will cause their efficiency and elegance to become apparent.

1.4 Digression on Notation and Mappings

Engineers and physicists are not the only ones to express concepts and ideas in words before translating them into the language of mathematics. Regarding the latter, some standards have evolved over time which are subsumed here under the term mathematical notation. The use of such a notation does not make ideas or concepts smarter and it does not change the principal informational content either. However, a standardized notation is not only more concise and encourages precision, but the information can also be accessed by a larger group of people and not just by a limited community. Furthermore, notation helps in discovering structural similarities between concepts which might have emerged in different contexts.

Although it is assumed that the reader does possess knowledge of applied mathematics at least at the undergraduate level, a few explanatory remarks seem appropriate at this point. The concept of a function is undoubtedly one of the most widely used in all areas of engineering and science. The notation commonly used in school or in undergraduate courses in applied mathematics at universities, for instance,

$$f(n) = \sqrt{n} \quad , \quad n \in \mathbb{N}\,, \tag{1.1}$$

is sufficient to understand that the function f takes a natural number as argument and that it delivers a real number. Therefore, f relates every element of the set of natural numbers $\mathbb{N}$ to an element of the set of real numbers $\mathbb{R}$. In other words, f is a mapping of $\mathbb{N}$ into $\mathbb{R}$. This concept also applies to sets, the elements of which can be objects of any kind.

Definition 1.3 (*Mapping*) Given two sets D and C. A mapping M from D to C, written as $M : D \rightarrow C$, relates every element of D to at least one element of C. D is called domain whereas C is referred to as co-domain.

The notation

$$f : \mathbb{N} \rightarrow \mathbb{R}\,, \tag{1.2}$$

emphasizes that f is only a particular case of a mapping. If more specific information about f is available, e.g., (1.1), then (1.2) is supplemented by it, i.e.,

$$n \mapsto a = \sqrt{n}$$

which leads to the use of the combined notation

$$\begin{aligned} f : \mathbb{N} &\rightarrow \mathbb{R} \\ n &\mapsto a = \sqrt{n} \end{aligned}$$

instead of (1.1).

Definition 1.4 (*Cartesian product*) The Cartesian product between two sets A and B, written as $A \times B$, is a set which contains all ordered pairs (a, b) with $a \in A$, $b \in B$.

For instance, a mapping

$$\begin{aligned} f : \mathbb{N} \times \mathbb{N} &\to \mathbb{R} \\ (m, n) &\mapsto a = \sqrt{m+n} \end{aligned}$$

takes two natural numbers and delivers a real number.

Furthermore, some essential properties of general mappings will be briefly revised next. These properties will become rather important in the next chapters.

Definition 1.5 (*Injective mapping*) A mapping $D \to C$ is called injective or one-to-one if every element of D is mapped to exactly one element of C.

Definition 1.6 (*Surjective mapping*) A mapping $D \to C$ is called surjective or onto if at least one element of D corresponds to every element of C.

Definition 1.7 (*Bijective mapping*) A mapping is called bijective if it is injective and surjective.

Often, it is not only one mapping but rather the composition of various mappings that is of interest. Consider the mappings

$$\begin{aligned} f : \quad \mathbb{R} &\to \mathbb{R} \times \mathbb{R} \\ t &\mapsto (\cos t, 2 \sin t) \end{aligned} \tag{1.3}$$

and

$$\begin{aligned} g : \quad \mathbb{R} \times \mathbb{R} &\to \mathbb{R} \\ (x, y) &\mapsto \sqrt{x^2 + y^2} \end{aligned} \tag{1.4}$$

as an example. The composition $h = g \circ f$ is a mapping

$$\begin{aligned} h : \quad \mathbb{R} &\to \mathbb{R} \\ \tau &\mapsto \sqrt{\cos^2 \tau + \sin^2 \tau}\,, \end{aligned} \tag{1.5}$$

and so-called commutative diagrams, as shown in Fig. 1.2, are commonly used to visualize such situations. The interested reader is referred to Valenza [3].

Before getting into the main matter, some general aspects of notation issues should be discussed. Ideally, notation should be as precise as possible. However, with increasing degrees of complexity, this can lead to formulas with symbols surrounded by clouds of labels, indices, etc. Despite the comprehensiveness of the information, it usually makes it almost impossible to capture the essence of a symbol or equation,

Fig. 1.2 Commutative diagram for the mappings given by (1.3)–(1.5)

$$\begin{tikzcd} \mathbb{R} \arrow[r, "g \circ f"] \arrow[d, "f"'] & \mathbb{R} \\ \mathbb{R} \times \mathbb{R} \arrow[ur, "g"'] & \end{tikzcd}$$

respectively. The other extreme is to keep notation as plain as possible, leaving it to the reader to deduce the specific meaning of symbols from the context. Here, as in most books, a reasonable compromise is intended.

References

1. BBC (2013) Non-euclidean geometry. http://www.youtube.com/watch?v=an0dXEImGHM
2. Holme A (2010) Geometry our cultural heritage, 2nd edn. Springer, Berlin
3. Valenza R (1993) Linear algebra: an introduction to abstract mathematics. Undergraduate texts in mathematics. Springer, Berlin

Chapter 2
Notes on Point Set Topology

Abstract The chapter provides a brief exposition of point set topology. In particular, it aims to make readers from the engineering community feel comfortable with the subject, especially with those topics required in latter chapters. The implicit appearance of topological concepts in the context of continuum mechanics is sketched first. Afterwards, topological concepts like interior and boundary of sets, continuity of mappings, etc., are discussed within metric spaces before the introduction of the concept of topological space.

2.1 Preliminary Remarks and Basic Concepts

Here, tensor calculus is not seen as an end in itself, but rather in the context of engineering mechanics, particularly continuum mechanics. As already discussed earlier, the latter first requires a model for space.

A first step in defining such a model is to interpret space as a set of points. However, if, for instance, different points should refer somehow to different locations, then a set of points, as such, is rather useless. Although sets already possess a kind of structure due to the notion of subsets and operations like union, intersection, etc., some additional structure is required by which a concept like location makes sense. Experience tells us that location is usually expressed with respect to a previously defined point of reference. This, on the other hand, requires relating different points with each other, for instance, by means of a distance.

Distance is primarily a quantitative concept. However, apart from quantitative characteristics like distances between points, length of a curve, etc., there are also others, commonly subsumed under the term topological characteristics. For instance, in continuum mechanics, the concept of a material body is usually introduced as follows. A material body becomes apparent to us by occupying a region in space at a certain instant in time, i.e., a subset of the set of points mentioned above. A look at the sketches commonly used to illustrate basic ideas of continuum mechanics reveals that it is tacitly assumed that this region is "connected" and usually has an "interior" and a "boundary."

U. Mühlich, *Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds*, Solid Mechanics and Its Applications 230,
DOI 10.1007/978-3-319-56264-3_2

Intuitively, we have no problem accepting "connected," "interior," and "boundary" as topological properties. However, there are other properties whose topological character is not at all that obvious. Therefore, the question arises as to how to define more precisely what topological properties actually are. The following experiment might give us some initial clues. A party balloon is inflated until it is possible to draw geometric figures on it. This might be called state one. Afterwards, the balloon is inflated further up to a state two. The figures can be inspected in order to work out those properties which do not change under the transition from state one to state two. Having this experiment in mind, one might be tempted to state that topological properties are those which do not change under a continuous transformation, i.e., a continuous mapping. However, without a precise notion of continuity, this is not really a definition but, at best, a starting point from which eventually to work out a rigorous framework. Furthermore, it turns out that not all continuous mappings are suitable for distinguishing unambiguously topological properties from others, but rather only so-called homeomorphisms, bijective mappings which are continuous in both directions, to be discussed in greater detail later.

The discipline which covers the problems sketched so far is called point set topology and it can be outlined in a variety of ways. Perhaps the most intuitive one is to start with metric spaces, i.e., a set of points together with a structure which allows us to measure distances between points. Most texts on topology in metric spaces focus first on convergence of series and continuity. Properties of sets like interior or boundary are addressed only afterwards in order to prepare the next level of abstraction, namely topological spaces. Here, however, we follow Geroch [2] and start with topological properties of sets in metric spaces.

2.2 Topology in Metric Spaces

The concept of a metric space is a generalization of the notion of distance. Measuring distances consists in assigning real numbers to pairs of points by means of some ruler, hence, it is essentially a mapping $d : X \times X \to \mathbb{R}$ where d is called a distance function. A second step toward a more general scheme is to abstain from interpreting the elements of X in a particular way. In the following, the elements of X can be objects of any kind. By combining the set X with a distance function d, a space (X, d) is generated. The most challenging part is to define a minimal set of rules that a distance function must obey such that certain concepts can be defined unambiguously and facts about (X, d) can be deduced by logical reasoning.

Defining this minimal set of rules is an iterative process for a variety of reasons. The design of a particular distance function has to take into account the specific nature of the elements of X. Furthermore, as in daily life, different rulers, hence different distance functions, should be possible for the same X. A minimal set of rules must cover all these cases and should therefore be rather general. In addition, the set of rules has to account for what should be agreed upon no matter which particular distance function is used to perform a measurement on a given set X. And last but not least, for cases in which we do not even need mathematics because things are simply obvious just by looking at them, facts about (X, d) compiled in respective theorems should be in line with our intuition. The final result of this process is the following definition.

Definition 2.1 (*Metric*) Given a point set X, a metric is a mapping

$$d : X \times X \to \mathbb{R}$$

with the properties:

(i) $d(p, q) > 0$,
(ii) $d(p, q) = 0$ implies $p = q$ and vice versa,
(iii) $d(p, q) = d(q, p)$,
(iv) $d(p, q) \le d(p, r) + d(r, q)$,

where $p, q, r \in X$.

The most common representatives of a metric are the absolute value of the difference of two real numbers in $\mathbb{R}$ and the euclidean metric in $\mathbb{R}^n$. However, there are other options (see Example 2.1). While the properties (i)–(iii) in Definition 2.1 are rather obvious, the triangle inequality (iv) is not. Although (iv) is true for absolute value and euclidean metric, it is not obvious at this point why a metric should possess this property in general. This will become clear only later (see Example 2.3).

Example 2.1 The following distance functions are commonly used for $X = \mathbb{R}^n$,

- $d_1(p, q) = \sum_{i=1}^{n} |p_i - q_i|$,
- $d_2(p, q) = \sqrt{\sum_{i=1}^{n} [p_i - q_i]^2}$,
- $d_\infty(p, q) = \max_{1 \le i \le n} |p_i - q_i|$,

where $|a|$ denotes the absolute value of the argument a. Regarding the notation, see Fig. 2.1 as well.

The following definitions refer to a metric space (X, d), specifically subsets $A, B \subset X$. By means of a metric, the interior of a set can now be defined unambiguously after introducing a so-called ε-neighborhood.

Definition 2.2 (*Open ε-ball*) The set of points q fulfilling $d(p, q) < \varepsilon$ is called the open ε-ball around p or ε-neighborhood of p.

Definition 2.3 (*Interior*) Given a point set A. The interior of A, int(A), is the set of all points p for which ***some*** ε-neighborhood exists which is completely in A.

It is important for further understanding to dissect Definition 2.3 thoroughly, since it illustrates a rather general idea. Although a metric allows for assigning numbers to pairs of points, what matters are not the specific values but only the possibility as such to assign numbers.

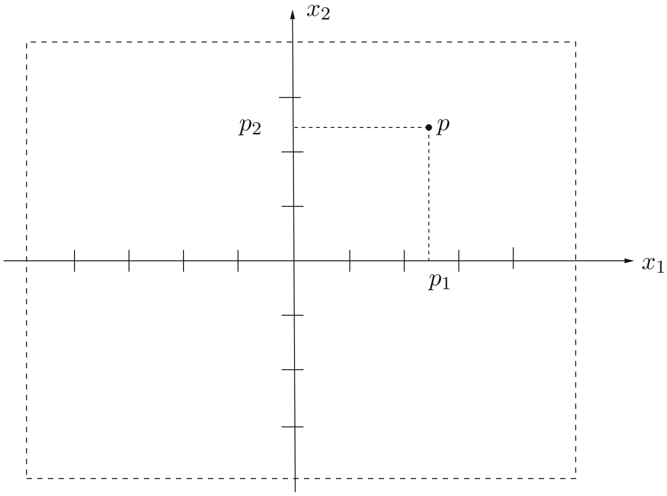

Fig. 2.1 Visualization of the $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$

Example 2.2 Given the interval $A = [0, 1] \subset \mathbb{R}$ and $d(p, q) = |p - q|$. Do the points $p_1 = 0.5$, $p_2 = 0.99$ and $p_3 = 1.0$ belong to the interior of A?
In order to check for p_1, some positive number ε has to be found such that all points q with $d(0.5, q) < \varepsilon$ belong to A. Any ε with $0 < \varepsilon \leq 0.5$ does the job. Since there exists some positive number according to Definition 2.3, p_1 belongs to the interior of A. The point p_2 belongs to the interior of A too, but now we have to chose an ε according to $0 < \varepsilon \leq 0.01$. However, p_3 does not belong to the interior of A.
More generally, the interior of A is the open interval $(0, 1)$.

Similarly, other properties of subsets of X can be defined precisely, for instance, the boundary of a subset of X.

Definition 2.4 (*Boundary*) Given a point set A. The boundary of A, bnd(A), is the set of all points p such that for ***every*** $\varepsilon > 0$ there exists a point q with $d(p, q) < \varepsilon$ in A ***and*** also a point q' with $d(p, q') < \varepsilon$ not in A.

Based on the properties of the distance function and the foregoing definitions, certain facts about sets in metric spaces can be deduced. Some of these are:

1. $\text{int}(A) \subset A$,
2. $\text{int}(\text{int}(A)) = \text{int}(A)$,
3. if $A \subset X$ then every point of X is either in $\text{int}(A)$ or $\text{int}(A^c)$ or $\text{bnd}(A)$ and no point of X is in more than one of these sets, where A^c is the complement of A.

Example 2.3 In order to illustrate the need for the triangle inequality in Definition 2.1, $\text{int}(\text{int}(A)) = \text{int}(A)$ is discussed in more detail. If a point p belongs to $\text{int}(A)$, then p has an ε-neighborhood of points which all belong to A. On the

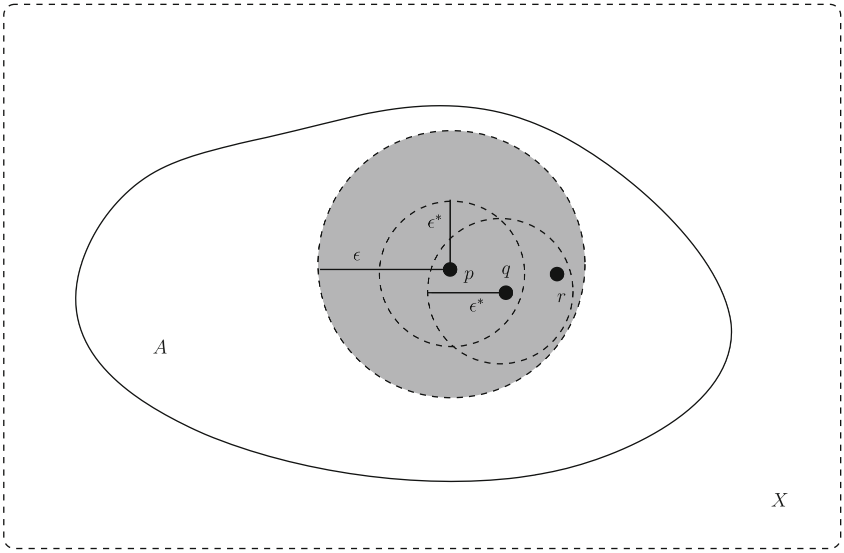

Fig. 2.2 Sketch related to Example 2.3

other hand, if p belongs to int(int(A)) there must exist an ε^*-neighborhood of p which belongs entirely to int(A). In other words, all points in an ε^*-neighborhood of p must have themselves an ε^*-neighborhood of points which belong to A. A sketch, see Fig. 2.2, based on subsets and points in a plane reveals that in this case, this is obviously true for any $p \in \text{int}(A)$. However, in order to ensure the functionality of a metric space (X, d) independently of the particular nature of the elements of X, a formal proof is required.

If $p \in \text{int}(A)$, and, for instance, $\varepsilon^* \leq \frac{1}{2}\varepsilon$ is used, then every point q with $d(p, q) \leq \varepsilon^*$ belongs to A and has itself an ε^*-neighborhood of points which are in A. The latter is ensured by the triangle inequality $d(p, r) \leq d(p, q) + d(q, r)$ where q is any point within the ε^*-neighborhood of p and r is a point within the ε^*-neighborhood of q. Since $d(p, q) \leq \varepsilon/2$ and $d(q, r) \leq \varepsilon/2$, the point r belongs to A, since $d(p, r) \leq \varepsilon$. According to our intuition, an operation which deletes the boundary of an argument should leave that argument, which no longer has a boundary, unchanged. Without triangle inequality, this cannot be assured in general, and such a space (X, d) would not fulfill its purpose.

As already mentioned, the concept of a continuous mapping plays a crucial role in topology. We first define continuity by means of a metric.

Definition 2.5 (*Continuous mapping: $\varepsilon - \delta$ version*) Given the metric spaces (X, d_X) and (Y, d_Y). A mapping $f : X \to Y$ is continuous at $a \in X$ if for ***every*** $\varepsilon > 0$, there exists a $\delta > 0$ such that $d_X(x, a) < \delta$ whenever $d_Y(f(x), f(a)) < \varepsilon$ with $x \in X$.

The following examples use cases, the reader might already know from analysis in $\mathbb{R}$, in the more general setting of this section.

Example 2.4 We consider $X = \mathbb{R}$, $Y = \mathbb{R}$, both with the absolute value distance function and $f = 2x + 14$. Is f continuous at $x = a$?
We have $d_Y(f(x), f(a)) = 2|x - a| < \varepsilon$ and $d_X(x, a) = |x - a| < \delta$. Choosing a particular value for ε, e.g., $\varepsilon = 0.01$, every $\delta < 0.05$ assures that the images of x and a are within a distance smaller than 0.01. However, this must work for every $\varepsilon > 0$, e.g., $\varepsilon = 0.001$, $\varepsilon = 0.0001$, etc. Since, in general we can set $\delta < \frac{\varepsilon}{2}$, a $\delta > 0$ can be found for every $\varepsilon > 0$, and hence, f is continuous at $x = a$. Since this is true for any a, f is continuous everywhere.

So far, this is just the very beginning of an exposition of the theory of metric spaces. However, since here the main objective is to provide a working knowledge in point set topology, we take a short cut. So far, all definitions rely on a distance function. However, a distance can hardly be a topological property, which is a rather unsatisfying situation. One alternative way of defining a minimal structure by which topological concepts can be discussed requires the definition of self-interior sets, also called open sets.

Definition 2.6 (*Open set*) A set A is said to be open or self-interior if $A = \text{int}(A)$.

Once open sets have been defined by means of a distance function d, the latter can be avoided in all subsequent definitions and theorems referring to concepts considered as purely topological. For instance, it can be shown that for metric spaces, the following definition is completely equivalent to Definition 2.7, see, e.g., Mendelson [4].

Definition 2.7 (*Continuous mapping: open sets version*) A mapping $f : X \to Y$ is continuous if for ***every*** open subset $U \subset Y$, the subset $f^{-1}(U) \subset X$ is open.

However, such a reformulation also makes use of the following properties of the collection τ of all open sets induced by the distance function:

1. The empty set $\emptyset$ and X are open.
2. Arbitrary unions of open sets are open.
3. The intersections of two open sets is open.

This leads to the conclusion that a distance function is not needed at all to impose a topological structure on a set. It is sufficient to provide an appropriate collection of sets which are defined formally as open. This is done in the following section. Prior to this, another observation which supports a metric independent generalization of point set topology should be mentioned. A closer look at Example 2.1 reveals that most distance functions are defined through the use of addition, subtraction, and multiplication. This requires the existence of a structure on the considered set which is already rather elaborated. This additional structure is by no means necessary for discussing fundamental topological concepts and it might even obscure the view as to what is important.

2.3 Topological Space: Definition and Basic Notions

Definition 2.8 (*Topological space*) A nonempty set X together with a collection τ of subsets of X with the properties:

(i) the empty set $\emptyset$ and X are in τ,
(ii) arbitrary unions of members of τ are in τ,
(iii) the intersection of any two members of τ is in τ,

is called a topological space (X, τ). The elements of τ are by definition open sets.

Remark 2.1 There is a number of equivalent definitions which can be obtained in a similar way as that sketched for open sets in the previous section. For instance, defining closed sets and working out their properties in a metric space eventually leads to the definition of a topological space based on the notion of closed sets. Another alternative can be worked out using so-called neighborhood systems.

Definition 2.9 (*Closed set*) Given a topological space (X, τ). A set $A \subset X$ is closed, if its complement A^{c} is open.

Remark 2.2 From Definitions 2.9 and 2.8 it follows that $\emptyset$ and X of (X, τ) are both open and closed since $X^{\mathrm{c}} = \emptyset$ and $\emptyset^{\mathrm{c}} = X$.

Example 2.5 The collection of all open intervals $\tau = \{(a, b)\}, a, b \in \mathbb{R}$ defines a topological space $(\mathbb{R}, \tau)$. On the other hand, the collection of all open intervals $\tau^* = \{(x - a, x + a)\}, a, b, x \in \mathbb{R}$ defines the same topological space.

Metric spaces are automatically topological spaces, because a distance function induces a topology. In this context, two questions arise. The first asks whether the reverse is true as well and the answer here is no. There are topological spaces which do not arise from metric spaces. The second question asks whether the topology of a metric space depends on the choice of a particular distance function. Here, the answer is yes, since different distance functions may induce different topologies. However, it turns out that the distance functions d_1, d_2 and d_∞ defined for $\mathbb{R}^n$ (see Example 2.1) induce the same topology, which is called the standard topology of $\mathbb{R}^n$. One way to prove this is to show that every set that is open in terms of one distance function is also open if one of the other distance functions is used, and vice versa.

Defining the notion of continuous mapping now consists essentially in repeating Definition 2.7 in the context of topological spaces according to Definition 2.8.

Definition 2.10 (*Continuous mapping*) Given two topological spaces (X, τ) and (Y, τ^*). A mapping $f : X \to Y$ is continuous if for ***every*** open subset $U \subset Y$, the subset $f^{-1}(U) \subset X$ is open.

It can be shown that the composition of continuous mappings is again continuous, which is a important fact, for obvious reasons.

Example 2.6 Consider the finite sets $X = Y = \{0, 1\}$. In order to check whether the mapping

$$\begin{aligned} f : X &\to Y \\ 0 &\mapsto 1 \\ 1 &\mapsto 0 \end{aligned}$$

is continuous or not, it has to be checked as to whether the domains of all open sets in Y are open sets in X. The first observation is that there is not enough information available to do this, since it is not clear which topologies are used. Given the topologies $\tau_1 = \{X, \emptyset\}, \tau_2 = \{X, \emptyset, \{0\}, \{1\}\}$, then $f : (X, \tau_1) \to (Y, \tau_2)$ is not continuous, since the domain of $\{1\}$ is $\{0\}$ but $\{0\}$ does not belong to τ_1.

Remark 2.3 (*Continuity*) Intuitively, one associates continuity with mappings between infinite sets like $\mathbb{R}$ or subsets of $\mathbb{R}$. But Example 2.6 illustrates that after introducing the concept of topological spaces, continuity makes sense for finite sets too. This, on the other hand, has reverse effects. For instance, it provides a reasoning for the transparent motivation of compactness, a far reaching concept to be discussed in more detail later.

Having defined what is meant by a continuous mapping, it is about time to make the notion of topological properties more precise. Intuitively accepting the property of being open as a topological property of a set, the standard example of $f = x^2$ shows that defining topological characteristics as those which are invariant under continuous mappings does not work properly. This can be seen by taking the open interval $(-1, 1)$, which is mapped by f to the half open interval $[0, 1)$, i.e., $f(\,(-1, 1)\,) = [0, 1)$. Although f is continuous, the property of being open is not preserved under f. The reason for this is that f is not bijective. This leads to the notion of homeomorphism.

Definition 2.11 (*Homeomorphism*) Given two topological spaces (X, τ_X) and (Y, τ_Y). A homeomorphism is a bijective mapping $f : X \to Y$ where f and f^{-1} are continuous.

Since a homeomorphism preserves what is accepted intuitively as topological characteristics, it might serve as a general criterion for distinguishing topological characteristics from others. Eventually, this hypothesis can be confirmed, and topological spaces which can be related by such a mapping are called topologically equivalent or homeomorphic.

One of the most familiar applications of homeomorphisms are coordinate charts, the definition of which also implies a notion of dimensionality of a topological space.

Definition 2.12 (*Coordinate chart*) Given a topological space (X, τ) and an open subset $U \subset X$. A homeomorphism $\phi : U \to V$ where V is an open subset of $\mathbb{R}^n$ is called a coordinate chart.

Definition 2.13 (*Dimension of a topological space*) Given a topological space (X, τ). X has dimension n if every open subset of X is homeomorphic to some open subset of $\mathbb{R}^n$.

2.4 Connectedness, Compactness, and Separability

After introducing the most basic notions, it is rather compulsory to ask under which conditions certain operations can be safely performed. In this context, a number of additional topological concepts will be discussed in the following section, starting with the idea of subspace.

Definition 2.14 (*Subspace*) A topological space (Y, τ_Y) is a subspace of the topological space (X, τ_X), if all members of τ_Y can be derived from $o = O \cap Y$ for $o \in \tau_X$ and some $O \in \tau_Y$. The topology τ_Y is called the relative topology on Y induced by τ_X.

In order to discuss the concept of connectedness, we start with the definition of a path in a topological space X.

Definition 2.15 (*Path*) A path is a continuous function f which maps a closed interval of $\mathbb{R}$ to a topological space X,

$$\begin{aligned} f : \mathbb{R} &\to X \\ [a, b] &\mapsto \Gamma \subset X \, , \end{aligned}$$

in this way connecting the start point $f(a)$ with the end point $f(b)$.

Definition 2.16 (*Path-connected*) A topological space is path-connected if for each pair of points $x, y \in X$, there is a path which connects them.

It can be shown that a topological space is connected if it is path-connected. According to Definition 2.16, finite spaces with more than one point can obviously not be connected. The connectedness of subsets of a topological space can be discussed by adapting Definitions 2.15 and 2.16 using the relative topology according to Definition 2.14.

Remark 2.4 In continuum mechanics, the configuration of a material body at an instant of time, e.g., $\tau = t_0$ can be encoded as a mapping $\kappa(t_0)$ which maps all points of the body to a subset of the space X. A motion is now described as a sequence of such mappings, one mapping for each instant of time $\tau = t$. It is supposed that these mappings and their inverses are continuous, i.e., the motion is assembled from homeomorphisms. This ensures automatically that a connected body remains connected in the course of the motion.

Certain properties of continuous mappings are of particular interest. This will be illustrated by means of continuous functions $f : A \to \mathbb{R}$ where $A \subset \mathbb{R}$. We are often particularly interested in where a function attains its maximum or minimum values. However, existence of such properties first requires that the considered function be bounded.

Definition 2.17 (*Bounded*) A function $f : A \to \mathbb{R}$ with $A \subset \mathbb{R}$ is bounded if $|f(x)| \leq M$ for $x \in A$ and $M \in \mathbb{R}$.

The standard example $f(x) = \frac{1}{x}$ indicates that being bounded depends somehow on the properties of the domain, since f is bounded on the intervals $[0.001, 1)$ and $[0.001, 1]$ but not on $(0, 1)$ or $(0, 0.5]$. Therefore, the question arises if there are conditions regarding the domain under which a continuous function is definitely bounded. At this point, the concept of compactness enters the scene. One possible line of argument to motivate compactness takes finite sets as a role model, since functions on finite sets are bounded in any case.

Remark 2.5 Care should be be taken here regarding the interpretation of $\pm\infty$ in the context of real analysis. There is no number $\frac{1}{0}$ in $\mathbb{R}$ and ∞ is no element of $\mathbb{R}$ but only an indication for something arbitrarily big. Therefore, $A = \{0, 1\}$ and $g = \frac{1}{x}, x \in A$ does not make sense, because the element 0 is not mapped to an element of $\mathbb{R}$. For the same reason, g together with $A = [-1, 1]$ is not a proper definition of a function according to Definition 1.3, since not every element of A has an image. On the other hand, g together with $A = (0, 1)$, for example, is correct.

If A is a finite set, then a function $f : A \to \mathbb{R}$ is locally bounded, since it assigns to every element of A some real number. One of these numbers will be the one with the largest absolute value, and f is bounded globally by the latter. This argument, trivial for finite sets, does not apply if A is infinite. Here, a similar line of argument is developed by means of the concept of open covers.

Definition 2.18 (*Open cover*) A collection of open sets is a cover of a set A if A is a subset of the union of these open sets.

Suppose that an infinite set can be covered by a finite number of open sets. If, in addition f is bounded on all these open sets, the same argument used for finite sets can be adapted. However, there are many possible open covers for a given set and the finiteness argument must not depend on the choice of a particular cover. This gives raise to the following definition of compactness.

Definition 2.19 (*Compactness*) A set A of a topological space is compact if ***every*** open cover of A has a finite sub-cover which covers A.

Example 2.7 The open interval $I = (0, 1)$ can, of course, be covered by a finite number of open sets, starting with I itself. However, this does not mean that I is compact. In order to show that I is not compact, it suffices to find at least one open cover which does not have a finite sub-cover. The collection $O_n = (\frac{1}{n}, 1 - \frac{1}{n})$ covers, for $n \to \infty$, the open interval I but there is no finite n for which I is covered completely. Therefore, I is not compact.

Of course, to check compactness for a topological space or some subspace by means of Definition 2.19 might be an option for a mathematician, but will not be for nonmathematicians. However, it turns out that, for instance, all closed intervals $[a, b]$ with $a, b \in \mathbb{R}$ are compact.

Last but not least, the so-called Hausdorff property ensures that if a sequence converges in a topological space X, it converges to exactly one point of X.

Definition 2.20 (*Hausdorff*) A topological space X is separable or Hausdorff if, for every $x, y \in X$ and $x \neq y$, there exist two disjoint open sets $A, B \subset X$ such that $x \in A$ and $y \in B$.

The concept of a basis for a topology is, among other things, useful for checking if a topological space is a Hausdorff-space.

Definition 2.21 (*Basis*) Given a topological spaces (X, τ). A collection τ_B of elements of τ is called a basis if every member of τ is a union of members of τ_B.

It can be shown that every space for which a countable basis can be constructed is a Hausdorff-space. All metric spaces in particular fulfill this condition. A possible basis for the standard topology in $\mathbb{R}$ consists of all open sets (a, b) with $a, b \in \mathbb{Q}$, where $\mathbb{Q}$ is the set of rational numbers. Since $\mathbb{Q}$ is countable, $\mathbb{R}$ is a Hausdorff-space.

2.5 Product Spaces and Product Topologies

Building a new set X from two sets X_1 and X_2 by means of the Cartesian product, i.e., $X = X_1 \times X_2$, is a common and most natural thing to do. If X_1 and X_2 are equipped with respective topologies τ_1 and τ_2, the question about the topology of X inevitably emerges. Since τ_1 is the collection of all open sets in X_1 and τ_2 is the collection of all open sets in X_2, it seems reasonable to try with the Cartesian product of these sets, specifically to define a collection

$$\mathcal{B} = \{U_1 \times U_2 \,|\, U_1 \in \tau_1 \text{ and } U_2 \in \tau_2\} \tag{2.1}$$

and to check if $\mathcal{B}$ may serve as a topology.

Remark 2.6 In the following, most examples refer to $\mathbb{R}^n$. Since we started with the intention of working out a minimal structure, suitable for discussing topological concepts, this might seem inconsistent, because $\mathbb{R}^n$ has a lot of additional structure. However, $\mathbb{R}$ and $\mathbb{R}^2$ especially are quite accessible to our intuition. Furthermore, a topology on a set X can be defined, for instance, by means of a bijective mapping $g : X \to \mathbb{R}^n$. Since the existence of a topology is assured in $\mathbb{R}^n$, a topology on X can be defined as the collection of the images $g^{-1}(U)$ of all open sets U of the standard topology of $\mathbb{R}^n$. This makes X a topological space and g a continuous mapping.

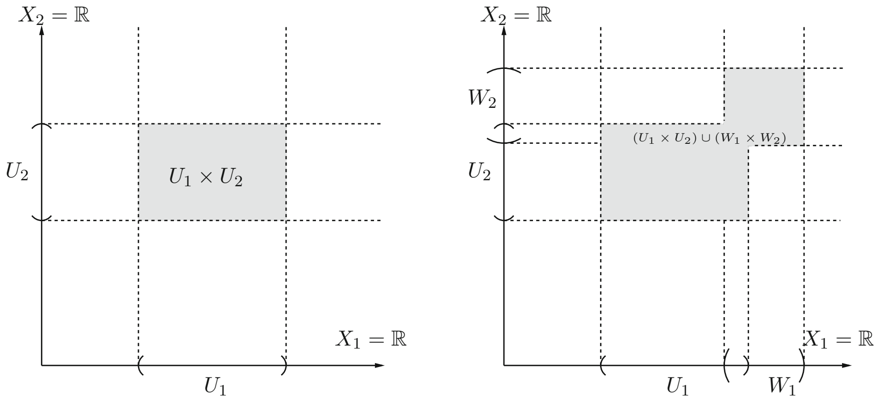

Fig. 2.3 Sketch illustrating for $X_1 = X_2 = \mathbb{R}$ that the union of members of $\mathcal{B}$, see (2.1), is not an element of $\mathcal{B}$. Therefore, $\mathcal{B}$ given by (2.1) can not be a topology

According to Definition 2.8, arbitrary unions and finite intersections of members of τ must result in an element of τ. Figure 2.3 illustrates that this does not hold for $\mathcal{B}$. Therefore, $\mathcal{B}$ is not a topology.

However, $\mathcal{B}$ can be used to construct a topology. To this end, the concept of the so-called sub-basis is introduced. The reader is encouraged to solve Exercise 2.3 after reading the following definition carefully.

Definition 2.22 (*Sub-basis*) A sub-basis $\mathcal{B}$ for a topological space (X, τ) is a collection of open sets in which the union of all members of $\mathcal{B}$ equals X. The topology of X is the collection of all unions of finite intersections of members of $\mathcal{B}$.

$\mathcal{B}$ as defined by (2.1) fulfills the requirements for being a sub-basis according to Definition 2.22. It turns out that for $X_1 = X_2 = \mathbb{R}$ the product topology coincides with the standard topology of $\mathbb{R}^2$. Considering the projections

$$\begin{aligned} \pi_1 : X_1 \times X_2 &\to X_1 \\ (x_1, x_2) &\mapsto x_1 \end{aligned} \qquad\qquad \begin{aligned} \pi_2 : X_1 \times X_2 &\to X_2 \\ (x_1, x_2) &\mapsto x_2 \,, \end{aligned}$$

more insight about the product topology can be gained. The sub-basis $\mathcal{B}$ can be written by means of the projections as follows:

$$\mathcal{B} = \left\{ \pi_1^{-1}(U_1) \cap \pi_2^{-1}(U_2) \mid U_1 \in \tau_1 \text{ and } U_2 \in \tau_2 \right\} ,$$

which shows that the product topology τ generated by $\mathcal{B}$ ensures that the projections are continuous. In fact, τ is the coarsest topology for which the projections are continuous. What coarse means depends on the topologies τ_1 and τ_2 (see Exercises 2.4 and 2.5). Furthermore, it can be shown that a function

$$\begin{aligned} f : Z &\to X_1 \times X_2 \\ z &\mapsto (\, f_1(z),\, f_2(z)\,) \end{aligned}$$

is continuous if $f_1 : Z \to X_1$ and $f_2 : Z \to X_2$ are continuous, and vice versa.

The step toward general finite products, i.e., $X_1 \times X_2 \times \cdots \times X_n$, is straight forward. Infinite products, however, require additional considerations.

2.6 Further Reading

A didactic introduction to topology, which nicely illustrates the creativity behind the formal “definition-theorem-proof” structure by showing how the formal language of topology evolves from first ideas accompanied by trial and error, is given in Geroch [2]. The classic book by Mendelson [4] is a compact and accessible introduction. Also recommendable are Conover [1] and Runde [5], the latter of which also contains a number of historical notes. For further steps in topology, we recommend, for example, Jänich [3]. In addition, summaries of point set topology can be found in most textbooks on differentiable manifolds.

Exercises

2.1 Check continuity of f in Example 2.6 for the following cases:

- $f : (X, \tau_2) \to (Y, \tau_1)$
- $f : (X, \tau_2) \to (Y, \tau_2)$

2.2 Check if (Y, τ_Y) with $Y = \{a, b\}$ and $\tau_Y = \{\emptyset, \{a\}, \{b\}\}$ is a subspace of (X, τ_X) with $X = \{a, b, c, d\}$ and $\tau_x = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}\}$.

2.3 Given $X_1 = \{0, 1\}$, $X_2 = \{a, b\}$ together with the topologies $\tau_1 = \{X_1, \emptyset, \{0\}, \{1\}\}$ and $\tau_2 = \{X_2, \emptyset, \{a\}, \{b\}\}$. Determine the sub-basis $\mathcal{B}$ for the product topology of $X = X_1 \times X_2$.

2.4 Given X_1, X_2 and τ_1 as in Exercise 2.3. Determine the sub-basis $\mathcal{B}$ for the topology τ of $X = X_1 \times X_2$ for $\tau_2 = \{X_2, \emptyset\}$. Determine τ and show that the projections are continuous.

2.5 Given X_1, X_2 and τ_2 as in Exercise 2.4. Determine the sub-basis $\mathcal{B}$ for the topology τ of $X = X_1 \times X_2$ for $\tau_1 = \{X_1, \emptyset\}$. Determine τ and show that the projections are continuous. Compare τ with the corresponding result of Exercise 2.4.

References

1. Conover R (1975) A first course in topology. The Wiliams & Wilkins Company, Baltimore
2. Geroch R (2013) Topology. Minkowski Institute Press, Montreal
3. Jänich K (1994) Topology. Undergraduate texts in mathematics. Springer, Berlin
4. Mendelson B (1975) Introduction to topology. Dover Publications, New York
5. Runde V (2000) A taste of topology. Universititext. Springer, Berlin

Chapter 3
The Finite-Dimensional Real Vector Space

Abstract The chapter introduces the notion of the finite-dimensional real vector space together with fundamental concepts like linear independence, vector space basis, and vector space dimension. The discussion of linear mappings between vector spaces prepares the ground for introducing the dual space and its basis. Finally, inner product space and reciprocal basis are contrasted with dual space and the corresponding dual basis.

3.1 Definitions

Within the previous chapter, the combination of a set with a topological structure has been discussed. Now the objective is to discuss a set in combination with an algebraic structure.

Algebraic structures consist of at least one set of objects, also called points in a more general sense, together with a set of operations defined between these objects satisfying a number of axioms. Specific meanings of objects and operations are certainly important, for instance, in order to discuss implications of certain results for technical applications. On the other hand, abstraction from specific details classifying problems by the underlying algebraic structure has proved to be extremely economic and flexible.

Within the hierarchy of algebraic structures, the vector space sits already at a rather high level, preceded by the concepts of groups, rings, and fields.

Definition 3.1 (*Real vector space*) A real vector space $\mathcal{V}$ is a set whose elements are called vectors, for which two binary operations are defined:

1. addition (inner composition "$\oplus$")
   The addition of two vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a}, \mathbf{b} \in V$, is a mapping from $\mathcal{V} \times \mathcal{V}$ to $\mathcal{V}$

$$\begin{aligned} \oplus : \mathcal{V} \times \mathcal{V} &\to \mathcal{V} \\ (\mathbf{a}, \mathbf{b}) &\mapsto \mathbf{a} \oplus \mathbf{b} \end{aligned}$$

U. Mühlich, *Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds*, Solid Mechanics and Its Applications 230,
DOI 10.1007/978-3-319-56264-3_3

with the properties:

(i) $\mathbf{a} \oplus \mathbf{b} = \mathbf{b} \oplus \mathbf{a}$
(ii) $(\mathbf{a} \oplus \mathbf{b}) \oplus \mathbf{c} = \mathbf{a} \oplus (\mathbf{b} \oplus \mathbf{c})$
(iii) there exists a neutral element $\mathbf{0}$ such that

$$\mathbf{a} \oplus \mathbf{0} = \mathbf{a}$$

holds for arbitrary $\mathbf{a} \in \mathcal{V}$
(iv) to every element $\mathbf{a}$, $\mathbf{a} \in \mathcal{V}$, there exists exactly one element $-\mathbf{a}$, such that

$$\mathbf{a} \oplus (-\mathbf{a}) = \mathbf{0}$$

holds.

2. multiplication with a real number (outer composition "$\odot$")
The multiplication of a vector $\mathbf{a} \in \mathcal{V}$ with a real number $\alpha \in \mathbb{R}$ is a mapping from $\mathbb{R} \times V$ to V

$$\begin{aligned} \odot : \mathbb{R} \times \mathcal{V} &\to \mathcal{V} \\ (\alpha, \mathbf{a}) &\mapsto \alpha \odot \mathbf{a} \end{aligned}$$

with the properties:

(I) $\alpha \odot (\beta \odot \mathbf{a}) = (\alpha \beta) \odot \mathbf{a}$
(II) $\alpha \odot (\mathbf{a} \oplus \mathbf{b}) = \alpha \odot \mathbf{a} \oplus \alpha \odot \mathbf{b}$
(III) $(\alpha + \beta) \odot \mathbf{a} = \alpha \odot \mathbf{a} \oplus \beta \odot \mathbf{a}$
(IV) there exists a neutral element 1, such that

$$1 \odot \mathbf{a} = \mathbf{a}$$

holds.

The inner composition combines two elements of $\mathcal{V}$, whereas the outer composition combines an element of $\mathcal{V}$ with an element of $\mathbb{R}$. Both operations are by definition closed with respect to $\mathcal{V}$, which means that the result is again an element of $\mathcal{V}$.

The use of $\oplus$ and $\odot$ indicates that these operations are not necessarily summation and scalar multiplication in the usual sense. On the contrary, they can even mean graphical operations, as defined in the Chap. 1.

In the geometrical setting sketched within Chap. 1, vectors have been defined by specific properties, namely magnitude and direction. Neither one of these two things appear in the algebraic definition. On the contrary, specific properties are of no importance in the algebraic definition but a vector is defined by its behavior under the considered operations. Furthermore, at this stage, it does not even make sense to talk about a magnitude, because of the lack of a way to measure length, which would require the definition of a metric first.

Be aware that the use of the ordinary "+" on the left-hand side of the third property of the multiplication is not a mistake, because it expresses the sum of two real numbers.

In order to simplify the notation, the "$\oplus$" is often replaced by "+" and $\lambda\mathbf{a}$ is used instead of $\lambda \odot \mathbf{a}$. This is commonly known as operator overloading.

Remark 3.1 Usually, operator overloading does not cause confusion; however, there are vector spaces for which it is better to drop this overloading of operation symbols and to return to the original notation, see, e.g. Exercise 3.1. In the follow-up, operation overloading will be used as long as it does not cause confusion.

Definition 3.1 provides the minimal amount of statements necessary to deduce the remaining properties of a vector space by logical reasoning. Two of them are as follows:

- The neutrals are uniquely determined.
- The operations "$\oplus$" and "$\odot$" give unique results.

Proving these statements falls under the duties of mathematicians and is beyond the scope here. The interested reader is referred, for example, to Valenza [3].

3.2 Linear Independence and Basis

Definition 3.1 provides an abstract frame for vector spaces. On the other hand, the interpretation of vectors as geometrical objects in the plane, as used in Chap. 1, is actually a particular realization of a vector space. Again, in this context, the addition of vectors and multiplication by scalar are graphical operations. Given a vector $\mathbf{a}$ and two other vectors $\mathbf{g}_1$ and $\mathbf{g}_2$, and let us assume that

$$\mathbf{a} = a^1\mathbf{g}_1 + a^2\mathbf{g}_2 \tag{3.1}$$

holds, where a^1 and a^2 are two real numbers. In this context, the superscripts also mean indices, not powers. The use of upper and lower indices may seem strange at first glance, but it provides the possibility of transmitting additional information just by notation. The advantage of this will become clearer later on. Furthermore, let us assume that for another vector $\mathbf{b}$,

$$\mathbf{b} = b^1\mathbf{g}_1 + b^2\mathbf{g}_2 \tag{3.2}$$

holds. According to Definition 3.1, the summation of $\mathbf{a}$ and $\mathbf{b}$ with result $\mathbf{c}$ can be written as

$$\mathbf{c} = \mathbf{a} + \mathbf{b} = c^1\mathbf{g}_1 + c^2\mathbf{g}_2 \tag{3.3}$$

with

$$\begin{aligned} c^1 &= a^1 + b^1 \\ c^2 &= a^2 + b^2 , \end{aligned}$$

which indicates a way to replace the graphical operations with computations, because the real numbers c^1 and c^2 can be determined just by adding other real numbers. Of course, the result refers to the use of $\mathbf{g}_1$ and $\mathbf{g}_2$. Without this information, the numbers c^1 and c^2 are completely meaningless. Obviously, the scheme works only if $\mathbf{a}$ and $\mathbf{b}$ can be expressed by the very same set of vectors $\{\mathbf{g}_1, \mathbf{g}_2\}$. Furthermore, in order for this scheme to be applicable in general, the following questions have to be answered. Is there always a set of vectors by which every other vector can be expressed uniquely? Are there always two vectors which have to be chosen for this, and if not, how many are necessary? Can these vectors be chosen arbitrarily, or are there restrictions? In order to answer these questions, the following definitions are required.

Definition 3.2 (*Linear independence*) N vectors $\mathbf{g}_i, i = 1, \ldots, N, \mathbf{g}_i \in \mathcal{V}, i \in \mathbb{N}$ are linearly independent if

$$\alpha^1 \mathbf{g}_1 + \alpha^2 \mathbf{g}_2 + \cdots + \alpha^N \mathbf{g}_N = \mathbf{0}$$

with $\alpha^i \in \mathbb{R}$ can be fulfilled only by $\alpha^i = 0, i = 1, \ldots, N$.

Definition 3.3 (*Maximal set of linearly independent vectors*) A set consisting of N vectors $\{\mathbf{g}_i\}, i = 1, \ldots, N, \mathbf{g}_i \in \mathcal{V}, i \in \mathbb{N}$ is called a maximal set of linearly independent vectors if it cannot be extended by any other element of $\mathcal{V}$ without violating Definition 3.2.

> For a finite-dimensional vector space $\mathcal{V}$, the size of all possible maximal sets according to Definition 3.3 is equal and uniquely determined.

Definition 3.4 (*Dimension of a vector space*) The dimension of a finite-dimensional vector space is defined as the size of the sets defined by Definition 3.3 and denoted by $\dim(\mathcal{V})$.

Before proceeding further, it seems appropriate to do something about notation. Definition 3.2 already indicates how tedious notation can become if N is just sufficiently large. Of course, the right-hand side in Definition 3.2 could be written shorter using the summation symbol

$$\alpha^1 \mathbf{g}_1 + \alpha^2 \mathbf{g}_2 + \cdots + \alpha^N \mathbf{g}_N = \sum_{i=1}^{N} \alpha^i \mathbf{g}_i , \tag{3.4}$$

but even this could be dropped if the following conventions, commonly subsumed under the term Einstein notation, are applied.

Definition 3.5 (*Einstein notation*) In Einstein notation, indexing obeys the following rules:

(i) an index appears, at most, twice within a term;
(ii) indices which appear only once per term must be the same in every term of an expression;
(iii) if an index appears twice, it must appear exactly once as upper index and once as lower index, and it indicates that this index is a summation index;
(iv) summation indices can be chosen freely as long as this does not conflict with (i)–(iii).

By employing Einstein notation, the right-hand side of the equation in Definition 3.2 can be written concisely as $\alpha^i \mathbf{g}_i$ without any ambiguity or loss of information. The range of summation is also clear from the context because it starts with one and the upper range is given by $\dim(\mathcal{V})$. An important advantage of such a concise notation is that the structure of a theory or a problem can be apprehended much more easily. Einstein notation is nothing difficult, it only requires some training to get used to it.

Definition 3.6 (*Basis of a vector space*) Every maximal set of linearly independent vectors according to Definition 3.3 forms a basis of the considered vector space. The members of such a set are called base vectors.

Based on the two foregoing definitions, answers to the questions raised at the beginning of this section can be provided.

Every element $\mathbf{b}$ of a vector space $\mathcal{V}$ can be uniquely represented by a basis $\{\mathbf{g}_i\}$ as follows:

$$\mathbf{b} = b^i \mathbf{g}_i \quad , \quad i = 1, \ldots, N$$

where the b^i are called the coordinates or the components of a vector with respect to the basis $\{\mathbf{g}_i\}$.

A vector space usually has at least dimension one, and it should be noted that the zero vector cannot belong to the set of base vectors, since $\alpha\mathbf{0} = \mathbf{0}$ for any $\alpha \in \mathbb{R}$. On the other hand, the set $\{\mathbf{0}\}$ endowed with vector space operations is a legitimate vector space. Difficulties which may be caused by this are avoided by defining $\dim(\{\{\mathbf{0}\}, \oplus, \odot\}) = 0$.

3.3 Some Common Examples for Vector Spaces

In the following, some vector spaces are defined which will be used to generate examples.

Definition 3.7 (*The vector space* $\mathbb{R}$) The set of real numbers, denoted by $\mathbb{R}$ together with the usual addition and multiplication defined for real numbers, is a vector space.

Definition 3.8 (*The vector space* $\mathbb{R}^N$) The elements of $\mathbb{R}^N$ are all N-tuples of real numbers

$$\mathbf{x} = (x^1, x^2, \ldots, x^N),$$

for which addition and scalar multiplication are defined as follows:

$$\begin{aligned} \mathbf{y} + \mathbf{z} &= (y^1 + z^1, y^2 + z^2, \ldots, y^N + z^N) \\ \alpha\mathbf{y} &= (\alpha y^1, \alpha y^2, \ldots, \alpha y^N), \end{aligned}$$

with $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^N$, $\alpha, x^i, y^i \in \mathbb{R}$. Keep in mind that just defining $\mathbb{R}^N$ as the set of all N-tuples of real numbers is not sufficient. $\mathbb{R}^N$ becomes a vector space only by defining the respective operations at the same time.

Definition 3.9 (*The vector space* $\mathcal{P}^M$) The elements of $\mathcal{P}^M$ are all polynomials of order M with real coefficients

$$p(x) = p^0 + p^1 \operatorname{pow}(x, 1) + p^2 \operatorname{pow}(x, 2) \cdots + p^M \operatorname{pow}(x, M),$$

where $\operatorname{pow}(x, k)$ stands for the k-th power of x in order to distinguish powers from upper indices. Addition and multiplication by scalar are defined by

$$\begin{aligned} p(x) + q(x) &= (p^0 + q^0) + (p^1 + q^1) \operatorname{pow}(x, 1) + \cdots + (p^M + q^M) \operatorname{pow}(x, M) \\ \lambda p(x) &= \lambda p^0 + \lambda p^1 \operatorname{pow}(x, 1) + \cdots + \lambda p^M \operatorname{pow}(x, M), \end{aligned}$$

where, subsequently, $\operatorname{pow}(x, k)$ and $(x)^{\mathrm{k}}$ are used synonymously. Please note that $\dim(\mathcal{P}^M) = M + 1$.

More examples can be found, e.g., in Winitzki [4]. While $\{(1, 2), (4, 1)\}$ is a suitable basis for $\mathbb{R}^2$, a possible basis for $\mathcal{P}^2$ is given by $\{1, x, (x)^2\}$. Seemingly unrelated objects like triples of real numbers and second-order polynomials possess the same underlying algebraic structure. This might be surprising at first glance and it already gives an idea regarding the power of the vector space concept.

Remark 3.2 $\mathbb{R}^N$ with $N = 2$ is a short-hand notation for $\mathbb{R} \times \mathbb{R}$, which can be extended easily for arbitrary $N \in \mathbb{N}$, as used in Definition 3.8. The latter indicates how new vector spaces can be generated by means of the Cartesian product.

3.4 Change of Basis

Given a vector space $\mathcal{V}$ and a basis $\{\mathbf{g}_i\}$. What happens if we want to use another basis, say $\{\hat{\mathbf{g}}_i\}$. Besides the importance of this question in general, it is a good start for practicing the Einstein notation.

Because every element of $\mathcal{V}$ can be expressed by means of the basis $\{\mathbf{g}_i\}$, this also holds for a vector $\hat{\mathbf{g}}_i$. For $\dim(\mathcal{V}) = 2$, we have

$$\begin{aligned}\hat{\mathbf{g}}_1 &= \hat{A}_1^1\,\mathbf{g}_1 + \hat{A}_1^2\,\mathbf{g}_2\\ \hat{\mathbf{g}}_2 &= \hat{A}_2^1\,\mathbf{g}_1 + \hat{A}_2^2\,\mathbf{g}_2\,,\end{aligned}$$

or for arbitrary dimension using Einstein notation,

$$\hat{\mathbf{g}}_i = \hat{A}_i^m\,\mathbf{g}_m\,, \tag{3.5}$$

where the $\hat{A}_i^j$ are real numbers. On the other hand, the $\mathbf{g}_i$ can be written in terms of the base vectors $\hat{\mathbf{g}}_i$ as follows:

$$\mathbf{g}_i = A_i^m\,\hat{\mathbf{g}}_m\,. \tag{3.6}$$

The very same vector $\mathbf{v}$ can be expressed using the basis $\{\mathbf{g}_i\}$ but also by means of $\{\hat{\mathbf{g}}_i\}$. The coordinates of both representations will differ, but

$$v^i\,\mathbf{g}_i = \hat{v}^i\,\hat{\mathbf{g}}_i \tag{3.7}$$

must hold, because the left- and right-hand sides are just different representations of the same vector. Evaluating (3.7) by replacing the $\hat{\mathbf{g}}_i$ on the right-hand side with (3.5) leads to

$$v^i\,\mathbf{g}_i - \hat{v}^i\,\hat{A}_i^m\,\mathbf{g}_m = \mathbf{0}\,. \tag{3.8}$$

By means of the so-called Kronecker symbol defined by

$$\delta_i^j = \begin{cases} 1 & \textbf{if} \quad i = j \\ 0 & \phantom{\textbf{if}} \quad i \neq j \end{cases}\,, \tag{3.9}$$

Equation (3.8) can be written as

$$\left[v^i\,\delta_i^m - \hat{v}^i\,\hat{A}_i^m\right]\mathbf{g}_m = \mathbf{0}\,.$$

Because the $\mathbf{g}_m$ are linearly independent, the above equation can be fulfilled only if all coefficients are equal to zero. Therefore, the term within the brackets has to vanish, from which

$$v^m = \hat{A}_i^m \hat{v}^i$$

is deduced. Applying the same steps, starting again from (3.7) but replacing the $\mathbf{g}_i$ on the left-hand side with (3.6), leads finally to

$$\hat{v}^m = A_i^m v^i \,,$$

which is left as an exercise. The relation governing a change of basis are summarized in Box 3.4.1.

Box 3.4.1 Change of basis

$$\hat{\mathbf{g}}_i = \hat{A}_i^m \mathbf{g}_m \qquad\qquad \mathbf{g}_i = A_i^m \hat{\mathbf{g}}_m$$

$$\hat{v}^i = A_i^m v^m \qquad\qquad \hat{v}^i = A_i^m v^m$$

$$\hat{A}_i^m A_m^k = \delta_i^k$$

3.5 Linear Mappings Between Vector Spaces

Definition 3.10 (*Linear mapping*) Given two vector spaces $\mathcal{V}$ and $\mathcal{W}$. A mapping $\varphi : \mathcal{V} \rightarrow \mathcal{W}$ is linear if

$$\varphi(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha\varphi(\mathbf{u}) + \beta\varphi(\mathbf{v})$$

holds for any $\alpha, \beta \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v} \in \mathcal{V}$.

Remark 3.3 Definition 3.10 makes use of operator overloading. Without the latter, and given $\mathcal{V} = \{\mathcal{A}, \oplus, \odot\}$ and $\mathcal{V} = \{\mathcal{B}, \boxplus, \boxdot\}$ where $\mathcal{A}$ and $\mathcal{B}$ are some sets, linearity reads as

$$\varphi(\alpha \odot \mathbf{u} \oplus \beta \odot \mathbf{v}) = \alpha \boxdot \varphi(\mathbf{u}) \boxplus \beta \boxdot \varphi(\mathbf{v}) \,,$$

which makes it rather difficult to capture the very essence of the definition quickly.

Like all mappings, linear mappings between vector spaces can be injective, surjective, or bijective (see Definitions 1.5–1.7). Bijective linear mappings are of particular importance, since their existence makes the corresponding vector spaces algebraically indistinguishable or “isomorphic”.

Definition 3.11 (*Isomorphism*) An isomorphism between vector spaces $\mathcal{V}$ and $\mathcal{W}$ is a bijective linear mapping. Vector spaces for which an isomorphism exists are called isomorphic, which is denoted by $\mathcal{V} \cong \mathcal{W}$.

Example 3.1 Let $\mathcal{V}$ be the vector space of geometric vectors in the plane, as discussed in Sect. 1.2. Although we might have graphical objects in mind, graphical operations can be replaced by operations with real numbers in $\mathbb{R}^2$ simply because $\mathcal{V} \cong \mathbb{R}^2$.

Important properties of a linear mapping φ follow from its two key characteristics, namely its kernel and its image.

Definition 3.12 (*Kernel of a linear mapping*) Given two vector spaces $\mathcal{V}$ and $\mathcal{W}$ and a mapping $\varphi : \mathcal{V} \to \mathcal{W}$. The kernel of φ, denoted by $\ker \varphi$, is the set of all $\mathbf{v} \in \mathcal{V}$ for which $\varphi(\mathbf{v}) = \mathbf{0}$ holds, specifically

$$\ker \varphi = \{\mathbf{v} \in \mathcal{V} | \varphi(\mathbf{v}) = \mathbf{0}\}$$

using set notation.

Definition 3.13 (*Image of a linear mapping*) Given two vector spaces $\mathcal{V}$ and $\mathcal{W}$ and a mapping $\varphi : \mathcal{V} \to \mathcal{W}$. The image of φ, denoted by $\operatorname{Im} \varphi$, is the set of all $\mathbf{w} \in \mathcal{W}$ for which there exist at least one $\mathbf{v} \in \mathcal{W}$ with $\varphi(\mathbf{v}) = \mathbf{w}$, specifically

$$\operatorname{Im} \varphi = \{\mathbf{w} \in \mathcal{W} | \mathbf{w} = \varphi(\mathbf{v}), \mathbf{v} \in \mathcal{V}\}$$

using set notation.

Considering two vector spaces $\mathcal{V}$ and $\mathcal{W}$ and a linear mapping $\varphi : \mathcal{V} \to \mathcal{W}$, the so-called rank-nullity theorem states that

$$\operatorname{rank} \varphi + \dim(\ker \varphi) = \dim(\mathcal{V}) \,, \tag{3.10}$$

with $\operatorname{rank} \varphi = \dim(\operatorname{Im} \varphi)$. Other important facts about linear mappings between vector spaces, some of them a direct consequence of (3.10), can be summarized as follows:

1. $\operatorname{Im} \varphi$ and $\ker \varphi$ are subspaces of their respective vector spaces.
2. If $\dim(\ker \varphi) = 0$, then φ is injective.
3. If $\dim(\mathcal{V}) = \dim(\mathcal{W})$, then φ is injective if it is surjective, and vice versa, which implies that finite vector spaces of equal dimensions are isomorphic.
4. The inverse of an isomorphism is an isomorphism.

Item one emphasizes the fact that the kernel and image of a linear mapping are not just subsets of the sets associated with the respective vector spaces but also vector spaces according to Definition 3.1. This has far reaching implications. It can be used, for instance, to filter out an important or unimportant part, respectively, of a vector space by means of a linear mapping. The second and third items allow for checking the bijectivity of a linear mapping rather easily. Furthermore, item three reveals that all finite-dimensional real vector spaces of dimension N are isomorphic to $\mathbb{R}^N$. That vector spaces are algebraically indistinguishable does not mean that the corresponding operations are equally comfortable to perform (see, e.g., Example 3.1). An isomorphism φ in the spirit of Example 3.1 is really advantageous only, if the results can be transferred back to the original vector space of interest. This requires the existence of the inverse φ^{-1} and the possibility of deriving φ^{-1} explicitly. Fortunately, this is the case which is the essence of item four above. Regarding the respective proofs, see, e.g., Axler [1] or Lang [2].

Remark 3.4 The concept of linearity is of outstanding importance. However, as with most fundamental concepts, the passage from an intuitive understanding to a rigorous algebraic definition is by no means entirely obvious from the very beginning. Definition 3.10 reveals that linearity cannot be defined without a notion of summation. In this context, it might be helpful to remember that multiplication with a real number is a generalization of adding things n times, $n \in \mathbb{N}$. The reader is encouraged to revise Definition 3.1 in order to see that it relies entirely on the concept of linearity, and hence vector space is a synonym for linear space.

3.6 Linear Forms and the Dual Vector Space

Definition 3.14 (*Linear form*) A linear form $\underset{\sim}{\omega}$ is a linear mapping $\underset{\sim}{\omega} : \mathcal{V} \rightarrow \mathbb{R}$.

The space $\mathcal{P}^M$ as defined in Definition 3.9 with $M = 2$ is used to discuss the implications of the concept in more detail. Evaluating the polynomial $p(x) \in \mathcal{P}^2$ with $p(x) = p^0 + p^1 x + p^2 (x)^2$ for some number $\hat{x}$

$$\mathrm{eval}_{\hat{x}}(p(x)) = p^0 + p^1 \hat{x} + p^2 \left(\hat{x}\right)^2$$

is a linear form but also integration with limits

$$\int_a^b \mathrm{d}x (p(x)) = p^0[b-a] + p^1 \frac{1}{2}[(b)^2 - (a)^2] + p^2 \frac{1}{3}[(b)^3 - (a)^3]$$

and differentiation at some $\hat{x}$

$$\left.\frac{\mathrm{d}}{\mathrm{d}x}\right|_{x=\hat{x}} (p(x)) = p^0 0 + p^1 + 2p^2 \hat{x}$$

are also linear forms over $\mathcal{P}^2$. The notation used above differs from the common notation for integrals and derivatives in order to stress explicitly that these operations require an argument. Keep in mind that $(y)^{\mathrm{k}}$ denotes the k-th power of y.

The example illustrates that there are infinitely many linear forms over a given vector space $\mathcal{V}$. Interpreting all these linear forms as members of one set, the question arises as to whether this set possesses an algebraic structure or if some structure can be imposed on it. As a first step, linearity in Definition 3.14 can be explored as

$$\underset{\sim}{\boldsymbol{\omega}}(\mathbf{v}) = \underset{\sim}{\boldsymbol{\omega}}(v^k \mathbf{g}_k) = v^k \underset{\sim}{\boldsymbol{\omega}}(\mathbf{g}_k) \tag{3.11}$$

for $\mathbf{v} \in \mathcal{V}$ and some basis $\{\mathbf{g}_i\}$ of $\mathcal{V}$. Since the $\underset{\sim}{\boldsymbol{\omega}}(\mathbf{g}_k)$ have to be real numbers,

$$\omega_k = \underset{\sim}{\boldsymbol{\omega}}(\mathbf{g}_k)$$

is used in the following. Hence, (3.11) reads as

$$\underset{\sim}{\boldsymbol{\omega}}(\mathbf{v}) = v^k \omega_k \,,$$

and the following observations can be made. First of all, any linear form can be represented by a sum of N products, where $N = \dim(\mathcal{V})$. Furthermore, different linear forms are obviously represented by different values for the ω_i and the question arises as to whether these representations are unique. In other words, can the same linear form be represented by more than one set of coefficients? Considering two linear forms $\underset{\sim}{\boldsymbol{\omega}}$ and $\underset{\sim}{\boldsymbol{\eta}}$, $\underset{\sim}{\boldsymbol{\omega}}(\mathbf{v}) = \underset{\sim}{\boldsymbol{\eta}}(\mathbf{v})$ leads to

$$v^k \omega_k = v^k \eta_k$$

and

$$v^k [\omega_k - \eta_k] = 0 \,,$$

respectively. Since this holds for any $\mathbf{v} \in \mathcal{V}$, two linear forms are equal only if their respective coefficients are equal. This, on the other hand, implies that the representation of a linear form $\underset{\sim}{\boldsymbol{\omega}}$ by the numbers ω_i is unique.

The next step consists in considering particular linear forms, namely those which read of the j-th component of a given vector. For reasons which will become clear later, we denote these linear forms by $\underset{\sim}{\boldsymbol{g}}^j$, and

$$\underset{\sim}{\boldsymbol{g}}^j(\mathbf{v}) = v^j \tag{3.12}$$

defines them explicitly. Using (3.12) in (3.11) reads as

$$\underset{\sim}{\boldsymbol{\omega}}(\mathbf{v}) = \omega_k \, \underset{\sim}{\boldsymbol{g}}^k(\mathbf{v}) \,, \tag{3.13}$$

the implications of which are illustrated for $\dim(\mathcal{V}) = 2$. In this case, (3.13) reads as

$$\underset{\sim}{\boldsymbol{\omega}}(\mathbf{v}) = \omega_1 \, \underset{\sim}{\boldsymbol{g}}^1(\mathbf{v}) + \omega_2 \, \underset{\sim}{\boldsymbol{g}}^2(\mathbf{v}) \,, \tag{3.14}$$

which in view of Sect. 3.2 indicates that the linear forms over $\mathcal{V}$ might form a vector space. In order to exploit this idea further, the forms have to be separated from their arguments. This, on the other hand, requires defining a summation of linear forms indirectly by means of (3.14). This also has to be done for multiplication of a linear form with a real number and is part of the following definition.

Definition 3.15 (*Dual space*) The set of all linear forms over a vector space $\mathcal{V}$, together with two operations "$\underset{\sim}{\oplus}$" and "$\underset{\sim}{\odot}$", is called the dual space of $\mathcal{V}$ and denoted by $\mathcal{V}^*$. The operations are defined according to

$$\left[\underset{\sim}{\omega}\underset{\sim}{\oplus}\underset{\sim}{\eta}\right](\mathbf{v}) = \underset{\sim}{\omega}(\mathbf{v}) + \underset{\sim}{\eta}(\mathbf{v}) \tag{3.15}$$

$$\left[\lambda\underset{\sim}{\odot}\underset{\sim}{\omega}\right](\mathbf{v}) = \underset{\sim}{\omega}(\lambda\mathbf{v}) \,, \tag{3.16}$$

and the elements of $\mathcal{V}^*$ are called co-vectors or dual vectors, respectively.

The meaning of the operations "$\underset{\sim}{\oplus}$" and "$\underset{\sim}{\odot}$" depends on $\mathcal{V}$. It is important in the following to distinguish between the mapping itself and the result when evaluated for the respective argument. That

$$\dim(\mathcal{V}) = \dim(\mathcal{V}^*)$$

is clear due to the observation discussed after (3.11) together with (3.13). Furthermore, (3.12) implies that

$$\underset{\sim}{g}^i(\mathbf{g}_j) = \delta^i_j$$

and the $\underset{\sim}{g}^i$ are called dual base vectors. Here, the advantage of using the index positions in order to transmit additional information by notation becomes apparent. That the index position alone is not sufficient to distinguish between base vectors and dual base vectors is due to the existence of the so-called reciprocal basis which will be discussed later.

In the following, the notations $\underset{\sim}{\omega} + \underset{\sim}{\nu} := \underset{\sim}{\omega}\underset{\sim}{\oplus}\underset{\sim}{\nu}$ and $\lambda\underset{\sim}{\omega} := \lambda\underset{\sim}{\odot}\underset{\sim}{\omega}$ are used as long as this operator overloading does not cause confusion.

The properties of $\underset{\sim}{\omega}(\mathbf{v})$ (see (3.11)) suggest an alternative notation which emphasizes the pairing of a vector and a co-vector

$$\left\langle\underset{\sim}{\omega}, \mathbf{v}\right\rangle := \underset{\sim}{\omega}(\mathbf{u}) \,, \tag{3.17}$$

where $\left\langle\underset{\sim}{\omega}, \mathbf{v}\right\rangle$ is called dual product since it is distributive regarding addition defined for $\mathcal{V}$ and $\mathcal{V}^*$. The coordinates of vectors and dual vectors can be expressed by means of (3.17), and the corresponding base vectors as

$$v^i = \left\langle \underset{\sim}{g}_i, \mathbf{v} \right\rangle \tag{3.18}$$

$$\omega_i = \left\langle \underset{\sim}{\omega}, \mathbf{g}_i \right\rangle , \tag{3.19}$$

which will be used extensively in the follow-up.

Example 3.2 Let $\mathcal{V} = \mathcal{P}^2$ with $\{\mathbf{g}_i\} = \{1, x, (x)^2\}$. The corresponding dual basis is given by $\left\{\text{eval}_0, \frac{\mathrm{d}}{\mathrm{d}x}\Big|_0, \frac{(\mathrm{d})^2}{\mathrm{d}(x)^2}\Big|_0\right\}$.

Remark 3.5 (*Importance of the dual space in mechanics*) Imagine the state of a system is described by a scalar function, e.g., work which depends on the actual position expressed by a vector $\mathbf{x}(\tau)$ where τ stands for time. Derivative of W wrt. time at a specific time t leads to the concept of power, which is actually a linearization concept. Although derivatives are not discussed more rigorously until Chap. 6, it is clear from previous education in analysis that a kind of chain rule will be involved by which the vector $\dot{\mathbf{x}}(t)$ appears. The result of the time derivative is a real number, obtained by pairing the vector $\dot{\mathbf{x}}(t)$ with something else, hence

$$\dot{W}(\mathbf{x}(t)) = \left\langle \frac{\mathrm{d}W}{\mathrm{d}\mathbf{x}}\bigg|_{\mathbf{x}(\tau=t)}, \dot{\mathbf{x}}(t) \right\rangle .$$

If $\dot{\mathbf{x}}$ is a vector paired with something and the result is a real number, then $\frac{\mathrm{d}W}{\mathrm{d}\mathbf{x}}\Big|_{\mathbf{x}(\tau=t)}$ is a co-vector. In mechanics, $\frac{\mathrm{d}W}{\mathrm{d}\mathbf{x}}\Big|_{\mathbf{x}(\tau=t)}$ is interpreted as a force. Nowadays, it is therefore understood that forces should be encoded mathematically as dual vectors rather than as vectors.

3.7 The Inner Product, Norm, and Metric

Definition 3.16 (*Inner product*) A mapping

$$\begin{aligned} \cdot :\ \mathcal{V} \times \mathcal{V} &\to \mathbb{R} \\ (\mathbf{a}, \mathbf{b}) &\mapsto \mathbf{a} \cdot \mathbf{b} \end{aligned}$$

with the properties:

(i) $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$
(ii) $(\alpha\mathbf{a} + \beta\mathbf{b}) \cdot \mathbf{c} = \alpha\mathbf{a} \cdot \mathbf{c} + \beta\mathbf{b} \cdot \mathbf{c}$
(iii) $\mathbf{a} \neq \mathbf{0} \Rightarrow \mathbf{a} \cdot \mathbf{a} > 0$

for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{V}$ und $\alpha, \beta \in \mathbb{R}$ is called an inner product. A vector space equipped with an inner product is called an inner product space or pre-Hilbert space.

Example 3.3 Here again, definitions are derived from studying different situations and extracting common patterns. This leads eventually to definitions providing a minimum amount of information from which other facts can be deduced by logical reasoning. For instance, that for $\mathbf{v} \neq \mathbf{0}$, $\mathbf{u} \cdot \mathbf{v} = 0$ implies $\mathbf{u} = \mathbf{0}$ is not part of the definition. However, using the fact that $\mathbf{u} + \mathbf{0} = \mathbf{u}$ leads to

$$\mathbf{u} \cdot \mathbf{v} + \mathbf{0} \cdot \mathbf{v} = 0 \,.$$

Since $\mathbf{u} \cdot \mathbf{v} = 0$, the only inner product involving $\mathbf{v}$ which evaluates to zero is the one with the zero vector. Hence, $\mathbf{u} = \mathbf{0}$.

According to Definition 3.16, the inner product $\mathbf{u} \cdot \mathbf{v}$ with $\mathbf{u}, \mathbf{v} \in \mathcal{V}$ takes two arguments and it is linear in both. This is referred to as bilinear mapping. If one argument is fixed, e.g., $\mathbf{u} = \hat{\mathbf{u}}$, then the inner product becomes a linear form. Because the dual space $\mathcal{V}^*$ is the space of all linear forms, there has to be a corresponding dual vector $\underset{\sim}{\eta} \in \mathcal{V}^*$ by which exactly the same result can be obtained, specifically

$$\left\langle \underset{\sim}{\eta}, \mathbf{v} \right\rangle = \hat{\mathbf{u}} \cdot \mathbf{v} \,.$$

In other words, a mapping

$$\begin{aligned} \mathcal{G} : \mathcal{V} &\rightarrow \mathcal{V}^* \\ \hat{\mathbf{u}} &\mapsto \underset{\sim}{\eta} = \mathcal{G}(\hat{\mathbf{u}}) \end{aligned} \tag{3.20}$$

must exist such that

$$\left\langle \mathcal{G}(\hat{\mathbf{u}}), \mathbf{v} \right\rangle = \hat{\mathbf{u}} \cdot \mathbf{v} \,. \tag{3.21}$$

Linearity of $\mathcal{G}$ is implied by the linearity of the inner product together with Definition 3.15. Bijectivity follows from $\ker \mathcal{G} = \underset{\sim}{0}$ and $\dim(\mathcal{V}) = \dim(\mathcal{V}^*)$. Therefore, the inner product induces an isomorphism between $\mathcal{V}$ and$\mathcal{V}^*$ (see also Exercise 3.4).

The importance of the inner product consists in the fact that it induces a norm. The latter is a generalization of measuring length.

Definition 3.17 (*Norm*) A norm $||.||$ on a vector space $\mathcal{V}$ is a mapping with the properties:

(i) $||\alpha\mathbf{v}|| = \alpha||\mathbf{v}||$
(ii) $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$
(iii) $||\mathbf{v}|| = 0$ implies $\mathbf{u} = \mathbf{0}$

for $\alpha \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v} \in \mathcal{V}$.

The norm of a vector $\mathbf{u}$ induced by the inner product is simply

$$||\mathbf{u}|| = \sqrt{\mathbf{u} \cdot \mathbf{u}} \,,$$

and a norm on the other hand induces a metric by means of the distance function

$$d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$$

(see also Definition 2.1). Therefore, $\mathcal{V}$ becomes a metric space in the presence of an inner product.

3.8 The Reciprocal Basis and Its Relations with the Dual Basis

Definition 3.18 (*Reciprocal basis*) Given a vector space $\mathcal{V}$ with inner product and some basis $\{\mathbf{g}_i\}$. The reciprocal basis $\{\mathbf{g}^i\}$ is defined by

$$\mathbf{g}_i \cdot \mathbf{g}^j = \delta_i^j$$

with $\mathbf{g}_i, \mathbf{g}^j \in \mathcal{V}$.

Box 3.8.1 Relation between inner product and dual product

vector space $\mathcal{V}$		**dual space** $\mathcal{V}^*$
$\mathbf{u} \in \mathcal{V}$	$\underset{\sim}{\omega} : \mathcal{V} \rightarrow \mathbb{R}$	$\underset{\sim}{\omega} \in \mathcal{V}^*$
basis $\{\mathbf{g}_i\}$		basis $\{\underset{\sim}{g}^i\}$

<u>common notations ($\alpha \in \mathbb{R}$):</u>

$$\begin{aligned} \alpha &= \underset{\sim}{\omega}(\mathbf{u}) \\ \alpha &= \left\langle \underset{\sim}{\omega}, \mathbf{u} \right\rangle \\ &\text{or} \quad \text{even} \\ \alpha &= \underset{\sim}{\omega} \cdot \mathbf{u} \end{aligned}$$

If $\mathcal{V}$ is an inner product space then the reciprocal basis $\{\mathbf{g}^j\}$ is defined by $\mathbf{g}_i \cdot \mathbf{g}^j = \delta_i^j$. Furthermore,

$$\left\langle \underset{\sim}{g}^i, \mathbf{v} \right\rangle = \mathbf{g}^i \cdot \mathbf{v}$$

holds.

Using the knowledge of the previous subsection about the isomorphism established by the existence of an inner product,

$$\left\langle \mathcal{G}(\mathbf{g}^i), \mathbf{v} \right\rangle = \mathbf{g}^i \cdot (v^k \mathbf{g}_k) = \mathbf{g}^i \cdot \mathbf{g}_k v^k = v^i \tag{3.22}$$

where the properties of the reciprocal basis were used. On the other hand,

$$\left\langle \underset{\sim}{g}^i, \mathbf{v} \right\rangle = v^i \tag{3.23}$$

holds, and comparison of (3.22) with (3.23) yields

$$\mathcal{G}(\mathbf{g}^i) = \underset{\sim}{g}^i \; ,$$

from which

$$\left\langle \underset{\sim}{g}^i, \mathbf{v} \right\rangle = \mathbf{g}^i \cdot \mathbf{u} \tag{3.24}$$

can be deduced. This does not mean that dual basis and reciprocal basis are the same thing. It just means that $\left\langle \underset{\sim}{g}^i, \mathbf{u} \right\rangle$ could be replaced by $\mathbf{g}^i \cdot \mathbf{u}$ because it gives the same result, the same real number, to be precise. The relations between inner product and dual product are summarized in Box 3.8.1 together with notations commonly employed for the dual product.

Remark 3.6 The result (3.24) is often used to abandon the concept of the dual space completely, which is by no means a good idea, because most of the underlying information becomes invisible without getting any benefit in exchange.

In the presence of an inner product, the so-called metric coefficients g_{ij} and g^{ij} can be defined by

$$g_{ij} = \mathbf{g}_i \cdot \mathbf{g}_j \tag{3.25}$$
$$g^{ij} = \mathbf{g}^i \cdot \mathbf{g}^j \; , \tag{3.26}$$

which will play an important role later. A basis $\{\mathbf{E}_i\}$ with the property

$$\mathbf{E}_i \cdot \mathbf{E}_j = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \phantom{\text{if}} \quad i \neq j \end{cases} .$$

is called orthonormal. In this case, base vectors and reciprocal vectors coincide. This subject will be enlarged upon later (see Sect. 6.3).

Exercises

3.1 Show that the set of all positive real numbers $\mathbb{R}^+$

$$\mathbb{R}^+ := \{x \mid x \in \mathbb{R} \;\text{ and }\; x > 0\}$$

together with the operations

$$a \oplus b := a \cdot b$$
$$\lambda \odot a := a^{\lambda}$$

with $a, b \in \mathbb{R}^{+}, \lambda \in \mathbb{R}$ forms a real vector space. Hints: show that $\mathbb{R}^{+}$ is closed under the given operations and that the operations obey the rules defined for vector spaces. Furthermore, determine explicitly the neutral element for the operation $\oplus$, the neutral element for the operation $\odot$ and the inverse with respect to the operation $\oplus$.

3.2 Consider $\mathcal{V} = \mathbb{R}^2$. The elements of $\mathcal{V}$ are pairs of real numbers, e.g., (x^1, x^2). Choose $\mathbf{g}_1$ and $\mathbf{g}_1$ and deduce the meanings of $\underset{\sim}{\oplus}$ and $\underset{\sim}{\otimes}$ for this case.

3.3 Determine the transformation relationship for the dual base vectors in case of a change of basis (see Box 3.4.1).

3.4 Show that $\mathcal{G}$ as defined by (3.20) and (3.21) is an isomorphism.

3.5 Show that $\int_0^1 \mathrm{d}x \; p(x)q(x)$ is a suitable definition for an inner product in $\mathcal{P}^2$.

3.6 Evaluate explicitly δ^i_i, $\delta^i_i \delta^j_j$, and $\delta^j_i \delta^i_j$ for $i, j, k = 1, 2, 3$.

3.7 Simplify $A^k_j \delta^j_k$, $\delta^j_i \delta^k_j$, and $\delta^j_i A_j - \delta^k_i A_k$.

3.8 Given the space $\mathcal{W} = \mathcal{V}^* \times \mathcal{V}$ with elements $(\underset{\sim}{\omega}, \mathbf{u})$ where $\underset{\sim}{\omega} \in \mathcal{V}^*$, $\mathbf{u} \in \mathcal{V}$. $\mathcal{W}$ becomes a vector space by defining

$$(\underset{\sim}{\omega}, \mathbf{u}) \oplus (\underset{\sim}{\nu}, \mathbf{v}) = (\underset{\sim}{\omega} + \underset{\sim}{\nu}, \mathbf{u} + \mathbf{v}) \,,$$
$$\lambda \odot (\underset{\sim}{\omega}, \mathbf{u}) = (\lambda \underset{\sim}{\omega}, \lambda \mathbf{u})$$

with $\lambda \in \mathbb{R}$, $\mathbf{v} \in \mathcal{V}$, $\underset{\sim}{\nu} \in \mathcal{V}^*$. Show that $\dim(\mathcal{V}) = 3$ implies $\dim(\mathcal{W}) = 6$.

Hint: Since $(\underset{\sim}{\omega}, \mathbf{u}) = \omega_i v^k (\underset{\sim}{g}^i, \mathbf{g}_k)$, the question can be answered by determining the number of linearly independent pairs $(\underset{\sim}{g}^i, \mathbf{g}_k)$.

3.9 Given a vector space $\mathcal{V}$ with inner product. Use the relationship

$$\mathbf{g}_i \cdot \mathbf{g}^j \equiv \delta_i^{\ j} \qquad (\mathbf{g}_i, \mathbf{g}^j \in \mathcal{V})$$

1. to determine the component matrix B_{ij} for the basis transformation $\mathbf{g}_i = B_{ij} \, \mathbf{g}^j$,
2. to prove the identity $g_{ik} \, g^{kj} = \delta_i^{\ j}$ where the $g^{ij} := \mathbf{g}^i \cdot \mathbf{g}^j$ and $g_{ij} := \mathbf{g}_i \cdot \mathbf{g}_j$ are the so-called metric coefficients.

References

1. Axler S (2010) Linear algebra done right. Undergraduate texts in mathematics. Springer, Berlin
2. Lang S (1997) Introduction to linear algebra. Undergraduate texts in mathematics. Springer, Berlin
3. Valenza R (1993) Linear algebra: an introduction to abstract mathematics. Undergraduate texts in mathematics. Springer, Berlin
4. Winitzki S (2010) Linear algebra via exterior products. www.lulu.com

Chapter 4
Tensor Algebra

Abstract Dyadic product and tensors are introduced in the context of bilinear forms before extending this scheme to arbitrary but finite dimensions. Afterwards, tensor product spaces are defined. The exterior product is motivated within this chapter by the aim to generalize the notion of volume for arbitrary dimensions and to overcome the limitations implied by the cross product of conventional vector calculus. Within this context, symmetric and skew-symmetric tensors, as well as a generalized version of the Kronecker symbol, are discussed. Furthermore, basic aspects of the so-called star-operator are examined. The latter relates spaces of alternating tensors of equal dimension based on the existence of an inner product.

4.1 Tensors and Multi-linear Forms

The coordinate representation of tensors involves systematic indexing, which makes the use of tensors of arbitrary order somewhat difficult at the beginning. Hence, the fundamental relations will be derived first for second order tensors.

Consider the vector space $\mathcal{V}$ and its dual $\mathcal{V}^*$ without any inner product together with the following linear mapping:

$$\mathsf{A} : \mathcal{V}^* \times \mathcal{V} \to \mathbb{R} . \tag{4.1}$$

Due to the linearity, applying the mapping A for two particular arguments $\underset{\sim}{\boldsymbol{\omega}} \in \mathcal{V}^*$ and $\boldsymbol{v} \in \mathcal{V}$ reads as

$$\alpha = \mathsf{A}(\underset{\sim}{\boldsymbol{\omega}}, \boldsymbol{v}) = \omega_i v^j \; \mathsf{A}(\underset{\sim}{\boldsymbol{g}}^i, \boldsymbol{g}_j) , \tag{4.2}$$

where α is the result of the operation, and according to the definition above, $\alpha \in \mathbb{R}$. Because the objects on both sides of (4.2) have to be real numbers, the $\mathsf{A}(\underset{\sim}{\boldsymbol{g}}^i, \boldsymbol{g}_j)$ must be real numbers. If the abbreviation

$$A^i_{\;j} = \mathsf{A}(\underset{\sim}{\boldsymbol{g}}^i, \boldsymbol{g}_j)$$

U. Mühlich, *Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds*, Solid Mechanics and Its Applications 230,
DOI 10.1007/978-3-319-56264-3_4

is used, (4.2) can be written as

$$\alpha = \mathsf{A}(\underset{\sim}{\boldsymbol{\omega}}, \boldsymbol{v}) = A^i_{\ j}\ \omega_i\ v^j\ ,$$

and it can be shown that the right hand side is invariant under a change of basis, which is left as an exercise to the reader. Again, the idea is to isolate the mapping (operation) from its arguments. This can be done by means of the dual product, which can be used to read off the coordinates of a vector or a covector, respectively. Application of (3.18) and (3.19) yields

$$\mathsf{A}(\underset{\sim}{\boldsymbol{\omega}}, \boldsymbol{v}) = A^i_{\ j}\ \langle \underset{\sim}{\boldsymbol{\omega}}, \boldsymbol{g}_i \rangle \left\langle \underset{\sim}{\boldsymbol{g}}^i, \boldsymbol{v} \right\rangle . \tag{4.3}$$

The term $\langle \underset{\sim}{\boldsymbol{\omega}}, \mathbf{g}_i \rangle \left\langle \underset{\sim}{\boldsymbol{g}}^i, \mathbf{v} \right\rangle$ is a bilinear form itself and written symbolically as

$$\langle \underset{\sim}{\boldsymbol{\omega}}, \mathbf{g}_i \rangle \left\langle \underset{\sim}{\boldsymbol{g}}^i, \mathbf{v} \right\rangle = \mathbf{g}_i \otimes \underset{\sim}{\boldsymbol{g}}^j (\underset{\sim}{\boldsymbol{\omega}}, \mathbf{v})\,, \tag{4.4}$$

where "$\otimes$" is the so-called dyadic product. The mapping (4.1) can therefore be represented by the object

$$\mathsf{A} = A^i_{\ j}\ \mathbf{g}_i \otimes \underset{\sim}{\boldsymbol{g}}^j\ ,$$

and $\mathsf{A}(\underset{\sim}{\boldsymbol{\omega}}, \boldsymbol{v})$ again yields (4.3) if the definition (4.4) is used. A is called a second order tensor.

Of course, the right hand side of (4.3) consists of real numbers, the order of which can be chosen arbitrarily. However, here the convention is used that the order of the arguments according to the definition (4.1) determines the order of the base vectors and dual base vectors, respectively, such that the n-th argument meets the n-th base vector or dual base vector. As long as no inner product is defined for $\mathcal{V}$, the operation can be performed only if a vector argument meets a dual base vector, and vice versa.

Obviously, this scheme can be used to encode other possible bilinear forms. The results are summarized as follows:

$$\begin{aligned}
&\mathsf{A}: \mathcal{V}^* \times \mathcal{V} \rightarrow \mathbb{R} \qquad && \mathsf{A} = A^i_{\ j}\ \boldsymbol{g}_i \otimes \underset{\sim}{\boldsymbol{g}}^j \\
&\mathsf{B}: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R} && \mathsf{B} = B_{ij}\ \underset{\sim}{\boldsymbol{g}}^i \otimes \underset{\sim}{\boldsymbol{g}}^j \\
&\mathsf{C}: \mathcal{V} \times \mathcal{V}^* \rightarrow \mathbb{R} && \mathsf{C} = C_i^{\ j}\ \underset{\sim}{\boldsymbol{g}}^i \otimes \boldsymbol{g}_j \\
&\mathsf{D}: \mathcal{V}^* \times \mathcal{V}^* \rightarrow \mathbb{R} && \mathsf{D} = D^{ij}\ \boldsymbol{g}_i \otimes \boldsymbol{g}_j\ .
\end{aligned}$$

Remark 4.1 Regarding the RHS of (4.4), the notation $\mathbf{g}_i\ \ \underset{\sim}{\boldsymbol{g}}^j\ \ (\underset{\sim}{\boldsymbol{\omega}}, \mathbf{v})$ might seem more natural at this point, and, in fact, it can be found in the literature as well. The appearance of $\otimes$ in (4.4) and calling $\mathbf{g}_i \otimes \underset{\sim}{\boldsymbol{g}}^j$ a product is justified by the distributive property of this construction, to be discussed in more detail later.

The generalization from bilinear forms discussed above to multi-linear forms is straight forward.

Definition 4.1 (*Tensor as multi-linear form*) A tensor $\mathsf{T}_{\binom{n}{m}}$ is a mapping

$$\mathsf{T}_{\binom{m}{n}} : \underbrace{\mathcal{V}^* \times \cdots \times \mathcal{V}^*}_{m\text{-times}} \times \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_{n\text{-times}} \to \mathbb{R}\,, \tag{4.5}$$

which is linear in all of its arguments.

Since m and n can be arbitrary natural numbers, it is advisable to work with so-called multi-indices. This will be illustrated by means of a $\mathsf{T}_{\binom{2}{2}}$ tensor. $\mathsf{T}_{\binom{2}{2}}$ can be expressed by

$$\mathsf{T}_{\binom{2}{2}} = T^{ij}_{kl}\, \mathbf{g}_i \otimes \boldsymbol{g}_j \otimes \underset{\sim}{\boldsymbol{g}}^k \otimes \underset{\sim}{\boldsymbol{g}}^l\,. \tag{4.6}$$

Performing the replacements

$$i \to i_1 \quad , \quad j \to i_2 \quad , \quad k \to j_1 \quad , \quad l \to j_2\,, \tag{4.7}$$

$\mathsf{T}_{\binom{2}{2}}$ reads as

$$\mathsf{T}_{\binom{2}{2}} = T^{i_1 i_2}_{j_1 j_2}\, \mathbf{g}_{i_1} \otimes \mathbf{g}_{i_2} \otimes \underset{\sim}{\boldsymbol{g}}^{j_1} \otimes \underset{\sim}{\boldsymbol{g}}^{j_2}\,.$$

As long as m and n are small, a representation like the one in (4.6) is preferable. However, for m, n large or arbitrary, the following representation of the tensor related to Definition 4.1 is possible:

$$\mathsf{T}_{\binom{m}{n}} = T^{i_1 \ldots i_m}_{j_1 \ldots j_n}\, \mathbf{g}_{i_1} \cdots \otimes \cdots \mathbf{g}_{i_m} \otimes \underset{\sim}{\boldsymbol{g}}^{j_1} \cdots \otimes \cdots \underset{\sim}{\boldsymbol{g}}^{j_n}.$$

Definition 4.1 considers only a vector space $\mathcal{V}$ and its dual. However, more general cases are not only possible but also rather important. Therefore, more general settings are also discussed subsequently.

4.2 Dyadic Product and Tensor Product Spaces

The intention of the following discussion is to generalize the operation "$\otimes$", which first appeared in the definition of the dyadic product (4.4),

$$\mathbf{g}_i \otimes \underset{\sim}{\boldsymbol{g}}^j (\underset{\sim}{\boldsymbol{\omega}}, \boldsymbol{v}) = \left\langle \underset{\sim}{\boldsymbol{\omega}}, \mathbf{g}_i \right\rangle \left\langle \underset{\sim}{\boldsymbol{g}}^i, \mathbf{v} \right\rangle\,,$$

with $\mathbf{g}_i \in \mathcal{V}$, $\underset{\sim}{\boldsymbol{g}}^i \in \mathcal{V}^*$, for arbitrary elements of $\mathcal{V}$ and $\mathcal{V}^*$.

The dual product is a linear mapping, hence (4.4) is a bilinear mapping, meaning linear in both arguments. Because the $\mathbf{g}_i$ and $\underset{\sim}{\boldsymbol{g}}^j$ are themselves just vectors and dual vectors, (4.4) is adapted as follows:

$$\mathbf{u} \otimes \underset{\sim}{\boldsymbol{\eta}}(\underset{\sim}{\boldsymbol{\omega}}, \mathbf{v}) = u^i \eta_k \;\; \mathbf{g}_i \otimes \underset{\sim}{\boldsymbol{g}}^k \;\; (\underset{\sim}{\boldsymbol{\omega}}, \mathbf{v}) \,,$$

which indicates that "$\otimes$" can be defined in a more general way that includes (4.4).

Definition 4.2 (*Tensor space* $\mathcal{V} \otimes \mathcal{W}$) Given two vector spaces $\mathcal{V}$ and $\mathcal{W}$ with $\dim(\mathcal{V}) = N$, $\dim(\mathcal{W}) = K$. The tensor product space $\mathcal{V} \otimes \mathcal{W}$ is defined by the following properties:

(i) There exists a mapping "$\otimes$" called tensor product or dyadic product such that

$$\begin{aligned} \otimes : \mathcal{V} \times \mathcal{W} &\to \mathcal{V} \otimes \mathcal{W} \\ (\mathbf{u}, \mathbf{W}) &\mapsto \mathbf{u} \otimes \mathbf{W} \,. \end{aligned}$$

(ii) If $\{\mathbf{g}_i\}$ is a basis of $\mathcal{V}$ and $\{\mathbf{G}_i\}$ a basis of $\mathcal{W}$, then the $\{\mathbf{g}_i \otimes \mathbf{G}_j\}$ form a basis of $\mathcal{V} \otimes \mathcal{W}$. Every element T of $\mathcal{V} \otimes \mathcal{W}$ can therefore be expressed by

$$\mathsf{T} = T^{ij} \mathbf{g}_i \otimes \mathbf{G}_j \,.$$

(iii) Summation and multiplication with scalar are defined for $\mathcal{V} \otimes \mathcal{W}$ by

$$\begin{aligned} \mathsf{A} + \mathsf{B} &= [A^{ij} + B^{ij}] \mathbf{g}_i \otimes \mathbf{G}_j \\ \lambda \mathsf{A} &= \lambda A^{ij} \mathbf{g}_i \otimes \mathbf{G}_j \,. \end{aligned}$$

(iv) The mapping "$\otimes$" has the following properties:

$$\begin{aligned} \mathbf{u} \otimes (\mathbf{V} + \mathbf{W}) &= \mathbf{u} \otimes \mathbf{V} + \mathbf{u} \otimes \mathbf{W} \\ (\mathbf{u} + \mathbf{v}) \otimes \mathbf{W} &= \mathbf{u} \otimes \mathbf{W} + \mathbf{v} \otimes \mathbf{W} \\ (\lambda \mathbf{u}) \otimes \mathbf{W} &= \mathbf{u} \otimes (\lambda \mathbf{W}) = \lambda (\mathbf{u} \otimes \mathbf{W}) \,, \end{aligned}$$

with $\mathbf{u}, \mathbf{v} \in \mathcal{V}$, $\mathbf{V}, \mathbf{W} \in \mathcal{W}$, $\lambda, T^{ij} \in \mathbb{R}$, and $i = 1 \ldots N$, $j = 1 \ldots K$.

Due to (iii), $\mathcal{V} \otimes \mathcal{W}$ is a vector space. Obviously, $\mathbf{v} \otimes \mathbf{W}$, $\mathbf{v} \in \mathcal{V}$, $\mathbf{W} \in \mathcal{W}$ is an element of $\mathcal{V} \otimes \mathcal{W}$. However, not all elements of $\mathcal{V} \otimes \mathcal{W}$ are of this form. This can be seen easily through simple examples (see Exercise 4.3).

Remark 4.2 Without operator overloading, (iv) of Definition 4.2 would take the form

$$\begin{aligned} \mathbf{u} \otimes (\mathbf{V} \boxplus \mathbf{W}) &= \mathbf{u} \otimes \mathbf{V} \hat{\boxplus} \mathbf{u} \otimes \mathbf{W} \\ (\mathbf{u} \oplus \mathbf{v}) \otimes \mathbf{W} &= \mathbf{u} \otimes \mathbf{W} \hat{\boxplus} \mathbf{v} \otimes \mathbf{W} \\ (\lambda \odot \mathbf{u}) \otimes \mathbf{W} &= \mathbf{u} \otimes (\lambda \boxdot \mathbf{W}) = \lambda \hat{\boxdot} (\mathbf{u} \otimes \mathbf{W}) \end{aligned}$$

for $\{\mathcal{V}, \oplus, \odot\}$, $\{\mathcal{W}, \boxplus, \boxdot\}$, and $\{\mathcal{V} \otimes \mathcal{W}, \hat{\boxplus}, \hat{\boxdot}\}$, which makes it hardly readable.

Defining the spaces $\mathcal{V} \otimes \mathcal{W}^*$, $\mathcal{W}^* \otimes \mathcal{V}$, etc., works analogously to Definition 4.2. Furthermore, the tensor space concept can be generalized in a straightforward manner for dyadic products of more than two vector spaces.

The tensor space $\mathcal{V} \otimes \mathcal{V}$ can be written more concisely as $\otimes_2 \mathcal{V}$. If the tensor product is applied k times, then $\otimes_k \mathcal{V}$ is commonly used for the corresponding tensor space.

In order to illustrate the relation between multi-linear forms and tensor product spaces, the bilinear form

$$\mathsf{F} : \mathcal{V}^* \times \mathcal{W} \rightarrow \mathbb{R}\,, \tag{4.8}$$

which corresponds to

$$\mathsf{F} = F_i^{\,j}\, \mathbf{G}_j \otimes \underset{\sim}{\boldsymbol{g}}^i\,, \tag{4.9}$$

is considered together with a mapping $\varphi : \mathcal{W} \otimes \mathcal{V}^* \rightarrow \mathbb{R}$. Evaluating φ with respect to the argument $\underset{\sim}{\boldsymbol{\Omega}} \otimes \mathbf{v}$ and standard manipulation yields

$$\varphi(\underset{\sim}{\boldsymbol{\Omega}} \otimes \mathbf{v}) = \Omega_i\, v^k \varphi(\underset{\sim}{\boldsymbol{G}}^i \otimes \mathbf{g}_k) = \Phi_k^{\,i}\, \mathbf{G}_i \otimes \underset{\sim}{\boldsymbol{g}}^k(\underset{\sim}{\boldsymbol{\Omega}}, \mathbf{v})\,,$$

which shows that φ can be converted into a bilinear form similar to F in (4.8). Feeding φ and F with the same arguments gives equal results if

$$F_i^{\,j} = \Phi_i^{\,j} \tag{4.10}$$

holds. Since (4.10) has a unique solution for either F or Φ given, there exists a unique φ for every F. This is a particular case reflecting the so-called universal property of the tensor product illustrated in Fig. 4.1. It justifies a definition of a tensor as an object in its own right, i.e., as an element of a tensor product space

$$\mathsf{T}_{\binom{m}{n}} \;\in\; \underbrace{\mathcal{W} \otimes \cdots \otimes \mathcal{W}}_{m\text{-times}} \otimes \underbrace{\mathcal{V}^* \otimes \cdots \otimes \mathcal{V}^*}_{n\text{-times}}\,. \tag{4.11}$$

Equivalence with Definition 4.1 is ensured by the universal property.

Definition 4.2 relies on the existence of a basis. Once the construction of the tensor space is established, it can be shown that this construction is independent of a particular choice for a basis by means of the transformation properties of tensors.

Fig. 4.1 Commutative diagram illustrating the universal property of the tensor product

$$\begin{tikzcd}
\mathcal{V}^* \times \mathcal{W} \arrow[rr, "\mathsf{F}"] \arrow[d, "\otimes"'] & & \mathbb{R} \\
\mathcal{W} \otimes \mathcal{V}^* \arrow[urr, "\varphi"'] & &
\end{tikzcd}$$

Alternatively, basis-free constructions can be performed, an action which has the advantage that a tensor product can be established for arbitrary sets that might not even have a basis. However, these constructions are more involved and work via the notion of the free vector space and an appropriate quotient space. The interested reader is referred, e.g., to Lee [2].

As long as all involved vector spaces have the same dimension, which is often the case, the universal property ensures that it is sufficient to define tensors by Definition 4.1, since by choosing one vector space as $\mathcal{V}$, all others can be related to $\mathcal{V}$ by proper isomorphisms.

4.3 The Dual of a Linear Mapping

The dual A^* of a linear mapping $\mathsf{A} : \mathcal{V} \to \mathcal{W}$ is defined as follows:

$$\left\langle \underset{\sim}{\boldsymbol{\Omega}}, \mathsf{A}(\mathbf{v})\right\rangle = \left\langle \mathsf{A}^*(\underset{\sim}{\boldsymbol{\Omega}}), \mathbf{v}\right\rangle , \tag{4.12}$$

with $\mathbf{v} \in \mathcal{V}$ and $\omega \in \mathcal{V}^*$. Since the dual product is only defined between a vector space and its own dual space, A^* has to be a mapping

$$\mathsf{A}^* : \mathcal{W}^* \to \mathcal{V}^* .$$

Using the bases $\{\mathbf{g}_i\} \in \mathcal{V}$ and $\{\mathbf{G}_i\} \in \mathcal{W}$ together with their duals, A and A^* read as

$$\mathsf{A} = A^i_k \, \mathbf{G}_i \otimes \underset{\sim}{\boldsymbol{g}}^k$$
$$\mathsf{A}^* = A^{*i}_{\ \ k} \, \underset{\sim}{\boldsymbol{g}}^k \otimes \mathbf{G}_i \, ,$$

and direct evaluation of (4.12) yields

$$A^i_k = A^{*i}_{\ \ k} \, .$$

While the linear mapping $A : \mathcal{V} \to \mathcal{W}$ establishes a relation between two vector spaces, its dual establishes a corresponding relation between the respective dual spaces.

4.4 Remarks on Notation and Inner Product Operations

Regarding the notation employed so far, there is not much space for ambiguity. Given $\mathsf{T}_{\binom{0}{1}}(\mathbf{u})$, $\mathsf{T}_{\binom{1}{1}}(\mathbf{u})$ or $\mathsf{T}_{\binom{0}{2}}(\mathbf{u})$, Definition 4.1 or (4.11) indicate the corresponding operations

$$\mathsf{T}_{\binom{0}{1}}(\mathbf{u}) = T_i \underset{\sim}{g}^i (u^k \mathbf{g}_k) = T_i u^i$$
$$\mathsf{T}_{\binom{1}{1}}(\mathbf{u}) = T_i^{\ j} \mathbf{g}_j \otimes \underset{\sim}{g}^i (u^k \mathbf{g}_k) = T_i^{\ j} u^i \mathbf{g}_j$$
$$\mathsf{T}_{\binom{0}{2}}(\mathbf{u}) = T_{ij} \underset{\sim}{g}^i \otimes \underset{\sim}{g}^j (u^k \mathbf{g}_k) = T_{ij} u^j \underset{\sim}{g}^i ,$$

but an expression like $\mathsf{T}(\mathbf{u})$, for instance, does not provide sufficient information. It is also clear that $\mathsf{T}_{\binom{1}{0}}(\mathbf{u})$ and $\mathsf{T}_{\binom{2}{0}}(\mathbf{u})$ do not make sense because the dual product is defined for a dual vector paired with a vector.

If, however, dual vectors are replaced by reciprocal vectors and operations are expressed by means of the inner product, additional conventions are commonly introduced in order to avoid ambiguity. Of course, this is only possible for inner product spaces. Since in

$$T_i^{\ j} \mathbf{g}_j \otimes \mathbf{g}^i \cdot (u^k \mathbf{g}_k) , \tag{4.13}$$

$\mathbf{g}^j \cdot \mathbf{g}_k$ as well as $\mathbf{g}^i \cdot \mathbf{g}_k$ make sense, it has to be agreed upon as to how the operation should be performed. Furthermore, inner products of arbitrary order, e.g., a double inner product $\mathsf{A} \cdot \cdot \mathsf{B}$, can be defined. Here, we define for a single inner product between two tensors that the last vector/reciprocal vector of the first tensor meets the first vector/reciprocal vector of the second tensor. Furthermore, a multiple inner product proceeds to the left regarding the first tensor and to the right with respect to the second tensor. For example,

$$\mathsf{T}_{\binom{1}{1}} \cdot \mathbf{u} = T^{j}_{\ i} u^k \mathbf{g}_j \ (\mathbf{g}^i \cdot \mathbf{g}_k) = T_i^{\ j} u^i \mathbf{g}_j$$
$$\mathsf{T}_{\binom{2}{0}} \cdot \mathbf{u} = T^{ij} u^k \mathbf{g}_i \ (\mathbf{g}_j \cdot \mathbf{g}_k) = T_i^{\ j} u^i g_{jk} \ \mathbf{g}_i$$
$$\mathsf{T}_{\binom{1}{1}} \cdot \cdot \mathsf{T}_{\binom{2}{0}} = \left[T^{j}_{\ i} \, \mathbf{g}_j \otimes \mathbf{g}^i\right] \cdot \cdot \left[T^{kl} \, \mathbf{g}_k \otimes \mathbf{g}_l\right] = T^{j}_{\ i} T^{ik} g_{jl} ,$$

where the g_{ij} are the metric coefficients discussed in Sect. 3.8. Please note that other conventions are also possible, and used in the literature.

Two very common operations when working with inner product spaces and using the reciprocal basis instead of the dual basis are raising and lowering indices. A vector $\mathbf{u}$ can be expressed either by $\{\mathbf{g}_i\}$ or $\left\{\mathbf{g}^i\right\}$. Taking the inner product of both sides of

$$u^i \, \mathbf{g}_i = u_k \, \mathbf{g}^k$$

with $\mathbf{g}_l$ and $\mathbf{g}^l$, respectively, yields

$$u_l \ = u^k g_{kl} \quad , \quad u^l \ = u_k g^{kl} ,$$

which means that

$$\mathbf{u} = u^k \, \mathbf{g}_k = u_k g^{kl} \, \mathbf{g}_l = u^k g_{kl} \, \mathbf{g}^l = u_k \, \mathbf{g}^k .$$

This scheme extends in a straightforward manner to tensors of arbitrary order.

4.5 The Exterior Product and Alternating Multi-linear Forms

An area element, i.e., a parallelogram, in $\mathbb{R}^2$ is spanned by two vectors. Computing the area of the parallelogram essentially means assigning a real number to two vectors, hence, it is, in general, a mapping

$$\mathsf{A} : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}\,,$$

where $\dim(\mathcal{V}) = 2$. Having the principles related to the computation of an area in mind, the following list of demands can be specified:

- If $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent, then $\mathsf{A}(\mathbf{u}, \mathbf{v})$ should give the value 0.
- A should be able to distinguish between different orientations. This can be accomplished by demanding $\mathsf{A}(\mathbf{u}, \mathbf{v}) = -\mathsf{A}(\mathbf{v}, \mathbf{u})$.
- A should be linear in both arguments.
- A should not depend on the particular choice for a basis.

It turns out that a determinant which fulfils these requirements can be constructed. In order to ensure coordinate invariance, the co-vectors $\underset{\sim}{\boldsymbol{\omega}}, \underset{\sim}{\boldsymbol{\mu}} \in \mathcal{V}^*$ and respective dual products are involved in the construction. The result reads as

$$\mathsf{A}(\mathbf{u}, \mathbf{v}) = \det \begin{pmatrix} \langle \underset{\sim}{\boldsymbol{\omega}}, \mathbf{u} \rangle & \langle \underset{\sim}{\boldsymbol{\omega}}, \mathbf{v} \rangle \\ \langle \underset{\sim}{\boldsymbol{\mu}}, \mathbf{u} \rangle & \langle \underset{\sim}{\boldsymbol{\mu}}, \mathbf{v} \rangle \end{pmatrix} = \langle \underset{\sim}{\boldsymbol{\omega}}, \mathbf{u} \rangle \left\langle \underset{\sim}{\boldsymbol{\mu}}, \mathbf{v} \right\rangle - \langle \underset{\sim}{\boldsymbol{\omega}}, \mathbf{v} \rangle \left\langle \underset{\sim}{\boldsymbol{\mu}}, \mathbf{u} \right\rangle \,. \tag{4.14}$$

The use of two co-vectors is due to the fact that for just one co-vector A would obviously always vanish.

The RHS of (4.14) defines as a new algebraic operation denoted by $\wedge$

$$\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}}(\mathbf{u}, \mathbf{v}) = \langle \underset{\sim}{\boldsymbol{\omega}}, \mathbf{u} \rangle \left\langle \underset{\sim}{\boldsymbol{\mu}}, \mathbf{v} \right\rangle - \langle \underset{\sim}{\boldsymbol{\omega}}, \mathbf{v} \rangle \left\langle \underset{\sim}{\boldsymbol{\mu}}, \mathbf{u} \right\rangle \,,$$

which is known as exterior product, and $\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}}$ is called a bi-(co)vector or 2-form defined as

$$\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}} = \underset{\sim}{\boldsymbol{\omega}} \otimes \underset{\sim}{\boldsymbol{\mu}} - \underset{\sim}{\boldsymbol{\mu}} \wedge \underset{\sim}{\boldsymbol{\omega}}\,. \tag{4.15}$$

It can be checked easily that

$$\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}}(\mathbf{u}, \mathbf{v}) = \frac{1}{2} \underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}}(\mathbf{u} \wedge \mathbf{v})$$

holds. Consistently, $\boldsymbol{u} \wedge \boldsymbol{v}$ is called a bi-vector, and it is interpreted geometrically as an oriented area element.

Of course, without an inner product, A cannot be an area as such. Therefore, the term content of a bi-vector is commonly used. However, by providing an inner product, setting

$$\underset{\sim}{\omega} = \underset{\sim}{g}^1 \quad , \quad \underset{\sim}{\mu} = \underset{\sim}{g}^2$$

and making use of the correspondence between dual base vectors and reciprocal base vectors (see (3.24)), A delivers the area of the parallelogram which corresponds to the bi-vector $\mathbf{u} \wedge \mathbf{v}$.

For $\dim(\mathcal{V}) = 2$, it can easily be checked by direct evaluation that

$$\mathbf{u} \wedge \mathbf{v} = \mathbf{u} \otimes \mathbf{v} - \mathbf{v} \otimes \mathbf{u} = -\mathbf{v} \wedge \mathbf{u} = T^{12} \mathbf{g}_1 \wedge \mathbf{g}_2 \,, \tag{4.16}$$

which shows that $\mathbf{u} \wedge \mathbf{v}$ is related to a totally skew-symmetric tensor T. The information content of the latter consists of a real number $T^{12} = u^1 v^2 - v^1 u^2$ together with an orientation.

The scheme can be generalized in a straightforward manner to arbitrary dimensions. For $\dim(\mathcal{V}) = 3$,

$$\mathsf{A}(\mathbf{u}, \mathbf{v}, \mathbf{w}) = (\underset{\sim}{\omega} \wedge \underset{\sim}{\mu} \wedge \underset{\sim}{\kappa})(\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w}) = \det \begin{pmatrix} \langle \underset{\sim}{\omega}, \mathbf{u} \rangle & \langle \underset{\sim}{\omega}, \mathbf{v} \rangle & \langle \underset{\sim}{\omega}, \mathbf{v} \rangle \\ \langle \underset{\sim}{\mu}, \mathbf{u} \rangle & \langle \underset{\sim}{\mu}, \mathbf{v} \rangle & \langle \underset{\sim}{\mu}, \mathbf{v} \rangle \\ \langle \underset{\sim}{\kappa}, u \rangle & \langle \underset{\sim}{\kappa}, \mathbf{v} \rangle & \langle \underset{\sim}{\kappa}, \mathbf{v} \rangle \end{pmatrix}$$

is the content of an oriented volume specified by the tri-vector $\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w}$. Direct evaluation similar to (4.16) yields

$$\begin{aligned} \mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w} = \; & \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} + \mathbf{w} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{w} \otimes \mathbf{u} \\ & -\mathbf{v} \otimes \mathbf{u} \otimes \mathbf{w} - \mathbf{w} \otimes \mathbf{v} \otimes \mathbf{u} - \mathbf{u} \otimes \mathbf{w} \otimes \mathbf{v} \\ = \; & T^{12} \mathbf{g}_1 \wedge \mathbf{g}_2 + T^{13} \mathbf{g}_1 \wedge \mathbf{g}_3 + T^{23} \mathbf{g}_2 \wedge \mathbf{g}_3 \,, \end{aligned}$$

which indicates that permutations may be helpful for a further generalization of this procedure.

The operation $\underset{\sim}{\omega} \wedge \underset{\sim}{\mu}(\mathbf{u}, \mathbf{v})$ is not limited to $\dim(\mathcal{V}) = 2$, but also works for $\dim(\mathcal{V}) = 3$ or higher (see Exercise 4.8). The content of a line element specified by $\mathbf{u}$ is, of course, just $\underset{\sim}{\omega}(\mathbf{u})$. Therefore, introducing the exterior product opens up a straight forward manner for generalizing the notion of volume, starting from one dimension up to arbitrary dimensions.

Remark 4.3 Traditional vector calculus ignores the concept of dual space and considers, from the very beginning, only vector spaces equipped with an inner product. The length of a line element specified by a vector $\mathbf{u}$ is computed by the norm induced by the inner product, i.e., $l = ||\mathbf{u}|| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$. The area of the parallelogram spanned by two vectors $\mathbf{u}$ and $\mathbf{v}$, on the other hand, is computed by $A = ||\mathbf{u} \times \mathbf{v}||$, where "$\times$" stands for the cross product of traditional vector calculus. Last and actually also least, the volume spanned by three vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ reads as $V = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$. This makes a rather convincing case for the exterior product.

According to Sect. 3.6, a dual vector $\underset{\sim}{\boldsymbol{\omega}}$ is a linear form. Consistently, $\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\eta}}$ is commonly called an alternating 2-form or just a 2-form. This scheme extends in a straightforward manner to n-forms.

This section indicates distinctly that a more general treatment of skew-symmetric tensors is necessary, which implies the need to consider symmetric tensors as well.

4.6 Symmetric and Skew-Symmetric Tensors

The following definitions refer to the definition of tensors in the context of multilinear forms (see Definition 4.1).

Definition 4.3 (*Symmetric tensors*) A $\mathsf{T}_{\binom{m}{0}}$ tensor is symmetric with respect to its k^{th} and l^{th} argument if

$$\alpha = \mathsf{T}_{\binom{m}{0}}(\ldots, \overset{(k)}{\mathbf{v}}, \ldots, \overset{(l)}{\mathbf{v}}, \ldots) = \mathsf{T}_{\binom{m}{0}}(\ldots, \overset{(l)}{\mathbf{v}}, \ldots, \overset{(k)}{\mathbf{v}}, \ldots)$$

holds. $\mathsf{T}_{\binom{m}{0}}$ is totally symmetric if, for every possible permutation of its arguments, α remains unchanged. Analogously, a $\mathsf{T}_{\binom{0}{n}}$ tensor is symmetric with respect to its k^{th} and l^{th} argument if

$$\alpha = \mathsf{T}_{\binom{0}{n}}(\ldots, \overset{(k)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots, \overset{(l)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots) = \mathsf{T}_{\binom{0}{n}}(\ldots, \overset{(l)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots, \overset{(k)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots)$$

holds. $\mathsf{T}_{\binom{0}{q}}$ is totally symmetric if, for every possible permutation of its arguments, α remains unchanged.

Definition 4.4 (*Skew-symmetric tensors*) A $\mathsf{T}_{\binom{0}{n}}$ tensor is skew symmetric with respect to its k^{th} and l^{th} argument if

$$\mathsf{T}_{\binom{0}{n}}(\ldots, \overset{(k)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots, \overset{(l)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots) = -\mathsf{T}_{\binom{0}{n}}(\ldots, \overset{(l)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots, \overset{(k)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots)$$

holds. It is totally skew symmetric if

$$\mathsf{T}_{\binom{0}{n}}(\overset{(i_1)}{\underset{\sim}{\boldsymbol{\omega}}}, \overset{(i_2)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots, \overset{(i_n)}{\underset{\sim}{\boldsymbol{\omega}}}) = q\ \mathsf{T}_{\binom{0}{n}}(\overset{(j_1)}{\underset{\sim}{\boldsymbol{\omega}}}, \overset{(j_2)}{\underset{\sim}{\boldsymbol{\omega}}}, \ldots, \overset{(j_n)}{\underset{\sim}{\boldsymbol{\omega}}})$$

with

$$q = \begin{cases} -1 & \\ & \text{for } j_1\, j_2 \,\ldots\, j_n \text{ distinct and } \begin{array}{c} \text{an odd} \\ \\ \text{an even} \end{array} \text{ permutation of } i_1\, i_2 \,\ldots\, i_n \\ 1 & \end{cases},$$

and $q = 0$ otherwise, holds. Skew-symmetry of a $\mathsf{T}_{\binom{m}{0}}$ tensor is defined analogously.

4.7 Generalized Kronecker Symbol

A closer look at the operations introduced in Sect. 4.5 reveals that $\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}}(\mathbf{u} \wedge \mathbf{v})$ can be expressed, as well as

$$\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}}(\mathbf{u} \wedge \mathbf{v}) = \omega_i \mu_k u^o v^p \begin{vmatrix} \delta^i_o & \delta^j_o \\ \delta^i_p & \delta^j_p \end{vmatrix} , \tag{4.17}$$

and similarly, for $\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}} \wedge \underset{\sim}{\boldsymbol{\kappa}}(\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w})$,

$$\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}} \wedge \underset{\sim}{\boldsymbol{\kappa}}(\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w}) = \omega_i \mu_j \kappa_l \, u^o v^p w^q \begin{vmatrix} \delta^i_o & \delta^j_o & \delta^l_o \\ \delta^i_p & \delta^j_p & \delta^l_p \\ \delta^i_q & \delta^j_q & \delta^l_q \end{vmatrix}$$

can be used, which suggests the following generalization of the Kronecker symbol.

Definition 4.5 (*Generalized Kronecker symbol*) The generalized Kronecker symbol is defined by means of the standard Kronecker symbol as follows:

$$\delta^{\,i_1 i_2 \ldots i_q}_{\,j_1 j_2 \ldots j_q} = \begin{vmatrix} \delta^{i_1}_{j_1} & \ldots & \delta^{i_1}_{j_q} \\ \vdots & & \vdots \\ \delta^{i_q}_{j_1} & \ldots & \delta^{i_q}_{j_q} \end{vmatrix} ,$$

with $i_1, \ldots, i_q; \, j_1, \ldots, j_q \in \mathbb{N}$.

In order to compute with the generalized Kronecker symbol, it is helpful to note that Definition 4.5 implies

$$\delta^{\,i_1 i_2 \ldots i_q}_{\,j_1 j_2 \ldots j_q} = \begin{cases} -1 & \text{for } j_1 \, j_2 \, \ldots \, j_p \ \text{an odd} \\ 1 & \phantom{\text{for } j_1 \, j_2 \, \ldots \, j_p}\ \text{an even} \end{cases} \text{permutation of } i_1 \, i_2 \, \ldots i_p \, ,$$

provided that the $j_1 \, j_2 \, \ldots \, j_p$ are distinct and $\delta^{\,i_1 i_2 \ldots i_q}_{\,j_1 j_2 \ldots j_q} = 0$ otherwise.

Furthermore, the exterior product can be expressed by the generalized Kronecker symbol

$$\overset{(i_1)}{\underset{\sim}{\boldsymbol{\omega}}} \wedge \cdots \wedge \overset{(i_q)}{\underset{\sim}{\boldsymbol{\omega}}} = \delta^{\,i_1 i_2 \ldots i_q}_{\,j_1 j_2 \, \ldots \, j_q} \, \overset{(j_1)}{\underset{\sim}{\boldsymbol{\omega}}} \otimes \cdots \otimes \overset{(j_q)}{\underset{\sim}{\boldsymbol{\omega}}} \, , \tag{4.18}$$

which establishes a relation between the dyadic product and the exterior product for arbitrary q.

4.8 The Spaces $\Lambda^k\mathcal{V}$ and $\Lambda^k\mathcal{V}^*$

The totally skew-symmetric tensors of a tensor space form a linear subspace of particular interest. In order to explore this, totally skew-symmetric projections are first discussed for the space $\mathcal{V}\otimes\mathcal{V}$. In this context, a totally skew-symmetric projection is a linear mapping

$$\begin{aligned} P:\ \mathcal{V}\otimes\mathcal{V} &\to \mathcal{V}\otimes\mathcal{V} \\ \mathsf{B} &\mapsto \mathsf{C} = P(\mathsf{B}) \end{aligned} \tag{4.19}$$

with the properties

$$P(\mathsf{C}) = \mathsf{C} \tag{4.20}$$

$$\mathsf{C}(\underset{\sim}{\omega}, \underset{\sim}{\eta}) = -\mathsf{C}(\underset{\sim}{\eta}, \underset{\sim}{\omega})\,, \tag{4.21}$$

with $\underset{\sim}{\omega}, \underset{\sim}{\eta} \in \mathcal{V}^*$. Property (4.20) classifies P as a projection, whereas (4.21) demands skew-symmetry.

Applying the procedure used in Sect. 4.2, it can be deduced from (4.19) that P is encoded by a forth order tensor P. The representation of P in terms of a basis $\{\mathbf{g}_i\}$ of $\mathcal{V}$ reads as

$$\mathsf{P} = P^{ij}_{kl}\ \mathbf{g}_i \otimes \mathbf{g}_j \otimes \mathbf{g}^k \otimes \mathbf{g}^l\,.$$

The coordinates P^{ij}_{kl} can be specified by means of (4.20) and (4.21). From (4.20),

$$P^{ij}_{kl} = -P^{ji}_{kl}$$

follows, and using this result together with (4.21) yields

$$P^{ij}_{kl} = \frac{1}{2}\left[\delta^i_k\delta^j_l - \delta^j_k\delta^i_l\right] = \frac{1}{2}\delta^{ij}_{kl} \tag{4.22}$$

(see Exercise 4.9). The result of the projection applied to a tensor B is then given by

$$\mathsf{C} = \mathsf{P}(\mathsf{B}) = \frac{1}{2}B^{ij}\,\mathbf{g}_i \wedge \mathbf{g}_j\,, \tag{4.23}$$

if (4.18) is used.

The skew-symmetric projection P generates a space $\Lambda^2\mathcal{V}$. The latter contains the zero element of $\mathcal{V}\otimes\mathcal{V}$, $P(\mathsf{0}) = \mathsf{0}$. Furthermore, $\Lambda^2\mathcal{V}$ is closed under addition and multiplication with scalar. Therefore, it is a subspace of $\mathcal{V}\otimes\mathcal{V}$.

Every result of P can be represented by $\mathbf{g}_i \wedge \mathbf{g}_j$ and respective coordinates. Therefore, a basis for $\Lambda^2\mathcal{V}$ is given by

$$\left\{\mathbf{g}_i \wedge \mathbf{g}_j\right\} \qquad i < j\,, \tag{4.24}$$

where the condition $i < j$ is necessary in order to assure that the elements of the basis are linearly independent, i.e., it accounts for $\mathbf{g}_i \wedge \mathbf{g}_j = -\mathbf{g}_j \wedge \mathbf{g}_i$. Furthermore, it excludes $\mathbf{g}_i \wedge \mathbf{g}_i$, since a zero element can never be part of a basis. The dimension of $\Lambda^2\mathcal{V}$ follows directly from the length of the set of base vectors. Constructing (4.24) for $\dim(\mathcal{V}) = N$ requires forming groups with two members each from N base vectors of $\mathcal{V}$ without repetitions, i.e.,

$$\dim(\Lambda^2\mathcal{V}) = \binom{N}{2} = \frac{N!}{2!(N-2)!} .$$

Generalization of (4.19) for $\otimes_k\mathcal{V}$ is straightforward. For $\dim(\mathcal{V}) = N$, the dimension of the corresponding subspace $\Lambda^k\mathcal{V}$ is given by

$$\dim(\Lambda^k\mathcal{V}) = \binom{N}{k} = \frac{N!}{k!(N-k)!} ,$$

which reveals that $\Lambda^k\mathcal{V}$ is isomorphic to the space $\Lambda^q\mathcal{V}$ if $q = N-k$. This result can be used to relate the algebra of multi-linear forms with the conventional approach, i.e., cross product, etc. The space $\Lambda^k\mathcal{V}^*$ is constructed analogously.

4.9 Properties of the Exterior Product and the Star-Operator

From the definition (4.15), the following properties of the exterior product can be deduced:

$$\begin{aligned}
(\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}}) \wedge \underset{\sim}{\boldsymbol{\nu}} &= \underset{\sim}{\boldsymbol{\omega}} \wedge (\underset{\sim}{\boldsymbol{\mu}} \wedge \underset{\sim}{\boldsymbol{\nu}}) \\
(\alpha\underset{\sim}{\boldsymbol{\omega}} + \beta\underset{\sim}{\boldsymbol{\nu}}) \wedge \underset{\sim}{\boldsymbol{\eta}} &= \alpha\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\eta}} + \beta\underset{\sim}{\boldsymbol{\nu}} \wedge \underset{\sim}{\boldsymbol{\eta}} \\
\underset{\sim}{\boldsymbol{\eta}} \wedge (\alpha\underset{\sim}{\boldsymbol{\omega}} + \beta\underset{\sim}{\boldsymbol{\nu}}) &= \alpha\underset{\sim}{\boldsymbol{\eta}} \wedge \underset{\sim}{\boldsymbol{\omega}} + \beta\underset{\sim}{\boldsymbol{\eta}} \wedge \underset{\sim}{\boldsymbol{\nu}}
\end{aligned}$$

and

$$\underset{\sim}{\boldsymbol{\omega}} \wedge \underset{\sim}{\boldsymbol{\mu}} = (-1)^{pq}\underset{\sim}{\boldsymbol{\mu}} \wedge \underset{\sim}{\boldsymbol{\omega}}$$

for $\underset{\sim}{\boldsymbol{\omega}} \in \Lambda^p\mathcal{V}^*$, $\underset{\sim}{\boldsymbol{\mu}} \in \Lambda^q\mathcal{V}^*$ and $\alpha, \beta \in \mathbb{R}$. These relations hold analogously for multi-vectors, i.e., elements of $\Lambda^k\mathcal{V}$.

The observation that $\Lambda^k\mathcal{V}$ and $\Lambda^{N-k}\mathcal{V}$ with $N = \dim(\mathcal{V})$ have the same dimension indicates that there exists a relation between classical vector calculus and calculus with alternating forms. This will be discussed subsequently for $N = 3$.

As sketched in Fig. 4.2, a parallelogram area in $\mathbb{R}^3$ can be characterized by a bi-vector $\mathbf{a} \wedge \mathbf{b}$ or by means of a surface normal vector $\mathbf{n}$, where $\mathbf{n} \in \Lambda^1\mathcal{V} = \mathcal{V}$ and $\mathbf{a} \wedge \mathbf{b} \in \Lambda^2\mathcal{V}$, which leads to the question as to whether there is a canonical way to relate these two spaces. Since $\dim(\mathcal{V}) = \dim(\Lambda^2\mathcal{V})$, the two spaces are

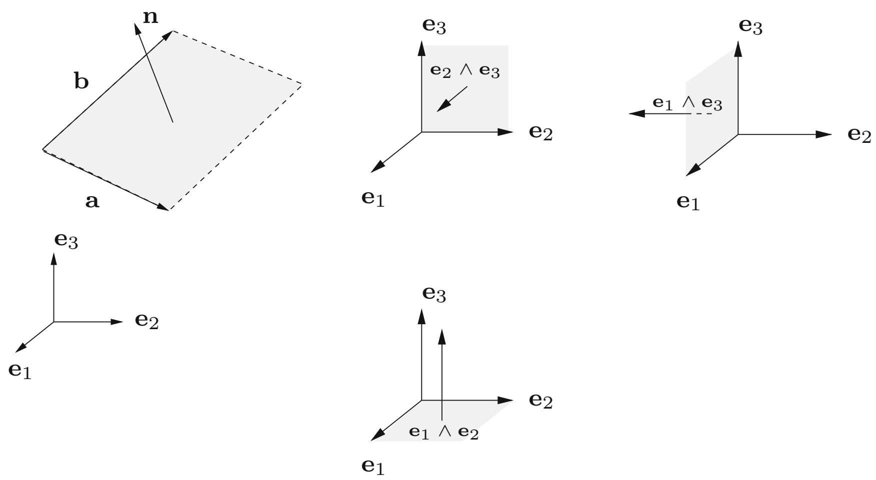

Fig. 4.2 A parallelogram area in $\mathbb{R}^3$ described by a bi-vector $\mathbf{a} \wedge \mathbf{b}$ or by a surface normal vector $\mathbf{n}$ (*left*) and normal vectors to surfaces defined by $\mathbf{e}_1 \wedge \mathbf{e}_3$, $\mathbf{e}_1 \wedge \mathbf{e}_2$, and $\mathbf{e}_2 \wedge \mathbf{e}_3$ according to the right hand rule (*right*)

isomorphic, i.e., $\mathcal{V} \cong \Lambda^2\mathcal{V}$. The question arises as to how to extract one particular isomorphism from all possible isomorphisms. A similar problem has already been discussed in Sect. 3.7, where the isomorphism was induced by the inner product. There, the isomorphism $\mathcal{G}$ was defined by the condition $\langle \mathcal{G}(\mathbf{v}), \mathbf{u} \rangle = \mathbf{u} \cdot \mathbf{v}$ for arbitrary $\mathbf{u}$. Choosing an orthonormal basis $\{\mathbf{e}_i\}$ for simplicity, we are looking for a bijective, linear mapping

$$\begin{aligned} \mathcal{H}:\ & \mathcal{V} \rightarrow \Lambda^2\mathcal{V} \\ & \mathbf{v} \mapsto w^{ij}\, \mathbf{e}_i \wedge \mathbf{e}_j\,, \end{aligned}$$

In order to specify $\mathcal{H}$, a condition is required. Since we want to state an equation with $\mathbf{u} \cdot \mathbf{v}$ on one side and $\mathcal{H}(\mathbf{v})$ paired with $\mathbf{u}$ on the other side, $\mathcal{H}(\mathbf{v})$ and $\mathbf{u}$ have to be connected by an operation. A natural way seems to be $\mathbf{u} \wedge \mathcal{H}(\mathbf{v})$. The result of the latter is a 3-vector. Therefore,

$$\mathbf{u} \wedge \mathcal{H}(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}\ \Omega \tag{4.25}$$

turns out to be a suitable condition. The 3-vector Ω, e.g.,

$$\Omega = \omega\ \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3\,, \tag{4.26}$$

on the right hand side is required for consistency, since the inner product returns a real number. Computation of (4.25) using the basis $\{\mathbf{e}_i\} \in \mathcal{V}$ yields

$$[u^1 w^{23} - u^2 w^{13} + u^3 w^{12}]\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3 = [u^1 v^1 + u^2 v^2 + u^3 v^2]\omega \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3\,, \tag{4.27}$$

and setting $\mathbf{v} = \mathbf{e}_1 = \mathbf{e}_1 + 0\,\mathbf{e}_2 + 0\,\mathbf{e}_3$ gives $\mathcal{H}(\mathbf{e}_1) = \omega\,\mathbf{e}_2 \wedge \mathbf{e}_3$ by equating the coefficients. It shows that the result depends on the choice for ω. Furthermore, it depends on the choice regarding the orientation, since, e.g., $\Omega = \omega\,\mathbf{e}_1 \wedge \mathbf{e}_3 \wedge \mathbf{e}_2$ could be used instead of (4.26). Commonly, $*\mathbf{v}$ is used in the literature as short hand notation for $\mathcal{H}(\mathbf{v})$ and called Hodge-star-operator. Evaluating (4.27) for all base vectors using (4.26) with $\omega = 1$ gives

$$
\begin{aligned}
*\mathbf{e}_1 &= \mathbf{e}_2 \wedge \mathbf{e}_3 \\
*\mathbf{e}_2 &= -\mathbf{e}_1 \wedge \mathbf{e}_3 \\
*\mathbf{e}_3 &= \mathbf{e}_1 \wedge \mathbf{e}_2\,,
\end{aligned}
$$

which looks rather convincing since it reflects orientation in terms of the well-known right hand rule sketched in Fig. 4.2. However, it has to be checked if the result of the Hodge-star-operator is independent of the choice for a basis and if it is really an isomorphism, i.e., linear and bijective. This is actually the case (see, e.g., Jänich [1]).

By means of a Hodge-star operation, the spaces $\Lambda^k\mathcal{V}$, $\Lambda^k\mathcal{V}^*$ can be related with the spaces $\Lambda^{N-k}\mathcal{V}$ and $\Lambda^{N-k}\mathcal{V}^*$, respectively. Formulas for arbitrary N and k exist, but they are rather cumbersome due to the need for multi-indexing. As long as the dimensions are small, we recommend to derive required formulas from scratch based on the methodology discussed above. Furthermore, modern computer algebra systems could be used as well.

Remark 4.4 In order to avoid limitations in terms of dimensions, induced, for instance, by the cross product in classical vector calculus, an existing theory or model could be reformulated in the language of exterior algebra. However, it is not always entirely clear what this reformulation should look like. The Hodge operator can give valuable hints in the course of the process of reformulation. This is one reason for its importance.

4.10 Relation with Classical Linear Algebra

Classical linear algebra takes place in $\mathbb{R}^N$ using an orthonormal basis $\{\mathbf{e}_i\}$ called the standard basis in $\mathbb{R}^N$. The i-th base vector is commonly written as a column where the i-th entry is one and all other entries are zero. If, for a vector space of dimension N and basis $\{\mathbf{g}_i\}$, a so-called basis isomorphism is used, i.e., a linear mapping

$$
\varphi : \mathcal{V} \to \mathbb{R}^N
$$

with the property

$$
\varphi(\mathbf{g}_i) = \mathbf{e}_i\,,
$$

then all computations reduce to classical linear algebra. For example, the representation of a tensor T with respect to the basis $\{\mathbf{e}_i\}$ reads, in general, as

$$\mathsf{T} = T^{ij}\mathbf{e}_i \otimes \mathbf{e}_j \,.$$

In classical linear algebra, the coordinates of T are arranged in a matrix, which reads, e.g., for $N = 2$, as

$$\left[T^{ij}\right] = \begin{bmatrix} T^{11} & T^{12} \\ T^{21} & T^{22} \end{bmatrix} ,$$

and tensor operations can be represented by classical matrix operations. Afterwards, the results can be transformed back to $\mathcal{V}$ by means of φ^{-1}. Therefore, classical linear algebra is a special case of general tensor algebra.

Exercises

4.1 Tensors have been introduced by their mapping properties. An equivalent, alternative definition of a tensor is given by the transformation behavior of its components under change of basis, i.e., every quantity that shows this particular transformation behavior can be considered the component-matrix of a tensor. Given a second order tensor $\mathbf{T} = T^{ij}\,\mathbf{g}_i \otimes \mathbf{g}_j$ and a basis $\{\mathbf{g}_i\} \in \mathcal{V}$. A second basis is given by

$$\widehat{\mathbf{g}}_i = \widehat{A}_i^{\;k}\mathbf{g}_k \,.$$

How do the components T^{ij} transform under a change of basis

$$\{\mathbf{g}_i\} \to \{\widehat{\mathbf{g}}_i\} \Rightarrow T^{ij} \to \widehat{T}^{ij} \,?$$

4.2 Use the transformation relationships under change of basis to show that the so-called metric coefficients $g^{ij} := \mathbf{g}^i \cdot \mathbf{g}^j$ and $g_{ij} := \mathbf{g}_i \cdot \mathbf{g}_j$ are the coordinates of a second order tensor

$$\mathbf{G} = g_{ij}\,\mathbf{g}^i \otimes \mathbf{g}^j = g^{ij}\,\mathbf{g}_i \otimes \mathbf{g}_j$$

called the fundamental tensor.

4.3 Consider the space $\mathcal{V} \otimes \mathcal{V}$, $\dim(\mathcal{V}) = 2$ and a given basis of $\mathcal{V}$, $\{\mathbf{g}_i\}$. Show that the tensor $\mathsf{T} = T^{ij}\mathbf{g}_i \otimes \mathbf{g}_j$ with $T^{11} = T^{22} = T^{21} = 1$, $T^{12} = 0$ cannot be obtained from a dyadic product of the form $\mathbf{u} \otimes \mathbf{v}$.

4.4 Given $\underset{\sim}{\omega} \wedge \underset{\sim}{\eta} = \underset{\sim}{\omega} \otimes \underset{\sim}{\eta} - \underset{\sim}{\eta} \otimes \underset{\sim}{\omega}$ and $\mathsf{A} = A^{ij}\mathbf{g}_i \otimes \mathbf{g}_j$. Show by explicit computation that $\underset{\sim}{\omega} \wedge \underset{\sim}{\eta}(\mathsf{A}) = 0$ if A is symmetric.

4.5 Show that $\underset{\sim}{\omega} \wedge \underset{\sim}{\eta}(\mathbf{u} \otimes \mathbf{v}) = \frac{1}{2}\underset{\sim}{\omega} \wedge \underset{\sim}{\eta}(\mathbf{u} \wedge \mathbf{v})$ holds.

4.6 Show that $(\mathbf{u} + \mathbf{v}) \wedge (\mathbf{v} + \mathbf{w}) = \mathbf{u} \wedge \mathbf{v} + \mathbf{u} \wedge \mathbf{w} + \mathbf{v} \wedge \mathbf{w}$ holds.

4.7 Show that the result of $\underset{\sim}{\omega} \wedge \underset{\sim}{\eta}(\mathbf{u} \otimes \mathbf{v})$ is independent of the choice of a particular basis.

4.8 Given $\mathcal{V} = \mathbb{R}^3$ with the standard basis $\{\mathbf{e}_i\}$. An area element can be represented by a bi-vector A

$$\mathsf{A} = \mathbf{a} \wedge \mathbf{b}$$

with $\mathbf{a}, \mathbf{b} \in \mathcal{V}$.

1. Express A by means of coordinates and base vectors.
2. Determine the coordinates $\bar{A}^{12}$, $\bar{A}^{13}$ and $\bar{A}^{23}$ related to the following representation of A:

$$\mathsf{A} = \bar{A}^{12}\mathbf{e}_1 \wedge \mathbf{e}_2 + \bar{A}^{13}\mathbf{e}_1 \wedge \mathbf{e}_3 + \bar{A}^{23}\mathbf{e}_2 \wedge \mathbf{e}_3 \,,$$

3. Give an interpretation of the coordinates $\bar{A}^{12}$, $\bar{A}^{13}$ and $\bar{A}^{23}$.

4.9 Show by means of (4.20) that the skew-symmetric projection defined by (4.19) corresponds to the tensor

$$\mathsf{P} = \frac{1}{2}[\delta^k_o \delta^l_p - \delta^l_o \delta^k_p]\, \mathbf{g}_k \otimes \mathbf{g}_l \otimes \underset{\sim}{\boldsymbol{g}}^o \otimes \underset{\sim}{\boldsymbol{g}}^p \,.$$

References

1. Jänich K (2001) Vector analysis. Undergraduate texts in mathematics. Springer, New York
2. Lee M (2002) Introduction to smooth manifolds. Graduate texts in mathematics. Springer, Berlin

Chapter 5
Affine Space and Euclidean Space

Abstract Affine and euclidean space are discussed primarily in view of their use as models for physical space. Affine mappings and coordinate charts generated by them are examined. Furthermore, topological aspects are addressed.

5.1 Definitions and Basic Notions

As mentioned earlier, the general idea of a model for physical space is to endow a set of points X with a structure which allows us to encode certain physical concepts. It seems obvious to try the simplest case first, namely a point set X together with a linear structure. This, on the other hand, is tantamount to combining X with a vector space $\mathcal{V}$. In addition, the model for space should allow for choosing an origin freely. Such a model is called affine space. It relies on the idea of assigning vectors to pairs of points by means of some mapping F. However, an affine structure on X requires more than just that, but such an assignment has to be consistent with the vector space operations. In order to achieve this, F has to obey certain rules which form part of the following definition.

Definition 5.1 (*Affine space*) An affine space $\mathcal{A} = (X, \mathcal{V}, F)$ consists of a point set X, a vector space $\mathcal{V}$ according to Definition 3.1 and a mapping F,

$$\begin{aligned} F : X \times X &\to \mathcal{V} \\ (p, q) &\mapsto F(p, q) = \boldsymbol{v}\,, \end{aligned}$$

with the following properties for $p_0, p, r, q \in X$ and $\boldsymbol{v} \in \mathcal{V}$:

(i) $F(p, q) = -F(q, p)$
(ii) $F(p, q) = F(p, r) + F(r, q)$
(iii) For all p_0 and $\mathbf{v}$, there exists a unique q, such that $F(p_0, q) = \mathbf{v}$.

U. Mühlich, *Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds*, Solid Mechanics and Its Applications 230,
DOI 10.1007/978-3-319-56264-3_5

Example 5.1 Points in $\mathbb{R}^2$ are represented by pairs of real numbers, e.g., the point p corresponds to (p^1, p^2) with $p^1, p^2 \in \mathbb{R}$. A mapping F is usually defined by

$$F(p, q) = \begin{bmatrix} q^1 - p^1 \\ q^2 - p^2 \end{bmatrix} ,$$

by which the point set $\mathbb{R}^2$ becomes an affine space.

The dimension of an affine space coincides with the dimension of the associated vector space. One of the most important properties of an affine space is that everything which can be interpreted as a result of F is an element of $\mathcal{V}$ and can, therefore, be added with any other element of $\mathcal{V}$ (see (ii) of Definition 5.1). Furthermore, an affine space is homogeneous and isotropic. The choice of a particular point p_0 (see Definition 5.1 (iii)) is equivalent to choosing an origin. After choosing an origin, an affine space is indistinguishable from $\mathcal{V}$.

Example 5.2 Figure 5.1 illustrates a possible exploration of space applying the affine space concept. It sketches a bird's eye view of a limited part of the surface of the earth. The filled polygons could be the roofs of houses. Some landmarks in terms of corners of these houses are chosen. One point, e.g., p_0, is defined as origin. Pairs of points are chosen and respective vectors are assigned to these pairs by means of a mapping F. The pairs (p_0, p) and (p_0, q) could be used to define a basis $\mathbf{g}_1 = F(p_0, p), \mathbf{g}_2 = F(p_0, q)$. A point z can be addressed now by a corresponding vector $\mathbf{z} = F(p_0, x)$, and $\mathbf{z}$ can be expressed by means of the base vectors and respective coordinates

$$\mathbf{z} = z^1\mathbf{g}_1 + z^2\mathbf{g}_2 \,.$$

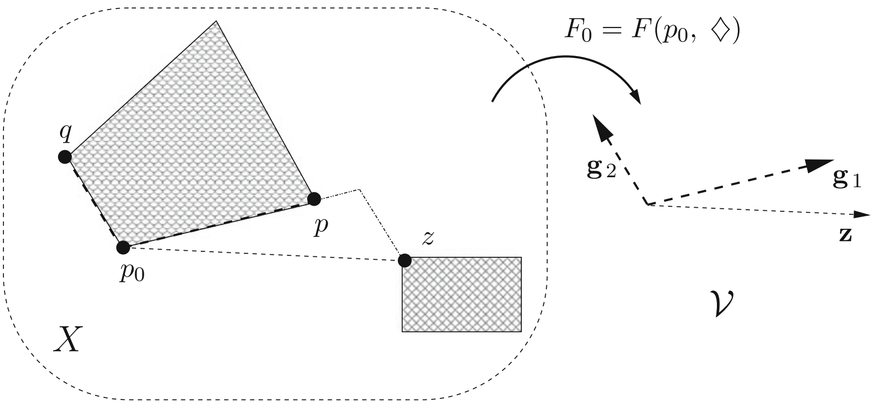

Fig. 5.1 Sketch regarding the exploration of the space of observation X based on the assumption that X is an affine space and choosing p_0 as the origin. The symbol $\diamondsuit$ indicates a place-holder

The example illustrates the limits of this concept as well. The reader is encouraged to reflect on the situation described here in terms of the validity of the concept in relation to the extension of the surface explored this way.

An affine space does not provide a structure which allows for measuring length, distance, or angles. A model for space, for which these concepts make sense, can be derived from an affine space in the following manner.

Definition 5.2 (*Euclidean space* $\mathcal{E}$) An affine space where $\mathcal{V}$ is an inner product space is an euclidean space.

5.2 Alternative Definition of an Affine Space by Hybrid Addition

Definition 5.3 (*Hybrid addition of points and vectors*) Given an affine space according to Definition 5.1 with $\{\mathcal{V}, \oplus, \odot\}$. Hybrid addition "$\boldsymbol{+}$" is a mapping

$$\begin{aligned} \boldsymbol{+} &: X \times \mathcal{V} \to X \\ (p, \mathbf{u}) &\mapsto q = p \boldsymbol{+} \mathbf{u} \end{aligned}$$

with the properties:

(i) $[p \boldsymbol{+} \mathbf{u}] \boldsymbol{+} \mathbf{v} = p \boldsymbol{+} [\mathbf{u} \oplus \mathbf{v}]$
(ii) $p \boldsymbol{+} \mathbf{u} = p$ implies $\mathbf{u} = \mathbf{0}$
(iii) For all p, q, there exists a unique $\mathbf{u}$ such that $q = p \boldsymbol{+} \mathbf{u}$.

The properties of the hybrid addition can be deduced from the properties of F in Definition 5.1, and vice versa (see e.g., Epstein [2]). This means that hybrid addition and F are equivalent. Therefore, an affine space can be defined either by $(X, \mathcal{V}, F)$ as in Definition 5.1 or by $(X, \mathcal{V}, \boldsymbol{+})$.

Hybrid addition is rather useful for encoding geometrical concepts transparently. A straight line L, for instance, can simply be expressed by

$$L = p \boldsymbol{+} \lambda \mathbf{u} \, .$$

Furthermore, L is parallel with respect to another straight line $\bar{L}$

$$\bar{L} = q \boldsymbol{+} \bar{\lambda} \mathbf{v} \, .$$

if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent. If, in addition, $p = q$, then $L = \bar{L}$.

5.3 Affine Mappings, Coordinate Charts and Topological Aspects

Within this section, mappings are discussed which preserve the affine structure. Two affine spaces $\mathcal{A}_1 = (X_1, \mathcal{V}_1, F_1)$ and $\mathcal{A}_2 = (X_2, \mathcal{V}_2, F_2)$ are considered. Due to the correspondence of X_1 and $\mathcal{V}_1$ established by F_1, respectively X_2 and $\mathcal{V}_2$ due to F_2, every affine mapping

$$\mu : X_1 \to X_2 ,$$

automatically induces a mapping $A : \mathcal{V}_1 \to \mathcal{V}_2$, and vice versa. Furthermore, every linear mapping between $\mathcal{V}_1$ and $\mathcal{V}_2$, as well as every global shift, preserves the affine structure. This is the essence of the following definition.

Definition 5.4 (*Affine mapping*) Given two affine spaces $(X_1, \mathcal{V}_1, F_1)$, $(X_2, \mathcal{V}_2, F_2)$ according to Definition 5.1 and some linear mapping $B : \mathcal{V}_1 \to \mathcal{V}_2$. A mapping $\mu : X_1 \to X_2$ with the property

$$B(\, F_1(p, q)\,) = F_2(\, \mu(p), \mu(q)\,)$$

is called an affine mapping.

The most general form of μ which ensures that μ is an affine mapping according to Definition 5.4 reads as

$$\begin{aligned} \mu : X_1 &\to X_2 \\ q &\mapsto Q \dotplus \mathsf{B}(F_1(p_0, q)) , \end{aligned}$$

where $Q \in X_2$ and $p_0 \in X_1$ can be chosen arbitrarily. The second-order tensor B encodes the mapping between the respective vector spaces.

Affine coordinate charts, also called affine coordinate systems, are common applications of the mappings discussed above. In the following, we consider an affine space $\mathcal{A} = (X, \mathcal{V}, \dotplus)$. Explanations will be given by means of examples corresponding to $\dim(\mathcal{A}) = \dim(\mathcal{V}) = 2$.

In X, there exist points, curves, e.g., $\Gamma : \mathbb{R} \to X$ and functions, e.g., $\Phi : \mathcal{A} \to \mathbb{R}$. In order to be able to perform computations with numbers, a pair of real numbers is assigned to each point by a mapping

$$\begin{aligned} \mu : X &\to \mathbb{R}^2 \\ p &\mapsto (x^1(p), x^2(p)) , \end{aligned}$$

where $x^i(p) : \mathcal{A} \to \mathbb{R}$ is called the i-th coordinate function because it assigns to a point p its i-th coordinate. The mapping μ generates a coordinate system, also called a coordinate chart. Different mappings of this kind establish different coordinate systems. As illustrated in Fig. 5.2, these systems differ in the meaning of a specific

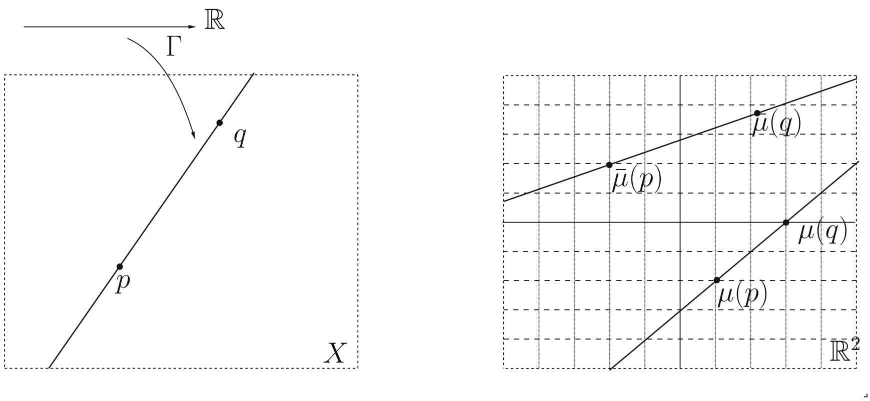

Fig. 5.2 Sketch illustrating the effect of two affine mappings μ and $\bar{\mu}$ on two points p and q, respectively, on a curve in X

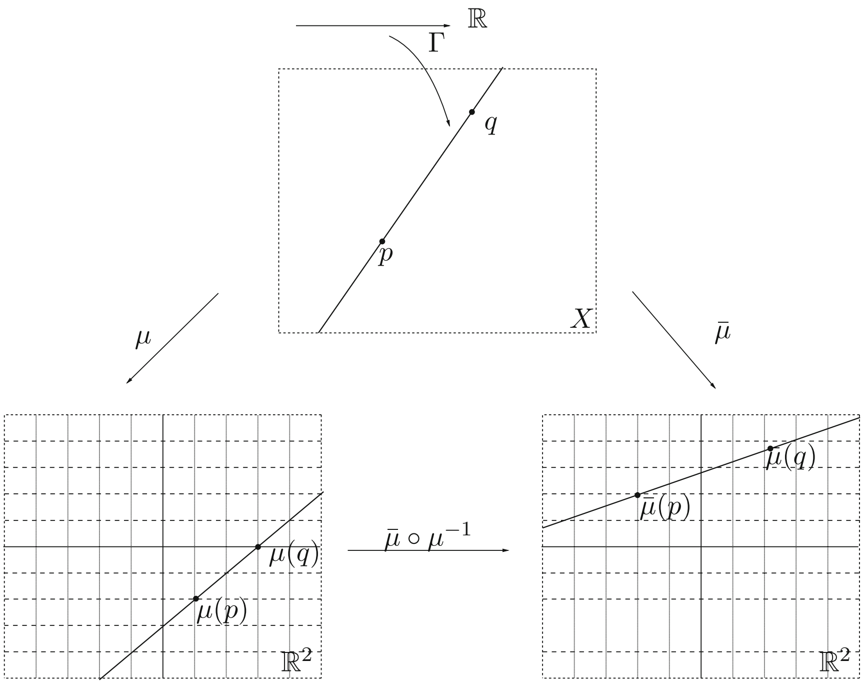

Fig. 5.3 Visualization of different coordinate charts established by different mappings and the transition between these charts

pair of real numbers. However, the visualization in Fig. 5.2 will become confusing as the complexity of the problem increases.

Therefore, a different visualization is commonly preferred which better accounts for the transition between different coordinate charts. This is illustrated in Fig. 5.3.

When referring to coordinate charts of $\mathcal{A}$ established by mappings μ and $\bar{\mu}$, the notations (X, μ) and $(X, \bar{\mu})$ are used in the following. In order to distinguish between the function Φ and its representation in a chart $f = \Phi \circ \mu^{-1}$, the notation

$$f = f(x^1, x^2) \tag{5.1}$$

with $f = \Phi \circ \mu^{-1}$ is employed in general, which reveals a common notational abuse (see e.g., Crampin and Pirani [1]), since in (5.1), the x^i are coordinates and not coordinate functions. Hence, (x^1, x^2) denotes either a point in $\mathbb{R}^2$, the coordinates of a point in X, or the mappings which generate the coordinate system (X, μ). Although the specific meaning of x^i is usually clear from the context, it requires a careful reading of the respective equations in order to avoid misinterpretations and mistakes. Similarly,

$$\bar{f} = \bar{f}(\bar{x}^1, \bar{x}^2)$$

expresses that $\bar{f} = \Phi \circ \bar{\mu}^{-1}$ is the coordinate representation of Φ in chart $(\mathcal{A}, \bar{\mu})$. The transition between $(\mathcal{A}, \mu)$ and $(\mathcal{A}, \bar{\mu})$ is expressed by the mapping

$$\begin{aligned} g = \bar{\mu} \circ \mu^{-1} : \mathbb{R}^2 &\to \mathbb{R}^2 \\ (x^1, x^2) &\mapsto (\bar{x}^1(x^1, x^2), \bar{x}^1(x^1, x^2)) \,, \end{aligned}$$

where the $\bar{x}^i$ are actually again coordinate functions, whereas the x^i are coordinates of points in $\mathbb{R}^2$.

In the following, we consider Example 5.2 together with a mapping from $\mathcal{V}$ into the $\mathbb{R}^2$, which just reads off the coordinates of $\mathbf{z}$ related to the basis $\{\mathbf{g}_i\}$

$$\begin{aligned} \varphi : \mathcal{V} &\to \mathbb{R}^2 \\ \mathbf{z} = z^1\mathbf{g}_1 + z^2\mathbf{g}_2 &\mapsto (x^1 = z^1, x^2 = z^2) \,. \end{aligned} \tag{5.2}$$

Its inverse is given by

$$\begin{aligned} \varphi^{-1} : \mathbb{R}^2 &\to \mathcal{V} \\ (z^1, z^2) &\mapsto \mathbf{z} = z^1\mathbf{g}_1 + z^2\mathbf{g}_2 \,. \end{aligned}$$

A topology can be defined for $\mathcal{V}$ by the collection of all sets U of the standard topology of $\mathbb{R}^2$ under φ^{-1}, i.e., $\varphi^{-1}(U)$. Obviously, φ and φ^{-1} are continuous for this topology, and since φ is a homeomorphism, $\phi = \varphi \circ F$, is a coordinate chart, according to Definition 2.12.

The generalization for arbitrary but finite dimensions is straightforward. Since $\mathcal{V}$ can be endowed with a topology, an affine space is a possible model for physical space, but again, it depends on the problem as to whether it is also a suitable model.

The composition $\mu = \varphi \circ F$ sketched in Fig. 5.4 is an affine mapping $\mu : X \to \mathbb{R}^2$ according to Definition 5.4 with $Q = (0, 0)$ and

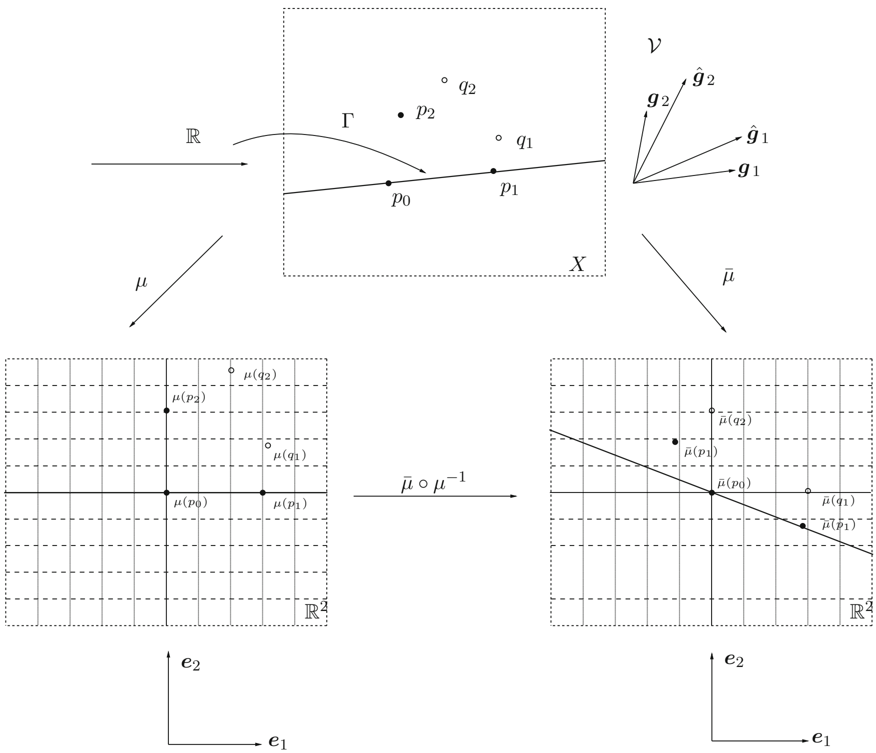

Fig. 5.4 Sketch of the affine coordinates which result from the mappings μ and $\bar{\mu}$, where these mappings differ with respect to the choice of base vectors in $\mathcal{V}$ and $\{\mathbf{e}_1, \mathbf{e}_2\}$ is the standard basis in $\mathbb{R}^2$

$$\mathsf{B} = \delta^i_k \mathbf{e}_i \otimes \underset{\sim}{g}^k \,, \tag{5.3}$$

where $\{\mathbf{e}_i\}$ is the standard basis in $\mathbb{R}^2$. This is an example for a basis isomorphism.

A second mapping $\bar{\mu}$ using a different origin $\bar{p}_0$ and a basis $\left\{\hat{\mathbf{g}}_i\right\}$ is considered. While a point $x \in X$ has coordinates (x^1, x^2) in the chart induced by μ, the same point is referred to by the coordinates $(\bar{x}^1, \bar{x}^2)$ in the chart induced by $\bar{\mu}$. In order to figure out how these coordinates are related, (iii) of Definition 5.1 is applied:

$$F(p_0, x) = F(p_0, \bar{p}_0) + F(\bar{p}_0, x)\,, \tag{5.4}$$

where the LHS corresponds to $\mathbf{x}$, the term $F(\bar{p}_0, x)$ equals $\bar{\mathbf{x}}$ and $F(p_0, \bar{p}_0)$ is a shift due to the change of the origin called $\mathbf{c}$ in the following. According to the transformation rules given in Box 3.4.1, the base vectors $\mathbf{g}_i$ and $\bar{\mathbf{g}}_j$ are related by

$$\hat{\mathbf{g}}_i = \hat{A}^k_i \mathbf{g}_k \,.$$

The vector $\mathbf{c}$ can either be expressed by $\mathbf{c} = c^i \mathbf{g}_i$ or $\mathbf{c} = \hat{c}^i \hat{\mathbf{g}}_i$. The coordinates x^i of a point x in chart (X, μ) result from reading off the coordinates of the vector $\mathbf{x} = x^i \mathbf{g}_i$, whereas the coordinates $\bar{x}^i$ originate from $\bar{x} = \bar{x}^i \hat{\mathbf{g}}_i$. It is important to note at this point that $\mathbf{x}$ and $\bar{\mathbf{x}}$ are different vectors.

Rewriting (5.4) and changing to the coordinate representation, applying the transformation rules as well, gives

$$\mathbf{x} - [\mathbf{c} + \bar{\mathbf{x}}] = \mathbf{0} = \left[x^i - \left(c^i + \hat{A}^i_k\, \bar{x}^k\right)\right] \mathbf{g}_i \,,$$

from which

$$x^i = c^i + \hat{A}^i_k\, \bar{x}^k$$

is deduced. It encodes the transition from the chart induced by p_0, $\{\mathbf{g}_i\}$ into the chart related to $\bar{p}_0$, $\left\{\bar{\mathbf{g}}_i\right\}$. Its inverse defines the opposite direction.

It should be noted that metric properties are in general not preserved by affine mappings. This is only possible if

$$\mathbf{g}_i \cdot \mathbf{g}_j = \mathbf{e}_i \cdot \mathbf{e}_j$$

and the only way to achieve this is by defining the basis $\{\mathbf{g}_i\}$ accordingly. The corresponding coordinates are called Cartesian coordinates.

Choosing an origin and a basis for an affine space is also called a framing. In this context, p_0, $\{\mathbf{g}_i\}$ and $\bar{p}_0$, $\left\{\hat{\mathbf{g}}_i\right\}$ define two different frames for the same affine space.

Remark 5.1 Compositions like $\psi = \hat{\psi} \circ \varphi$, where φ is given by (5.2) and $\hat{\psi} : \mathbb{R}^2 \to \mathbb{R}^2$ with $(x^1, x^1) \mapsto (\bar{x}^1 = x^1, \bar{x}^2 = x^1 x^2)$, are not affine mappings and are usually associated with the term "curvilinear coordinates." Such mappings are discussed in the next chapter.

Exercises

5.1 Consider the two affine spaces $S_1 = (X_1, \mathcal{V}_1, F_1)$ and $S_2 = (X_2, \mathcal{V}_2, F_2)$ with $X_1 = X_2 = \mathbb{R}^2$, $\mathcal{V}_1 = \mathcal{V}_2 = \mathbb{R}^2$ and $F_1 = F_2 = F$ according to Example 5.1 but in two dimensions. Furthermore, consider three points p_0, q_0, r_0 of S_1 and a given vector $\mathbf{v} \in \mathcal{V}_1$. The points p, q, r are determined such that

$$F(p_0, p) = F(q_0, q) = F(r_0, r) = \mathbf{v}\,.$$

Given the mapping

$$\begin{aligned} \mu : S_1 &\to S_2 \\ (s^1, s^2) &\mapsto (s^1 s^2, s^1 + s^2) \quad s \in S_1 \ : \end{aligned}$$

(a) Determine the images of the points p_0, p, q_0, q, r_0, r in S_2 under μ.
(b) Determine the vectors

$$\mathbf{a} = F(\mu(P_0), \mu(P)) \ , \ \mathbf{b} = F(\mu(Q_0), \mu(Q)) \ , \ \mathbf{c} = F(\mu(R_0), \mu(R)) \ ,$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{V}_2$.
(c) The vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ can be related by a linear mapping which is an isomorphism (1-1 and onto), but is there a unique mapping which maps $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{V}\in$ to $\mathbf{v} \in \mathcal{V}_1$?

5.2 Perform Exercise 5.1 for

$$\begin{aligned} \mu : S_1 &\to S_2 \\ (s^1, s^2) &\mapsto (s^1 + s^2, s^1 - s^2) \quad s \in S_1 \ . \end{aligned}$$

References

1. Crampin M, Pirani F (1987) Applicable differential geometry, vol 58. London mathematical society lecture note series. Cambridge University Press, Cambridge
2. Epstein M (2010) The geometrical language of continuum mechanics. Cambridge University Press, Cambridge

Chapter 6
Tensor Analysis in Euclidean Space

Abstract After briefly reviewing differentiability in $\mathbb{R}$, a generalization of differentiability based on the directional derivative in $\mathbb{R}^N$ is established. Gradients of scalar and vector fields are first discussed in $\mathbb{R}^N$ before adapting these concepts for general euclidean spaces in combination with global charts. Afterwards, nonlinear chart relations, also known as curvilinear coordinates, are examined, and the concept of tangent space at a point is introduced. In this context, the covariant derivative is derived, insinuating its character as special case of the covariant derivative for smooth manifolds. Aspects of integration based on differential forms are discussed together with the exterior derivative and Stoke's theorem in $\mathbb{R}^N$.

6.1 Differentiability in $\mathbb{R}$ and Related Concepts Briefly Revised

The derivative from the left of a function $f(x)$ at some $x = x_0$ is defined by

$$\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0-} = \lim_{h \to 0^-} \frac{1}{h}\left[f(x_0 + h) - f(x_0)\right] , \tag{6.1}$$

whereas the derivative from the right may be written as

$$\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0+} = \lim_{h \to 0^+} \frac{1}{h}\left[f(x_0 + h) - f(x_0)\right] . \tag{6.2}$$

These derivatives may or may not exist. If they exist, they may be equal or not. The existence and equality of both derivatives is referred to as a continuous derivative. It implies continuity of f at $x = x_0$ and

$$\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} = \lim_{h \to 0} \frac{1}{h}\left[f(x_0 + h) - f(x_0)\right] \tag{6.3}$$

U. Mühlich, *Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds*, Solid Mechanics and Its Applications 230,
DOI 10.1007/978-3-319-56264-3_6

is sufficiently precise. The reverse does not hold, a fact that can be seen easily from the standard example $f(x) = |x|$, which is continuous at $x = 0$ but not continuously differentiable. The derivative at some arbitrary x is indicated by $\frac{\mathrm{d}f}{\mathrm{d}x}$, and performing this operation yields a function of x. That way, derivatives of arbitrary order can be defined for which the notation

$$\left.\frac{\mathrm{d}^k f}{\mathrm{d}x^k}\right|_{x_0} = \lim_{h\to 0}\frac{1}{h}\left[\frac{\mathrm{d}^{k-1}}{\mathrm{d}x^{k-1}}f(x_0+h) - \frac{\mathrm{d}^{k-1}}{\mathrm{d}x^{k-1}}f(x_0)\right] \quad k = 1,\ldots,N$$

is commonly employed together with the definition

$$\frac{\mathrm{d}^0}{\mathrm{d}x^0}f := f\,.$$

This allows for a classification of functions by the following definition.

Definition 6.1 (*Set of continuously differentiable functions*) The set of all functions defined on an interval $[a, b]$ or $[a, b)$, etc., having continuous derivatives up to order k is called $C^k([a, b])$, or $C^k([a, b))$, respectively, etc. If the interval is $(-\infty, +\infty)$, the notation $C^k(\mathbb{R})$ is used.

The scheme sketched so far is rather intuitive and successful. However, it is not obvious how we might transfer the concept of being continuously differentiable in a straightforward manner to higher dimensions, and more general objects, such as tensor fields over $\mathbb{R}^N$. In order to obtain hints as to what a more general scheme might look like, it is useful to dissect (6.3) first. If f is continuously differentiable at x_0, then (6.3) exists and can be written as

$$\lim_{h\to 0}\frac{1}{h}\left[f(x_0+h) - f(x_0) - \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} h\right] = 0\,. \tag{6.4}$$

It reveals that being continuously differentiable implies the existence of a unique linear approximation of f at x_0,

$$\mathrm{d}|_{x_0} f(h) = \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} h\,, \tag{6.5}$$

which is the differential of f at x_0 and usually denoted by $\mathrm{d}f|_{x_0}$. The reason for employing a more expressive notation will become clear shortly. If f is continuously differentiable at x_0, then the differential is unique, because in this case, there exists, by definition, exactly one number $\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0}$. In one dimension, this also implies linearity of (6.5). The differential (6.5) can also be written as

$$\mathrm{d}_h|_{x_0} f = \lim_{t\to 0}\frac{1}{t}\left[f(x_0+th) - f(x_0)\right]\,, \tag{6.6}$$

which is the so-called directional derivative in one dimension. It has a straightforward generalization to higher dimensions just by replacing $x_0 + th$ by, e.g., $p_0 \boldsymbol{+} t\mathbf{h}$. Furthermore, general tensor functions on $\mathbb{R}^N$, e.g., $\mathsf{A}(p)$, do not induce any conceptual difficulty either. However, (6.6) must be handled with care, since it can also be computed for functions which are not continuously differentiable (see Exercise 6.2). The key is that continuous differentiability not only implies the existence of (6.6), but also its linearity wrt. h, or—vice versa—the existence of (6.6) implies continuous differentiability only if (6.6) is also linear wrt. h. Hence, the following three statements are equivalent:

1. the function $f(x)$ is continuously differentiable at $x = x_0$,
2. the derivatives (6.1) and (6.2) exist and are equal,
3. there exists a unique linear approximation of $f(x)$ at $x = x_0$,

where the second statement is tantamount to (6.3).

In the following, only functions belonging to C^∞ are considered in order to keep the discussion simple. Then, the right hand side of (6.6) can be interpreted as a linear mapping from C^∞ into $\mathbb{R}$,

$$\mathrm{d}_h|_{x_0} \;: C^\infty(\mathbb{R}) \to \mathbb{R}$$

or as a linear mapping from $\mathbb{R}$ into $\mathbb{R}$

$$\mathrm{d}|_{x_0} f \;: \mathbb{R} \to \mathbb{R}\,,$$

depending on whether a function f or a real number h is considered as the argument of (6.6).

If a function f is differentiable at a certain $x = x_0$, it can be approximated by a Taylor series

$$f(x_0 + t) \approx f(x_0) + \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} t + \frac{1}{2}\left.\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}\right|_{x_0} t^2 + \cdots$$

in the vicinity of x_0 exclusively using information about f at x_0.

In what follows, we consider a function $f(x)$ together with a mapping

$$\begin{aligned} g : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto y = \alpha(x)\,, \end{aligned}$$

where α is a function of x, e.g., $\alpha = 2x$. By means of its inverse

$$\begin{aligned} g^{-1} : \mathbb{R} &\to \mathbb{R} \\ y &\mapsto x = \beta(y) \end{aligned}$$

the function

$$\Phi = f \circ g^{-1} = f(\beta(y))$$

can be defined. With a slight abuse of notation, $\Phi(y) = f(x(y))$ is commonly used, as long as it does not cause confusion. The derivative of Φ with respect to y at $y = y_0$ reads as

$$\begin{aligned} \left.\frac{\mathrm{d}\Phi}{\mathrm{d}y}\right|_{y_0} &= \lim_{t \to 0} \frac{1}{t} \left[\Phi(y_0 + t) - \Phi(y_0)\right] \\ &= \lim_{t \to 0} \frac{1}{t} \left[f(x(y_0 + t)) - f(x(y_0))\right] . \end{aligned} \tag{6.7}$$

Linearizing first $x(y_0 + t)$ by applying Taylor series expansion

$$x(y_0 + t) \approx x(y_0) + \left.\frac{\mathrm{d}x}{\mathrm{d}y}\right|_{y_0} t$$

and taking into account that $x(y_0) = x_0$, a subsequent linearization can be performed:

$$f(x(y_0 + t)) \approx f(x_0) + \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} \left.\frac{\mathrm{d}x}{\mathrm{d}y}\right|_{y_0} t . \tag{6.8}$$

Using (6.8), (6.7) finally gives

$$\left.\frac{\mathrm{d}\Phi}{\mathrm{d}y}\right|_{y_0} = \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} \left.\frac{\mathrm{d}x}{\mathrm{d}y}\right|_{y_0} \tag{6.9}$$

which is known as chain rule in $\mathbb{R}$.

6.2 Generalization of the Concept of Differentiability

A powerful generalization of (6.6) is provided by the so-called directional derivative.

Definition 6.2 (*Directional derivative*) Given some finite-dimensional euclidean space $\mathcal{E}^n = (X, \mathcal{V}, \boldsymbol{+})$ and a mapping $\mathcal{F} : U \to \mathcal{V}$ where U is an open subset of X and $\mathcal{V}$ is some vector space. The directional derivative of $\mathcal{F}$ at point p in the direction determined by $\mathbf{h} \in \mathcal{V}$ is defined by

$$\underset{\sim}{\boldsymbol{d}}|_p\, \mathcal{F}(\mathbf{h}) = \lim_{t \to 0} \frac{1}{t} \left[\mathcal{F}(p \boldsymbol{+} t\mathbf{h}) - \mathcal{F}(p)\right] .$$

Because Definition 6.2 refers to the difference of images of $\mathcal{F}$, $\mathcal{V}$ has to be a vector space. Since the limit is performed with respect to a real number, Definition 6.2 is extremely flexible in terms of $\mathcal{F}$. Possible examples for $\mathcal{F}$ are $\mathcal{F} : U \to \mathbb{R}$, $\mathcal{F} : U \to \mathcal{W}$ or $\mathcal{F} : U \to \mathcal{W} \otimes \mathcal{W}$, etc., and $\mathcal{V}$ and $\mathcal{W}$ do not have to be equal. The directional derivative forms part of the following definition of differentiability.

Definition 6.3 (*Differentiability in $\mathcal{E}^n$*) Given some finite dimensional euclidean space $\mathcal{E}^n = (X, \mathcal{V}, \dotplus)$ and a mapping $\mathcal{F} : U \to \mathcal{V}$ where U is an open subset of X and $\mathcal{V}$ is some vector space. $\mathcal{F}$ is differentiable at p if

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{||\mathbf{h}||}\left[\mathcal{F}(p \dotplus \mathbf{h}) - \mathcal{F}(p) - \underset{\sim}{\boldsymbol{d}}|_p\, \mathcal{F}(\mathbf{h})\right] = 0$$

exists, and in addition, $\mathrm{d}|_p\, \mathcal{F}(\mathbf{h})$ is linear wrt. $\mathbf{h} \in \mathcal{V}$.

Definition 6.3 is based on the idea of relating differentiability to the existence of a linear approximation, as discussed within the previous section. It is worth noting that Definition 6.3 relies only on the norm defined for $\mathcal{E}^n$. Of course, the meaning of the RHS in Definition 6.3 depends on the meaning of $\mathcal{F}$.

It can be shown that if the directional derivative of $\mathcal{F}$ is linear with respect to $\mathbf{h}$, then $\mathcal{F}$ is differentiable in the sense of Definition 6.3. The type of linear mapping which ensures differentiability of $\mathcal{F}$ under consideration can be read off immediately from Definition 6.3. This allows for computing the linear approximation at a specific point and is discussed in the following.

It should be noted that Definitions 6.2 and 6.3 are also known as the Gateaux differential and the Fréchet derivative, respectively, in the much wider context of functional analysis in Banach spaces. Without going into detail, we remark that finite dimensional euclidean spaces are Banach spaces. However, the latter have been designed in the first place for infinite dimensional vector spaces, e.g., the space of all functions of a certain type defined on the interval $[a, b] \subset \mathbb{R}$.

Functions were characterized in Sect. 1.4 as mappings which assign a real number to every point of a given set. On the other hand, functions are also called scalar fields. Consistently, mappings which assign a vector to every point of a set are also known as vector fields. In the following, these terms are also used, because the term tensor field is simpler than talking about a mapping which assigns a tensor to every point.

6.3 Gradient of a Scalar Field and Related Concepts in $\mathbb{R}^N$

In $\mathbb{R}^N$, every point is addressed by an N-tuple of real numbers (see Definition 3.8). As already pointed out in Sect. 4.10, the standard basis $\{\mathbf{e}_i\}$ in $\mathbb{R}^N$ corresponds to columns with N entries. The k^{th} base vector $\mathbf{e}_k$ consists of zeros, except for the k^{th} entry, which equals one.

For $\mathbb{R}^N$ endowed with the usual inner product, $\left\langle \underset{\sim}{\mathbf{e}}^i, \mathbf{e}_j \right\rangle = \mathbf{e}_i \cdot \mathbf{e}^j$ holds (see Sect. 3.8). This allows for replacing dual vectors with reciprocal vectors, and correspondingly dual products with inner products. Here, we will make use of this eventually in order to provide those operators commonly used in conventional vector analysis.

Since $\{\mathbf{e}_i\}$ is an orthonormal basis, base vectors and reciprocal base vectors even coincide, i.e., $\mathbf{e}_i = \mathbf{e}^i$. Therefore, provided that summation convention is adapted accordingly, the distinction between vectors and reciprocal vectors, and correspondingly lower and upper indices, could be dropped as well. However, this implies the abandonment of a simple and efficient control mechanism regarding index notation. More importantly, all underlying structure becomes invisible without any real benefit in return. Therefore, we continue with the notation used so far.

For a given scalar function $f : \mathbb{R}^N \to \mathbb{R}$ to be continuously differentiable, $\underset{\sim}{\boldsymbol{d}}|_p \, \mathcal{F}(\mathbf{h})$ in (6.3) has to be a linear mapping

$$\underset{\sim}{\boldsymbol{d}}|_p \, f : \mathcal{V} \to \mathbb{R} \,, \tag{6.10}$$

with $\mathcal{V} = \mathbb{R}^N$. Please note the subtlety in the meaning of $\mathbb{R}^N$. While in the definition of f, $\mathbb{R}^N$ means just the set of points, (6.10) refers to $\mathbb{R}^N$ as a vector space according to Definition 3.8. Linearity is explored by applying the very same chain of steps widely used in Chap. 4,

$$\underset{\sim}{\boldsymbol{d}}|_p \, f(\mathbf{h}) = h^i \; \underset{\sim}{\boldsymbol{d}}\Big|_p \; f(\mathbf{e}_i) \,, \tag{6.11}$$

where the $\underset{\sim}{\boldsymbol{d}}\Big|_p \; f(\mathbf{e}_i)$ have to be real numbers. The abbreviation

$$\lambda_i := \underset{\sim}{\boldsymbol{d}}\Big|_p \; f(\mathbf{e}_i)$$

is used in the following. The remaining steps aim to separate the object which encodes the mapping (6.10) from the argument. This yields

$$\underset{\sim}{\boldsymbol{d}}|_p \, f(\mathbf{h}) = h^i \lambda_i = \lambda_i \left\langle \underset{\sim}{\boldsymbol{e}}^i , \mathbf{h} \right\rangle = \lambda_i \; \underset{\sim}{\boldsymbol{e}}^i(\mathbf{h}) \,,$$

and therefore

$$\underset{\sim}{\boldsymbol{d}}|_p \, f = \lambda_i \; \underset{\sim}{\boldsymbol{e}}^i \,. \tag{6.12}$$

This is the so-called gradient of f at p, which is a dual vector, i.e., $\underset{\sim}{\boldsymbol{d}}|_p \, f \in \mathcal{V}^*$. Computation of the coordinates λ_i requires the notion of partial derivatives.

Definition 6.4 (*Partial derivatives in* $\mathbb{R}^N$) The partial derivatives of a function $f(x^1, x^2, \ldots, x^N)$ at a point $p = (p^1, p^2, \ldots, p^N)$ with respect to the coordinate x^k, where $k = 1 \ldots N$, are defined by

$$\left.\frac{\partial f}{\partial x^k}\right|_{p^-} = \lim_{t \to 0^-} \frac{1}{t} \left[f(p^1, \ldots, p^{k-1}, p^k + t, p^{k+1}, \ldots, p^N) - f(p^1, p^2, \ldots, p^N) \right]$$
$$\left.\frac{\partial f}{\partial x^k}\right|_{p^+} = \lim_{t \to 0^+} \frac{1}{t} \left[f(p^1, \ldots, p^{k-1}, p^k + t, p^{k+1}, \ldots, p^N) - f(p^1, p^2, \ldots, p^N) \right] \,,$$

and if both derivatives exist and are equal,

$$\left.\frac{\partial f}{\partial x^k}\right|_p = \lim_{t\to 0}\frac{1}{t}\left[f(p^1,\ldots,p^{k-1},p^k+t,p^{k+1},\ldots,p^N) - f(p^1,p^2,\ldots,p^N)\right]$$

is used.

For the sake of simplicity, we first consider the case $N = 2$ in (6.10). According to Definition 6.4, there are two partial derivatives in $\mathbb{R}^2$:

$$\left.\frac{\partial f}{\partial x^1}\right|_p = \lim_{t\to 0}\frac{1}{t}\left[f(p^1+t,p^2) - f(p^1,p^2)\right]$$
$$\left.\frac{\partial f}{\partial x^2}\right|_p = \lim_{t\to 0}\frac{1}{t}\left[f(p^1,p^2+t) - f(p^1,p^2)\right].$$

Comparing Definition 6.4 with Eq. 6.11, taking into account hybrid addition in $\mathbb{R}^2$, reveals that

$$\left.\frac{\partial f}{\partial x^1}\right|_p = \underset{\sim}{\boldsymbol{d}}|_p\, f(\mathbf{e}_1) \quad , \quad \left.\frac{\partial f}{\partial x^2}\right|_p = \underset{\sim}{\boldsymbol{d}}|_p\, f(\mathbf{e}_2)\,,$$

since $p \boldsymbol{+} t\mathbf{e}_1 = (p^1, p^2) \boldsymbol{+} t\begin{pmatrix}1\\0\end{pmatrix} = (p^1+t, p^2)$, and similarly for $p \boldsymbol{+} t\mathbf{e}_2$. This extends to higher dimensions, i.e.,

$$\left.\frac{\partial f}{\partial x^i}\right|_p = \underset{\sim}{\boldsymbol{d}}|_p\, f(\mathbf{e}_i)\,. \tag{6.13}$$

On the other hand,

$$\underset{\sim}{\boldsymbol{d}}|_p\, f(\mathbf{e}_i) = \lambda_k \left\langle \underset{\sim}{\boldsymbol{e}}^k, \mathbf{e}_i \right\rangle = \lambda_i\,, \tag{6.14}$$

and therefore

$$\underset{\sim}{\boldsymbol{d}}|_p\, f = \left.\frac{\partial f}{\partial x^i}\right|_p \underset{\sim}{\boldsymbol{e}}^i \tag{6.15}$$

results. Furthermore, (6.15) shows that in $\mathbb{R}^N$, Definition 6.3 is equivalent to the existence of partial derivatives. Therefore, an alternative definition of being continuously differentiable can be given.

Definition 6.5 (*Set of continuously differentiable functions in* $\mathbb{R}^N$) The set of all functions defined on $\mathbb{R}^N$ having continuous partial derivatives up to order k is called $C^k(\mathbb{R}^N)$.

Not only are scalar fields usually of interest, but vector fields are as well. Next, we consider a vector field

$$\mathbf{f} : \mathbb{R}^N \to \mathcal{W}$$

with $\mathcal{W} = \mathbb{R}^N$. According to Definition 6.3, $\mathbf{f}$ is continuously differentiable at p if there exists a linear mapping

$$\underset{\sim}{\boldsymbol{d}}|_p \mathbf{f} : \mathcal{V} \rightarrow \mathcal{W}$$

with $\mathcal{V} = \mathbb{R}^N$. According to Definition 6.2, $\underset{\sim}{\boldsymbol{d}}\big|_p \mathbf{f}(\mathbf{h})$ is given explicitly by

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \mathbf{f}(\mathbf{h}) = \lim_{t \to 0} \frac{1}{t} \left[\mathbf{f}(p \dotplus t\mathbf{h}) - \mathbf{f}(p)\right] , \tag{6.16}$$

which, by using $\mathbf{f} = f^i \mathbf{e}_i$, can be written as

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \mathbf{f}(\mathbf{h}) = \left[\lim_{t \to 0} \frac{1}{t} \left[f^i(p \dotplus t\mathbf{h}) - f^i(p)\right]\right] \mathbf{e}_i . \tag{6.17}$$

Within the brackets, we find the directional derivatives of the scalar coordinate functions, hence

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \mathbf{f}(\mathbf{h}) = \left[\underset{\sim}{\boldsymbol{d}}\Big|_p f^i(\mathbf{h})\right] \mathbf{e}_i . \tag{6.18}$$

Taking into account (6.15), the final result reads as

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \mathbf{f}(\mathbf{h}) = \left[\left.\frac{\partial f^i}{\partial x^k}\right|_p \mathbf{e}_i \otimes \underset{\sim}{\boldsymbol{e}}^k\right](\mathbf{h})$$

where the term in brackets is the gradient of $\mathbf{f}$.

This scheme extends to general tensor fields in $\mathbb{R}^N$, and there is a number of possibilities for designing a more concise notation. For instance, a gradient operator

$$\nabla := \frac{\partial}{\partial x^i} \mathbf{e}^i$$

can be defined by taking into account the relation between inner product and dual product. This leads eventually to the following scheme:

$$\begin{aligned} \operatorname{grad}\Diamond &= \Diamond \otimes \nabla \\ \operatorname{div}\Diamond &= \Diamond \cdot \nabla \\ \operatorname{curl}\Diamond &= \Diamond \times \nabla , \end{aligned}$$

where $\Diamond$ is a place holder for a general tensor object, "·" denotes the inner product, and "×" is the cross product defined for $\mathbb{R}^3$. Due to the notation used so far, ∇ is defined here as the gradient from the right. Making the corresponding adjustments, it is also possible to work with a gradient from the left.

6.4 Differentiability in Euclidean Space Supposing Affine Relations

Analysis in $\mathbb{R}^N$ usually means treating some real world problem employing an appropriate coordinate chart. In what follows, the aim is to discuss how to relate analysis regarding physical space with the results obtained within the previous section, provided all involved mappings between charts and physical space are affine mappings. Such charts are called affine charts in the following.

For a given scalar field Φ defined over an euclidean space $\mathcal{E} = (X, \mathcal{V}, \dotplus)$ and continuously differentiable at point p, $\underset{\sim}{\boldsymbol{d}}\big|_p\ \mathcal{F}(\mathbf{h})$ in (6.3) has to be a linear mapping

$$\underset{\sim}{\boldsymbol{d}}\Big|_p\ \Phi : \mathcal{V} \to \mathbb{R}\,. \tag{6.19}$$

By now it should be clear that $\underset{\sim}{\boldsymbol{d}}\big|_p\ \Phi \in \mathcal{V}^*$ (see Sect. 3.6). Choosing a basis $\{\mathbf{g}_i\}$ for $\mathcal{V}$ and exploring linearity again eventually yields

$$\underset{\sim}{\boldsymbol{d}}\Big|_p\ \Phi = L_i\ \underset{\sim}{g}^i\,. \tag{6.20}$$

However, this result is only useful if the coordinates L_i of $\underset{\sim}{\boldsymbol{L}}$ can be computed. Computation, on the other hand, requires a chart. Since affine mappings are global homeomorphisms, i.e., they are invertible everywhere, we can work with global charts as long as the latter are generated by affine mappings. We consider a chart (X, μ) induced by a mapping $\mu : X \to \mathbb{R}^N$. The image of Φ in the chart is given by $f = \Phi \circ \mu^{-1}$, and regarding the corresponding $\underset{\sim}{\boldsymbol{d}}|_{\mu(p)} f$, we just have to adapt (6.15) accordingly

$$\underset{\sim}{\boldsymbol{d}}\Big|_{\mu(p)}\ f = \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} \underset{\sim}{\boldsymbol{e}}^i\,,$$

where $\mu(p)$ is the image of p in the chart. This is sketched for a two-dimensional problem in Fig. 6.1.

The value of Φ at a point p coincides with the value of f at the image of p in the chart, i.e., $\Phi(p) = f(\mu(p))$. The operation $\underset{\sim}{\boldsymbol{d}}\big|_p\ \Phi(\mathbf{H})$ refers to a limit process passing through points in X addressed by $p \dotplus t\mathbf{H}$. If the limit process in the chart, $\underset{\sim}{\boldsymbol{d}}\big|_{\mu(p)}\ f(\mathbf{h})$, passes through the images of the very same points, then both limit processes yield identical results.

Since μ is an affine mapping according to Definition 5.4 with an associated second order tensor B, then $\mu(q) = \mu(p) \dotplus \mathsf{B}(\mathbf{V})$ holds for two points p and q in X related by $q = p \dotplus \mathbf{V}$. Therefore,

$$\underset{\sim}{\boldsymbol{d}}\Big|_p\ \Phi(\mathbf{H}) = \underset{\sim}{\boldsymbol{d}}\Big|_{\mu(p)}\ f(\mathbf{h}) \tag{6.21}$$

holds if $\mathbf{h} = \mathsf{B}(\mathbf{H})$. The tensor B encodes a mapping

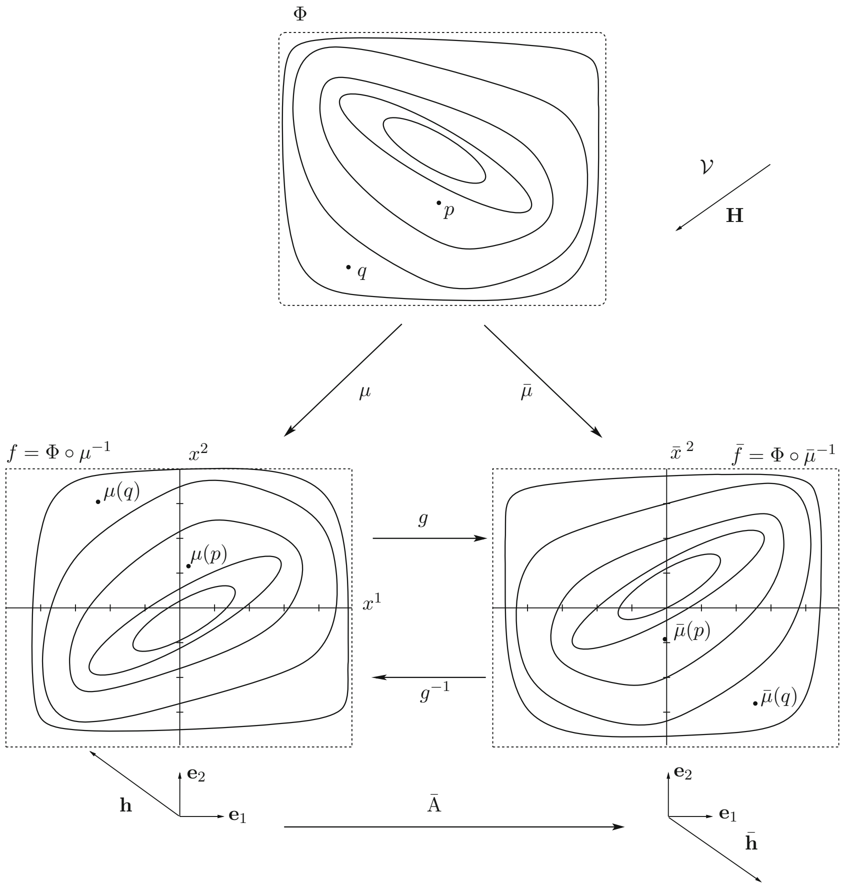

Fig. 6.1 Contour plot of a function Φ in a two-dimensional euclidean spaces together with two possible charts induced by affine mappings

$$\mu_\star : \mathcal{V} \rightarrow \mathbb{R}^N \tag{6.22}$$

called the push forward under the mapping μ. It relates the two vector spaces $\mathcal{V}$ and $\mathbb{R}^N$ based on (6.21), i.e., it establishes a correspondence between $\mathbf{H}$ and $\mathbf{h}$ in the sense that $\mathbf{H}$ acts on Φ as $\mathbf{h}$ acts on the image of Φ in the chart induced by μ. Its dual, defined by (see Sect. 4.3)

$$\left\langle \mu^\star(\underset{\sim}{\boldsymbol{\eta}}), \mathbf{H} \right\rangle = \left\langle \underset{\sim}{\boldsymbol{\eta}}, \mu_\star(\mathbf{H}) \right\rangle , \tag{6.23}$$

relates the corresponding dual spaces

$$\mu^\star : \mathbb{R}^{N*} \rightarrow \mathcal{V}^* \tag{6.24}$$

and is called the pull back under μ. The reason for assigning specific names to the mappings (6.22) and (6.24) lies in the importance of the underlying concepts. This is discussed in more detail later.

Evaluating (6.21) for the corresponding bases and taking into account the correspondence between $\mathbf{h}$ and $\mathbf{H}$ gives

$$L_k H^k = \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} h^i = \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} B^i_k H^k \,, \tag{6.25}$$

from which

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \Phi = \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} B^i_k \, \underset{\sim}{g}^k = \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} \underset{\sim}{\boldsymbol{e}}^i \tag{6.26}$$

follows. Depending on which coordinate base vector representation for the gradient of Φ is used, the corresponding argument, either $\mathbf{H} = H^k \mathbf{g}_k$ or $\mathbf{h} = h^k \mathbf{e}_k$, has to be provided when computing the directional derivative.

In the particularly simple case of affine coordinates, $B^i_k = \delta^i_k$ holds and, therefore,

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \Phi = \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} \underset{\sim}{g}^i = \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} \underset{\sim}{\boldsymbol{e}}^k \,,$$

which means that, at first glance, there is no difference to analysis in $\mathbb{R}^N$ except for the fact that f means the image of some function Φ. But for transferring results obtained through computations performed in charts back into the real world, the way physical space has been explored must be taken into account, i.e., eventually the meaning of the $\mathbf{g}_i$ matters.

In order to check that the scheme discussed above does not depend on the choice for a chart, another affine mapping $\overline{\mu}$ with its associated second order tensor $\overline{\mathsf{B}}$ is considered. It induces a chart $(X, \overline{\mu}) \subset \mathbb{R}^N$ with coordinates $\overline{x}^k$, and the image of Φ in this chart is $\overline{f} = \Phi \circ \overline{\mu}^{-1}$. Analogously to the case discussed above, we demand

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \Phi(\mathbf{H}) = \underset{\sim}{\boldsymbol{d}}\Big|_{\overline{\mu}(p)} \overline{f}(\overline{\mathbf{h}}) \,, \tag{6.27}$$

which is fulfilled if $\overline{\mathbf{h}} = \overline{\mathsf{B}}(\mathbf{H})$.

The aim in what follows is to show that the result for $\underset{\sim}{\boldsymbol{d}}\Big|_p \Phi(\mathbf{H})$ does not depend on the choice for a particular chart by showing that the right hand sides of (6.21) and (6.27) coincide, i.e.,

$$\underset{\sim}{\boldsymbol{d}}\Big|_{\mu(p)} f(\mathbf{h}) = \underset{\sim}{\boldsymbol{d}}\Big|_{\overline{\mu}(p)} \overline{f}(\overline{\mathbf{h}}) \,.$$

This requires some preparatory work. Using the identity

$$\Phi \circ \overline{\mu}^{-1} = \Phi \circ \mu^{-1} \circ \mu \circ \overline{\mu}^{-1} \,, \tag{6.28}$$

shows that $\overline{f} = f \circ g^{-1}$ where $g^{-1} = \mu \circ \overline{\mu}^{-1}$ encodes the transition between the charts (X, μ) and $(X, \overline{\mu})$. Since the composition of affine mappings is itself an affine mapping, g and g^{-1} are affine mappings provided that μ and $\overline{\mu}$ are affine. Therefore,

$$\begin{aligned} g : (X, \mu) &\to (X, \overline{\mu}) \\ (x^1, x^2, \ldots, x^N) &\mapsto (\overline{x}^1, \overline{x}^2, \ldots, \overline{x}^N) \text{ with } \overline{x}^i = \overline{c}^i + \overline{A}^i_k x^k \end{aligned} \tag{6.29}$$

and

$$\begin{aligned} g^{-1} : (X, \overline{\mu}) &\to (X, \mu) \\ (\overline{x}^1, \overline{x}^2, \ldots, \overline{x}^N) &\mapsto (x^1, x^2, \ldots, x^N) \text{ with } x^i = c^i + A^i_k \overline{x}^k , \end{aligned} \tag{6.30}$$

where the A^i_k and $\overline{A}^i_k$ are the coordinates of the tensors A and $\overline{\mathsf{A}}$ which encode the push forwards $g_\star$ and $(g^{-1})_\star$ between the vector spaces associated with the charts. It follows from (6.29) and (6.30) that

$$A^i_k = \frac{\partial x^i}{\partial \overline{x}^k} \quad \text{and} \quad \overline{A}^i_k = \frac{\partial \overline{x}^i}{\partial x^k} . \tag{6.31}$$

The vectors $\mathbf{h}$ and $\overline{\mathbf{h}}$ are related in the context of a push forward if $\mathbf{h} = \mathsf{A}(\overline{\mathbf{h}})$ and $\overline{\mathbf{h}} = \overline{\mathsf{A}}(\mathbf{h})$. The tensors A and $\overline{\mathsf{A}}$ read as

$$\mathsf{A} = \frac{\partial x^i}{\partial \overline{x}^k} \mathbf{e}_i \otimes \underset{\sim}{\boldsymbol{e}}^k \quad , \quad \overline{\mathsf{A}} = \frac{\partial \overline{x}^i}{\partial x^k} \mathbf{e}_i \otimes \underset{\sim}{\boldsymbol{e}}^k , \tag{6.32}$$

which reveals that $\overline{\mathsf{A}} = \mathsf{A}^{-1}$. The matrix of the coordinates of the second order tensor $\overline{\mathsf{A}}$ is also known as the Jacobian of the mapping g, whereas the matrix of the coordinates of A is the Jacobian of g^{-1}. The corresponding pull back operations $g^\star$ and $(g^{-1})^\star$ are encoded by the tensors $\underset{\sim}{\mathsf{A}}$ and $\underset{\sim}{\overline{\mathsf{A}}}$:

$$\underset{\sim}{\mathsf{A}} = \frac{\partial \overline{x}^i}{\partial x^k} \underset{\sim}{\boldsymbol{e}}^k \otimes \mathbf{e}_i \quad , \quad \underset{\sim}{\overline{\mathsf{A}}} = \frac{\partial x^i}{\partial \overline{x}^k} \underset{\sim}{\boldsymbol{e}}^k \otimes \mathbf{e}_i . \tag{6.33}$$

It is important to note that although push forward and pull back relations look similar to those which encode a change of basis introduced in Sect. 3.4, the concepts behind them differ. A push forward relates in general different vectors based on their action on functions, whereas a change of basis considers different representations of the very same vector.

After these preparatory considerations, we apply (6.28) to the rhs of (6.27),

$$\underset{\sim}{\boldsymbol{d}}\big|_{\overline{\mu}(p)} \overline{f}(\overline{\mathbf{h}}) = \left.\frac{\partial (f \circ g^{-1})}{\partial x^i}\right|_{\overline{\mu}(p)} (\overline{\mathbf{h}}) ,$$

which shows that a generalization of the chain rule (6.9) is needed to proceed. This generalization is not really challenging and is left as an exercise (see Exercise 6.4). Applying chain rule, also taking into account $\overline{\mathbf{h}} = \overline{\mathsf{A}}(\mathbf{h})$, gives

$$\underset{\sim}{\boldsymbol{d}}\Big|_{\overline{\mu}(p)} \overline{f}(\overline{\mathbf{h}}) = \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} \frac{\partial x^i}{\partial \overline{x}^k} \frac{\partial \overline{x}^k}{\partial x^m} h^m = \left.\frac{\partial \overline{f}}{\partial x^i}\right|_{\mu(p)} h^i ,$$

which reveals that (6.20) does not depend on the choice of a particular chart.

> The directional derivative of a function defined in an N-dimensional euclidean space $\mathcal{E} = (X, \mathcal{V}, \boldsymbol{+})$ can be computed by using a chart $(X, \mu) \subset \mathbb{R}^N$ obtained from an affine mapping $\mu : X \to \mathbb{R}^N$.

Similarly, the directional derivative $\underset{\sim}{\boldsymbol{d}}\big|_p \mathcal{F}(\mathbf{h})$ in (6.3) for a vector field $\mathbf{W}$ defined over an euclidean space $\mathcal{E} = (X, \mathcal{V}, \boldsymbol{+})$ is a linear mapping

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \mathbf{W} : \mathcal{V} \to \mathcal{V}$$

if $\mathbf{W}$ is continuously differentiable. Adapting (6.16) and (6.17) accordingly yields

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \mathbf{W}(\mathbf{H}) = \left[\underset{\sim}{\boldsymbol{d}}\Big|_p W^i(\mathbf{H})\right] \mathbf{g}_i$$

for a basis $\{\mathbf{g}_i\} \in \mathcal{V}$. After choosing a chart (X, μ) with μ according to Definition 5.4, (6.18) can be evaluated at a point $p \in X$ by

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \mathbf{W}(\mathbf{H}) = \left.\frac{\partial\, w^i}{\partial x^k}\right|_{\mu(p)} B^k_m H^m \,\mathbf{g}_i$$

with $w^i = W^i \circ \mu^{-1}$. Separation of the mapping from the argument can be achieved, as usual, by using $H^m = \left\langle \underset{\sim}{g}^m, \mathbf{H}\right\rangle$:

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \mathbf{W}(\mathbf{H}) = \left[\left.\frac{\partial\, w^i}{\partial x^k}\right|_{\mu(p)} B^k_m \,\mathbf{g}_i \otimes \underset{\sim}{g}^m\right](\mathbf{H}) ,$$

where the term within brackets is the gradient of $\mathbf{W}$ at p with reference to the bases $\{\mathbf{g}_i\} \in \mathcal{V}$. Using affine coordinates, the gradient of $\mathbf{W}$ reads as

$$\underset{\sim}{\boldsymbol{d}}\Big|_p \mathbf{W} = \left.\frac{\partial\, w^i}{\partial x^k}\right|_{\mu(p)} \mathbf{g}_i \otimes \underset{\sim}{g}^k = \left.\frac{\partial\, w^i}{\partial x^k}\right|_{\mu(p)} \mathbf{e}_i \otimes \underset{\sim}{\boldsymbol{e}}^k ,$$

and there is again no difference to analysis in $\mathbb{R}^N$ except for the use of images of functions originally defined in some euclidean space.

So far, the change of scalar and vector fields in direction of given vectors has been expressed by means of the directional derivative. In what follows, the change of such fields along general curves $\Gamma : \mathbb{R} \to X$ is examined. But first, Γ itself is discussed by means of

$$\dot{\Gamma}(t_0) = \lim_{t\to 0} \frac{1}{t} \left[\Gamma(t_0 + t) - \Gamma(t_0)\right]$$

with $\Gamma(t_0) = p$. Using an affine chart (X, μ), we have

$$\dot{\gamma}(t_0) = \lim_{t\to 0} \frac{1}{t} \left[\gamma(t_0 + t) - \gamma(t_0)\right] ,$$

where $\gamma = \mu \circ \Gamma$ is the image of Γ in (X, μ). The term within brackets is a difference between points, and therefore corresponds to some vector. The result of the limit process $\dot{\gamma}(t_0)$ is the vector

$$\mathbf{t}|_{\mu(p)} = \dot{\gamma}(t_0) \tag{6.34}$$

tangent to γ at $\gamma(t_0) = \mu(p)$. Its push forward under μ^{-1} is the vector tangent to Γ at p

$$\mathbf{T}|_p = \mu_\star^{-1}(\mathbf{t}|_{\mu(p)}) .$$

The change of a scalar field Φ along Γ is examined by means of

$$\lim_{t\to 0} \frac{1}{t} \left[\Phi \circ \Gamma(t_0 + t) - \Phi \circ \Gamma(t_0)\right] .$$

Once again using a chart (X, μ) and $\Phi \circ \Gamma = \Phi \circ \mu^{-1} \circ \mu \circ \Gamma$, we obtain

$$\lim_{t\to 0} \frac{1}{t} \left[f \circ \gamma(t_0 + t) - f \circ \gamma(t_0)\right] ,$$

where $f = \Phi \circ \mu^{-1}$ is the image of Φ in (X, μ) and $\mu \circ \Gamma$ is the corresponding image of Γ. Applying the chain rule yields

$$\lim_{t\to 0} \frac{1}{t} \left[f \circ \gamma(t_0 + t) - f \circ \gamma(t_0)\right] = \lim_{t\to 0} \frac{1}{t} \left[f \circ \gamma(t_0) + \left.\frac{\partial f}{\partial x^k}\right|_{\mu(p)} \dot{\gamma}^k(t_0) t - f \circ \gamma(t_0) \right] ,$$

from which it follows that

$$\lim_{t\to 0} \frac{1}{t} \left[f \circ \gamma(t_0 + t) - f \circ \gamma(t_0)\right] = \left.\underset{\sim}{\boldsymbol{d}}\right|_{\mu(p)} f(\mathbf{t}|_{\mu(p)}) = \left.\underset{\sim}{\boldsymbol{d}}\right|_p \Phi(\mathbf{T}|_p) . \tag{6.35}$$

Similarly, the change of a vector field $\mathbf{W} : X \to \mathcal{V}$ along a curve Γ is examined by means of

$$\lim_{t\to 0} \frac{1}{t} \left[\mathbf{W} \circ \Gamma(t_0 + t) - \mathbf{W} \circ \Gamma(t_0)\right] = \lim_{t\to 0} \frac{1}{t} \left[W^i \circ \Gamma(t_0 + t) - W^i \circ \Gamma(t_0)\right] \mathbf{g}_i \tag{6.36}$$

regarding a basis $\{\mathbf{g}_i\} \in \mathcal{V}$. Based on previous analysis within this section, it is immediately clear that

$$\lim_{t\to 0} \frac{1}{t} \left[\mathbf{W} \circ \Gamma(t_0 + t) - \mathbf{W} \circ \Gamma(t_0)\right] = \underset{\sim}{d}\Big|_p \mathbf{w}(\mathbf{t}|_{\mu(p)}) = \underset{\sim}{d}\Big|_p \mathbf{W}(\mathbf{T}|_p) \tag{6.37}$$

holds, extending in a straightforward manner to tensor fields of arbitrary order. Therefore, the results (6.35) and (6.37) can be summarized as follows.

The derivative of a tensor field along a curve Γ at some point p coincides with the directional derivative in the direction of the vector tangent to Γ at p.

Since it has been shown already that the result of the directional derivative is independent of the choice of a particular affine chart, the derivative along a curve is also chart independent.

Finally, in view of the remaining sections of this chapter, we explore an interpretation of the derivative of vector fields along curves which, in addition, reveals an important subtlety of the push forward of vector fields. Interpreting $f = \Phi \circ \mu^{-1}$ as the push forward of Φ under μ and $\mu \circ \Gamma$ as the respective push forward of Γ, the LHS of (6.35) encodes the limit process along the push forward of Γ using the push forward of Φ. Similarly, one can think of performing the limit process along the push forward of Γ considering the push forward of the vector field $\mathbf{W}$ under μ and pushing forward the result using $\mu_\star^{-1}$. The RHS of (6.36) can be written as

$$\lim_{t\to 0} \frac{1}{t} \left[W^i \circ \mu^{-1} \circ \mu \circ \Gamma(t_0 + t) - W^i \circ \mu^{-1} \circ \mu \circ \Gamma(t_0)\right] \mathbf{g}_i \, . \tag{6.38}$$

The push forward of the above equation reads as

$$\lim_{t\to 0} \frac{1}{t} \left[W^i \circ \mu^{-1} \circ \mu \circ \Gamma(t_0 + t) - W^i \circ \mu^{-1} \circ \mu \circ \Gamma(t_0)\right] \mu_\star(\mathbf{g}_i) \, , \tag{6.39}$$

which reveals that the push forward of a vector field $\mathbf{W}$ under a mapping μ should be defined as the push forward of $\mathbf{W} \circ \mu^{-1}$,

$$\mu_\star(\mathbf{W} \circ \mu^{-1}) = \mu_\star(W^i \circ \mu^{-1}\mathbf{g}_i) = W^i \circ \mu^{-1}\mu_\star(\mathbf{g}_i) \, , \tag{6.40}$$

in contrast to the push forward of an ordinary vector $V = V^i\mathbf{g}_i$ where the V^i are just given numbers. Although, $\mu_\star$ does not depend on location, the push forward operation (6.40) does, because of μ^{-1}.

The result of (6.39) is given by

$$\left.\frac{\partial w^i}{\partial x^k}\right|_{\mu(p)} t^k\big|_{\mu(p)}\, \mu_\star(\mathbf{g}_i)\,,$$

which is an ordinary vector and its push forward under μ^{-1} equals (6.37).

Defining the push forward of vector fields according to (6.40) ensures that the result of the push forward is a well-defined vector field in $\mathbb{R}^N$ in the sense that $\mu_\star(\mathbf{W}\circ\mu^{-1}) : \mathbb{R}^N \to \mathbb{R}^N$ and not a mapping from X to $\mathbb{R}^N$.

6.5 Characteristic Features of Nonlinear Chart Relations

As discussed within the previous section, one global chart (X, μ) induced by an affine mapping is sufficient to perform computations related to $\mathcal{E} = (X, \mathcal{V}, \boldsymbol{+})$. However, it can be more convenient to adapt to the geometry of a considered problem by using charts generated by nonlinear mappings, called nonlinear charts from here on in.

A number of problems emerge in the case of nonlinear chart relations. For instance, nonlinear mappings are usually not global homeomorphisms, but there exist, in general, points where either the mapping itself or its inverse are not well defined. Therefore, the use of nonlinear charts is limited to subsets, e.g., $U \subset X$, for which the corresponding mappings exist and are invertible. The following example is used to illustrate the key ideas discussed in the remaining part of this chapter.

Example 6.1 Given an euclidean space $\mathcal{E} = (X, \mathcal{V}, \boldsymbol{+})$ and $U \subset X$. In addition, there are two charts $(X, \mathcal{V})$ and (U, μ). The mapping $\mu : X \to \mathbb{R}^2$ is an affine mapping, whereas $\mu : U \to \mathbb{R}^2$ is nonlinear. The example is depicted in Fig. 6.2. The coordinates in $(X, \mathcal{V})$ are labeled by (x^1, x^2) whereas $(\overline{x}^1, \overline{x}^2)$ are used for (U, μ). The image of a function $\Phi : U \to \mathbb{R}$ in chart $(X, \mathcal{V})$ is given by

$$\begin{aligned} f : (\overline{X}, \mathcal{V}) &\to \mathbb{R} \\ (x^1, x^2) &\mapsto x^1 + x^2 \,. \end{aligned}$$

The charts are related by the mapping g,

$$\begin{aligned} g : (\overline{X}, \mathcal{V}) &\to (U, \mu) \\ (x^1, x^2) &\mapsto \left(\overline{x}^1 = \sqrt{\left(x^1\right)^2 + \left(x^2\right)^2}, \overline{x}^2 = \arctan\frac{x^2}{x^1}\right)\,, \end{aligned}$$

or, respectively, its inverse,

$$\begin{aligned} g^{-1} : (U, \mu) &\to (\overline{X}, \mathcal{V}) \\ (\overline{x}^1, \overline{x}^2) &\mapsto \left(x^1 = \overline{x}^1 \cos\overline{x}^2, x^2 = \overline{x}^1 \sin\overline{x}^2\right)\,, \end{aligned}$$

which should be familiar to the reader at the latest when setting $\overline{x}^1 = r$ and $\overline{x}^2 = \varphi$.

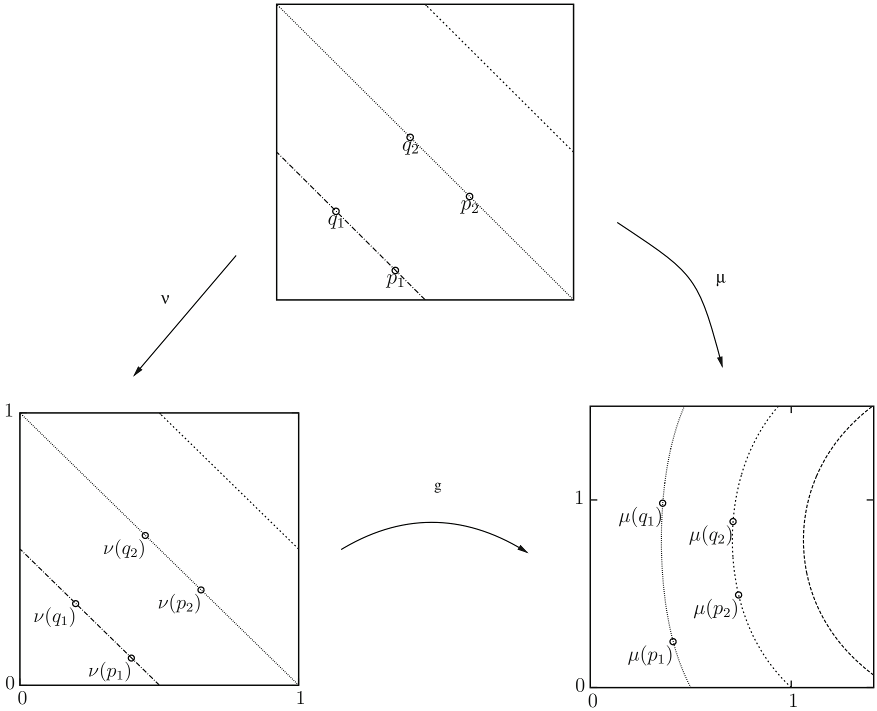

Fig. 6.2 Graphical representation of Example 6.1 showing four points in $U \subset X$ and their images in the charts (U, ν) and (U, μ) together with contour lines of $\Phi : U \to \mathbb{R}$, where the representation of the latter in (U, ν) is given by $f = \Phi \circ \nu^{-1} = x^1 + x^2$

The visualization of Example 6.1 in Fig. 6.2 reveals the essential difference between affine and nonlinear chart relations.

For affine charts, push forward and pull back operations are valid globally. In the case of nonlinear chart relations, however, push forward and pull back depend on position.

The dependence of push forward and pull back operations on location has severe consequences regarding the computation of directional derivatives. As discussed in detail within the previous section, such derivatives rely on evaluating differences, hence multiplication by scalar and summation. For scalar fields, no difficulties arise in the case of nonlinear charts in terms of adding values associated with different points.

The interpretation related to the computation of directional derivatives of vector fields at the end of the previous section can give us an impression regarding the difficulties to be expected if nonlinear charts are used. We consider an euclidean space $\mathcal{E} = (X, \mathcal{V}, \boldsymbol{+})$ and vectors $\mathbf{W}|_p$ and $\mathbf{W}|_q$ assigned to different points p and q in X. In order to see what happens if we work with a nonlinear chart (U, μ), we apply the respective push forwards which now depend on location, i.e.,

$$\mathbf{w}|_{\mu(p)} = \mu_\star|_p \left(\mathbf{W}|_p\right) \quad , \quad \mathbf{w}|_{\mu(q)} = \mu_\star|_q \left(\mathbf{W}|_q\right).$$

Computing the differences $\mathbf{W}|_p - \mathbf{W}|_q$ and $\mathbf{w}|_{\mu(p)} - \mathbf{w}|_{\mu(q)}$ reveals that this scheme will definitely fail. For instance, $\mathbf{W}|_p = \mathbf{W}|_q$ does not necessarily mean that the difference of the push forwards also vanishes. Furthermore, whatever the meaning of the result in the chart might be, we do not even know if we shall use $\mu_\star^{-1}$ at $\mu(p)$ or at $\mu(q)$ in order to bring it back to $\mathcal{E}$.

Before giving a solution for this problem, the concept of tangent space at a point is introduced within the next section because it is more suitable for attacking the problem.

6.6 Partial Derivatives as Vectors and Tangent Space at a Point

The dependence of push forwards and pull backs on location in the chart can be dealt with in at least two ways. The first consists in working with global vector spaces associated with the charts and taking into account the dependence on location in the respective computations.

However, there is an alternative concept eventually generating a notational machinery that is also more suitable in view of Chap. 7. This concept is described in the following. First of all, an alternative interpretation of $\underline{\boldsymbol{d}} f(\mathbf{h})$ at a point $\mu(p)$ in some chart (U, μ) is employed in the sense that $\underline{\boldsymbol{d}} f(\mathbf{h})$ at $\mu(p)$ is the result of the action of a vector on a function f,

$$\begin{aligned} \mathbf{h} : C^\infty(\mathbb{R}^N) &\to \mathbb{R} \\ f &\mapsto \lim_{t \to 0} \frac{1}{t}\left[f(\mu(p) \boldsymbol{+} t\mathbf{h}) - f(\mu(p))\right], \end{aligned}$$

or, respectively,

$$f \mapsto \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} h^i \tag{6.41}$$

due to (6.15). This mapping is linear and it obeys the product rule, also known as the Leibniz rule

$$\mathbf{h}(fg) = f(\mu(p))\mathbf{h}(g) + \mathbf{h}(f)g(\mu(p))$$

for $\mathbf{h}$ applied at p to the product of two functions f and g. This motivates the following definition.

Definition 6.6 (*Derivation at a point in* $\mathbb{R}^N$) A derivation at a point $p \in \mathbb{R}^N$ is a linear mapping

$$D : C^\infty(\mathbb{R}^N) \to \mathbb{R} \tag{6.42}$$

which obeys the Leibniz rule.

For every derivation D, $D(c) = 0$ if c is a constant. Furthermore, $f(p) = 0$ and $g(p) = 0$ implies $D(fg) = 0$. The Leibniz rule additionally ensures that derivations are always of the form (6.41) (see Exercise 6.5). Therefore, the space of all derivations at a point $\mu(p)$, subsequently designated tangent space at $\mu(p)$ denoted by $T_{\mu(p)}\mathbb{R}^N$, is isomorphic to $\mathbb{R}^N$. Detailed proofs can be found, e.g., in Lee [4] or Loring [5].

In light of the fact that elements of $\mathbb{R}^N$ can be identified with derivations at a point, partial derivatives can be interpreted as vectors, e.g.,

$$\begin{aligned} \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(x)} &: C^\infty(\mathbb{R}^N) \to \mathbb{R} \\ & f \mapsto \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(x)} f = \left.\frac{\partial f}{\partial x^i}\right|_{\mu(p)} , \end{aligned}$$

where the operator character is emphasized here by notation. Since $\mathbf{e}_i$ in $\underline{\boldsymbol{d}}\big|_p f(\mathbf{e}_i)$ reads off the i-th partial derivative, $\left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(x)}$ can be identified with $\mathbf{e}_i$ if necessary. This identification implies that the set $\left\{ \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(x)} \right\}$, $i = 1 \dots N$, defines a basis for $T_{\mu(p)}\mathbb{R}^N$.

Interpreting partial derivative operators as vectors requires the existence of summation and multiplication by scalar. Therefore, we define

$$\begin{aligned} \left[u^i \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(x)} \oplus v^i \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(x)} \right] f &= [u^i + v^i] \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(x)} f, \\ \left[\lambda \odot u^i \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(x)} \right] f &= \lambda u^i \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(x)} f . \end{aligned}$$

A tangent space always refers to a corresponding point. The chart is implied by the coordinates, e.g., x^i or $\overline{x}^i$. Therefore, unless there is possible confusion, explicit indication of the point in the chart is omitted in the following with the understanding that we refer always to a point and its respective tangent space.

For two charts (U, μ) and $(U, \overline{\mu})$ related by the mapping g, the basis in $T_{g(p)}\mathbb{R}^N$ is related to the corresponding basis in $T_p\mathbb{R}^N$ by

$$g_\star\left(\frac{\partial}{\partial \boldsymbol{x}^i}\right) = \frac{\partial \overline{x}^k}{\partial x^i}\frac{\partial}{\partial \overline{\boldsymbol{x}}^k} \quad \text{and} \quad g_\star^{-1}\left(\frac{\partial}{\partial \overline{\boldsymbol{x}}^i}\right) = \frac{\partial x^k}{\partial \overline{x}^i}\frac{\partial}{\partial \boldsymbol{x}^k}\,. \tag{6.43}$$

Furthermore, given a vector $\overline{\mathbf{u}} \in T_{g(p)}\mathbb{R}^N$ as a push forward of a vector $\mathbf{u} \in T_p\mathbb{R}^N$ and considering

$$g_\star(\mathbf{u}) = u^i g_\star\left(\frac{\partial}{\partial \boldsymbol{x}^i}\right) = u^i \frac{\partial \overline{x}^k}{\partial x^i}\frac{\partial}{\partial \overline{\boldsymbol{x}}^k} = \overline{u}^i \frac{\partial}{\partial \overline{\boldsymbol{x}}^i}$$

yields

$$\overline{u}^i = \frac{\partial \overline{x}^i}{\partial x^k} u^k$$

for the transformation of the coordinates.

In order to find a corresponding representation for elements of the dual space $T_p^*\mathbb{R}^k$, we evaluate the gradient of the i-th coordinate function according to (6.15):

$$\underset{\sim}{\boldsymbol{d}} x^i = \frac{\partial x^i}{\partial x^k}\underset{\sim}{\boldsymbol{e}}^k = \delta^i_k\, \underset{\sim}{\boldsymbol{e}}^k = \underset{\sim}{\boldsymbol{e}}^i\,.$$

Therefore,

$$\underset{\sim}{\boldsymbol{d}} x^i : T_p\mathbb{R}^k \rightarrow \mathbb{R}\,,$$

where $\underset{\sim}{\boldsymbol{d}} x^i \in T_p^*\mathbb{R}^k$ can be identified with $\underset{\sim}{\boldsymbol{e}}^i$ if necessary. The identification of the base vectors and dual base vectors with $\mathbf{e}_i$ and $\underset{\sim}{\boldsymbol{e}}^i$, respectively, implies

$$\left\langle \underset{\sim}{\boldsymbol{d}} x^i, \frac{\partial}{\partial \boldsymbol{x}^k}\right\rangle = \delta^i_k\,,$$

and since, analogously,

$$\left\langle \underset{\sim}{\boldsymbol{d}} \overline{x}^i, \frac{\partial}{\partial \overline{\boldsymbol{x}}^k}\right\rangle = \delta^i_k$$

holds, the pull back relations for the corresponding dual vectors

$$(g^{-1})^\star\left(\underset{\sim}{\boldsymbol{d}} x^i\right) = \frac{\partial x^i}{\partial \overline{x}^k}\,\underset{\sim}{\boldsymbol{d}}\,\overline{x}^k \quad , \quad g^\star\left(\underset{\sim}{\boldsymbol{d}} \overline{x}^i\right) = \frac{\partial \overline{x}^i}{\partial x^k}\,\underset{\sim}{\boldsymbol{d}} x^k$$

can be deduced by taking into account (6.43). Finally, $\underset{\sim}{\boldsymbol{d}} f$ can be encoded by

$$\underset{\sim}{\boldsymbol{d}} f = \frac{\partial f}{\partial x^k}\underset{\sim}{\boldsymbol{e}}^i = \frac{\partial f}{\partial x^k}\underset{\sim}{\boldsymbol{d}} x^i\,.$$

Box 6.6.1 Summary regarding push forward and pull back for charts (U, μ) with coordinates x^i and $(U, \overline{\mu})$ with coordinates $\overline{x}^i$ related by g.

$$g_\star\left(\frac{\partial}{\partial \boldsymbol{x}^i}\right) = \frac{\partial \overline{x}^k}{\partial x^i}\frac{\partial}{\partial \overline{\boldsymbol{x}}^k} \qquad g_\star^{-1}\left(\frac{\partial}{\partial \overline{\boldsymbol{x}}^i}\right) = \frac{\partial x^k}{\partial \overline{x}^i}\frac{\partial}{\partial \boldsymbol{x}^k}$$

$$g^\star\left(\underset{\sim}{\boldsymbol{d}}\,\overline{x}^i\right) = \frac{\partial \overline{x}^i}{\partial x^k}\,\underset{\sim}{\boldsymbol{d}}x^k \qquad (g^{-1})^\star\left(\underset{\sim}{\boldsymbol{d}}x^i\right) = \frac{\partial x^i}{\partial \overline{x}^k}\,\underset{\sim}{\boldsymbol{d}}\,\overline{x}^k$$

$$\left\langle \frac{\partial}{\partial \boldsymbol{x}^i}, \underset{\sim}{\boldsymbol{d}}x^k \right\rangle = \delta_i^k \qquad \left\langle \frac{\partial}{\partial \overline{\boldsymbol{x}}^i}, \underset{\sim}{\boldsymbol{d}}\overline{x}^k \right\rangle = \delta_i^k$$

Given $\mathcal{A} = (X, \mathcal{V}, \boldsymbol{+})$ and $U \subset X$. The tangent space at a point p in U, $T_p\mathcal{A}$ is related to the tangent space at the image of p in a chart (X, U) by the push forward under μ^{-1} at $\mu(p)$. A basis in $T_{\mu(p)}\mathbb{R}^N$ induces a basis $\boldsymbol{\partial}_i|_p$ in $T_p\mathcal{A}$ given by

$$\boldsymbol{\partial}_i|_p = \mu_\star^{-1}\left(\left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(p)}\right). \tag{6.44}$$

From $\mathbb{R}^N \cong \mathcal{V}$ and $T_{\mu(p)}\mathbb{R}^N \cong \mathbb{R}^N$, we deduce $T_p\mathcal{A} \cong \mathcal{V}$.

At the end of this section, a property of affine respectively euclidean spaces called absolute parallelism is discussed in terms of tangent spaces at points.

Definition 6.7 (*Absolute parallelism*) Given an affine space $\mathcal{A} = (X, \mathcal{V}, \boldsymbol{+})$. Two vectors $\mathbf{u}$ and $\mathbf{v}$ where $\mathbf{u} \in T_p\mathcal{A}$ and $\mathbf{v} \in T_q\mathcal{A}$ are parallel if their coordinates with respect to a basis $\{\mathbf{g}_i\} \in \mathcal{V}$ are equal.

6.7 Curvilinear Coordinates and Covariant Derivative

Within this section, the following situation is considered. Given a euclidean space $\mathcal{E} = (X, \mathbf{V}, \boldsymbol{+})$ and a global affine chart (X, μ). To adapt, for instance, to the geometry of a specific problem, a nonlinear chart $(U, \overline{\mu})$ with $U \subset X$ is used in addition. This procedure is known as curvilinear coordinates since the images of the coordinate lines of $(U, \overline{\mu})$ in (X, μ) appear as curved lines. In $\mathcal{E}$, a vector field $\mathbf{W}$ exists, and the respective push forwards in the charts are denoted by $\mathbf{w}$ and $\overline{\mathbf{w}}$.

The aim is to determine the derivative of $\mathbf{W}$ along a curve $\Gamma : \mathbb{R} \to X$ at $p = \Gamma(t_0)$ by means of the respective computation in chart $(U, \overline{\mu})$. The images of Γ in the charts are γ and $\overline{\gamma}$.

$\mathcal{E}$, as well as (X, μ) and $(U, \overline{\mu})$, disposes of an absolute parallelism. However, while $\mathcal{E}$ and (X, μ) are compatible in this respect, the parallelism of $(U, \overline{\mu})$ is incompatible with the remaining two due to the nonlinearity of $\overline{\mu}$. Therefore, we define the derivative

$$\nabla_{\overline{\gamma}} \overline{\mathbf{w}} = \lim_{t \to t_0} \frac{1}{t} \left[(\overline{\mathbf{w}} \circ \overline{\gamma}(t_0 + t))_{||} - \overline{\mathbf{w}} \circ \overline{\gamma}(t_0) \right] , \tag{6.45}$$

where $()_{||}$ stands for "transporting in parallel" from $T_{\overline{\gamma}(t_0+t)}\mathbb{R}^N$ to $T_{\overline{\gamma}(t_0)}\mathbb{R}^N$, but in terms of the parallelism in $\mathcal{E}$. Using the representation of vectors by means of coordinates and basis, it is sufficient to discuss the parallel transport of $\frac{\partial}{\partial \overline{x}^i}$ from $T_{\overline{\gamma}(t_0+t)}\mathbb{R}^N$ to $T_{\overline{\gamma}(t_0)}\mathbb{R}^N$. This should yield a relation

$$\left(\frac{\partial}{\partial \overline{\boldsymbol{x}}^i} \bigg|_{\overline{\gamma}(t_0+t)} \right)_{||} = a_i^k(t_0 + t) \, \frac{\partial}{\partial \overline{\boldsymbol{x}}^k} \bigg|_{\overline{\gamma}(t_0)} . \tag{6.46}$$

In order to determine the a_i^k, the following scheme is used:

1. push forward $\frac{\partial}{\partial \overline{x}^i}$ at $\overline{\gamma}(t_0 + t)$ under g^{-1}
2. parallel transport of the result of step 1 from $T_{\gamma(t_0+t)}\mathbb{R}^N$ to $T_{\gamma(t_0)}\mathbb{R}^N$ by the identity map Id
3. express the result of step 2 in terms of the push forwards of the base vectors $\frac{\partial}{\partial \overline{x}^i}$ under g^{-1} at $\overline{\gamma}(t_0)$
4. push forward the result of step 3 under g at $\gamma(t_0)$

which is also sketched in Fig. 6.3.

Performing step 1 and denoting the result of the respective push forward by $\mathbf{G}_i|_{\gamma(t_0+t)}$, we have

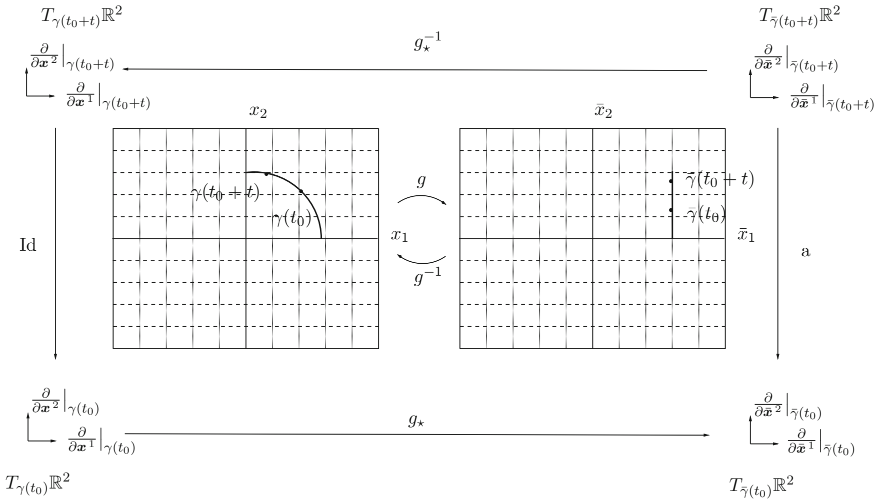

Fig. 6.3 Sketch of parallel transport in $(U, \overline{\mu})$ according to the notion of parallelism in $\mathcal{E}$

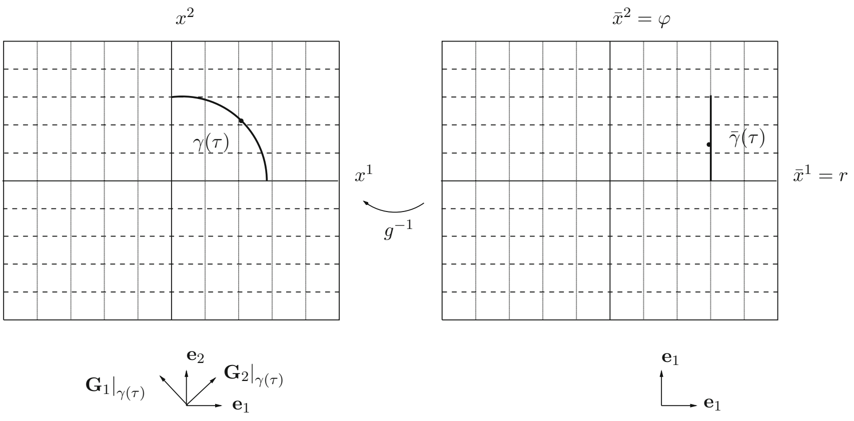

Fig. 6.4 Meaning of the vectors $\mathbf{G}_i$ for Example 6.1. Regarding chart (X, μ) (LHS), the $\mathbf{e}_i$ can be identified with $\frac{\partial}{\partial \boldsymbol{x}^i}$ whereas they are identified with respect to chart $(U, \overline{\mu})$ (RHS) with $\frac{\partial}{\partial \overline{\boldsymbol{x}}^i}$

$$\mathbf{G}_i|_{\gamma(t_0+t)} = g_\star^{-1}\left(\frac{\partial}{\partial \overline{\boldsymbol{x}}^i}\right)_{\overline{\gamma}(t_0+t)} = \left.\frac{\partial x^m}{\partial \overline{x}^i}\right|_{\overline{\gamma}(t_0+t)} \frac{\partial}{\partial \boldsymbol{x}^m} .$$

Regarding the meaning of $\mathbf{G}_i$, see also Fig. 6.4. Step 2 is trivial, and for step 3,

$$\mathbf{G}_i|_{\gamma(t_0)} = g_\star^{-1}\left(\frac{\partial}{\partial \overline{\boldsymbol{x}}^i}\right)_{\overline{\gamma}(t_0)} = \left.\frac{\partial x^m}{\partial \overline{x}^i}\right|_{\overline{\gamma}(t_0)} \frac{\partial}{\partial \boldsymbol{x}^m}$$

is required. From the above equation,

$$\frac{\partial}{\partial \boldsymbol{x}^m} = \left.\frac{\partial \overline{x}^j}{\partial x^m}\right|_{\overline{\gamma}(t_0)} \mathbf{G}_j|_{\gamma(t_0)}$$

can be deduced, by which (6.7) can be written as

$$\mathbf{G}_i|_{\gamma(t_0+t)} = \left.\frac{\partial x^m}{\partial \overline{x}^i}\right|_{\overline{\gamma}(t_0+t)} \left.\frac{\partial \overline{x}^j}{\partial x^m}\right|_{\gamma(t_0)} \mathbf{G}_j|_{\gamma(t_0)} = a_i^j\,(\overline{\gamma}(t_0+t))\;\; \mathbf{G}_j|_{\gamma(t_0)} .$$

Performing step 4 with the result above as input yields (6.46).

Based on the results obtained above, (6.45) reads as

$$\nabla_{\overline{\gamma}} \overline{\mathbf{w}} = \lim_{t\to t_0} \frac{1}{t}\left[\overline{w}^i \circ \overline{\gamma}(t_0+t) a_i^m(\overline{\gamma}(t_0+t)) - \overline{w}^m \circ \overline{\gamma}(t_0)\right] \frac{\partial}{\partial \overline{\boldsymbol{x}}^m} .$$

Taylor expansion of the terms within brackets and performing the limit process results eventually in

$$\nabla_{\overline{\gamma}}\overline{\mathbf{w}} = \left[\frac{\partial \overline{w}^i}{\partial \overline{x}^j}\dot{\overline{\gamma}}^j + \overline{w}^k \Gamma^i_{kq}\dot{\overline{\gamma}}^q\right]\frac{\partial}{\partial \overline{\boldsymbol{x}}^i}, \tag{6.47}$$

which is called the covariant derivative of a vector field along a curve $\overline{\gamma}$ evaluated at $\overline{\gamma}(t_0)$.

The so-called Christoffel-symbols at $\overline{\gamma}(t_0)$ are given by

$$\Gamma^m_{ik}\Big|_{\overline{\gamma}(t_0)} = \frac{\partial \alpha^m_i}{\partial \overline{x}^k}\bigg|_{\overline{\gamma}(t_0)} = \frac{\partial^2 x^j}{\partial \overline{x}^i \partial \overline{x}^k}\bigg|_{\overline{\gamma}(t_0)} \frac{\partial \overline{x}^m}{\partial x^j}\bigg|_{\gamma(t_0)}.$$

According to the point of view introduced within this section, the α^m_i in (6.46) relate, i.e. connect, tangent spaces of different points with each other. The Christoffel symbols are a linearization of this connection valid in the vicinity of a considered point. Therefore, they are also known as coefficients of a so-called affine connection. It should be noted that the Γ^m_{ik} are not the coordinates of a third order tensor. For affine chart relations, the Γ^m_{ik} vanish. Hence, the covariant derivative generalizes the directional derivative for nonlinear chart relations and contains affine chart relations as a special case.

The $\dot{\overline{\gamma}}^k$ in (6.47) are the coordinates of the tangent vectors of $\overline{\gamma}$ at $\overline{\gamma}(t_0)$. This indicates that (6.47) can be generalized as covariant derivative of a vector field $\overline{\mathbf{w}}$ with respect to another vector field $\overline{\mathbf{v}}$,

$$\nabla_{\overline{\mathbf{v}}}\overline{\mathbf{w}} = \left[\frac{\partial \overline{w}^i}{\partial \overline{x}^j}\overline{v}^j + \overline{w}^k \Gamma^i_{kq}\overline{v}^q\right]\frac{\partial}{\partial \overline{\boldsymbol{x}}^i}. \tag{6.48}$$

The following properties of the covariant derivative can be deduced from its definition (6.48):

$$\begin{aligned}
\nabla_{\Phi\mathbf{u}+\Psi\mathbf{v}}\mathbf{W} &= \Phi\,\nabla_{\mathbf{u}}\mathbf{W} + \Psi\,\nabla_{\mathbf{v}}\mathbf{W}\\
\nabla_{\mathbf{u}}(\alpha\mathbf{W} + \beta\mathbf{V}) &= \alpha\nabla_{\mathbf{u}}(\mathbf{W}) + \beta\nabla_{\mathbf{u}}(\mathbf{V})\\
\nabla_{\mathbf{u}}(\Phi\mathbf{W}) &= \nabla_{\mathbf{u}}(\Phi)\mathbf{W} + \Phi\nabla_{\mathbf{u}}(\mathbf{W}),
\end{aligned}$$

where Φ and Ψ are scalar fields and α and β are real numbers.

Remark 6.1 Inspired by Gründeman [2], the derivation of the covariant derivative within this section differs from expositions often found in the context of curvilinear coordinates (see, e.g., Itskov [3], Fleisch [1] or Neuenschwander [6]). This is motivated by the objective of showing how this subject is related to Chap. 7. First of all, computation of the connection coefficients within this section was only possible due to the information that the space in question is euclidean. Furthermore, the information about parallel transport defines the connection coefficients Γ^i_{kq}, and vice versa. On the other hand, a covariant derivative defines the parallel transport, etc., which already indicates at this point the equivalence of these concepts. Furthermore, it is important to note that normed versions of the $\mathbf{G}_i$ are used most often in the literature, which has to be taken into account when comparing results.

6.8 Differential Forms in $\mathbb{R}^N$ and Integration

According to Sect. 4.5, the gradient of a function f, $\underset{\sim}{\boldsymbol{d}} f$, defined on $\mathbb{R}^N$, is a 1-form. In order to take account of the specific nature of $\underset{\sim}{\boldsymbol{d}} f$, it is called a differential 1-form. Given two differential 1-forms, $\underset{\sim}{\boldsymbol{d}} f$ and $\underset{\sim}{\boldsymbol{d}} g$, a differential 2-form

$$\underset{\approx}{\boldsymbol{\omega}} = \underset{\sim}{\boldsymbol{d}} f \wedge \underset{\sim}{\boldsymbol{d}} g$$

can be constructed by means of the exterior product. However, more general objects are possible, and the set of all possible $\underset{\sim}{\boldsymbol{\omega}} = \omega_i \, \underset{\sim}{\boldsymbol{d}} x^i$ forms a space Ω^1, whereas the $\underset{\approx}{\boldsymbol{\omega}} = \omega_{ij} \, \underset{\sim}{\boldsymbol{d}} x^i \wedge \underset{\sim}{\boldsymbol{d}} x^j$ form a space Ω^2, etc. The distinction between Λ^k introduced in Sect. 4.5 and Ω^k is that Ω^k is related to the tangent space at a point $T_p\mathbb{R}^N$.

It turns out that differential forms are the proper objects for integration, since integration defined by means of forms is applicable in more general contexts. Therefore, in what follows, a reformulation of integration based on differential forms is discussed. We start with an extremely simple case in $\mathbb{R}^2$.

Given a two form

$$\underset{\approx}{\boldsymbol{\omega}} = f(x^1, x^2) \, \underset{\sim}{\boldsymbol{d}} x^1 \wedge \underset{\sim}{\boldsymbol{d}} x^2 \tag{6.49}$$

at every point p which belongs to a rectangular domain $[0, 1] \times [0, 1]$ in $\mathbb{R}^2$. The function $f(x^1, x^2)$ is supposed to be continuous. Performing at some point $p = (p^1, p^2)$

$$\underset{\approx}{\boldsymbol{\omega}} \left(\Delta x^1 \frac{\partial}{\partial \boldsymbol{x}^1}, \Delta x^2 \frac{\partial}{\partial \boldsymbol{x}^2} \right) = f(p^1, p^2) \Delta x^1 \Delta x^2$$

with $\Delta x^1, \Delta x^2 \in \mathbb{R}$ not only indicates how integration based on differential forms should be designed, it also reveals the relation with traditional integration in $\mathbb{R}^N$.

Technically, we define a grid of points $p_{[m,n]}$ with coordinates

$$p_{[m,n]} = ((m-1)\Delta x^1, (n-1)\Delta x^2) \, .$$

For simplicity, we use

$$\Delta x^1 = \Delta x^2 = \frac{1}{N} \, ,$$

and define

$$\int\limits_{[0,1]\times[0,1]} \underset{\approx}{\omega} = \lim_{N\to\infty} \sum_{m=1}^{N+1} \sum_{n=1}^{N+1} \underset{\approx}{\omega}_{[m,n]} \left(\Delta x^1 \frac{\partial}{\partial \boldsymbol{x}^1}, \Delta x^2 \frac{\partial}{\partial \boldsymbol{x}^2} \right)$$
$$= \lim_{N\to\infty} \sum_{m=1}^{N+1} \sum_{n=1}^{N+1} f(p_{[m,n]}) \Delta x^1 \Delta x^2 ,$$

which ensures that regarding $\mathbb{R}^N$ integration by means of forms coincides with standard integration, e.g.,

$$\int\limits_D \underset{\approx}{\omega} = \int\limits_D f(x^1, x^2) \mathrm{d}x^1 \mathrm{d}x^2 \tag{6.50}$$

for a 2-form (6.49) in $\mathbb{R}^2$.

A rather simple approach to handling more complicated domains of integration D is to replace the argument $\underset{\approx}{\omega}$ by $\mathcal{D}\underset{\approx}{\omega}$ with

$$\mathcal{D} = \begin{cases} 1 & p \in D \\ 0 & p \notin D \end{cases} ,$$

the integration $\int \mathcal{D}\underset{\approx}{\omega}$ then being performed with respect to a domain which includes D.

In order to ensure that the result of integration does not depend on the choice of a particular chart, the pull back operation already discussed for 1-forms (see (6.24)) has to be generalized for differential n-forms. We consider the setting sketched in Fig. 6.3 and a 2-form in chart $(U, \overline{\mu})$,

$$\underset{\approx}{\omega} = \overline{\omega}(\overline{x}^1, \overline{x}^2)\, \underset{\sim}{\boldsymbol{d}}\, \overline{x}^1 \wedge \underset{\sim}{\boldsymbol{d}}\, \overline{x}^2 .$$

The pull back of $\underset{\approx}{\omega}$ under g is defined by

$$g^\star \underset{\approx}{\omega}(\mathbf{u}, \mathbf{v}) = \underset{\approx}{\omega}\left(g_\star(\mathbf{u}), g_\star(\mathbf{v})\right) , \tag{6.51}$$

i.e., the pull back $g^\star \underset{\approx}{\omega}$ should give the same result when evaluated for the vectors $\mathbf{u}$ and $\mathbf{v}$ as $\underset{\approx}{\omega}$ when evaluated for the respective push forwards. The push forwards of $\mathbf{u}$ and $\mathbf{v}$ read as (see Box 6.6.1)

$$g_\star(\mathbf{u}) = \left[u^1 \beta^1_1 + u^2 \beta^1_2\right] \frac{\partial}{\partial \overline{\boldsymbol{x}}^1} + \left[u^1 \beta^1_2 + u^2 \beta^2_2\right] \frac{\partial}{\partial \overline{\boldsymbol{x}}^2}$$
$$g_\star(\mathbf{v}) = \left[v^1 \beta^1_1 + v^2 \beta^1_2\right] \frac{\partial}{\partial \overline{\boldsymbol{x}}^1} + \left[v^1 \beta^1_2 + v^2 \beta^2_2\right] \frac{\partial}{\partial \overline{\boldsymbol{x}}^2}$$

with

$$\beta_i^j = \frac{\partial \overline{x}^j}{\partial x^i} \,.$$

If $g^\star \underset{\approx}{\omega}$ is written as

$$g^\star \underset{\approx}{\omega} = \omega(x^1, x^2) \, \underset{\sim}{d}x^1 \wedge \underset{\sim}{d}x^2 \,,$$

Equation (6.51), after evaluation, reads as

$$\omega(x^1, x^2)[u^1 v^2 - u^2 v^1] = \overline{\omega}(\overline{x}^1, \overline{x}^2)[u^1 v^2 - u^2 v^1][\beta_1^1 \beta_2^2 - \beta_1^2 \beta_2^1] \,,$$

from which

$$g^\star \underset{\approx}{\omega} = \det J \; \overline{\omega} \circ g \tag{6.52}$$

follows with the determinant of the Jacobian of the mapping g

$$\det J = [\beta_1^1 \beta_2^2 - \beta_1^2 \beta_2^1] \,,$$

which indicates the importance of the sign of $\det J$. As already discussed in Sect. 4.5, the sign of $\det J$ is related to orientation. Therefore, mappings can be divided into those which preserve orientation and those which do not. The result implies

$$\int\limits_D g^\star \underset{\approx}{\omega} = \int\limits_{g(D)} \underset{\approx}{\omega} \quad \begin{cases} (+1) \\ (-1) \end{cases} \text{if } g \text{ is} \quad \begin{matrix} \text{orientation preserving} \\ \text{not orientation preserving,} \end{matrix}$$

and the scheme can be extended towards general differential k-forms.

6.9 Exterior Derivative and Stokes' Theorem in Form Language

The importance of Stokes' theorem is well known from traditional vector calculus. It relates integration over a volume with integration over its boundary surface or the integral over an area with integration along its boundary curve or, last but not least, the integral on a line with values taken at its start and end points. More generally, Stokes' theorem relates integration regarding an $k-1$-dimensional object with integration of a corresponding object of dimension k for $k = 1, 2$ or 3.

However, the objects of integration in traditional vector calculus are vector fields and not forms. In order to answer the question as to whether Stokes' theorem or something similar can also be found if integration is based on differential forms, an operation is needed to generate a $k+1$-form from a k-form. This operation is the exterior derivative.

The operation $\underset{\sim}{\boldsymbol{d}} f$ generates a 1-form from a function. Interpreting a function as a 0-form, there is a certain temptation to proceed this way. However, evaluating $\underset{\sim}{\boldsymbol{d}}(\underset{\sim}{\boldsymbol{d}} f)$ yields

$$\underset{\sim}{\boldsymbol{d}}(\underset{\sim}{\boldsymbol{d}} f) = \frac{\partial^2 f}{\partial x^i \partial x^j} \underset{\sim}{\boldsymbol{d}} x^i \otimes \underset{\sim}{\boldsymbol{d}} x^j ,$$

which is not an alternating 2-form, and therefore not suitable as an object of integration. However, the result of an operation which takes a differential form as its argument and is defined as

$$\mathrm{d} := \underset{\sim}{\boldsymbol{d}} \wedge \tag{6.53}$$

is again an alternating differential form. For instance,

$$\mathrm{d}(\underset{\sim}{\boldsymbol{d}} f) = \mathrm{d}\left(\frac{\partial f}{\partial x^i} \underset{\sim}{\boldsymbol{d}} x^i\right) = \frac{\partial^2 f}{\partial x^i \partial x^k}\, \underset{\sim}{\boldsymbol{d}} x^k \wedge \underset{\sim}{\boldsymbol{d}} x^i .$$

By means of the exterior derivative (6.53), Stokes' theorem can be stated in form language as follows:

$$\int\limits_{\partial D} \underset{\sim}{\boldsymbol{\omega}} = \int\limits_{D} \mathrm{d}\underset{\sim}{\boldsymbol{\omega}} \tag{6.54}$$

for a differential form defined on the domain D where ∂D is the boundary of D. In order to see how (6.54) works, we consider the square $[0, 1] \times [0, 1]$ in $\mathbb{R}^2$ and a differential 1-form

$$\underset{\sim}{\boldsymbol{\omega}} = \omega_1(x^1, x^2)\underset{\sim}{\boldsymbol{d}} x^1 + \omega_2(x^1, x^2)\underset{\sim}{\boldsymbol{d}} x^2 .$$

The exterior derivative of $\underset{\sim}{\boldsymbol{\omega}}$ reads as

$$\mathrm{d}\underset{\sim}{\boldsymbol{\omega}} = \frac{\partial}{\partial x^k} \underset{\sim}{\boldsymbol{d}} x^k \wedge \underset{\sim}{\boldsymbol{\omega}} = \frac{\partial \omega_1}{\partial x^2} \underset{\sim}{\boldsymbol{d}} x^2 \wedge \underset{\sim}{\boldsymbol{d}} x^1 + \frac{\partial \omega_2}{\partial x^1} \underset{\sim}{\boldsymbol{d}} x^1 \wedge \underset{\sim}{\boldsymbol{d}} x^2$$

or, respectively,

$$\mathrm{d}\underset{\sim}{\boldsymbol{\omega}} = \left[\frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2}\right] \underset{\sim}{\boldsymbol{d}} x^1 \wedge \underset{\sim}{\boldsymbol{d}} x^2 .$$

Integrating according to (6.50), taking into account

$$\int\limits_{x^2=0}^{1} \int\limits_{x^1=0}^{1} \frac{\partial}{\partial x^1} \omega_2(x^1, x^2) \mathrm{d}x^1 \mathrm{d}x^2 = \int\limits_{x^2=0}^{1} \left[\omega_2(1, x^2) - \omega_2(0, x^2)\right] \mathrm{d}x^2$$

eventually yields the integral of $\underset{\sim}{\boldsymbol{\omega}}$ along the boundary of the square according to the orientation in Fig. 6.5 (left).

Stokes' theorem in form language reveals one of the advantages of integration based on differential forms. There is one general formula no matter the dimension

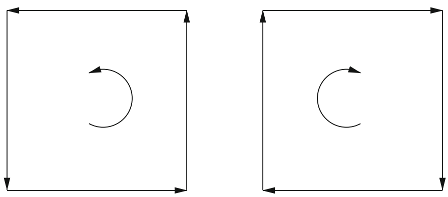

Fig. 6.5 Two possible orientations for the unit square

of the problem. General proofs of Stokes' theorem can be found, e.g., in Lee [4] and many other texts on differentiable manifolds.

The exterior derivative is linear in both arguments and has, in addition, the following properties:

$$\begin{aligned} \mathrm{d}(\underset{\sim}{\omega} \wedge \underset{\sim}{\eta}) &= \mathrm{d}\underset{\sim}{\omega} \wedge \underset{\sim}{\eta} + (-1)^{rq} \underset{\sim}{\omega} \wedge \mathrm{d}\underset{\sim}{\eta}\,, \\ \mathrm{d}(\mathrm{d}\underset{\sim}{\omega}) &= 0\,, \end{aligned}$$

for $\underset{\sim}{\omega} \in \Omega^r$ and $\underset{\sim}{\eta} \in \Omega^q$ at some point p. Furthermore, the exterior derivative commutes with the pull back, i.e.,

$$\mathrm{d}(g^* \underset{\sim}{\omega}) = g^*(\mathrm{d}\underset{\sim}{\omega})\,.$$

For a 0-form f and a 1-form $\underset{\sim}{\eta}$, it can be shown that $\mathrm{d}f = \underset{\sim}{\boldsymbol{d}} f$, $*\mathrm{d}\underset{\sim}{\eta}$, and $*\mathrm{d} * \underset{\sim}{\eta}$ correspond to the operations grad, curl and div of classical vector calculus.

Remark 6.2 The operations curl and div mentioned at the end of Sect. 6.3 are defined for vector fields. Vector fields, on the other hand, are the objects of integration in traditional vector calculus. However, if integration is performed entirely by means of differential forms, curl and div of vector fields are of no use any more.

Exercises

6.1 Compute $\underset{\sim}{\boldsymbol{d}}|_{x_0} f$ for $f = x^2$ according to (6.6).

6.2 Compute $\underset{\sim}{\boldsymbol{d}}|_{x_0=0} f$ for $f = |x|$ according to (6.6) for $h = -1$ and $h = 1$. Is $\underset{\sim}{\boldsymbol{d}}|_0 |x|$ linear wrt. h?

6.3 Show that (6.6) is equivalent to

$$\underset{\sim}{\boldsymbol{d}}|_{x_0} f = \left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \left[f(x_0 + \varepsilon h) - f(x_0) \right] \right]_{\varepsilon=0} .$$

Hint: use Taylor series expansion.

6.4 Derive a generalization of the chain rule (6.9) for $\mathbb{R}^N$.

6.5 Given two functions Φ and Ψ defined in $\mathbb{R}^2$. Show that the mapping

$$f \mapsto \lim_{t\to 0} \frac{1}{t}[f(p \dotplus t\mathbf{h}) - f(p)]$$

with $f = \Phi\Psi$ obeys the Leibniz rule by applying Taylor series expansion.

6.6 Show that a derivation D at a point $p \in \mathbb{R}^N$ is always of the form

$$D(f) = \left.\frac{\partial f}{\partial x^k}\right|_p v^k$$

for any $f \in C^\infty(\mathbb{R}^N)$.
Hint: Use Taylor series with remainder

$$f(x) = f(p) + \left.\frac{\partial f}{\partial x^i}\right|_p [x^i - p^i] + R(x)[x^i - p^i]$$

for representing f at points x close to p and apply D. Take into account that $R(p) = 0$.

6.7 For the Example 6.1, determine the push forward of the vector field

$$\mathbf{w} = x^1 \frac{\partial}{\partial \boldsymbol{x}^1} + x^2 \frac{\partial}{\partial \boldsymbol{x}^2}$$

from (U, μ) to $(U, \overline{\mu)}$ at $x^1 = x^2 = 1$.

6.8 Given $\underset{\sim}{\omega} = \omega_1(x^1, x^2)\underset{\sim}{\boldsymbol{d}}x^1 + \omega_2(x^1, x^2)\underset{\sim}{\boldsymbol{d}}x^2$ in $\mathbb{R}^2$. Evaluate $\int_{\partial D} \underset{\sim}{\omega} = \int_D \mathrm{d}\underset{\sim}{\omega}$ with D the unit square using the orientation sketched in Fig. 6.5 (right).

References

1. Fleisch D (2011) A student's guide to vectors and tensors. Cambridge University Press, Cambridge
2. Gründeman H (2015) HS Mittweida, private communication
3. Itskov M (2012) Tensor algebra and tensor analysis for engineers: with applications to continuum mechanics. Mathematical engineering. Springer, Berlin
4. Lee M (2002) Introduction to smooth manifolds. Graduate texts in mathematics. Springer, Berlin
5. Loring WT (2011) An introduction to manifolds, 2nd edn. Universitext. Springer, New York
6. Neuenschwander D (2014) Tensor calculus for physics: a concise guide. Johns Hopkins University Press, Baltimore

Chapter 7
A Primer on Smooth Manifolds

Abstract Within this chapter, a generalization of the tangent space concept, as well as the notion of a smooth atlas, is introduced in the context of analysis on curves in $\mathbb{R}^2$ and analysis on surfaces in $\mathbb{R}^3$. Implications of a complete avoidance of an embedding space, the last step in the transition to smooth manifolds, are discussed, focusing on abstraction level and topology. Furthermore, the notion of the tangent bundle is introduced in the context of vector fields defined on smooth manifolds. After introducing the Lie derivative, a guideline for studying the subject further is provided. Eventually, a selection of further literature is proposed.

7.1 Introduction

We consider a function Φ defined on the unit circle S^1 in $\mathbb{R}^2$ as illustrated in Fig. 7.1 in order to illustrate some of the key issues regarding analysis on manifolds. The unit circle is given by

$$S^1 = \left\{ (x^1, x^2) \mid \left(x^1\right)^2 + \left(x^2\right)^2 = 1 \right\} . \tag{7.1}$$

For Φ, we choose

$$\begin{aligned} \Phi : \ & S^1 \to \mathbb{R} \\ & (x^1, x^2) \in S^1 \mapsto e^{(x^1 + 2x^2)} \end{aligned}$$

as an example. Analysis eventually relies on two main concepts, namely linear approximation and performing limit processes. However, it is not clear how to derive an optimal linear approximation of Φ at some point p, because the partial derivatives defined in the ambient space $\mathbb{R}^2$ do not exist for Φ. On the other hand, the problem sketched so far is actually one-dimensional, because the coordinates x^1, x^2 cannot

U. Mühlich, *Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds*, Solid Mechanics and Its Applications 230,
DOI 10.1007/978-3-319-56264-3_7

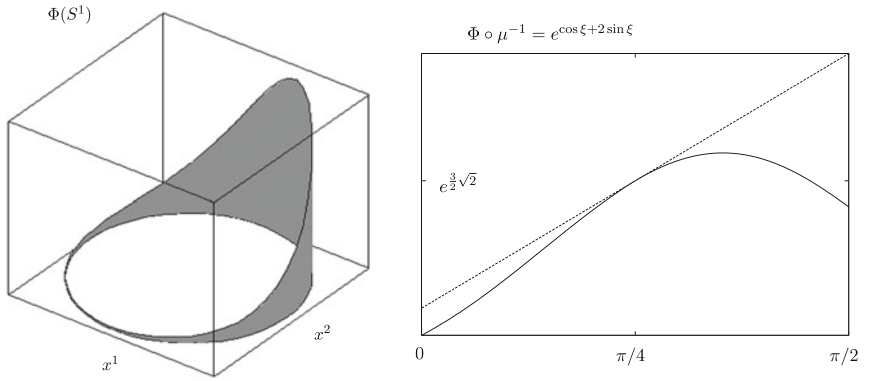

Fig. 7.1 Function defined on a curve in $\mathbb{R}^2$ (lhs) and its image $\Phi \circ \mu^{-1}$ for $\xi \in [0, \pi/2]$ in the chart (U, μ) (rhs)

be chosen independently due to the constraint imposed by (7.1). Therefore, a first step toward a solution scheme consists in using charts, which means mapping at least parts of S^1 into $\mathbb{R}$.

Alternatively to (7.1), S^1 can be represented by some parametrization. The most common is

$$\gamma : \mathbb{R} \to \mathbb{R}^2 \tag{7.2}$$

$$\xi \mapsto (\cos\xi, \sin\xi)\,, \tag{7.3}$$

but other parameterizations, e.g.,

$$\bar{\gamma} : \mathbb{R} \to \mathbb{R}^2 \tag{7.4}$$

$$\bar{\xi} \mapsto \left(\sqrt{1 - (\bar{\xi})^2},\ \bar{\xi}\right)$$

are possible.

The inverse of a parametrization is interpreted in the following as a mapping which generates a chart in $\mathbb{R}$. Now, another problem becomes apparent. Either the parametrization itself or its inverse fail to be invertible for every point of S^1. This reflects a more general issue, namely that a single chart is usually not sufficient for a general curve in $\mathbb{R}^2$ or a general surface in $\mathbb{R}^3$. In order to circumvent this problem, we focus only on that part U of S^1 for which $x^1, x^2 \geq 0$ holds. In the language of charts, we have

$$\mu = \gamma^{-1} : U \to \mathbb{R}$$

$$(x^1, x^2) \in U \mapsto \xi = \arctan\frac{x^2}{x^1}$$

and

$$\bar{\mu} = \bar{\gamma}^{-1} : U \to \mathbb{R}$$
$$(x^1, x^2) \in U \mapsto \bar{\xi} = \frac{1}{\sqrt{2}}\sqrt{1 - \left(x^1\right)^2 + x^2}\,,$$

together with $\mu^{-1} = \gamma$ and $\bar{\mu}^{-1} = \bar{\gamma}$ given by (7.2) and (7.4). The image of Φ in the chart (U, μ) reads as

$$f = \Phi \circ \mu^{-1} = e^{(\cos \xi + 2 \sin \xi)}\,,$$

whereas the image of Φ in the chart $(U, \bar{\mu})$ is given by

$$\bar{f} = \Phi \circ \bar{\mu}^{-1} = e^{\sqrt{1-(\bar{\xi})^2}+\bar{\xi}}\,.$$

Since f and $\bar{f}$ are functions of one variable, the derivatives of these functions at the point in the charts which correspond to $p = (\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2})$ can be determined, and we obtain

$$\left.\frac{\mathrm{d}}{\mathrm{d}\xi} f\right|_{\xi=\pi/4} = \frac{1}{2}\sqrt{2}e^{\frac{3}{2}\sqrt{2}}\,, \tag{7.5}$$

and

$$\left.\frac{\mathrm{d}}{\mathrm{d}\bar{\xi}} \bar{f}\right|_{\bar{\xi}=\frac{1}{2}\sqrt{2}} = e^{\frac{3}{2}\sqrt{2}}\,. \tag{7.6}$$

Since $\Phi(p) = f(\mu(p)) = \bar{f}(\bar{\mu}(p))$, one might be tempted to estimate the value of Φ at a point q, e.g. $q = (0, 1)$ based on a linear approximation of the image of Φ at p in some chart. For the chart (U, μ) the linear approximation of f reads as

$$\mathrm{d}_{\mu(p)} f(h) = \frac{1}{2}\sqrt{2}e^{\frac{3}{2}\sqrt{2}}\,h\,,$$

and setting $h = \mu(q) - \mu(p) = \frac{\pi}{4}$ yields the estimate

$$f(\mu(q)) \approx f(\mu(p)) + \frac{\pi}{8}\sqrt{2}e^{\frac{3}{2}\sqrt{2}}\,.$$

For the chart $(U, \bar{\mu})$, the estimate

$$\bar{f}(\bar{\mu}(q)) \approx \bar{f}(\bar{\mu}(p)) + e^{\frac{3}{2}\sqrt{2}}$$

is obtained, from which it can be seen that results generated this way are chart dependent, and therefore completely useless. This chart dependence is just a manifestation of a deeper problem.

Standard analysis is designed for functions defined in $\mathbb{R}^N$, i.e., functions defined in linear spaces or affine spaces. However, defining a function Φ on some general curve in $\mathbb{R}^2$ essentially means defining a function on a nonlinear object, i.e., in a nonlinear space.

The nonlinearity of the problem becomes apparent immediately by looking at the equations used to describe curves implicitly, e.g., (7.1). Therefore, simply using charts is insufficient; it has to be combined with a concept to approximate S^1 locally by means of a linear or affine space. This local approximation is achieved through the tangent space to S^1 at a point $p \in S^1$ denoted by $T_p S^1$.

Considering the point $p = (\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2})$ on S^1, taking into account that p corresponds to $\xi = \frac{\pi}{4}$ and $\bar{\xi} = \frac{1}{2}\sqrt{2}$, respectively, corresponding tangent vectors can be determined. Regarding the parametrization $\mu^{-1} = \gamma$, the tangent vector $\boldsymbol{\partial}$,

$$\boldsymbol{\partial} = \left.\frac{\mathrm{d}\cos\xi}{\mathrm{d}\xi}\right|_{\xi=\frac{\pi}{4}} \mathbf{e}_1 + \left.\frac{\mathrm{d}\sin\xi}{\mathrm{d}\xi}\right|_{\xi=\frac{\pi}{4}} \mathbf{e}_2 = -\frac{1}{2}\sqrt{2}\mathbf{e}_1 + \frac{1}{2}\sqrt{2}\mathbf{e}_2\,,$$

is obtained, whereas for $\bar{\mu}^{-1} = \bar{\gamma}$,

$$\bar{\boldsymbol{\partial}} = -\mathbf{e}_1 + \mathbf{e}_2$$

results. The vectors $\boldsymbol{\partial}$ and $\bar{\boldsymbol{\partial}}$ belong to the same one-dimensional subspace of $\mathbb{R}^2$ called the tangent space at the point p, $T_p S^1$. An evident but important observation at this point is that different parameterizations yield different elements of the tangent space. Furthermore, $T_p S^1$ induces an affine space with origin p and written as $p \boldsymbol{+} T_p S^1$.

The correct interpretation of the results is as follows. The derivative (7.5) actually means to apply $\underline{\boldsymbol{d}}$ to Φ, which leads to an estimate for Φ at the point indicated by $\underline{\boldsymbol{d}}$ in the affine space induced by $T_p S^1$, i.e.,

$$\Phi(p \boldsymbol{+} \boldsymbol{\partial}) \approx \Phi(p) + \left.\frac{\mathrm{d}}{\mathrm{d}\xi} f\right|_{\mu(p)}\,,$$

where $p \boldsymbol{+} \boldsymbol{\partial} \in p \boldsymbol{+} T_p S^1$. The tangent vector $\bar{\boldsymbol{\partial}}$, together with (7.6), generates the very same approximation. This situation is illustrated in Fig. 7.2.

A chart $\mu : \mathbb{R}^2 \to \mathbb{R}$ around a point p of a curve S induces a basis for the tangent space $T_p S$ at p. Given a function Φ defined on S. The derivative of the image of Φ in the chart evaluated at the image of p generates a linear approximation of Φ at p. This approximation is not defined on S, but rather in the affine space induced by the vector space $T_p S$.

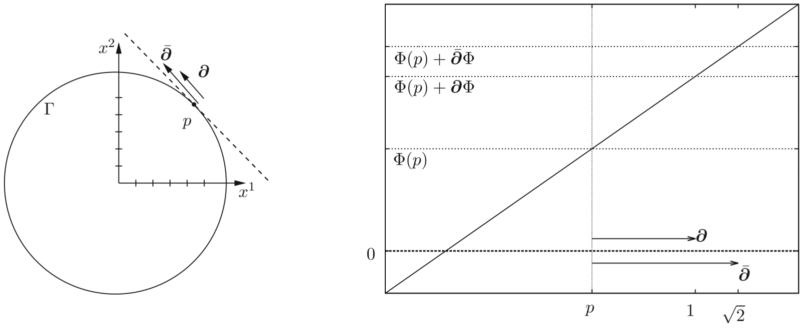

Fig. 7.2 Curve S^1 in $\mathbb{R}^2$, together with the tangent space at point p, T_pS^1 (LHS). Different base vectors of T_pS^1 induced by different charts are sketched as well. Furthermore, the linear approximation of a function Φ in the affine space induced by T_pS^1 is plotted (RHS). See also Fig. 7.1

The example already illustrates two of the key ideas of smooth manifolds, namely charts and tangent space. However, due to its simplicity, not everything can be illustrated properly by it. Therefore, we proceed with analysis on surfaces in $\mathbb{R}^3$.

7.2 Basic Concepts Regarding Analysis on Surfaces in $\mathbb{R}^3$

Within this section, functions defined on surfaces in $\mathbb{R}^3$ are considered. S is used to indicate a general surface and Φ denotes a function defined on it. The aim is to discuss decisive concepts like charts, atlases, and tangent space to a point of S in this context. Since we can still rely on the ambient space $\mathbb{R}^3$, continuity, and convergence can still be discussed, for instance, using the metric of the ambient space together with the implicit representation of S. Therefore, topological aspects do not have to be discussed yet.

The unit sphere S^2 in $\mathbb{R}^3$

$$S^2 = \left\{ (y^1, y^2, y^3) \mid \left(y^1\right)^2 + \left(y^2\right)^2 + \left(y^3\right)^2 - 1 = 0 \right\} \tag{7.7}$$

serves as specific example to illustrate certain concepts. S^2 can be represented by two-dimensional charts. A possible collection of charts (U, μ) and $(U, \bar{\mu})$ is generated by the mappings

$$\begin{aligned} &\mu : S^2 \to \mathbb{R}^2 \\ &\quad (y^1, y^2, y^3 \neq -1) \in S^2 \mapsto \left(x^1 = \frac{y^1}{1 + y^3}, x^2 = \frac{y^2}{1 + y^3} \right) \\ &\bar{\mu} : S^2 \to \mathbb{R}^2 \\ &\quad (y^1, y^2, y^3 \neq +1) \in S^2 \mapsto \left(\bar{x}^1 = \frac{y^1}{1 - y^3}, \bar{x}^2 = \frac{y^2}{1 - y^3} \right), \end{aligned}$$

which are commonly known as projections from the north pole and south pole, respectively. Such a collection is called an atlas. The interrelation between the two charts (U, μ) and $(\bar{U}, \bar{\mu})$ is given by

$$g : \mathbb{R}^2 \to \mathbb{R}^2$$
$$(x^1, x^2) \mapsto \left(\bar{x}^1 = \frac{x^1}{(x^1)^2 + (x^2)^2}, \bar{x}^2 = \frac{x^2}{(x^1)^2 + (x^2)^2}\right),$$

and

$$g^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$$
$$(\bar{x}^1, \bar{x}^2) \mapsto \left(x^1 = \frac{\bar{x}^1}{\left(\bar{x}^1\right)^2 + \left(\bar{x}^2\right)^2}, x^2 = \frac{\bar{x}^2}{\left(\bar{x}^1\right)^2 + \left(\bar{x}^2\right)^2}\right)$$

which indicates that the ambient space, in our example, $\mathbb{R}^3$, can be by-passed using an appropriate atlas.

The example reveals some important aspects of this approach from which a kind of program can be deduced to be worked through:

- In general, a single chart is not sufficient to describe a surface completely. A rather simple object like the surface of a sphere already requires at least two charts, since $y^3 = -1$ has to be excluded from (U, μ), whereas $y^3 = +1$ is not contained in $(U, \bar{\mu})$. It is not hard to imagine that, for more complicated surfaces, the minimum number of charts will be even higher.
- Since more than one chart is usually needed to cover S, most likely a switch between charts will be necessary in the course of a particular computation. Therefore, charts should fulfil certain requirements in order to allow for a smooth transition between them.
- In general, there is no unique collection of charts which cover S, i.e., there is not just one possible atlas. Therefore, it is crucial to make sure that results do not depend on the choice of a particular atlas.

The above observations give raise to the following definitions.

Definition 7.1 (*Diffeomorphism*) A k-times continuously differentiable mapping with a k-times continuously differentiable inverse is called k-diffeomorphic.

Definition 7.2 (*Atlas*) Given a set Σ. A collection of charts (U_i, μ_i), according to Definition 2.12, covering Σ entirely, i.e.,

$$\Sigma = \bigcup_{(i)} U_i \,,$$

is called an atlas (Fig. 7.3).

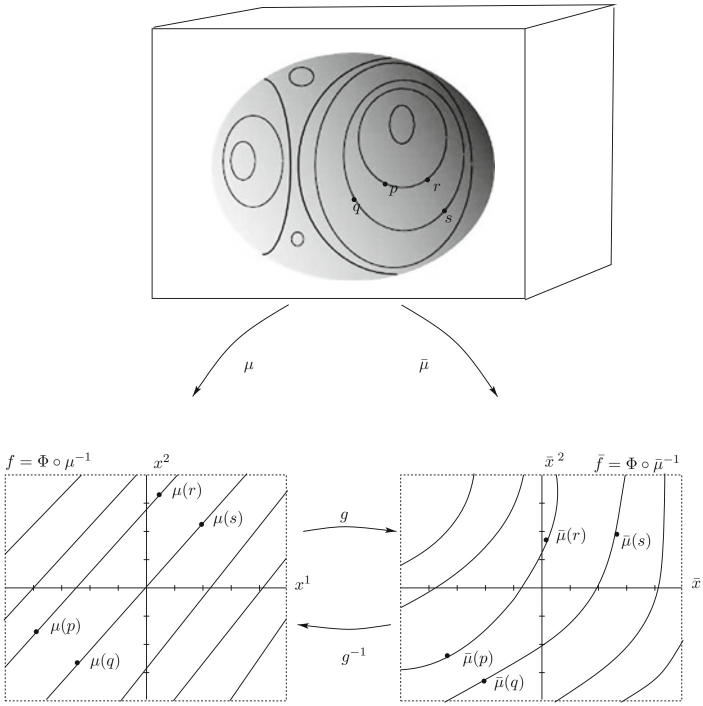

Fig. 7.3 Surface with a function defined on it and two possible charts of the neighborhood of a point p on the surface

Definition 7.3 (C^k*-Atlas*) If for any two charts (U_i, μ_i) and (U_k, μ_k) of an atlas $\mathcal{A}$, the mappings

$$\mu_k \circ \mu_i^{-1}$$

are k-diffeomorphic, then $\mathcal{A}$ is a C^k-atlas. A smooth atlas is a C^∞-atlas.

Definition 7.4 (*Set of continuously differentiable functions on a set Σ*) The set of all functions $f : \Sigma \to \mathbb{R}$ whose images in the charts have continuous partial derivatives up to order k is called $C^k(\Sigma)$.

Please note that the above definitions are not limited to sets which form surfaces in $\mathbb{R}^n$. In consideration of the topics discussed within the subsequent sections, a more general point of view is already employed here.

As already discussed within the previous section, the concept of a linearization of S at a point p on S is required in order to proceed. Tangent vectors at points of a surface can be defined as vectors tangent to curves passing through p. Given a curve Γ,

$$\begin{aligned}\Gamma : \mathbb{R} &\to S \\ \tau &\mapsto \left(y^1 = \Gamma^1(\tau),\ y^2 = \Gamma^2(\tau),\ y^3 = \Gamma^3(\tau)\right)\end{aligned}$$

with $\Gamma(\tau = t_0) = p \in S$, a vector $\mathbf{X}$ tangent to S at p is given by

$$\mathbf{v} = \dot{\Gamma}(t_0)\,,$$

as already discussed in Sect. 6.4 (see (6.34)). The vector $\mathbf{v}$ is called a geometric tangent vector and belongs to the space tangent to S at p labeled by $T_p^{(G)}S$. Since different curves correspond to the same $\mathbf{v}$, tangent vectors can be defined by means of equivalence classes of curves.

On the other hand, the directional derivative along Γ can be interpreted as the action of a vector on a function Φ defined on S in the vicinity of p. This reads analogously to Sect. 6.6,

$$\mathbf{X} : C^\infty(S) \to \mathbb{R} \tag{7.8}$$

with

$$\mathbf{X}(\Phi) = \left[\frac{\mathrm{d}}{\mathrm{d}\tau}\Phi \circ \Gamma\right]_{t_0}\,.$$

Considering the chart (U, μ), using the identity $\Phi \circ \Gamma = \Phi \circ \mu^{-1} \circ \mu \circ \Gamma$ together with the chain rule yields

$$\mathbf{X}(\Phi) = \left.\frac{\partial \Phi \circ \mu^{-1}}{\partial x^i}\right|_{\mu(p)} \left.\frac{\partial x^i}{\partial y^k}\right|_p \dot{\Gamma}^k(t_0)\,. \tag{7.9}$$

Rewriting the above equation as

$$\mathbf{X}(\Phi) = \dot{\Gamma}^k(p) \left.\frac{\partial x^i}{\partial y^k}\right|_p \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(p)} (\Phi \circ \mu^{-1})$$

reveals its push forward character, i.e., $\mathbf{X}(\Phi) = \mu_\star(\mathbf{X})(\Phi \circ \mu^{-1})$, with

$$\mu_\star(\mathbf{X}) = \dot{\Gamma}^k(p) \left.\frac{\partial x^i}{\partial y^k}\right|_p \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(p)}\,. \tag{7.10}$$

Applying $\mu_\star^{-1}$ on both sides of (7.10) yields

$$\mathbf{X} = \mu_\star^{-1}\left(\dot{\Gamma}^k(p) \left.\frac{\partial x^i}{\partial y^k}\right|_p \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(p)}\right) = \dot{\Gamma}^k(p) \left.\frac{\partial x^i}{\partial y^k}\right|_p \mu_\star^{-1}\left(\left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(p)}\right)\,.$$

Since $\mathbf{X}$ is an element of a vector space, $\mathbf{X} = X^k \boldsymbol{\partial}_k$ for some basis $\{\boldsymbol{\partial}_i\} \in T_pS$, and we can identify

$$\boldsymbol{\partial}_i = \mu_\star^{-1}\left(\left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(p)} \right) \quad , \quad X^i = \dot{\Gamma}^k(p) \left.\frac{\partial x^i}{\partial y^k}\right|_p .$$

Using another chart $(U, \bar{\mu})$, the same line of arguments yields the identification

$$\bar{\boldsymbol{\partial}}_i = \bar{\mu}_\star^{-1}\left(\left.\frac{\partial}{\partial \bar{\boldsymbol{x}}^i}\right|_{\bar{\mu}(p)} \right) \quad , \quad \bar{X}^i = \dot{\Gamma}^k(p) \left.\frac{\partial \bar{x}^i}{\partial y^k}\right|_p ,$$

i.e., different charts induce different bases in T_pS. Regarding a basis $\{\boldsymbol{\partial}_i\}$ in T_pS induced by the chart (U, μ) with coordinates x^i, the notation

$$\left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_p = \boldsymbol{\partial}_i$$

is also often found in literature. It has the advantage of making the origin of the basis explicit.

In the previous considerations, we tacitly assumed that addition and multiplication in T_pS are defined, because a representation by means of coordinates and base vectors would not otherwise be possible. However, analogously to Sect. 6.6, proper definitions have to be provided. Hence, we define

$$[\mathbf{X} \oplus \mathbf{Y}]\,(\Phi) = \mathbf{X}(\Phi) + \mathbf{Y}(\Phi) \tag{7.11}$$

$$[\lambda \odot \mathbf{X}]\,(\Phi) = \lambda\mathbf{X}(\Phi) = \mathbf{X}(\lambda\Phi) \tag{7.12}$$

for $\mathbf{X}, \mathbf{Y} \in T_pS$, $\Phi \in C^\infty(S)$ and $\lambda \in \mathbb{R}$. The definition is only appropriate if it can be shown that the results of $\mathbf{X} + \mathbf{Y}$ and $\lambda \odot \mathbf{X}$ are again elements of T_pS.

Furthermore, it can be shown that tangent vectors defined according to (7.8) obey the Leibniz rule

$$\mathbf{X}(\Phi\Psi) = \Phi(p)\mathbf{X}(\Psi) + \Psi(p)\mathbf{X}(\Phi)$$

for $\mathbf{X}, \mathbf{Y} \in T_pS$.

The methodology applied here for the purpose of proceeding toward general smooth manifolds is somehow similar to the one used in Chap. 2 regarding the step from metric spaces to topological spaces. For surfaces in an ambient space $\mathbb{R}^n$, it can be shown that T_pS is isomorphic to the geometric tangent space at a point p. Just like geometric tangent vectors, the elements of T_pS obey the Leibniz rule. The transition to smooth manifolds now consist, roughly speaking, in defining a manifold by the existence of a smooth atlas and tangent vectors at a point p as derivations, hence objects according to (7.8), (7.11), and (7.12) which obey the Leibniz rule, and no reference to some ambient space is required. This is discussed in more details in the remaining part of this chapter.

7.3 Transition to Smooth Manifolds

Within the previous sections, we have seen that the ambient space can be bypassed by means of an appropriate atlas, i.e., a suitable collection of charts. The total avoidance of an embedding, i.e., the complete absence of an ambient space, however, also implies the absence of a topology on the set of points S meant to serve as a model for physical space. At this point, it is worth noting that without being at least aware of the concept of topological space, it would be difficult to proceed further.

From now on, S refers to a general set, and not just to a set which forms a surface in $\mathbb{R}^3$. S can be endowed with a topology by means of an atlas $\mathcal{A}$, since the latter consist of charts which are by definition homeomorphisms.

Definition 7.5 (*Topological manifold*) A topological manifold $M = (S, \mathcal{A})$ is a set S together with an atlas, according to Definition 7.2.

Unfortunately, it is not possible now to define a differentiable or smooth manifold just by means of a differentiable atlas, according to Definition 7.3. The reason is that it can not be shown that the topology induced on S by the collection of charts makes S a Hausdorff-space (see, e.g., Crampin and Pirani [1]). Therefore, the Hausdorff-property must form part of the definition.

Definition 7.6 (*Smooth manifold*) A smooth manifold $M = (S, \mathcal{A})$ is a Hausdorff-space according to Definition 2.20, equipped with a C^∞ atlas defined by Definition 7.3.

In view of the last section, it is more or less obvious that the next step toward calculus on manifolds will be the definition of a tangent space at a point on a manifold M.

Definition 7.7 (*Tangent space at a point p of a smooth manifold*) A derivation at a point p on M is a linear mapping $\mathbf{X} : C^\infty(M) \rightarrow \mathbb{R}$ which obeys the Leibniz rule. The set of all derivations at p, together with

$$
\begin{aligned}
[\mathbf{X} \oplus \mathbf{Y}]\,(\Phi) &= \mathbf{X}(\Phi) + \mathbf{Y}(\Phi) \\
[\lambda \odot \mathbf{X}]\,(\Phi) &= \lambda \mathbf{X}(\Phi) = \mathbf{X}(\lambda \Phi)
\end{aligned}
$$

and $\Phi \in C^\infty(M)$, forms a vector space called $T_p M$.

Two goals are achieved by Definition 7.7. Firstly, a linearization of M at a point p is defined without reference to an ambient space. Secondly, if there is an ambient space, then the well-known case of geometric tangent vectors with all its potential for graphical illustration is recovered due to $T_p^{(G)} M \cong T_p M$. Alternative ways for constructing a tangent space at a point of M exist that are all equivalent. The interested reader is referred, e.g., to Jänich [3].

Consistently, the space of all linear mappings $\underset{\sim}{\mathbf{X}} : T_p M \rightarrow \mathbb{R}$ is called the dual tangent space $T_p^* M$ or the cotangent space at p.

Given two manifolds M and N together with a mapping $F : M \to N$, the push forward $F_\star$ of a vector $\mathbf{X} \in T_p M$ is defined analogously to Sect. 6.4 by the condition

$$\mathbf{X}(\Phi) = F_\star(\mathbf{X})(\Phi \circ F^{-1})$$

with $F_\star(\mathbf{X}) \in T_{F(p)} N$ and $\Phi \in C^\infty(M)$. The corresponding condition that defines the pull back operation,

$$\left\langle F^\star(\underset{\sim}{\boldsymbol{\Omega}}), \mathbf{X} \right\rangle = \left\langle \underset{\sim}{\boldsymbol{\Omega}}, F_\star(\mathbf{X}) \right\rangle ,$$

also shows the analogy to Sect. 6.4.

Due to the avoidance of an embedding, tangent vectors and cotangent vectors become rather abstract objects. If M is a smooth manifold according to Definition 7.6, it is certain that these objects exist. There are operations defined for these objects, e.g., summation, multiplication by a real number, push forward and pull back, but that is all. Computations have to be performed in charts which looks the same as in the previous section.

Charts are also used in the context of proofs through the performing of all necessary steps in a chart and showing afterwards that the result does not depend on the choice for a particular chart.

The concept of differentiable/smooth manifolds provides a unifying language for analysis in topological spaces with differentiable/smooth structure. Analysis in euclidean space is included as a special case for which one global chart is sufficient for performing analysis.

7.4 Tangent Bundle and Vector Fields

Regarding some euclidean space $\mathcal{E} = (X, \mathcal{V}, \dotplus)$, vector fields can be simply defined as mappings which assign to every point $p \in X$ an element of the associated global vector space $\mathcal{V}$. A basis $\{\mathbf{g}_i\}$ for $\mathcal{V}$ induces a global chart (X, μ) and continuity and differentiability of a vector field $\mathbf{V} = V^i(p)\mathbf{g}_i$ can be discussed in terms of the functions $V^i \circ \mu^{-1}$ in the global chart.

For a general smooth manifold M, this scheme is not applicable due to the absence of a global vector space. At this point, the concept of the tangent bundle enters the scene. The idea resides in defining a set consisting of the manifold M and all its tangent spaces $T_p M$, i.e.,

$$TM = \left\{ (p, \mathbf{v}) \mid p \in M, \mathbf{v} \in T_p M \right\} .$$

In order to illustrate certain aspects of this approach, the unit circle S^1 in $\mathbb{R}^2$ is again used as an example, i.e., $M = S^1$. The tangent bundle TS^1 is one of the

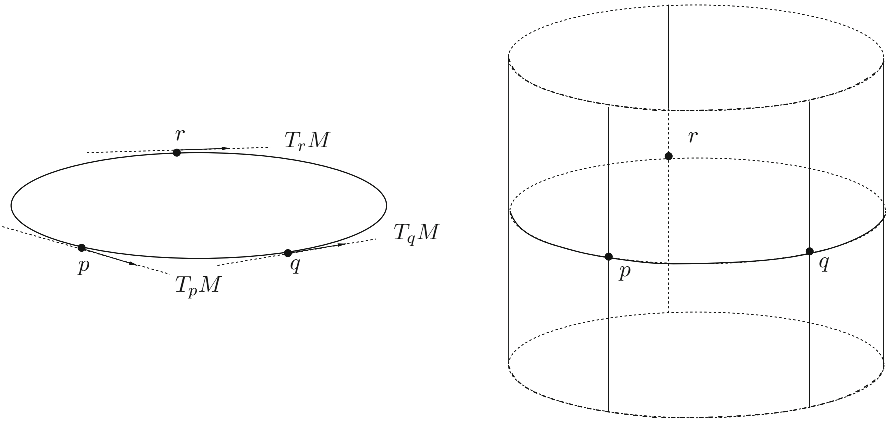

Fig. 7.4 Visualization of the tangent bundle for $M = S^1$. Since the visualization shown in the *left picture* causes spurious intersections of *lines* which indicate tangent spaces at different points, the visualization given in the *right picture* is commonly preferred. The latter allows in addition for interpretating tangent spaces as fibers attached to the manifold M (see, e.g., Schutz [10])

few cases for which a graphical illustration is possible. As shown in Fig. 7.4, TS^1 can be visualized as an infinite cylinder formed by S^1, together with a vertical fiber $\mathbb{R} \cong T_pS^1$ attached to every point of S^1. A smooth vector field on S^1 can now be interpreted as a smooth function on a cylinder, as sketched in Fig. 7.6. However, converting these ideas into a concept which can be applied safely requires some formal work.

Figure 7.4 suggests that TS^1 is a product space, i.e., $TS^1 = S^1 \times \mathbb{R}$, because there exist two natural projections, $\pi : TS^1 \to S$ and $\pi_* : TS^1 \to \mathbb{R}$, which are continuous. The inverse of π reads off the fiber at p as

$$\pi^{-1}(p) = \{p\} \times \mathbb{R}\,,$$

whereas the inverse of π_* provides a circle shifted vertically,

$$\pi_*^{-1}(z) = S^1 \times \{z\}$$

with $p \in S^1$ and $z \in \mathbb{R}$. Since S^1 and $\mathbb{R}$ are topological spaces, TS^1 can be endowed with a topology according to Sect. 2.5. Moreover, since S^1 and $\mathbb{R}$ can both be treated as smooth manifolds, TS^1 is even a product manifold, $TS^1 = S^1 \times \mathbb{R}$. But unfortunately, this does not hold in general, since, for instance, TS^2 is not a product manifold.

In order to discuss the implications of the preceding discussion in more detail, while still being able to supplement the discussion with graphical visualizations, we consider the more general concept of fiber bundles. As before, S^1 is used as the base manifold. A cylinder $\mathcal{C}$ of finite height and a Möbius band $\mathcal{M}$ as sketched in Fig. 7.5 are constructed by combining S^1 with the fiber $F = (-1, 1)$ in different

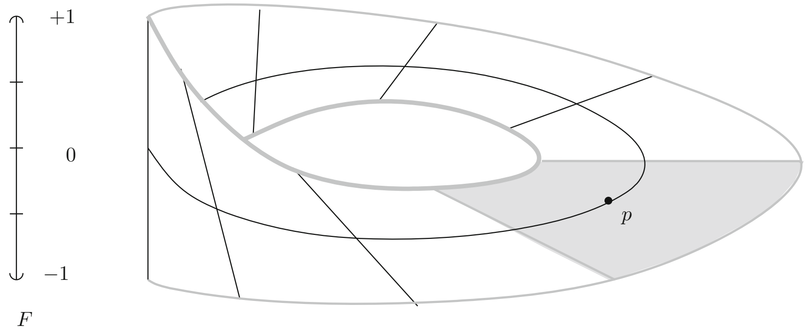

Fig. 7.5 Sketch of the Möbius band generated from S^1 and the fiber F

ways. Both constructions are fiber bundles over the same base manifold. In both cases, a continuous projection π exists and $\pi^{-1}(p)$ reads off the fiber associated with p, which is isomorphic to F. However, while for $\mathcal{C}$, the projection π_* is continuous, there is no such continuous projection for $\mathcal{M}$ except for $z = 0$, $z \in F$. For instance, starting at some $p \in S^1$, moving along S^1 and choosing $z = -\frac{1}{2}$ in every fiber eventually delivers a curve with a jump of magnitude 1 at p.

This leads to the conclusion that a common framework for $\mathcal{C}$ and $\mathcal{M}$ cannot rely on a projection π_* in a global sense. A projection π, i.e., the projection onto the base manifold, on the other hand, does not introduce any difficulties. Its inverse at some $p \in S^1$ delivers, in both cases, the fiber corresponding to p isomorphic to F. Hence, π can be used globally. Furthermore, although globally not a product manifold, locally, i.e., in the vicinity of a given point $p \in S^1$, $\mathcal{M}$ indeed possesses the properties of a product manifold as sketched in Fig. 7.5.

Summarizing the discussion so far, a fiber bundle consists at least of the following ingredients: a space E, a base manifold M, a fiber F and a projection $\pi : E \to M$ with $\pi^{-1}(p) \cong F$.

Already when introducing the smooth manifold concept, we encountered the idea of giving up certain concepts globally but retaining them locally. Smooth manifolds are spaces which cannot be treated globally as affine or euclidean. But, as long as they can be endowed at least locally with an affine/euclidean structure, a smooth manifold is, roughly speaking, obtained by gluing together small affine/euclidean pieces. A similar approach is used in the context of fiber bundles. As long as the concept of a product manifold applies locally, a fiber bundle is obtained by gluing together individual product manifolds. The last point indicates that the preliminary summary with respect to fiber bundles within the previous paragraph is not complete yet, because it does not contain any recipe as to how to glue local parts together.

At this point, we return to our original problem, the tangent bundle TM. The tangent bundle is a rather special case of a fiber bundle in the sense that the gluing procedure is already implied by the following definition.

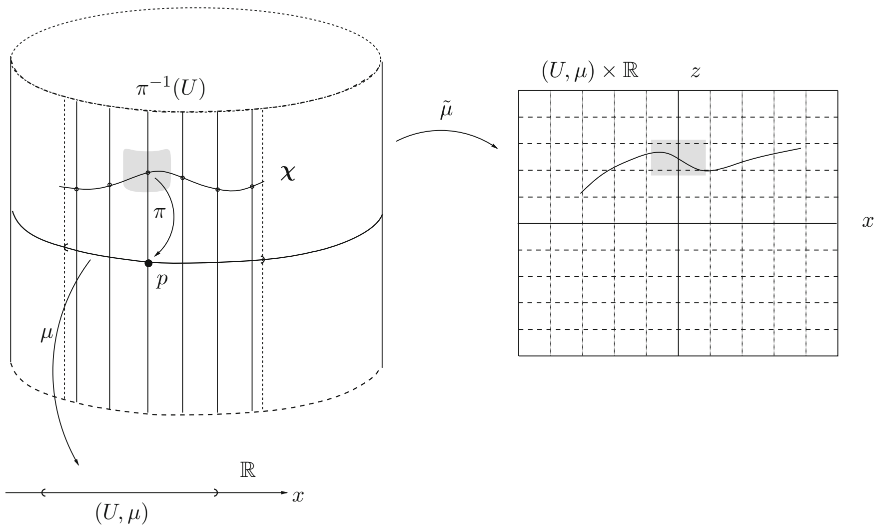

Fig. 7.6 Illustration of a vector field χ using the idea of tangent bundle together with a local chart for TM

Definition 7.8 (*Tangent bundle TM*) Given a smooth manifold M of dimension N. The set TM

$$TM = \left\{(p, \mathbf{v}) \mid p \in M, \mathbf{v} \in T_pM\right\}$$

together with the projection

$$\begin{aligned} \pi : TM &\rightarrow M \\ (p, \mathbf{v}) &\mapsto p\,, \end{aligned}$$

where $\pi^{-1}(p) = \{p\} \times T_pM \cong \mathrm{R}^N$, is called the tangent bundle of M.

An atlas for TM can be constructed from the atlas of M together with π. Let $U \subset M$ and (U, μ) be a corresponding chart with coordinates x^i. The mapping

$$\tilde{\mu} : \pi^{-1}(U) \rightarrow (U, \mu) \times \mathbb{R}^N$$

generates a chart for $\pi^{-1}(U) \subset TM$, as sketched for TS^1 in Fig. 7.6. For a point p and a vector $\mathbf{V} \in T_pM$, $\tilde{\mu}$ reads as

$$\tilde{\mu}(\mathbf{V}) = \tilde{\mu}(x^1(p), \ldots, x^N(p), v^1, \ldots, v^N)\,, \tag{7.13}$$

where the $x^i(p)$ are the coordinates of p in chart (U, μ). The v^i, on the other hand, are the coordinates of the vector $\mathbf{v}$ which represents $\mathbf{V}$ in chart (U, μ), i.e.,

$$\mathbf{v} = v^i \left.\frac{\partial}{\partial \boldsymbol{x}^i}\right|_{\mu(p)} .$$

Since M is supposed to be a smooth manifold, the atlas for M is a collection of charts (U_k, μ_k) with smooth transition functions (see Definition 7.3). Consequently, an atlas for TM can be defined as the collection of charts $(\pi^{-1}(U_k), \tilde{\mu}_k)$ which naturally raises the question regarding the transition between these charts. As (7.13) shows, this transition involves coordinates of points and vectors. With respect to the former, a smooth transition is ensured by the atlas of M. More interesting is the question regarding the v^i in (7.13) because it eventually yields the gluing procedure.

Unfortunately, discussing the effect of a chart transition on the v^i in (7.13) can make notation quite cumbersome. We consider two subsets U_1 and U_2 of M with overlap $U_1 \bigcap U_2$ and a vector $\mathbf{V} \in T_p M$ with $p \in U_1 \bigcap U_2$. In addition, there are two charts (U_1, μ_1) and (U_2, μ_2). The vector which represents $\mathbf{V}$ in chart (U_1, μ_1) reads as

$$\mathbf{v}_1 = v_1^i \left.\frac{\partial}{\partial \boldsymbol{x}_1^i}\right|_{\mu_1(p)} ,$$

whereas for chart (U_2, μ_2), $\mathbf{V}$ is represented by

$$\mathbf{v}_2 = v_2^i \left.\frac{\partial}{\partial \boldsymbol{x}_2^i}\right|_{\mu_2(p)} .$$

A change from chart (U_1, μ_1) to (U_2, μ_2) at p requires a push forward of $\mathbf{v}_1$ under the transition map $\mu_2 \circ \mu_1^{-1}$ which gives

$$v_2^i = \left.\frac{\partial x_2^i}{\partial x_1^k}\right|_{\mu_2(p)} v_1^k = (t_{12})^i_k \, v_1^k$$

(see as well Box 6.6.1). The $(t_{12})^i_k$ are the entries of an invertible quadratic matrix $\underset{=}{t}_{12}$ of dimension N. Such matrices belong to a specific Lie group, namely the general linear group $GL(N, \mathbb{R})$.

This scheme discussed above extends in a straightforward manner to a general collection of charts $\{(U_n, \mu_n)\}$, where transitions with regard to fibers are given by matrices $\underset{=}{t}_{mn}$ with the following properties:

$$\begin{aligned} \underset{=}{t}_{mm} &= \underset{=}{I} \\ \underset{=}{t}_{mn} &= \underset{=}{t}_{mn}^{-1} \end{aligned}$$

with the identity matrix $\underset{=}{I}$.

Similar to the situation in euclidean space, the aim is to define vector fields in the sense that there is an assignment of a vector to every point $p \in M$. The existence of vectors is ensured by the tangent space concept. But, since the individual tangent spaces are actually unrelated, it does not make sense to assign to a point p a vector which does not belong to the tangent space defined at p.

Definition 7.9 (*Vector field on a manifold*) A vector field χ on a manifold is a mapping

$$\begin{aligned} \chi : M &\to TM \\ p &\mapsto (p, \mathbf{v})\,, \end{aligned}$$

which fulfils the condition $\pi \circ \chi = \mathrm{Id}$.

For a particular point $p \in M$, the condition imposed on vector fields in Definition 7.9 reads as $\pi(\chi(p)) = p$. This ensures that a vector field indeed assigns to every point $p \in M$ not just some vector, but a member of the corresponding tangent space $T_p M$, as demanded earlier.

The definition of vector fields implies that a vector field $\mathbf{V}$ is smooth if the coordinates of $\mathbf{V}$ are smooth functions in the respective smooth charts (see, e.g., in Lee [4]).

7.5 Flow of Vector Fields and the Lie Derivative

Due to the avoidance of an ambient space in the definition of smooth manifolds and their local linearization, tangent spaces at different points are completely unrelated at this stage. Without relations between tangent spaces, however, analysis on a smooth manifold cannot be performed.

Objects available to work with on a smooth manifold M so far are functions and curves defined on M. On the other hand, curves can be interpreted as integral curves induced by a vector field, and vice versa. Exploring this idea further leads to the notion of the flow of a vector field and the Lie derivative.

The key ideas are explained in the following by means of an elementary example in $\mathbb{R}^2$. Given a vector field

$$\mathbf{v} = v^1\,\mathbf{e}_1 + v^2\,\mathbf{e}_2 \tag{7.14}$$

with $v^1 = a$, $v^2 = b$, where a and b are just real numbers. An integral curve Γ, $\tau \to (x^1(\tau), x^2(\tau))$ of this vector field fulfils the condition

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}\tau} = \dot{\Gamma} = \mathbf{v}|_{\Gamma(\tau)}\,,$$

or, in coordinate representation,

$$\begin{aligned} \dot{\Gamma}^1 &= v^1\big|_{\Gamma(\tau)} \\ \dot{\Gamma}^2 &= v^2\big|_{\Gamma(\tau)} \end{aligned}$$

for arbitrary values of τ within a given interval. This is a system of ordinary differential equations which can be solved by direct integration, the solution containing N integration constants. N coincides with the dimension of the space which, on the other hand, equals the number of equations. Focusing on a point p using $p = \Gamma(\tau = 0)$ as the initial condition to determine the integration constants, yields a particular curve referring to p. For the example (7.5), this curve reads as

$$\Gamma_p = (a\tau + p^1, b\tau + p^2)\,,$$

and a common interpretation of Γ_p as the trajectory of a material point which at $\tau = 0$ has position p is rather instructive for what follows.

Instead of focusing only on one point, we could also consider a set U of points occupying a connected region in space, or even the whole space. Regarding $\mathbb{R}^2$, the points belonging to U can be identified by their coordinates, x^1, x^2. In this case, $\chi(\tau, x^1, x^2)$ is called the flow of the vector field defined on U. In continuum mechanics, χ is usually called the motion. For the example (7.14), χ is simply

$$\chi(\tau, x^1, x^2) = (a\tau + x^1, b\tau + x^2)\,.$$

Evaluating χ for a particular $\tau = t$ is then a mapping $\chi_t : \mathbb{R}^2 \to \mathbb{R}^2$. Here, we consider only a small neighborhood $U_{(p)}$ of a point p, and therefore $\chi_t : U_{(p)} \to U_{(q)}$, where $U_{(p)}, U_{(q)} \in \mathbb{R}^2$. This is illustrated in Fig. 7.7 for a motion a bit more complex than the one in the example. In the language of continuum mechanics, χ_t is called a configuration and χ_0 usually serves as a reference configuration for the so-called Lagrangian description.

For a particular point p, $\chi_t(p) = \Gamma_p(t) = q$ and $\chi_{-t}(q) = \Gamma_q(-t) = p$. On the other hand, $p = \Gamma_p^{-1}(t)$ and therefore $\Gamma_p^{-1}(t) = \Gamma_q(-t)$, and for a set of points

$$\chi_t^{-1}(U_{(q)}) = \chi_{-t}(U_{(q)})\,.$$

Analogously, for the respective linearized mappings,

$$(\chi_{-t})_\star(U_{(q)}) = (\chi_t^{-1})_\star(U_{(q)}) \tag{7.15}$$

holds.

The mapping χ_t can now be used to relate the tangent spaces at $p = \Gamma_p(0)$ and $q = \Gamma_p(t)$ by means of the push forwards induced by χ_t. This, on the other hand, allows for defining a directional derivative of a vector field $\mathbf{w}$ along the integral curves of another vector field $\mathbf{v}$,

$$\mathcal{L}_\mathbf{v}\mathbf{w}|_p = \lim_{t\to 0}\frac{1}{t}\left[(\chi_t^{-1})_\star(\mathbf{w}|_{\Gamma_p(t)}) - \mathbf{w}|_{\Gamma_p(0)}\right]\,, \tag{7.16}$$

known as the Lie derivative. Using the standard basis, we write the push forward $(\chi_s^{-1})_\star(\mathbf{w}|_{\Gamma_p(t)})$ as

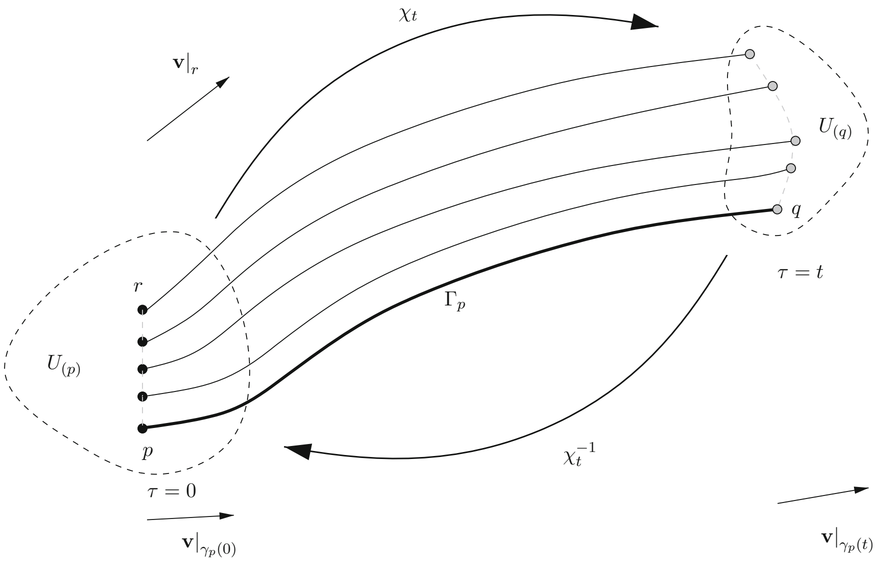

Fig. 7.7 Illustration of the flow of a vector field

$$(\chi_t^{-1})_\star(\mathbf{w}|_{\Gamma_p(t)}) = \alpha_i^k w^i(\Gamma(t))\,\mathbf{e}_k\,,$$

where α_i^k are the coordinates of the push forward and $w^k(\Gamma(t))$ takes into account that the coordinates of $\mathbf{w}$ change along Γ_p. Taylor series expansion of $w^k(\Gamma(t))$ yields

$$w^k(\Gamma(t)) \approx w^k(p) + \left.\frac{\partial w^k}{\partial x^m}\right|_p \dot{\Gamma}^m(0)\,t = w^k(p) + \left.\frac{\partial w^k}{\partial x^m}\right|_p v^m\,t\,. \tag{7.17}$$

For small t, $\Gamma_p(t)$ can be approximated by

$$\Gamma_p(t) \approx p + t\;\mathbf{v}|_p\;,$$

but regarding χ_t, it has to be taken into account that $\mathbf{v}$ changes in the vicinity of p. Therefore, χ_t is approximated for small t by

$$\chi_t = \xi + t\;\mathbf{v}|_\xi\;,$$

where $\xi \in U_{(p)}$. The directional derivative

$$\underset{\sim}{\boldsymbol{d}}\Big|_p\,\chi_t(\mathbf{h}) = \lim_{s\to 0}\frac{1}{s}\,[\xi \boldsymbol{+} s\mathbf{h} + t\mathbf{v}(\xi \boldsymbol{+} s\mathbf{h}) - \xi - t\mathbf{v}(\xi)] = \left[\delta_i^k + t\,\left.\frac{\partial v^k}{\partial x^i}\right|_p\right] h^i\,\mathbf{e}_k$$

yields the coordinates of the push forward of χ_t

$$\bar{\alpha}^k_i = \delta^k_i + t \left. \frac{\partial v^k}{\partial x^i} \right|_p ,$$

which can also be obtained just by differentiating the coordinate representation of χ_t

$$\begin{aligned} \chi^1_t &= x^1 + t v^1(x^1, x^2) \\ \chi^2_t &= x^2 + t v^2(x^1, x^2) \end{aligned}$$

accordingly. Applying (7.15), the coordinates of $(\chi_{-t})_\star$ read as

$$\alpha^k_i = \delta^k_i - t \left. \frac{\partial v^k}{\partial x^i} \right|_p \tag{7.18}$$

Using (7.17) and (7.18) in (7.16) eventually yields

$$\mathcal{L}_{\mathbf{v}} \mathbf{w}|_p = \left[\left. \frac{\partial w^k}{\partial x^m} \right|_p v^m(p) - \left. \frac{\partial v^k}{\partial x^m} \right|_p w^m(p) \right] \mathbf{e}_k = [\mathbf{v}, \mathbf{w}]_p \,, \tag{7.19}$$

where $[\mathbf{v}, \mathbf{w}]_p$ is the so-called Lie bracket evaluated at p.

The computation above can be interpreted as a computation performed in a chart (U, μ). Afterwards, it has to be shown that the result is chart independent. For details, see, e.g., Lee [4].

The Lie bracket is bilinear, and has, in addition, the following properties:

$$\begin{aligned} [\mathbf{V}, \mathbf{W}] &= -[\mathbf{W}, \mathbf{V}] \\ [\mathbf{V}, [\mathbf{W}, \mathbf{Z}]] + [\mathbf{W}, [\mathbf{Z}, \mathbf{V}]] + [\mathbf{Z}, [\mathbf{V}, \mathbf{W}]] &= 0 \,, \end{aligned}$$

which imply the following properties of the Lie derivative:

$$\begin{aligned} \mathcal{L}_{\mathbf{V}} \mathbf{W} &= -\mathcal{L}_{\mathbf{W}} \mathbf{V} \\ \mathcal{L}_{\mathbf{V}} [\mathbf{W}, \mathbf{Z}] &= [\mathcal{L}_{\mathbf{W}} \mathbf{V}, \mathbf{Z}] + [\mathbf{W}, \mathcal{L}_{\mathbf{V}} \mathbf{Z}] \\ \mathcal{L}_{f\mathbf{V}} \mathbf{W} &= (\mathbf{V} f) \mathbf{W} + f \mathcal{L}_{\mathbf{V}} \mathbf{W} \\ F_\star(\mathcal{L}_{\mathbf{V}} \mathbf{W}) &= \mathcal{L}_{F_\star(\mathbf{V})} F_\star(\mathbf{W}) \,, \end{aligned}$$

where $(\mathbf{V} f)$ is the directional derivative of the function f with respect to $\mathbf{V}$ and $F_\star$ is the push forward of a mapping $F : M \to N$ provided that F is diffeomorphic. It should be noted that the Lie derivative extends to general tensor fields (again see, e.g., Lee [4]).

For a vector field $\mathbf{W}$ which depends on time τ, the nonautonomous Lie derivative

$$\hat{\mathcal{L}}_{\mathbf{V}}\mathbf{W} = \frac{\partial \mathbf{W}}{\partial \tau} + \mathcal{L}_{\mathbf{V}}\mathbf{W}$$

measures the change of $\mathbf{W}$ over time with respect to the flow induced by $\mathbf{V}$. It is widely used, e.g., in continuum mechanics in order to measure the change of a quantity over time in case the support of this quantity changes in time as well, e.g, stress in a moving and deforming body.

7.6 Outlook and Further Reading

The objective of this section is to provide some guidance to the reader regarding the further exploration of the smooth manifold concept. We intend to give a short overview in terms of topics the reader will most likely be confronted with in this endeavor and to recommend appropriate literature. Regarding the latter, a large number of excellent textbooks exist, many of them in the context of general relativity or theoretical physics in a broader sense. The selection made here is surely biased by the accessibility of specific textbooks in libraries and bookshops. It is also a matter of taste, and therefore, rather subjective. In addition, literature is selected with a mind toward an audience interested in engineering mechanics, respectively, continuum mechanics.

As discussed earlier, tangent spaces on a manifold M are unrelated as long as no additional structure is imposed. Given a vector field on M, tangent spaces can be related by the Lie derivative. Another option consists in defining a connection (see also Sect. 6.5, Remark 6.1). One of the most popular connections is surely the Riemann connection for manifolds with a given metric tensor field. Contemporary expositions regarding vector fields on manifolds, connections, etc., rely on the concept of vector bundles. As already indicated at the end of Sect. 7.4, a deeper understanding of the latter requires knowledge of the theory of Lie groups.

Balance equations in continuum mechanics rely on integration. Integration on manifolds is defined by means of differential forms. Here again, additional structure has to be provided in terms of a volume form or a density defined on the manifold. Manifolds with given volume form are called symplectic. A general treatment of integration has to take into account manifolds with boundaries, as well as the orientation of manifolds. Furthermore, a concept called partition of unity is required in order to deal with overlaps of different charts due to the transition between charts in the course of the integration process.

The program is surely challenging, but it is feasible, given dedication and, of course, the appropriate literature. The book by Epstein [2] relates concepts of differential geometry directly to problems of continuum mechanics. This makes it especially suitable for those who prefer such a direct link from the very beginning. We recommend, in particular, the chapters about integration and connections. A direct relation between the manifold concept and continuum mechanics is also established in Marsden and Hughes [6]. However, we recommend this book only for a later stage of the learning process.

For engineers and engineering students familiar with the topics we have laid out here, Lee [4] is surely accessible. The author provides a comprehensive treatment of the subject, and therefore the book is also useful as a compendium. We also recommend Crampin and Pirani [1] which is a comprehensive exposition as well with a title that is taken rather seriously by the authors. Other texts we find especially suitable for beginners are McInerney [7] and Loring [5].

Regarding books from the physics community, we especially recommend Schutz [10], Warner [12] and O'Neill [9], as well as, at a more advanced level Nakahara [8]. Spivak [11] is very popular is as well.

Differences between the mathematical perspective and the point of view taken by physicists are discussed in Jänich [3], and the reader will most likely find a number of comments in this book especially enlightening. Furthermore, the author contrasts conventional vector analysis with calculus on manifolds.

References

1. Crampin M, Pirani F (1987) Applicable differential geometry, vol 58. London mathematical society lecture note series. Cambridge University Press, Cambridge
2. Epstein M (2010) The geometrical language of continuum mechanics. Cambridge University Press, Cambridge
3. Jänich K (2001) Vector analysis. Undergraduate texts in mathematics. Springer, New York
4. Lee M (2002) Introduction to smooth manifolds. Graduate texts in mathematics. Springer, Berlin
5. Loring WT (2011) An introduction to manifolds, 2nd edn. Universitext. Springer, New York
6. Marsden E, Hughes J (1993) Mathematical foundations of elasticity. Dover Publication, New York
7. McInerney A (2013) First steps in differential geometry. Undergraduate texts in mathematics. Springer, Berlin
8. Nakahara M (2003) Geometry, topology and physics, 2nd edn. IOP Publishing, Bristol
9. O'Neill B (2006) Elementary differential geometry. Academic Press, Dublin
10. Schutz B (1999) Geometrical methods of mathematical physics. Cambridge University Press, Cambridge
11. Spivak M (1999) A comprehensive introduction to differential geometry, vol 1, 3rd edn. Publish or Perish, Inc, Houston
12. Warner FW (1983) Foundations of differentiable manifolds and Lie groups, vol 94. Graduate texts in mathematics. Springer, New York

Solutions for Selected Problems

In the following, parts of the solutions for selected exercises are given. The full solution is only given for a limited number of exercises.

2.3 Taking into account that $A \times \emptyset = \emptyset$, hence $\emptyset \times \emptyset = \emptyset$, $\mathcal{B}$ is given by

$$\begin{aligned}\mathcal{B} = \{&X, \emptyset, X_1 \times \{a\}, X_1 \times \{b\}, \{0\} \times X_2, \{1\} \times X_2,\\ &\{0\} \times \{a\}, \{0\} \times \{b\}, \{1\} \times \{a\}, \{1\} \times \{b\}\}\end{aligned}$$

or,

$$\begin{aligned}\mathcal{B} = \{&X, \emptyset, \{(0,a),(1,a)\}, \{(0,b),(1,b)\}, \{(0,a),(0,b)\}, \{(1,a),(1,b)\},\\ &\{(0,a)\}, \{(0,b)\}, \{(1,a)\}, \{(1,b)\}\},\end{aligned}$$

and intersections of elements of $\mathcal{B}$ do not generate new sets. Therefore, the topology for X consists of all possible unions of elements of $\mathcal{B}$.

2.4 The result for τ reads as $\tau = \{X_1 \times X_2, \emptyset, \{(0,a),(0,b)\}, \{(1,a),(1,b)\}\}$. The projection π_2 and its inverse are given explicitly by

$$\begin{aligned}\pi_2 : X_1 \times X_2 &\to X_2 & \pi_2^{-1} : X_2 &\to X_1 \times X_2\\ (0,a) &\mapsto a & a &\mapsto \{(0,a),(1,a)\}\\ (0,b) &\mapsto b & b &\mapsto \{(0,b),(1,b)\},\\ (1,a) &\mapsto a & &\\ (1,b) &\mapsto b & &\end{aligned}$$

and since $\pi_2^{-1}(\emptyset) = \emptyset \in \tau$ and $\pi_2^{-1}(X_2) = X_1 \times X_2 \in \tau$, π_2 is continuous.

2.5 The result for τ reads $\tau = \{X_1 \times X_2, \emptyset\}$.

U. Mühlich, *Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds*, Solid Mechanics and Its Applications 230,
DOI 10.1007/978-3-319-56264-3

3.4 Isomorphism induced by $\mathcal{G}$
(a) $\mathcal{G}$ is linear.
For two vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$, we explore the relation between $\mathcal{G}$ and the inner product

$$\left[\hat{\mathbf{u}} + \hat{\mathbf{v}}\right] \cdot \mathbf{v} = \hat{\mathbf{u}} \cdot \mathbf{v} + \hat{\mathbf{v}} \cdot \mathbf{v}$$
$$\left\langle \mathcal{G}(\hat{\mathbf{u}} + \hat{\mathbf{v}}), \mathbf{v} \right\rangle = \left\langle \mathcal{G}(\hat{\mathbf{u}}), \mathbf{v} \right\rangle + \left\langle \mathcal{G}(\hat{\mathbf{u}}), \mathbf{v} \right\rangle$$

and use the definition for addition of dual vectors

$$\left\langle \mathcal{G}(\hat{\mathbf{u}}), \mathbf{v} \right\rangle + \left\langle \mathcal{G}(\hat{\mathbf{v}}), \mathbf{v} \right\rangle = \left\langle \mathcal{G}(\hat{\mathbf{u}}) + \mathcal{G}(\hat{\mathbf{v}}), \mathbf{v} \right\rangle ,$$

hence

$$\left\langle \mathcal{G}(\hat{\mathbf{u}} + \hat{\mathbf{v}}), \mathbf{v} \right\rangle = \left\langle \mathcal{G}(\hat{\mathbf{u}}) + \mathcal{G}(\hat{\mathbf{v}}), \mathbf{v} \right\rangle ,$$

which shows that $\mathcal{G}$ is linear.
(b) $\mathcal{G}$ is injective.
The kernel of $\mathcal{G}$ consists of all vectors $\hat{\mathbf{a}}$ for which $\mathcal{G}(\hat{\mathbf{a}}) = \underset{\sim}{\mathbf{0}}$. In this case,

$$\left\langle \mathcal{G}(\hat{\mathbf{a}}), \mathbf{v} \right\rangle = \left\langle \underset{\sim}{\mathbf{0}}, \mathbf{v} \right\rangle = \hat{\mathbf{a}} \cdot \mathbf{v} = 0 ,$$

which implies $\hat{\mathbf{a}} = \mathbf{0}$, and therefore $\ker \mathcal{G} = \{\mathbf{0}\}$, which means $\dim(\ker \mathcal{G}) = 0$.
(c) $\mathcal{G}$ is bijective.
Since $\dim(\mathcal{V}) = \dim(\mathcal{V}^*)$, $\mathcal{G}$ is bijective, hence $\mathcal{G}$ is an isomorphism.

3.6 $\delta^i_i = 3$, $\delta^i_i \delta^j_j = 9$, $\delta^j_i \delta^i_j = 3$

3.7 $A^k_j \delta^j_k = A^k_k$, $\delta^j_i \delta^k_j = \delta^k_i$, and $\delta^j_i A_j - \delta^k_i A_k = 0$.

3.8 Because, e.g., $(\underset{\sim}{\mathbf{g}}^1, \mathbf{g}_2) = (\underset{\sim}{\mathbf{g}}^1, \mathbf{g}_1) + (\underset{\sim}{\mathbf{g}}^2, \mathbf{g}_2) - (\underset{\sim}{\mathbf{g}}^2, \mathbf{g}_1)$, $\dim(\mathcal{W}) = 6$.

4.8 $\bar{A}^{12} = a^1 b^2 - a^2 b^1$, $\bar{A}^{13} = a^1 b^3 - a^3 b^1$, $\bar{A}^{23} = a^2 b^3 - a^3 b^2$ and e.g., $\bar{A}^{12}$ is the projection of $\mathbf{a} \wedge \mathbf{b}$ onto the plane defined by $\mathbf{e}_1 \wedge \mathbf{e}_2$.

6.5 Taking into account $D(f(p)) = 0$, $R(p) = 0$, and, in addition, that $R(x)$ $[x^i - p^i]$, as well as $\left.\frac{\partial f}{\partial x^i}\right|_p [x^i - p^i]$, are products of two functions,

$$D(f) = \left.\frac{\partial f}{\partial x^i}\right|_p D(x^i)$$

results, which shows that derivations are always of the form (6.41).

6.7 $\bar{w} = \frac{\sqrt{2}}{2} \bar{x}^1 \left[\sin \bar{x}^2 + \cos \bar{x}^2\right] \frac{\partial}{\partial \bar{x}^1} + \frac{1}{2} \bar{x}^1 \left[\sin \bar{x}^2 - \cos \bar{x}^2\right] \frac{\partial}{\partial \bar{x}^2}$

Index

U. Mühlich, *Fundamentals of Tensor Calculus for Engineers with a Primer on Smooth Manifolds*, Solid Mechanics and Its Applications 230,
DOI 10.1007/978-3-319-56264-3