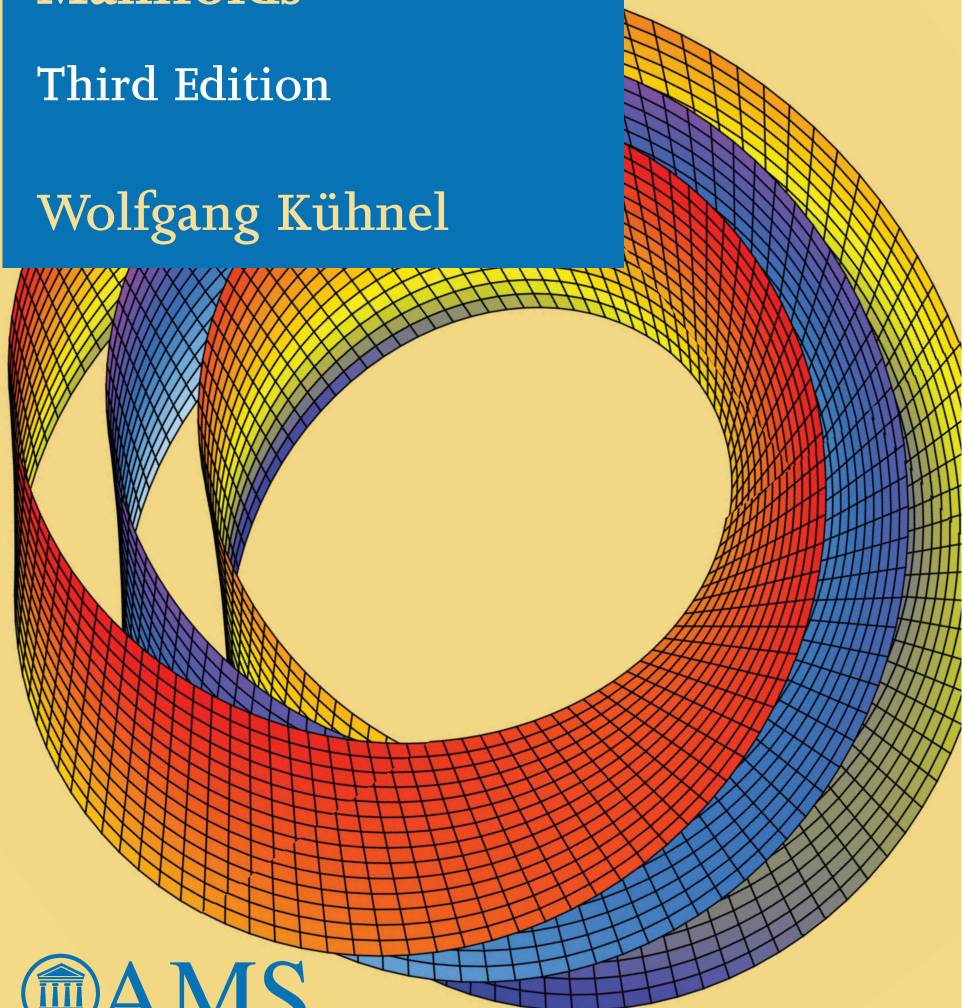
STUDENT MATHEMATICAL LIBRARY
Volume 77

Differential Geometry

Curves – Surfaces – Manifolds

Third Edition

Wolfgang Kühnel



AMS
American Mathematical Society

Differential Geometry

Curves–Surfaces–Manifolds

STUDENT MATHEMATICAL LIBRARY
Volume 77

Differential Geometry

Curves–Surfaces–Manifolds

Third Edition

Wolfgang Kühnel

Translated by
Bruce Hunt

American Mathematical Society
Providence, Rhode Island

Translation from German language edition: *Differentialgeometrie* by Wolfgang Kühnel, ©2013 Springer Vieweg | Springer Fachmedien Wiesbaden GmbH JAHR (formerly Vieweg+Teubner). Springer Fachmedien is part of Springer Science+Business Media. All rights reserved.

Translated by Bruce Hunt, with corrections and additions by the author.

Front and back cover image by Mario B. Schulz.

2010 *Mathematics Subject Classification*. Primary 53-01.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-77

Page 403 constitutes an extension of this copyright page.

Library of Congress Cataloging-in-Publication Data

Kühnel, Wolfgang, 1950–
[Differentialgeometrie. English]
Differential geometry : curves, surfaces, manifolds / Wolfgang Kühnel ; translated by Bruce Hunt.— Third edition.
pages cm. — (Student mathematical library ; volume 77)
Includes bibliographical references and index.
ISBN 978-1-4704-2320-9 (alk. paper)
1. Geometry, Differential. 2. Curves. 3. Surfaces. 4. Manifolds (Mathematics) I. Hunt, Bruce, 1958– II. Title.

QA641.K9613 2015
516.3′6—dc23

2015018451

10 9 8 7 6 5 4 3 2 1 20 19 18 17 16 15

Contents

Preface to the English Edition ix

Preface to the German Edition xi

Chapter 1. Notations and Prerequisites from Analysis 1

Chapter 2. Curves in $I\!R^n$ 7

2A Frenet curves in $I\!R^n$ 7

2B Plane curves and space curves 14

2C Relations between the curvature and the torsion 20

2D The Frenet equations and the fundamental theorem of the local theory of curves 27

2E Curves in Minkowski space $I\!R^3_1$ 33

2F The global theory of curves 37

Exercises 50

Chapter 3. The Local Theory of Surfaces 55

3A Surface elements and the first fundamental form 56

3B The Gauss map and the curvature of surfaces 66

3C Surfaces of rotation and ruled surfaces 77

3D Minimal surfaces 96

3E Surfaces in Minkowski space $I\!R^3_1$ 113

3F Hypersurfaces in $\mathbb{R}^{n+1}$ 122
Exercises 125

Chapter 4. The Intrinsic Geometry of Surfaces 133
4A The covariant derivative 134
4B Parallel displacement and geodesics 140
4C The Gauss equation and the Theorema Egregium 145
4D The fundamental theorem of the local theory of surfaces 152
4E The Gaussian curvature in special parameters 157
4F The Gauss-Bonnet Theorem 165
4G Selected topics in the global theory of surfaces 180
Exercises 192

Chapter 5. Riemannian Manifolds 197
5A The notion of a manifold 198
5B The tangent space 205
5C Riemannian metrics 212
5D The Riemannian connection 218

Chapter 6. The Curvature Tensor 233
6A Tensors 233
6B The sectional curvature 242
6C The Ricci tensor and the Einstein tensor 250

Chapter 7. Spaces of Constant Curvature 265
7A Hyperbolic space 266
7B Geodesics and Jacobi fields 276
7C The space form problem 291
7D Three-dimensional Euclidean and spherical space forms 296
Exercises 306

Chapter 8. Einstein Spaces 309
8A The variation of the Hilbert-Einstein functional 312
8B The Einstein field equations 321

8C Homogenous Einstein spaces 325
8D The decomposition of the curvature tensor 331
8E The Weyl tensor 341
8F Duality for four-manifolds and Petrov types 350
Exercises 358
Solutions to selected exercises 361
Bibliography 391
List of notation 395
Index 397

Preface to the English Edition

The German original was intended for courses on differential geometry for students in the middle of their academic education, that is, in the second or third year. In the Anglo-American system of university education, the contents of this textbook corresponds to an undergraduate course in elementary differential geometry (Chapters 1 – 4), followed by a beginning course in Riemannian geometry (Chapters 5 – 8). This led to the idea of having a translation of the German original into English.

I am very glad that the American Mathematical Society supported this project and published the present English version. I thank the translator, Bruce Hunt, for the hard work he had spent on the translation. From the beginning he was surprised by the quantity of text, compared to the quantity of formulas. In addition he had to struggle with complicated and long paragraphs in German. One of the major problems was to adapt the terminology of special notions in the theory of curves and surfaces to the English language. Another problem was to replace almost all references to German texts by references to English texts, in particular, all references to elementary textbooks on calculus, linear algebra, geometry, and topology. Ultimately all these problems could be solved, at least to a certain approximation. The bibliography contains only books in English, with just three exceptions. Therefore, the English version can be used as a textbook for third-year undergraduates and beginning graduate students.

Furthermore, I am grateful to Edward Dunne from the AMS who was extremely helpful at all stages of the project, not only for editorial and technical matters, but also for questions concerning the terminology and the tradition of notations. He pointed out that the ordinary spherical coordinates on the sphere, denoted by φ, ϑ in this book, are denoted ϑ, φ (that is, the other way around) in many English textbooks on calculus. We hope that this does not lead to major confusions.

In the second English edition a number of errors were corrected and a number of additional figures were added, following the second German edition. Most of the additional figures were provided by Gabriele Preissler and Michael Steller. The illustrations play an important rôle in this book. Hopefully they make the book more readable. The concept of having boxes around important statements was kept from the German original, even though now we have a few very large boxes covering major parts of certain pages.

Stuttgart, June 2005 *W. Kühnel*

The present third edition is a corrected and updated version that incorporates the development of altogether six editions in German, the last one from 2013. Each of these German editions was corrected, extended and improved in several directions. As an example, a number of proofs were made more precise if they turned out to be too short in the first edition. In comparison to the second English edition, the third edition includes many improvements, there are more figures and more exercises, and - as a new feature - at the end a number of solutions to selected exercises are given.

Stuttgart, July 2014 *W. Kühnel*

Preface to the German Edition

This book arose from courses given on the topic of "Differential geometry", which the author has given several times in different places. The amount of material corresponds roughly to a course in classical differential geometry of one semester length (Chapters 1-4 of the book), followed by a second one-semester course on Riemannian geometry (Chapters 5-8). The prerequisites are the standard courses in calculus (including several variables) and linear algebra. Only in section 3D (on minimal surfaces) do we assume some familiarity with complex function theory. For this reason the book is appropriate for a course in the latter part of the undergraduate curriculum, not only for students majoring in mathematics, but also those majoring in physics and other natural sciences. Accordingly, we do not present any material which could in any way be considered original. Instead, our intent is to present the basic notions and results which will enable the interested student to go on and study the masters. Especially in the introductory chapters we will take particular care in presenting the material with emphasis on the geometric intuition which is so characteristic of the topic of differential geometry; this is supported by a large number of figures in this part of the book. The results which the author considers particularly important are placed in boxes to emphasize them. These results can be thought of as a kind of skeleton of the theory.

This book wouldn't have been possible without the generous help of my students and colleagues, who found numerous mistakes in the distributed notes of the first version of this book. In particular I would like to mention Gunnar Ketelhut, Eric Sparla, Michael Steller and Gabriele Preissler, who spent considerable time and effort in reading the original notes. G. Ketelhut also supplied numerous suggestions for improvements in the text, as well as writing Section 8F himself. Martin Renner provided almost all the figures, which were produced with the computer algebra system MAPLE. Marc-Oliver Otto provided some figures for Chapter 7, and Ilva Maderer typed the original version in LaTeX. Finally, Michael Grüter accompanied the whole production process with helpful suggestions, as well as giving me personal support in several ways. The work and insistence of Dr. Ulrike Schmickler-Hirzebruch is responsible for the speed with which these lectures were nonetheless accepted for the series "Vieweg-Studium Aufbaukurs Mathematik" and then also appeared almost on time. My thanks goes to all of them.

Stuttgart, June 1999 *W. Kühnel*

Chapter 1

Notations and Prerequisites from Analysis

The differential geometry which is introduced in the following Chapters 2 and 3 (also referred to as *Euclidean differential geometry*) is based on Euclidean space $\mathbb{E}^n$ as the ambient space. The most important algebraic structures on this space are on the one hand the structure of vector space, and on the other hand the Euclidean inner product. In addition we use the topological structure in the form of limits, open sets, differentiation and integration. By fixing a preferred point as the origin one can identify Euclidean space $\mathbb{E}^n$ with $\mathbb{R}^n$, which will be implicitly done in this book. For basic notions from linear algebra we refer to [**31**], and for the basic notions of analysis (including ordinary differential equations) we refer to [**27**]. This Chapter 1 is only meant as a list of some basic notions. The sections 1.1 – 1.3 will be used thoughout the book. The implicit function theorem 1.4 is useful but not absolutely necessary, and the notion of a submanifold in 1.5 is helpful but is not really a necessary prerequisite. Therefore, the reader may skip directly from Sections 1.1-1.3 to Chapter 2.

1.1. $\mathbb{R}^n$ as a vector space with an inner product.

$\mathbb{R}^n$ is defined as the set of all n-tuples of real numbers, which are written $x = (x_1, \dots, x_n)$. Given the componentwise addition

$$x + y = (x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$$

as well as the scalar multiplication by real numbers

$$a \cdot (x_1, \dots, x_n) = (ax_1, \dots, ax_n),$$

the space $\mathbb{R}^n$ is an $\mathbb{R}$-vector space. On this vector space we have the *Euclidean inner product* (bilinear form) defined as

$$x, y \longmapsto \big\langle x, y \big\rangle = x_1 y_1 + \dots + x_n y_n.$$

Properties of the inner product are the following

1. $\big\langle x, y \big\rangle = \big\langle y, x \big\rangle,$ (*symmetry*)
2. $\big\langle a_1 x_1 + a_2 x_2, y \big\rangle = a_1 \big\langle x_1, y \big\rangle + a_2 \big\langle x_2, y \big\rangle,$ (*bilinearity*)
3. $\big\langle x, x \big\rangle > 0$ for all $0 \neq x \in \mathbb{R}^n$. (*positive definiteness*)

This allows us to define the *length* of vectors by the *norm*

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

as well as introducing the *angle* φ between two vectors $x, y \neq 0$ by

$$\cos \varphi = \Big\langle \frac{x}{||x||}, \frac{y}{||y||} \Big\rangle = \frac{\langle x, y \rangle}{||x|| \cdot ||y||}.$$

The (metric) *distance* between two points x, y is then defined as the norm of the difference vector $y - x$. This makes $\mathbb{R}^n$ a *normed vector space* on the one hand and a *metric space* on the other. Euclidean geometry can then be based on these notions of lengths and angles.

1.2. $\mathbb{R}^n$ as a topological space.

The topology of $\mathbb{R}^n$ is strongly based on the notion of an open ε-neighborhood of a point x: $U_\varepsilon(x) = \{y \in \mathbb{R}^n \mid ||x - y|| < \varepsilon\}$. With the help of these neighborhoods, the notion of *convergence* and of the *limit of a function* are defined. Moreover, a subset O is said to be *open*, if for every point $x \in O$ there is a certain ε-neighborhood U_ε contained in O (for an appropriately chosen $\varepsilon > 0$, depending on x). Then the *topology* of $\mathbb{R}^n$ is defined as the system of all open sets (including the empty set). A set A is called *closed*, if its complement $\mathbb{R}^n \setminus A$ is open.

1.3. Differentiation in $\mathbb{R}^n$.

The most important notion for the contents of this book (which is also the source of the name "differential geometry") is that of *derivative* or *differentiation* of real-valued functions which are defined on some open set $U \subset \mathbb{R}^n$ or, more generally, of maps defined on open sets $U \subset \mathbb{R}^n$ to $\mathbb{R}^m$. To say that a function is differentiable is to say that it can be linearized up to terms of second order. More precisely a map $F\colon U \to \mathbb{R}^m$ is said to be *differentiable* at a point $x \in U$, if there is a linear map $A_x\colon \mathbb{R}^n \to \mathbb{R}^m$ such that in a neighborhood $U_\epsilon(x)$ one has

$$F(x+\xi) = F(x) + A_x(\xi) + o(||\xi||).$$

Here, the symbol $o(||\xi||)$ means that the terms indicated by it tend to zero as $\xi \to 0$, even after previous division by $||\xi||$. Then A_x is the linear map described by the *Jacobi matrix* or the *Jacobian* of f:

$$J_xF = \Big(\frac{\partial F_i}{\partial x_j}\Big|_x\Big)_{i,j}.$$

The *rank* of the map F at the point x is then defined as the rank of the Jacobian. For our purposes the most important case is the one in which F is differentiable at every point $x \in U$ and the rank is everywhere maximal. In this case, one calls the map F an *immersion* (in case $n \leq m$) or a *submersion* (in case $n \geq m$). A immersion (resp. submersion) is characterized by the fact that the Jacobian represents an injective (resp. surjective) linear mapping, the linearization of F. The importance of this property becomes quite clear in the implicit function theorem.

1.4. The implicit function theorem.

An *implicit function* is given for example by an equation $F(x, y) = 0$. One views here either y as a function of x or vice versa. If F is linear, then this is just a question of the rank of F. If F is not linear, but continuously differentiable, then such an implicit description can hold in general at most locally (which is already seen on the very simple equation $x^2 + y^2 - 1 = 0$). Moreover, y can only be a differentiable function of x (resp. vice versa) if $\frac{\partial F}{\partial y} \neq 0$ (resp. $\frac{\partial F}{\partial x} \neq 0$). This holds in all dimensions, and can be formulated in the following way, compare [**27**], Ch. XVIII.

Let $U_1 \subset \mathbb{R}^k$ and $U_2 \subset \mathbb{R}^m$ be open sets and let $F\colon U_1 \times U_2 \to \mathbb{R}^m$ be a continuously differentiable mapping, which we write as $(x, y) \mapsto F(x, y)$ for $x \in U_1, y \in U_2$. Let $(a, b) \in U_1 \times U_2$ be a point for which $F(a, b) = 0$ and for which the square matrix

$$\frac{\partial F}{\partial y} := \Big(\frac{\partial F_i}{\partial y_j}\Big)_{i,j=1,\ldots,m}$$

is invertible. Then there are open neighborhoods $V_1 \subset U_1$ of a, $V_2 \subset U_2$ of b, and a continuously differentiable mapping

$$g\colon V_1 \to V_2$$

such that for all $(x, y) \in V_1 \times V_2$ the implicit equation $F(x, y) = 0$ *holds if and only if the* explicit equation $y = g(x)$ *is satisfied.*

The most important assumption in the formulation of this theorem is the rank of the mapping F (that is, the rank of the Jacobian). One can say that locally, a continuously differentiable mapping of maximal rank behaves like a linear mapping of maximal rank. One consequence is the *theorem on inverse mappings* ([**27**], Ch. XVIII), which can be stated as follows.

Let U be an open set in $\mathbb{R}^n$ and let $f\colon U \to \mathbb{R}^n$ be a continuosuly differentiable mapping with the property that the Jacobian at a fixed point u_0 is invertible. Then there is a neighborhood V with $u_0 \in V \subset U$ on which the mapping f is also invertible, i.e., $f\big|_V\colon V \to f(V)$ is a diffeomorphism.

Special cases. (Curves, surfaces)
For a given function F of two real variables x, y the equation $F(x, y) = 0$ describes a “curve” whenever the gradient of F does not vanish, that is to say if $\frac{\partial F}{\partial x} \neq 0$ or $\frac{\partial F}{\partial y} \neq 0$ at every point satisfying $F(x, y) = 0$. If this assumption is satisfied, then this curve can always be parametrized locally as a *regular parametrized curve* in the sense of Definition 2.1 below.

For a given function F of three variables x, y, z the equation $F(x, y, z) = 0$ describes a “surface” whenever the gradient of F does not vanish, i.e., if $\frac{\partial F}{\partial x} \neq 0$ or $\frac{\partial F}{\partial y} \neq 0$ or $\frac{\partial F}{\partial z} \neq 0$. If this assumption on the gradient is satisfied, then this surface can always be parametrized locally as a *parametrized surface element* in the sense of Definition 3.1 below.

If we generalize this concept to the situation of arbitrarily many variables and arbitrarily many real functions simultaneously, we obtain, directly and naturally, the notion of a *submanifold.*

1.5. Definition. (Submanifold)

A *k-dimensional submanifold* (of class C^α) $M \subset \mathbb{R}^n$ is defined by the condition that M is given locally as the zero set $F^{-1}(0)$ of an (α-times) continuously differentiable mapping

$$\mathbb{R}^n \supseteq U \xrightarrow{F} \mathbb{R}^{n-k}$$

with maximal rank, i.e., $\mathrm{rank}(J_x F) = \mathrm{rank}\big(\frac{\partial F}{\partial x}\big|_x\big) = n - k$ for every $x \in M \cap U$, where $M \cap U = F^{-1}(0)$ holds for an appropriately chosen neighborhood U of every point of M. Locally, one can also describe M as the image of an *immersion* of class C^α

$$\mathbb{R}^k \supseteq V \xrightarrow{f} M \subset \mathbb{R}^n$$

for which $\mathrm{rank}(Df) = k$. Such an f is said to be a *local parametrization*, while f^{-1} is called a *chart* of M. On the other hand, the image of an immersion is not always a submanifold, not even when f is injective. The number k is the *dimension*, $n-k$ the *codimension* of M. See also [**27**, p.531],

As special cases we recognize the cases $k = 1$ (*curves* in $\mathbb{R}^n$, which are treated in detail in Chapter 2), $k = 2$ and $n = 3$ (*surfaces* in $\mathbb{R}^3$, the most classical topic of differential geometry, studied in Chapter 3) and $k = n-1$ (*hypersurfaces* in $\mathbb{R}^n$, see Section 3F below).

In physics and other sciences it is required to distinguish between *points* on the one hand and *vectors* on the other. If in calculations both are regarded as elements of $\mathbb{R}^n$, confusion can sometimes occur. In order to avoid such confusion, points and vectors have to be declared as different objects in a formal definition. This leads to the notion of a tangent space and tangent bundle as follows:

1.6. Definition. (Tangent bundle of $\mathbb{R}^n$)

$T\mathbb{R}^n := \mathbb{R}^n \times \mathbb{R}^n$ is called the *tangent bundle* of $\mathbb{R}^n$. For every fixed point $x \in \mathbb{R}^n$ the space

$$T_x\mathbb{R}^n := \{x\} \times \mathbb{R}^n$$

is called the *tangent space* at the point x (= space of all tangent vectors at the point x). By means of this formal definition a clear distinction is made between *points* and *vectors* of $\mathbb{R}^n$. Moreover the tangent spaces $T_x\mathbb{R}^n$ and $T_y\mathbb{R}^n$ are disjoint by definition whenever $x \neq y$. The *derivative* (or *differential*) Df of a differentiable mapping f is defined for every x as the mapping

$$Df|_x \colon T_x\mathbb{R}^k \longrightarrow T_{f(x)}\mathbb{R}^n \quad \text{with} \quad (x,v) \mapsto (f(x), J_xf(v)).$$

For simplicity one also writes $Df|_x \colon \mathbb{R}^k \longrightarrow \mathbb{R}^n$ for this if there is no danger of confusion. Then $Df|_x$ can be viewed as a linear map between ordinary vector spaces, described just by the Jacobi matrix. In accordance with 1.3, one has the short expansion

$$f(x + \varepsilon \cdot v) = f(x) + \varepsilon \cdot J_xf(v) + o(\varepsilon).$$

1.7. Definition. (Tangent space to a submanifold)

Let $M \subset \mathbb{R}^n$ be a k*-dimensional submanifold*, and let $p \in M$. The tangent space to M at the point p is the vector subspace $T_pM \subset T_p\mathbb{R}^n$, which is defined by

$$T_pM := Df|_u(\{u\} \times \mathbb{R}^k) = Df|_u(T_u\mathbb{R}^k)$$

for a parametrization $f : U \to M$ with $f(u) = p$, where $U \subseteq \mathbb{R}^k$ is an open set. The vector space T_pM is k*-dimensional* and does not depend on the choice of f. The collection of tangent spaces

$$TM := \dot{\bigcup}_{p\in M} T_pM$$

is called the *tangent bundle of* M. It comes equipped with the *projection* $\pi \colon TM \longrightarrow M$, which is defined as $\pi(p, V) = p$. Note that there is a difference between TM and $M \times \mathbb{R}^k$, see [**39**].

1.8. Definition. (Normal space along a submanifold)

Let $M \subset \mathbb{R}^n$ be a k*-dimensional submanifold.* The *normal space* to M at the point $p \in M$ is the vector subspace $\perp_p M \subset T_p\mathbb{R}^n$, which is the *orthogonal complement* of T_pM:

$$T_p\mathbb{R}^n = \underbrace{T_pM}_{k\text{-dim.}} \oplus \underbrace{\perp_p M}_{(n-k)\text{-dim.}}$$

Here $\oplus$ denotes the orthogonal direct sum with respect to the Euclidean inner product.

Chapter 2

Curves in $\mathbb{R}^n$

In the practical world, curves arise in many different ways, for example as the profile curves or contours of technical objects. On white drawing paper, curves appear as the trace of the pencil or other drawing medium used to draw it. For physicists, curves arise naturally in the *motion of a particle* in time t. From this point of view the association of the parameter t to the position $c(t)$ is important, and this process is called a *parametrization* of the curve; the curve is then called a *parametrized curve*. This notion is the most appropriate for a formal mathematical treatment of curves. In this formulation, one passes from the real-world notion of a "thin" object to one which has no width whatsoever: a one-dimensional or "infinitely thin" object. Here both the parametrization and the curve are supposed to have reasonable properties, which allow an acceptable mathematical treatment. A short introduction to the theory of curves can be found in [**27**], Ch. X, §5, but we will not assume any familiarity with this on the part of the reader.

2A Frenet curves in $\mathbb{R}^n$

Mathematically one can define a *curve* most easily as a continuous mapping from an interval $I \subseteq \mathbb{R}$ to $\mathbb{R}^n$. Unfortunately, the assumption of continuity is so weak that curves defined in this manner can look very complicated and have unexpected (pathological) properties.

There are continuous curves which cover a whole square in the plane. Thus it is natural to take the point of view of analysis and require differentiability in addition to continuity. But still this assumption is not quite the right one. Differentiability of a map just means that it can be linearly approximated. For the image set, however, this no longer needs to be the case. From a geometrical point of view it makes sense to require that the image curve can be approximated by a line at each point, i.e., to require that the image curve has a tangent as a geometrical linearization at every point. This means that the derivative of the map from I to $I\!R^n$ must be non-vanishing. One calls a map with this property an *immersion*. This simply means that the derivative of the parametrization always has the highest possible rank, which in our case, where the domain is an interval, is one.

2.1. Definition. A *regular parametrized curve* is a continuously differentiable immersion $c\colon I \longrightarrow I\!R^n$, defined on a real interval $I \subseteq I\!R$. This means that $\dot{c} = \frac{dc}{dt} \neq 0$ holds everywhere.

The vector
$$\dot{c}(t_0) = \frac{dc}{dt}\Big|_{t=t_0}$$
is called the *tangent vector* to c at t_0, and the line spanned by this vector through $c(t_0)$ is called the *tangent* (line) to c at this point. This is a geometric approximation of the first order in a neighborhood of the point with $c(t_0 + t) = c(t_0) + t \cdot \dot{c}(t_0) + o(t)$.

A *regular curve* is an equivalence class of regular parametrized curves, where the equivalence relation is given by regular (orientation preserving) parameter transformations
$$\varphi\colon [\alpha, \beta] \longrightarrow [a, b], \quad \varphi' > 0, \quad \varphi \ \begin{array}{l}\text{bijective and}\\ \text{continuously differentiable;}\end{array}$$
c and $c \circ \varphi$ are then considered to be *equivalent*. The *length* of the curve
$$\int_a^b \left|\left|\frac{dc}{dt}\right|\right| dt$$
is invariant under the parameter transformations as just described. In the sciences one can view a curve as the motion of a particle, with the trajectory of the particle as a function of time. It is regular if the instantaneous speed $||\dot{c}||$ never vanishes.

2.2. Lemma. Every regular curve can be parametrized by its arc length (in other words, the tangent vector at every point has unit length).

PROOF: Let a curve $c: [a,b] \to \mathbb{R}^n$ be given, of total length $L = \int_a^b ||\frac{dc}{dt}||dt$. We then set $[\alpha, \beta] = [0, L]$ and introduce the *arc length parameter* s by the relation

$$s(t) := \psi(t) = \int_a^t \left\| \frac{dc}{dt}(\tau) \right\| d\tau.$$

This defines a map $\psi: [a,b] \to [0,L]$. Then one has $\frac{ds}{dt} = \frac{d\psi}{dt} = ||\frac{dc}{dt}|| \neq 0$, and consequently there is an inverse function $\varphi := \psi^{-1}$ such that $c \circ \varphi = c \circ \psi^{-1}$ is parametrized by arc length. Two distinct such parametrizations $c(s)$ and $c(\sigma)$ differ only by a parameter transformation $s \mapsto \sigma(s)$ with $\frac{d\sigma}{ds} = 1$. Therefore we have $\sigma = s + s_0$ with a constant s_0. Hence this parametrization is unique up to a translation $s \mapsto s + s_0$ Passing through the curve backwards is considered as another curve c^-, e.g. $c^-(s) = c(L - s)$. In this case the arc length parameter is transformed by $s \mapsto s_0 - s$ with a cnstant s_0. □

We will use the following notations in the sequel:

$c(t)$	denotes an arbitrary regular parametrization,
$c(s)$	denotes the parametrization by arc length,
$\dot{c} = \frac{dc}{dt}$	denotes the tangent vector,
$c' = \frac{dc}{ds}$	denotes the unit tangent vector.

In particular one then has $\dot{c} = \frac{ds}{dt}c' = ||\dot{c}||c'$ and $||c'|| = 1$.

2.3. Examples.

1. $c(t) = (at, bt)$, a *line* in standard parametrization. Since $\dot{c} = (a, b)$, the parameter is the arc length if and only if $a^2 + b^2 = 1$. The parametrization $c(t) = (at^3, bt^3)$ describes exactly the same line, but it is not regular for $t = 0$.
2. $c(t) = \frac{1}{2}(\cos 2t, \sin 2t)$, a *circle* of radius $\frac{1}{2}$. Since of course $\dot{c}(t) = (-\sin 2t, \cos 2t)$ one has $||\dot{c}|| = 1$. Hence t is the arc

length, i.e., $t = s$. See also Figure 2.1. The circle passed in the opposite direction is given by $c(t) = \frac{1}{2}\big(\cos(-2t), \sin(-2t)\big) = \frac{1}{2}(\cos 2t, -\sin 2t)$. This is parametrized by arc length as well.

3. $c(t) = (a\cos(\alpha t), a\sin(\alpha t), bt)$ with constants α, a, b. This is called a *(circular) helix*. Since
$$\dot{c}(t) = (-\alpha a \sin(\alpha t), \alpha a \cos(\alpha t), b),$$
one has $||\dot{c}|| = \sqrt{\alpha^2 a^2 + b^2}$. Therefore c is parametrized by arc length up to a constant multiple of t, i.e., one has $s = t \cdot \sqrt{\alpha^2 a^2 + b^2}$. Geometrically, the curve c arises as the trajectory of a point $(a, 0, 0)$ under the following one-parameter group of *screw-motions* or *helicoidal motions*:
$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \longmapsto \underbrace{\begin{pmatrix} \cos(\alpha t) & -\sin(\alpha t) & 0 \\ \sin(\alpha t) & \cos(\alpha t) & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{rotation}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \underbrace{\begin{pmatrix} 0 \\ 0 \\ bt \end{pmatrix}}_{\text{translation}} .$$
For appropriately chosen parameters, a motion of this kind maps every point on the curve to an arbitrary other point. Thus one expects that from a geometric point of view this curve will have good properties (a certain homogeneity in all scalar quantities which are geometrically relevant). The special case $b = 0$ leads back to a circle.

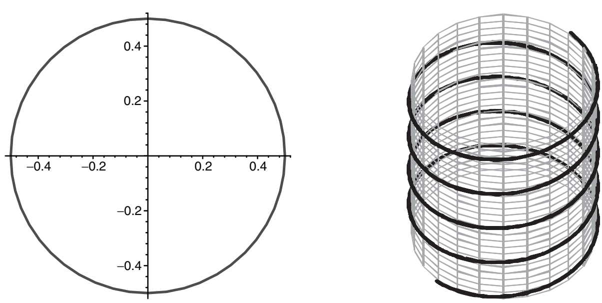


Figure 2.1. Circle, (circular) helix

4. $c(t) = (t^2, t^3)$, the so-called *Neil parabola* or *semicubical parabola*. The tangent vector is $\dot{c}(t) = (2t, 3t^2)$ with $\dot{c}(0) = (0, 0)$, hence

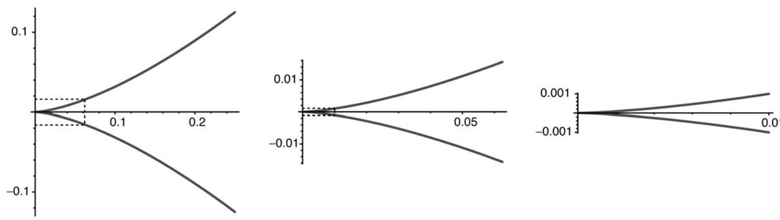


Figure 2.2. Neil parabola

at $t = 0$ there is no regular parametrization. In fact the curve doesn't have a tangent touching it at the point, as the curve has a "bend" by an angle π. This is no contradiction to the differentiability of the map c.

5. $c(t) = (t, a \cosh \frac{t}{a})$ with a constant a, the *catenary*. This curve arises as the stable position of a (heavy but infinitely supple) chain strung between two fixed points. Since $\dot{c}(t) = (1, \sinh \frac{t}{a})$, t is not the arc length.

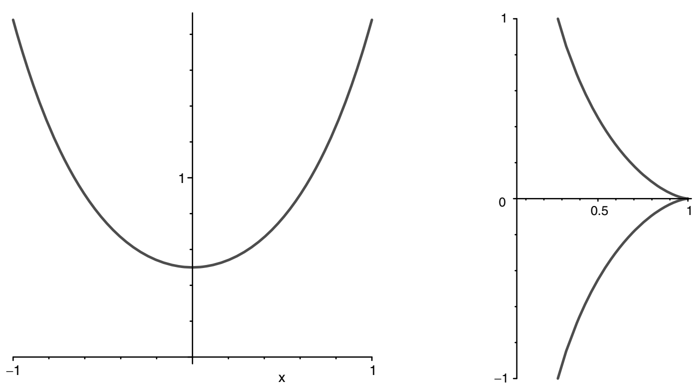


Figure 2.3. Catenary, tractrix

6. The *tractrix* is characterized by the property that from every point p the tangent meets a fixed line (for example the y-axis) at a constant distance. For the case where the fixed line is

the y-axis and the constant distance is 1, one can choose the parametrization $c(t) = \big(\exp(-t), \int_0^t \sqrt{1-\exp(-2x)}dx\big)$ for the upper part and $c(t) = \big(\exp(-|t|), \int_0^t \sqrt{1-\exp(-2|x|)}dx\big)$ for both parts together, see Figure 2.3.

REMARK: The local behavior of a curve which has been parametrized by arc length can be studied by means of its *Taylor expansion*:

$$c(s) = c(0) + sc'(0) + \frac{s^2}{2}c''(0) + \frac{s^3}{6}c'''(0) + o(s^3).$$

The linearization $c(0) + sc'(0)$ is a line, which is the *tangent* of c at $s = 0$ (since $c'(0) \neq 0$). The quadratic part of the expansion, $c(0) + sc'(0) + \frac{s^2}{2}c''(0)$, is a parabola (if $c''(0) \neq 0$) which is referred to as the (Euclidean) *osculating conic*. It has contact of second order with the curve. Note that c'' is perpendicular to c', as can be seen by differentiating $\langle c', c'\rangle = 1$: $0 = \langle c', c'\rangle' = 2\langle c'', c'\rangle$. This is further explained and extended in the following definition.

One says that two curves $c_1(s)$ and $c_2(s)$ (both assumed to be parametrized by arc length) are said to *have contact of the kth order* if

$$c_1(0) = c_2(0),\ c_1'(0) = c_2'(0),\ c_1''(0) = c_2''(0),\ \ldots,\ c_1^{(k)}(0) = c_2^{(k)}(0);$$

that is, if the Taylor expansions of the two curves coincide up to terms of the kth order. This obviously is related to the phenomenon of the two curves touching each other. One could also say that a curve touches another to the kth order. For example, the osculating conic above touches the curve to the second order, at the apex of the parabola. At a point other than the apex, the parabola can touch a given curve to even third order (cf. Exercise 2 at the end of the chapter). Similarly, one can look for cubic and quartic curves which have contact with a given curve of the highest possible order. For example, cubic splines are an important tool in the computer treatment of curves.

In three-dimensional space and all the more in spaces of higher dimensions, one requires an adequate system of coordinates to describe curves, one which is adapted to the curve. Here one would expect that the vectors $c', c'', c''', \ldots$ describe the local behavior of the curve, at least as long as they do not vanish or – even better – if they are linearly independent. This motivates the following definition. Recall that an *n-frame* in Euclidean n-space is a basis of orthonormal vectors $e_1, \ldots, e_n$, in a specific order. For curves in n-space we take advantage of an adapted n-frame as follows.

2.4. Definition. (Frenet curve)
Let $c(s)$ be a *regular curve* in $\mathbb{R}^n$, which is parametrized by arc length and n-times continuously differentiable. Then c is called a *Frenet curve*, if at every point the vectors $c', c'', \ldots, c^{(n-1)}$ are linearly independent. The *Frenet n-frame* $e_1, e_2, \ldots, e_n$ is then uniquely determined by the following conditions:

(i) $e_1, \ldots, e_n$ are orthonormal and positively oriented.

(ii) For every $k = 1, \ldots, n-1$ one has $Lin(e_1, \ldots, e_k) = Lin(c', c'', \ldots, c^{(k)})$, where Lin denotes the linear span.

(iii) $\langle c^{(k)}, e_k \rangle > 0$ for $k = 1, \ldots, n-1$.

Note: In the case discussed most often, $n = 3$, the only restrictive condition on a Frenet curve is $c'' \neq 0$. This excludes only inflection points. For $n = 2$ there are no actual restrictions, cf. 2.5.

One obtains $e_1, \ldots, e_{n-1}$ from $c', \ldots, c^{(n-1)}$ by means of the *Gram – Schmidt orthogonalization procedure* as follows:

$$\begin{aligned}
e_1 &:= c', \\
e_2 &:= c''/ \| c'' \|, \\
e_3 &:= \Big(c''' - \langle c''', e_1\rangle e_1 - \langle c''', e_2\rangle e_2\Big) \Big/ \| \cdots \|, \\
&\ \vdots \\
e_j &:= \Big(c^{(j)} - \sum_{i=1}^{j-1} \langle c^{(j)}, e_i\rangle e_i\Big) \Big/ \| \cdots \|, \\
&\ \vdots \\
e_{n-1} &:= \Big(c^{(n-1)} - \sum_{i=1}^{n-2} \langle c^{(n-1)}, e_i\rangle e_i\Big) \Big/ \| \cdots \| .
\end{aligned}$$

The missing vector e_n is then uniquely determined by condition (i) in the above definition. One can say that every Frenet curve uniquely induces through its Frenet n-frame a curve in the Stiefel manifold of all n-frames in $\mathbb{R}^n$. The converse does not hold in general since, for example, for $n \geq 3$ a constant n-frame cannot correspond to any Frenet curve.

2B Plane curves and space curves

2.5. Plane curves. For $n = 2$ every regular curve is a *Frenet curve*, provided it is twice continuously differentiable. The *tangent vector* is $e_1 = c'$, the *normal vector* is e_2, which – if the orientation is positive – is the rotation by an angle of $\pi/2$ to the left of the vector e_1. From $0 = \langle c', c'\rangle' = 2\langle c', c''\rangle = 2\langle e_1, c''\rangle$, it follows that c'' and e_2 are linearly dependent, hence $c'' = \kappa e_2$ with some function κ. This function κ is said to the (oriented) *curvature* of c. Its sign indicates in which direction the curve (resp. its tangent) is rotating. Here $\kappa > 0$ indicates that the tangent goes to the left, while $\kappa < 0$ indicates that it rotates to the right. At an *inflection point* one has $\kappa = 0$, and the direction of the tangent is stationary.

One has the following equations for the derivatives, in which the second follows from the first, since e_2 and e_1 differ by a rotation of $\pi/2$:

$$e_1' = c'' = \kappa e_2, \qquad e_2' = -\kappa e_1,$$

or, using matrix notation,

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}' = \begin{pmatrix} 0 & \kappa \\ -\kappa & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}.$$

Note that the matrix on the right is skew-symmetric, which follows already from the relation $0 = \langle e_1, e_2\rangle' = \langle e_1', e_2\rangle + \langle e_2', e_1\rangle$. These equations are also called the *Frenet equations*.

Exercise: If one describes a curve in an adapted coordinate system by $c(t) = (t, y(t))$ (t is not the arc length here), then one has

$$y(0) = \dot{y}(0) = 0, \quad \dot{c}(0) = (1, 0), \quad \ddot{c}(0) = (0, \ddot{y}(0)) = (0, \kappa(0)).$$

The curvature $\kappa(0) = \ddot{y}(0)$ hence coincides with the opening of the osculating parabola $t \mapsto (t, \frac{\ddot{y}(0)}{2}t^2)$, which is just the quadratic part of the Taylor expansion of c. In general two plane curves have contact of the *kth* order if and only if at the intersection point they have the same tangent and the same quantities $\kappa, \kappa', \kappa'', \ldots, \kappa^{(k-2)}$.

2.6. Theorem. (Plane curves with constant curvature)
A regular curve in $I\!R^2$ has constant curvature κ if and only if it is part of a circle of radius $\frac{1}{|\kappa|}$ (if $\kappa \neq 0$) or a line segment (if $\kappa = 0$).

PROOF: The proof follows directly from the Frenet equations. Assume first that $\kappa(s_0) \neq 0$ for a fixed s_0. Obviously the expression $c(s) + \frac{1}{\kappa(s_0)} e_2(s)$ is constant if and only if $c(s)$ is part of a circle of radius $|\frac{1}{\kappa(s_0)}|$, since the difference vector has constant length $|\frac{1}{\kappa(s_0)}|$. This is equivalent to the constancy $\kappa = \kappa(s_0)$ everywhere, because $c' + \frac{1}{\kappa(s_0)} e_2' = e_1 - \frac{1}{\kappa(s_0)} \kappa e_1$.
The fact that $\kappa \equiv 0$ only holds for line segments follows from $e_2' = -\kappa e_1$: The condition $0 = \kappa e_2(s) = e_1'(s) = c''(s)$ directly implies $c'(s) = \mathbf{a}$ and $c(s) = s\mathbf{a} + \mathbf{b}$ with constant $\mathbf{a}, \mathbf{b} \in I\!R^2$. $\square$

2.7. Remarks 1. For every regular curve in the plane with non-vanishing curvature the circle centered at $c(s_0) + \frac{1}{\kappa(s_0)} e_2(s_0)$ with radius $|\frac{1}{\kappa(s_0)}|$ is called the *osculating circle* of c in the point $c(s_0)$. It has contact of order two with the curve and is uniquely determined by this property. The curve which is formed by all of the centers of these circles,

$$s \mapsto c(s) + \frac{1}{\kappa(s)} e_2(s),$$

is called the *evolute* or the *focal curve* of c. This curve is not necessarily regular. Typically one has cusps like that occuring in the Neil parabola. In fact, the evolute of the catenary has such a cusp, and the evolute of an ellipse has four such cusps corresponding to the points with extremal curvature, compare exercise 3 at the end of the chapter.

2. Not only does every plane curve uniquely determine its curvature function $\kappa(s)$, but also conversely, the curvature function κ also determines the curve, up to translations and rotations, i.e., up to the prescription of a point on the curve and the tangent of the curve at

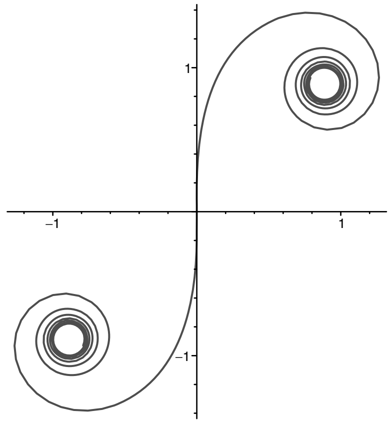


Figure 2.4. Cornu spiral with constant κ/s

that point. We even have the following *explicit determination* of the curve in terms of its curvature. Let the curvature function $\kappa(s)$ be given. Then one can set

$$e_1 = \big(\cos(\alpha(s)), \sin(\alpha(s))\big)$$

with a function $\alpha(s)$ which is to be found. Necessarily one has

$$e_2 = \big(-\sin(\alpha(s)), \cos(\alpha(s))\big).$$

The Frenet equation says that $\kappa e_2 = e_1' = \alpha' e_2$, hence $\kappa = \alpha'$. By a judicious choice of adapted coordinate system we can assume that for $s = 0$, the curve passes through the origin with $e_1 = (1, 0)$; then $\alpha(0) = 0$, and hence $\alpha(s) = \int_0^s \kappa(t)dt$. The sought-for curve is then given by the relation

$$x(s) = \int_0^s \cos\Big(\int_0^\sigma \kappa(t)dt\Big)d\sigma, \quad y(s) = \int_0^s \sin\Big(\int_0^\sigma \kappa(t)dt\Big)d\sigma.$$

For constant κ this again leads to the solutions we already met in Theorem 2.6. If κ is a linear function[1] of s, then we obtain the so-called *Cornu spiral*, see Figure 2.4.

[1]Pictures of the curves for which κ is quadratic in s can be found for example in F. Dillen, *The classification of hypersurfaces of a Euclidean space with parallel higher order fundamental form*, Math. Zeitschrift **203**, 635–643 (1990).

2.8. Space curves. For $n = 3$ a regular three-times continuously differentiable curve is called a *Frenet curve*, if $c'' \neq 0$ everywhere. The accompanying three-frame is then given by

$$\begin{aligned} e_1 &= c', && (\textit{tangent vector}) \\ e_2 &= \frac{c''}{\|c''\|}, && (\textit{principal normal vector}) \\ e_3 &= e_1 \times e_2. && (\textit{binormal vector}) \end{aligned}$$

The function $\kappa := ||c''||$ is called the *curvature* of c. By assumption this number is always positive. The equations for the derivatives are

$$\begin{aligned} e_1' &= c'' = \kappa e_2, \\ e_2' &= \langle e_2', e_1\rangle e_1 + \underbrace{\langle e_2', e_2\rangle}_{=0} e_2 + \langle e_2', e_3\rangle e_3 \\ &= \langle -e_2, e_1'\rangle e_1 + \underbrace{\langle e_2', e_3\rangle}_{=:\tau} e_3 \\ &= -\kappa e_1 + \tau e_3, \\ e_3' &= \langle e_3', e_1\rangle e_1 + \langle e_3', e_2\rangle e_2 + \underbrace{\langle e_3', e_3\rangle}_{=0} e_3 \\ &= -\underbrace{\langle e_3, e_1'\rangle}_{=0} e_1 - \underbrace{\langle e_3, e_2'\rangle}_{=\tau} e_2 \\ &= -\tau e_2. \end{aligned}$$

The function $\tau := \langle e_2', e_3\rangle$ is called the *torsion* of c. It describes how the (e_1, e_2)-plane changes along the curve. These three equations for the derivatives are called the *Frenet equations*, and in matrix notation they take the following form:

$$\begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}' = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}.$$

REMARK: A *plane curve* (viewed as a space curve) with $c'' \neq 0$ is also a Frenet curve in $\mathbb{R}^3$. The torsion of this curve is $\tau \equiv 0$, because e_3 is constant. The converse of this is also true: $\tau \equiv 0$ implies that e_3 is constant, and in addition that c lies in a plane which is perpendicular to e_3. This follows easily from the Frenet equations. If $c''(s) = 0$

at a single point, one gets a Frenet three-frame from the left and a different one from the right, with an orthogonal "jump matrix" at the point where $c''(s) = 0$. On the other hand, it is possible to join two curves lying in different planes in such a way that $\tau \equiv 0$ still holds everywhere, with the exception of this one point, see Problem 24 at the end of this chapter. If $\tau \neq 0$, then the sign of τ indicates a certain sense of rotation of the curve (in the sense of orientation). In earlier times, differential geometers had special names for these orientations ("weinwendig" and "hopfenwendig" in German because of the different growth behavior of grapevine and hops). For more on space curves with constant curvature and constant torsion, see Section 2.12.

2.9. Corollary. (Taylor expansion of the accompanying three-frame) The ordinary Taylor expansion around the point $s = 0$ is

$$c(s) = c(0) + sc'(0) + \frac{s^2}{2}c''(0) + \frac{s^3}{6}c'''(0) + o(s^3)$$

and can be translated into an expansion for the three-frame of the following form:

$$c(s) = c(0) + \alpha(s)e_1(0) + \beta(s)e_2(0) + \gamma(s)e_3(0) + o(s^3)$$

with unknown coefficients α, β, γ. This is seen as follows.

First one has, by the Frenet equations,

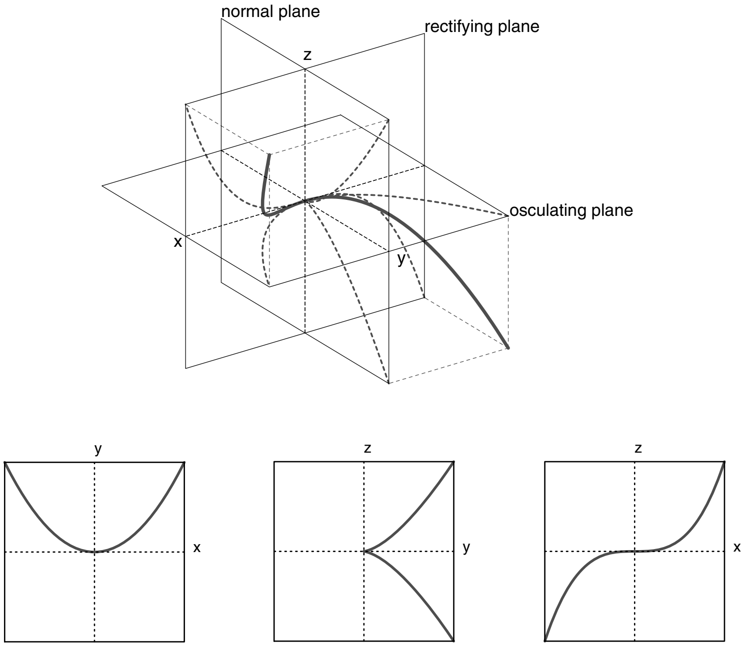
$$\begin{aligned} c' &= e_1, \\ c'' &= e_1' &&= \kappa e_2, \\ c''' &= (\kappa e_2)' &&= \kappa' e_2 + \kappa e_2' &&= \kappa' e_2 + \kappa(-\kappa e_1 + \tau e_3), \end{aligned}$$

which implies

$$c(s) = c(0) + se_1 + \frac{s^2}{2}\kappa e_2 + \frac{s^3}{6}\Big(\kappa' e_2 - \kappa^2 e_1 + \kappa\tau e_3\Big) + o(s^3)$$

$$= c(0) + \Big(s - \frac{s^3\kappa^2}{6}\Big)e_1 + \Big(\frac{s^2\kappa}{2} + \frac{s^3\kappa'}{6}\Big)e_2 + \frac{s^3\kappa\tau}{6}e_3 + o(s^3),$$

where κ, κ' and τ are evaluated at $s = 0$. The projections in the various (e_i, e_j)-planes are the following, see Figure 2.5:



(a) Osculating plane (b) Normal plane (c) Rectifying plane

Figure 2.5. Three projections of the space curve $xe_1 + ye_2 + ze_3$

(e_1, e_2)-*plane* (*osculating plane*):

$$c(s) = c(0) + se_1(0) + \frac{s^2\kappa(0)}{2} e_2(0) + o(s^2)$$

The projection onto the osculating plane has the form of a *parabola* (up to $o(s^2)$).

(e_2, e_3)-*plane* (*normal plane*):

$$c(s) = c(0) + \Big(\frac{s^2\kappa(0)}{2} + \frac{s^3\kappa'(0)}{6}\Big) e_2(0) + \frac{s^3\kappa(0)\tau(0)}{6} e_3(0) + o(s^3).$$

The projection onto the normal plane has the form of a *Neil parabola* in case $\tau(0) \neq 0$ (up to $o(s^3)$).

(e_1, e_3)*-plane* (*rectifying plane*):

$$c(s) = c(0) + \Big(s - \frac{s^3\kappa^2(0)}{6}\Big)e_1(0) + \frac{s^3\kappa(0)\tau(0)}{6}e_3(0) + o(s^3).$$

The projection onto the rectifying plane is of the type of a *cubical parabola*, in case $\tau(0) \neq 0$ (up to $o(s^3)$).

2C Relations between the curvature and the torsion

We have seen in Section 2.6 that a Frenet curve in $I\!R^3$ with constant κ and vanishing τ is necessarily an arc of a circle (because it is contained in a plane). A *circular helix* has constant κ and τ, since it is a trajectory of a fixed point under a one-parameter group of helicoidal rotations or screw-motions, see Section 2.3. On the other hand, every Frenet curve with constant κ and τ is such a helix, as will be shown in Section 2.12. More generally, one would expect that every equation between the curvature and the torsion will lead to a similar characterization of the corresponding curve. Conversely one can attempt to classify the different possible classes of curves by means of the equations which hold between the curvature and torsion of these classes of curves. This is particularly interesting for spherical curves, i.e., curves which lie entirely on a sphere.

2.10. Theorem. (Osculating sphere and spherical curves)

(i) Let c be a Frenet curve in $I\!R^3$ with $\tau(s_0) \neq 0$. Then the surface of the sphere centered at the point

$$c(s_0) + \frac{1}{\kappa(s_0)}e_2(s_0) - \frac{\kappa'(s_0)}{\tau(s_0)\kappa^2(s_0)}e_3(s_0),$$

which passes through the point $c(s_0)$, has a point of contact with the curve at the point s_0 of the third order. This sphere is uniquely determined by these properties and is called the *osculating sphere*.

(ii) Let c be a Frenet curve of class C^4 in $\mathbb{R}^3$ with $\tau \neq 0$ everywhere. Then c lies on a sphere if and only if the following equation holds:
$$\frac{\tau}{\kappa} = \Big(\frac{\kappa'}{\tau\kappa^2}\Big)'.$$
(iii) Let c be a C^3-curve that is parametrized by arc length and whose image lies on the unit sphere $S^2 \subset \mathbb{R}^3$. Set $J := \mathrm{Det}(c, c', c'')$. Then c is a Frenet curve with curvature $\kappa = \sqrt{1+J^2}$ and torsion $\tau = J'/(1+J^2)$. The great circles are characterized by the condition $J \equiv 0$, other circles by constant J.

PROOF: For part (i) we start with the center $m(s_0)$ of the hypothetical osculating sphere

$$m(s_0) = c(s_0) + \alpha e_1(s_0) + \beta e_2(s_0) + \gamma e_3(s_0),$$

with coefficients α, β, γ which are to be determined. For the function $r(s) = \big\langle m - c(s), m - c(s)\big\rangle$ we calculate the derivatives:

$$\begin{aligned} r' &= -2\big\langle m - c(s), c'(s)\big\rangle, \\ r'' &= -2\big\langle m - c(s), c''(s)\big\rangle + 2\big\langle c'(s), c'(s)\big\rangle, \\ r''' &= -2\big\langle m - c(s), c'''(s)\big\rangle. \end{aligned}$$

The optimal contact of the sphere with the curve simply means that as many derivatives of $r(s)$ as possible vanish at the point $s = s_0$:

$$\begin{aligned} r'(s_0) = 0 \quad &\Longleftrightarrow \big\langle m - c(s_0), c'(s_0)\big\rangle = 0 \\ &\Longleftrightarrow \big\langle m - c(s_0), e_1(s_0)\big\rangle = 0 \Longleftrightarrow \alpha = 0, \\ r''(s_0) = 0 \quad &\Longleftrightarrow \big\langle m - c(s_0), c''(s_0)\big\rangle - \big\langle c'(s_0), c'(s_0)\big\rangle = 0 \\ &\Longleftrightarrow \beta\kappa - 1 = 0 \Longleftrightarrow \beta = \frac{1}{\kappa(s_0)}, \\ r'''(s_0) = 0 \quad &\Longleftrightarrow \big\langle m - c(s_0), c'''(s_0)\big\rangle = 0 \\ &\Longleftrightarrow \big\langle m - c(s_0), \kappa' e_2 - \kappa^2 e_1 + \kappa\tau e_3\big\rangle = 0 \\ &\Longleftrightarrow \frac{\kappa'}{\kappa} + \kappa\tau\gamma = 0 \Longleftrightarrow \gamma = -\frac{\kappa'(s_0)}{\kappa^2(s_0)\tau(s_0)}. \end{aligned}$$

Part (ii) follows similarly, if one considers $m(s)$ for variable s and puts on it the condition that $m(s)$ is constant, i.e., $m' \equiv 0$. This condition is

$$(m(s))' = \Big(c(s) + \frac{1}{\kappa(s)} e_2(s) - \frac{\kappa'(s)}{\tau(s)\kappa^2(s)} e_3(s)\Big)' = \left[\frac{\tau}{\kappa} - \Big(\frac{\kappa'}{\tau\kappa^2}\Big)'\right] e_3(s);$$

hence $m(s)$ is constant if and only if the differential equation in (ii) is satisfied. Then one also has $r'(s) = 0$, and the statement of part (ii) follows from this.

It is not surprising that for the condition just considered a differential equation in only the two variables κ and τ arises. Still it is interesting that the property in question can be verified just from this differential equation, without even knowing the position of the sphere.

(iii) By assumption the vectors $c, c', c \times c'$ form an orthonormal three-frame along the curve. From this fact we get

$$c'' = \langle c'', c\rangle c + \langle c'', c'\rangle c' + \langle c'', c \times c'\rangle c \times c'.$$

Now one has $\langle c'', c\rangle = -\langle c', c'\rangle = -1$, hence $c'' = -c + Jc \times c'$ and from this

$$\kappa^2 = \langle c'', c''\rangle = 1 + J^2 > 0.$$

By differentiating the equation $\langle c'', c\rangle = -1$ one obtains $\langle c''', c\rangle = 0$. Moreover one has $e_2 = \frac{1}{\kappa} c''$, $e_3 = c' \times e_2$ from which it follows that

$$\begin{aligned} \tau = -\langle e_3', e_2\rangle &= -\Big\langle \big(\frac{1}{\kappa} c' \times c''\big)', \frac{1}{\kappa} c''\Big\rangle \\ &= -\frac{1}{\kappa^2}\Big\langle c' \times c''', c''\Big\rangle + \frac{\kappa'}{\kappa^3}\Big\langle c' \times c'', c''\Big\rangle \\ &= -\frac{1}{\kappa^2}\Big\langle c' \times c''', -c + Jc \times c'\Big\rangle = \frac{J'}{\kappa^2}. \end{aligned}$$

Here the last equality follows from the fact that c''' is perpendicular to c (see the discussion above) and consequently $c' \times c'''$ is perpendicular to $c' \times c$. $\square$

Remarks: 1. The determinant J is itself an interesting invariant of the curve, which is just the curvature inside of the sphere. The vector $c \times c'$ is the unit vector which is perpendicular to the curve but tangent to the sphere (as it is perpendicular to the vector in space determined by the points of c). Then $J = \langle c'', c \times c' \rangle$ is the part of c'' which is tangent to the sphere. One also calls this the *geodesic curvature* of the curve; see in this respect also 4.37 in Chapter 4. One has $J = 0$ precisely for the great circles, and J is a non-vanishing constant for all other circles (exercise), compare Figure 4.1.

2. Condition (ii) yields an equation between the curvature and the torsion, with the help of which it is easy to check if a curve lies on a sphere. If the condition is satisfied, then it is in principle clear that only one of the functions is necessary for a complete description of the curve, the other being itself a function of the first. If one prescribes κ, then (iii) gives an explicit way of expressing this dependency by introducing the function $J = \pm\sqrt{\kappa^2 - 1}$. To see this, one considers the system of equations

$$\kappa^2 = 1 + J^2, \quad \tau\kappa^2 = J'.$$

Even the case $\tau = 0$ is taken account of in this relation; this case can only occur in conjunction with the relation $\kappa' = 0$, for example the great or lesser circles. The case $\kappa = 0$ cannot occur. As a test of this statement, set $\kappa = \sqrt{1 + J^2}$ and $\tau = J'/(1+J^2)$ in the equation in (ii). It follows that the equation in (ii) makes sense even if there are points with $\tau = 0$ because in any case the quantity $\kappa'/(\tau\kappa^2) = J/\sqrt{1 + J^2}$ is well defined. On an interval with $\tau = 0$ or, equivalently, $J' = 0$ this quantity is constant. Therefore the converse direction in (ii) holds even for $\tau = 0$, and in (iii) any choice of a function J leads to a spherical curve.

2.11. Theorem. (Slope lines)
For a Frenet curve in $\mathbb{R}^3$, the following conditions are equivalent:

(i) There is a vector $v \in \mathbb{R}^3 \setminus \{0\}$ with the property that $\langle e_1, v \rangle$ is constant.

(ii) There is a vector $v \in \mathbb{R}^3 \setminus \{0\}$ with $\langle e_2, v \rangle = 0$.

(iii) There is a vector $v \in \mathbb{R}^3 \setminus \{0\}$ such that $\langle e_3, v \rangle$ is constant.

(iv) The quotient $\dfrac{\tau}{\kappa}$ is constant.

In particular, in this case all rectifying planes contain a fixed vector v. Such curves are called *slope lines*, because they run up or down a

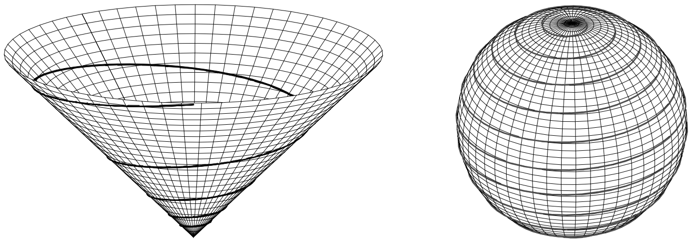


Figure 2.6. Slope line in a cone and in a sphere

surface with a constant slope. Also spherical curves can be slope lines – see Figure 2.6 as well as the exercises at the end of the chapter.

PROOF: If $\tau = 0$ holds on the entire interval I, then c is a plane curve, and the assertion is trivial with the choice $v = e_3$. So let us assume that there is a point with $\tau \neq 0$.

(i) $\Leftrightarrow$ (ii): $0 = \langle e_1, v\rangle' = \langle e_1', v\rangle$ implies that $\langle e_2, v\rangle = 0$, since $e_1' = \kappa e_2$ and conversely (note that $\kappa \neq 0$).

(iii) $\Leftrightarrow$ (ii) follows similarly from $0 = \langle e_3, v\rangle' = -\tau\langle e_2, v\rangle$ on any interval with $\tau \neq 0$, and (ii) $\Rightarrow$ (iii) holds in any case.

(i), (ii), (iii) together imply $v = \alpha e_1 + \beta e_3$ with constants α, β and with $\beta \neq 0$ since $\beta = 0$ would imply that e_1 is constant (not a Frenet curve). Since in addition v is constant, one has $0 = \alpha e_1' + \beta e_3' = (\alpha\kappa - \beta\tau)e_2$, hence $\frac{\tau}{\kappa} = \frac{\alpha}{\beta}$. In particular $\frac{\tau}{\kappa}$ is constant on a maximal interval $I' \subseteq I$ such that $\tau(s) \neq 0$ for any $s \in I'$. Such an I' must be open and closed, hence $I' = I$. So either we have $\tau = 0$ on the entire interval or $\tau \neq 0$ on the entire interval: A mixed case is impossible. Consequently we have (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii) $\Rightarrow$ (iv) along the entire curve.

Conversely, if $\frac{\tau}{\kappa}$ is constant, this implies that also

$$v := \frac{\tau}{\kappa} e_1 + e_3$$

is constant, because $v' = \frac{\tau}{\kappa} e_1' + e_3' = \frac{\tau}{\kappa}\kappa e_2 - \tau e_2 = 0$. This implies (i), (ii), (iii) by taking the inner product with the e_i because $\langle e_2, v\rangle = 0$.

The vector $\kappa v = \tau e_1 + \kappa e_3$, which points in the same direction, is also interesting for other curves, and is called the *Darboux rotation vector*, see 2.12. $\square$

In particular, a curve is a slope line whenever both κ and τ are constant. This case can be completely classified as follows.

2.12. Example. (Curves with constant Frenet curvature in $I\!R^3$)
For given constants a, b, α, the circular helix

$$c(t) = \big(a\cos(\alpha t), a\sin(\alpha t), bt\big)$$

is a Frenet curve in $I\!R^3$ when $a > 0, \alpha \neq 0$, see Figure 2.1. This curve is parametrized by arc length when

$$1 = \alpha^2 a^2 + b^2.$$

It then has constant Frenet curvature κ and constant torsion τ with

$$\begin{aligned} \kappa^2 &= \alpha^4 a^2, \\ \tau^2 &= \alpha^2 b^2. \end{aligned}$$

Conversely, for given constants κ, τ the system of the three equations above has the unique solution

$$\begin{aligned} \alpha^2 &= \kappa^2 + \tau^2, \\ a &= \kappa/(\kappa^2 + \tau^2), \\ b^2 &= \tau^2/(\kappa^2 + \tau^2). \end{aligned}$$

Consequence: Every Frenet curve in $I\!R^3$ with constant curvature κ and constant torsion τ is a part of a circular helix. The special case in which $\tau = 0$ is the case of a circle.

REMARKS: 1. The angular velocity α occurs in the normal form of the skew-symmetric matrix

$$K = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix};$$

one can calculate $-\alpha^2$ as the unique non-zero eigenvalue of the squared (and hence symmetric) matrix

$$K^2 = \begin{pmatrix} -\kappa^2 & 0 & \kappa\tau \\ 0 & -\kappa^2 - \tau^2 & 0 \\ \kappa\tau & 0 & -\tau^2 \end{pmatrix}.$$

The characteristic polynomial of this matrix is

$$\mathrm{Det}(K^2 - \lambda \cdot \mathrm{Id}) = -\lambda(\kappa^2 + \tau^2 + \lambda)^2,$$

hence $-\alpha^2 = -(\kappa^2 + \tau^2)$ is the sole non-vanishing eigenvalue. This determines α (up to sign), and we obtain from the above equations the result $a = \kappa/\alpha^2$ and $b = \pm\tau/\alpha$.

We can also see the normal form of the Frenet matrix K as follows:

$$\begin{pmatrix} 0 & \alpha & 0 \\ -\alpha & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

One clearly sees that the rank of the matrix is always 2. The Darboux rotation vector D is contained in the kernel of K, see Exercise 19 at the end of this chapter.

2. For every Frenet curve in $\mathbb{R}^3$ and every point p on that curve, there is a uniquely determined *accompanying helix* such that both curves have the same Frenet three-frame at the point p as well as having the same curvature and torsion. The screw-motion itself can be viewed as an accompanying motion to the curve. The *Darboux rotation vector*

$$D = \tau e_1 + \kappa e_3$$

should be seen in this context as well. It is contained in the rectifying plane and describes the accompanying screw-motion given by its direction (this is the axis of motion) and its length (this is the angular velocity of the motion). The *Darboux equations*

$$e_i' = D \times e_i \quad \text{for } i = 1, 2, 3$$

are then just a variant of the Frenet equations. See the exercises at the end of the chapter. For the helix considered above with $\alpha > 0$, the value of D is given by $D = (0, 0, \sqrt{\kappa^2 + \tau^2}) = (0, 0, \alpha)$, as is easily verified.

2D The Frenet equations and the fundamental theorem of the local theory of curves

2.13. Theorem and definition. (Frenet equations in $\mathbb{R}^n$)
Let c be a Frenet curve in $\mathbb{R}^n$ with Frenet n-frame $e_1 \dots, e_n$. Then there are functions $\kappa_1, \dots, \kappa_{n-1}$ defined on that curve with $\kappa_1, \dots, \kappa_{n-2} > 0$, so that every κ_i is $(n-1-i)$-times continuously differentiable and

$$\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ \vdots \\ e_{n-1} \\ e_n \end{pmatrix}' = \begin{pmatrix} 0 & \kappa_1 & 0 & 0 & \cdots & 0 \\ -\kappa_1 & 0 & \kappa_2 & 0 & \ddots & \vdots \\ 0 & -\kappa_2 & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 & \kappa_{n-1} \\ 0 & \cdots & \cdots & 0 & -\kappa_{n-1} & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ \vdots \\ e_{n-1} \\ e_n \end{pmatrix}.$$

κ_i is called the *i-th Frenet curvature* and the equations are called the *Frenet equations.*

Proof: We consider the components of $e_i' = \sum_{j=1}^n \langle e_i', e_j \rangle e_j$ in the Frenet n-frame. For every $i \leq n-1$, e_i lies in the linear subspace spanned by the vectors $c', c'', \dots, c^{(i)}$, so that e_i' lies in the subspace spanned by $c', \dots, c^{(i+1)}$. This is the same as the subspace spanned by $e_1, \dots, e_{i+1}$; hence

$$\langle e_i', e_{i+2} \rangle = \langle e_i', e_{i+3} \rangle = \dots = \langle e_i', e_n \rangle = 0.$$

Set $\kappa_i := \langle e_i', e_{i+1} \rangle$. Then one has $\kappa_1, \dots, \kappa_{n-2} > 0$, since by construction of the Frenet n-frame, the sign of $\langle e_i', e_{i+1} \rangle$ is for $i \leq n-2$ the same as for $\langle c^{(i+1)}, e_{i+1} \rangle$, and this is positive. The skew-symmetry of the matrix is a consequence of the equation $0 = \langle e_i, e_j \rangle' = \langle e_i', e_j \rangle + \langle e_j', e_i \rangle$. □

Consequence: A Frenet curve in $\mathbb{R}^n$ is contained in a hyperplane if and only if $\kappa_{n-1} \equiv 0$. This is equivalent to the requirement that e_n is a constant vector, which is then perpendicular to this hyperplane. Therefore κ_{n-1} is also called the *torsion.*

2.14. Lemma. The Frenet curvatures and the Frenet n-frame are invariant under all Euclidean motions.

More precisely, this means the following. Let c be a Frenet curve in $\mathbb{R}^n$, $B\colon \mathbb{R}^n \to \mathbb{R}^n$ a (proper) Euclidean motion and $B(x) = Ax + b$, where A is an orientation-preserving rotation, that is, $A^{-1} = A^T$ and $\mathrm{Det}(A) = 1$. Then $B \circ c$ is a Frenet curve as well. If $e_1, \ldots, e_n$ is the n-frame of c, then $Ae_1, \ldots, Ae_n$ is the n-frame of $B \circ c$, and the Frenet curvatures of $B \circ c$ and c are equal.

The proof of 2.14 consists of observing that on the one hand $(B\circ c)' = Ac', (B \circ c)'' = Ac'', \ldots, (B \circ c)^{(n)} = Ac^{(n)}$ while on the other hand

$$(Ae_i)' = A(e_i') = A(-\kappa_{i-1}e_{i-1}+\kappa_i e_{i+1}) = -\kappa_{i-1}(Ae_{i-1})+\kappa_i(Ae_{i+1}).$$

2.15. Theorem. (Fundamental theorem of the local theory of curves)
Let $\kappa_1, \ldots, \kappa_{n-1}\colon (a,b) \longrightarrow \mathbb{R}$ be given C^∞-functions with $\kappa_1, \ldots, \kappa_{n-2} > 0$. For a fixed parameter $s_0 \in (a,b)$, suppose we have been given a point $q_0 \in \mathbb{R}^n$ as well as an n-frame $e_1^{(0)}, \ldots, e_n^{(0)}$. Then there is a unique C^∞-Frenet-curve $c\colon (a,b) \longrightarrow \mathbb{R}^n$ parametrized by arc length and satisfying the following three conditions:

1. $c(s_0) = q_0$,
2. $e_1^{(0)}, \ldots, e_n^{(0)}$ is the Frenet n-frame of c at the point q_0,
3. $\kappa_1, \ldots, \kappa_{n-1}$ are the Frenet curvatures of c.

The assumption $\kappa_i \in C^\infty$ can be weakened as follows. Let κ_i be $(n-1-i)$ times continuously differentiable. Then the curve c is n times continuously differentiable.

Proof: We first set $F(s) = (e_1(s), \ldots, e_n(s))^T$, viewed as a matrix-valued function. Since the e_i form an orthonormal n-frame, F is automatically an orthogonal matrix with $\mathrm{Det}\,(F) = 1$. The Frenet equations are then equivalent to the matrix equation $F' = K \cdot F$, which is just a *system of linear differential equations of first order for* F, if one views the matrix

$$K(s) = \begin{pmatrix} 0 & \kappa_1(s) & 0 & 0 & \cdots & 0 \\ -\kappa_1(s) & 0 & \kappa_2(s) & 0 & \ddots & \vdots \\ 0 & -\kappa_2(s) & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 & \kappa_{n-1}(s) \\ 0 & \cdots & \cdots & 0 & -\kappa_{n-1}(s) & 0 \end{pmatrix}$$

as given. The proof of the Fundamental Theorem 2.15 is now based on the theory of solutions of differential equations of this kind, as well as on the following consideration. A matrix-valued curve $F(s)$ is orthogonal for all s if and only if the product $F' \cdot F^{-1}$ is skew-symmetric for all s and $F(s_0)$ is orthogonal for some s_0. This can also be expressed in the following way: the tangent space of the submanifold

$$\mathbf{SO}(n) \subset \mathbb{R}^{n^2}$$

at the "point" corresponding to the identity matrix is the set of skew-symmetric matrices.

Step 1: For a given matrix-valued function $K(s)$ with given initial conditions $F(s_0) = (e_1^{(0)}, \ldots, e_n^{(0)})^T$, the linear differential equation $F' = K \cdot F$ has a unique solution $F(s)$ which is defined for all $s \in (a, b)$. This follows from the existence and uniqueness theorem for solutions of linear differential equations ([**27**], Chapter XIX).

Step 2: The Frenet equations $F' = KF$ imply

$$(FF^T)' = F'F^T + F(F^T)' = F'F^T + F(F')^T = KFF^T + FF^TK^T.$$

The differential equation $(FF^T)' = K(FF^T) + (FF^T)K^T$, viewed as a differential equation for the unknown function FF^T, has a unique solution for given initial conditions $F(s_0)(F(s_0))^T = E$ (here E denotes the identity matrix). Now E is a constant function and as such surely a solution of the previous differential equation by virtue of the relation $0 = K + K^T$. This is just an expression of the skew-symmetry of the matrix K. Because of the uniqueness of the solution, one must have $FF^T = E$ on the entire interval (on which we are considering the differential equation), hence $F(s)$ is an orthogonal matrix. Because of the continuity of the determinant one also has $\mathrm{Det}\,(F) = 1$.

Step 3: The matrix $F(s)$ therefore determines a unique vector-valued function $e_1(s)$. For given initial conditions $c(s_0) = q_0$, we can find a unique curve $c(s)$ with $c' = e_1$ by setting $c(s) = q_0 + \int_{s_0}^{s} e_1(t)dt$. Moreover, from the relation $e_1' = \kappa_1 e_2 \neq 0$ and $\kappa_1 > 0$, we see that the e_2 which is defined by F must coincide with the second vector of the Frenet n-frame of c at every point, and analogously for te other e_i. Thus, $F(s)$ represents the Frenet n-frame of c at each point, and because $F' = KF$, the given functions κ_i coincide with the Frenet curvatures of c. In particular, c is a Frenet curve, which follows from the fact that

$$c' = e_1,\ c'' = \kappa_1 e_2,\ c''' = (\kappa_1 e_2)' = (-\kappa_1^2 e_1 + \kappa_1' e_2) + \kappa_1\kappa_2 e_3$$

and similarly, for every $i = 1, \ldots, n-1$,

$$c^{(i)} = (\text{linear combination of } \ e_1, \ldots, e_{i-1}) + \kappa_1 \cdot \kappa_2 \cdot \cdots \cdot \kappa_{i-1} e_i.$$

Finally, from our assumption $\kappa_1, \ldots, \kappa_{n-2} > 0$ we obtain the result that $c', c'', \ldots, c^{(n-1)}$ are linearly independent. □

In this proof we see (up to the choice of an initial point q_0) a one-to-one association $c \mapsto F$. That means that we may view a Frenet curve as a curve in the Stiefel manifold of all orthogonal n-frames, and conversely, the first row of such a vector-valued curve can be integrated to arrive back (up to translations) at the original curve in $\mathbb{R}^n$ that we started with. Note, however, that now every curve in the Stiefel manifold leads to a Frenet curve in $\mathbb{R}^n$.

2.16. Remark. (Explicit solutions)

For the *explicit* reconstruction of the curve from its Frenet curvatures, Theorem 2.15 is not necessarily the most convenient method, in contrast with the situation in two dimensions, where 2.6 is quite sufficient. Compare the Exercises 8 and 28. However, under the assumption that all κ_i are constant, i.e., if

$$K = \begin{pmatrix} 0 & \kappa_1 & 0 & 0 & \cdots & 0 \\ -\kappa_1 & 0 & \kappa_2 & 0 & \ddots & \vdots \\ 0 & -\kappa_2 & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 & \kappa_{n-1} \\ 0 & \cdots & \cdots & 0 & -\kappa_{n-1} & 0 \end{pmatrix}$$

is a constant matrix, then the differential equation $F' = KF$ can be explicitly solved by means of the exponential series

$$F(s) = \exp(sK) := \sum_{i=0}^{\infty} \frac{(sK)^i}{i!}.$$

The initial condition $F(s_0) = F_0$ is satisfied by setting $F(s_0 + s) = \exp(sK)F_0$. In order to calculate this series more precisely, one needs the eigenvalues of the symmetric matrix K^2. The special case $\kappa_{n-1} = 0$ is not interesting, since one can then view the curve as living in $\mathbb{R}^{n-1}$. Therefore we may assume that all κ_i are non-vanishing and that consequently the rank of the matrix K^2 is at least $n-1$. The rank is necessarily an even number $2m$, because the rank of K has to be even. The eigenvalues of K^2 are certain negative numbers $-\alpha_1^2, \ldots, -\alpha_m^2$, each with multiplicity two. Compare this with the case of $m = 1$ in Sections 2.6 and 2.12. The normal form of the matrix K consists of blocks of the form

$$\begin{pmatrix} 0 & \alpha_j \\ -\alpha_j & 0 \end{pmatrix}, \quad j = 1, \ldots, m,$$

along the diagonal and is zero otherwise, see [**32**], 8.16. Every such block yields an exponential series

$$\sum_{i=0}^{\infty} \frac{1}{i!} \begin{pmatrix} 0 & s\alpha_j \\ -s\alpha_j & 0 \end{pmatrix}^i = \begin{pmatrix} \cos(s\alpha_j) & \sin(s\alpha_j) \\ -\sin(s\alpha_j) & \cos(s\alpha_j) \end{pmatrix}.$$

In the coordinate system which is adapted to this normal form of K one finds after some more calculations the curves

$$c(s) = \Big(a_1 \sin(\alpha_1 s), a_1 \cos(\alpha_1 s), \ldots, a_m \sin(\alpha_m s), a_m \cos(\alpha_m s)\Big)$$

for even $n = 2m$ and

$$c(s) = \Big(a_1 \sin(\alpha_1 s), a_1 \cos(\alpha_1 s), \ldots, a_m \sin(\alpha_m s), a_m \cos(\alpha_m s), bs\Big)$$

for odd $n = 2m+1$. At this point it is easily seen that all the α_j are distinct numbers, because otherwise the curve would not be a Frenet curve. At any rate the Frenet curves with constant curvature are orbits under the action of a one-parameter group of rotations (if n is even) or of screw-motions (if n is odd). Compare this with the case $n = 3$ (the helix in 2.3 and 2.12) as well as the following case $n = 4$. For given constants a, b, α, β, the curve

$$c(t) = \big(a\cos(\alpha t), a\sin(\alpha t), b\cos(\beta t), b\sin(\beta t)\big)$$

is a Frenet curve in $I\!R^4$, in case $a, b \neq 0, \alpha \neq \beta \neq 0$. The curve c is clearly the trajectory of a particle under a rotation of $I\!R^4$, where this rotation is given to us in normal form. The curve is parametrized by arc length, provided

$$1 \quad = \quad \alpha^2 a^2 + \beta^2 b^2.$$

It then has constant Frenet curvatures $\kappa_1, \kappa_2, \kappa_3$ satisfying

$$\begin{aligned}
\kappa_1^2 &= \alpha^4 a^2 + \beta^4 b^2, \\
\kappa_1^2\kappa_2^2 &= \alpha^6 a^2 + \beta^6 b^2 - \kappa_1^4, \\
\kappa_1^2\kappa_2^2\kappa_3^2 &= \alpha^8 a^2 + \beta^8 b^2 - \kappa_1^2(\kappa_1^2 + \kappa_2^2)^2.
\end{aligned}$$

If, conversely, $\kappa_1, \kappa_2, \kappa_3$ are non-vanishing constants, one can solve this system of four equations for a, b, α, β (exercise with the hint: $-\alpha^2$ and $-\beta^2$ are eigenvalues of the matrix K^2, compare 2.12).

Remark: If one sets $a = b = 1$ and chooses $\alpha = p$, $\beta = q$ to be integers which are relatively prime, then this curve is actually closed and is known as a *torus knot of type $T_{p,q}$ in the three-sphere*. More precisely, this curve lies on what is known as the *Clifford torus*, given by the equation

$$\{(x_1, x_2, y_1, y_2) \in I\!R^4 \bigm| x_1^2 + x_2^2 = y_1^2 + y_2^2 = 1\},$$

which can be viewed as a part of the three-sphere of radius $\sqrt{2}$. Three-dimensional pictures of this can be obtained after a stereographic projection for example from the north pole $(\sqrt{2}, 0, 0, 0)$; one such picture is depicted in Figure 2.7.

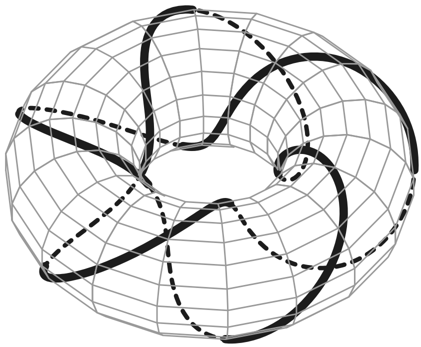


Figure 2.7. The torus knot $T_{2,5}$

2E Curves in Minkowski space $I\!R^3_1$

Up to now we have been considering Euclidean space as our ambient space. The Euclidean inner product $\langle X, Y\rangle = \sum_{i=1}^3 x_i y_i$ implies among other things that the length $||\dot{c}||$ of the tangent on a regular curve $c(t)$ never vanishes. However, there are good reasons for allowing more general "inner products" which are not necessarily positive definite. In the special theory of relativity, for example, one works in a space-time of $3+1$ dimensions, where time is viewed as a dimension. In the direction of this coordinate, the inner product has a negative sign. Similarly, one can consider three-dimensional space as a space of dimension $2+1$, treating some of the dimensions differently from the others. One can interpret this as a "toy model" for the special theory of relativity, but in physics other theories are considered which live in $2+1$ dimensions.

2.17. Definition. (Minkowski space)

The space $I\!R^3_1$ is defined as a space to be the usual three-dimensional $I\!R$-vector space consisting of vectors $\{(x_1, x_2, x_3) \mid x_1, x_2, x_3 \in I\!R\}$, but endowed with the inner product

$$\langle X, Y\rangle_1 = -x_1y_1 + x_2y_2 + x_3y_3.$$

This space is called the *Minkowski space* or *Lorentz space*. Tangent vectors are defined precisely as in the case of Euclidean space $I\!R^3$. A vector X is said to be:

space-like, if	$\langle X, X\rangle_1 > 0$,
time-like, if	$\langle X, X\rangle_1 < 0$,
light-like or *isotropic* or a *null vector*, if	$\langle X, X\rangle_1 = 0$, but $X \neq 0$.

The set of all null vectors of $I\!R^3_1$ forms what is called the *light-cone*[2], in coordinates:

$$\{(x_1, x_2, x_3) \mid x_1^2 = x_2^2 + x_3^2, x_1 \neq 0\}.$$

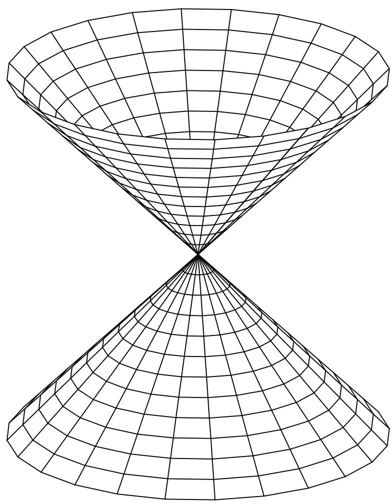


Figure 2.8. Light-cone in Minkowski space with vertical x_1-axis

In $I\!R^3_1$, the rules of calculus remain the same as in Euclidean space $I\!R^3$, so that we can speak of *immersions* or *regular curves* just as in the Euclidean case.

[2]If the inner product $\langle X, X\rangle_1$ is written in the form $-\gamma^2 t^2 + x_2^2 + x_3^2$ where t denotes the time parameter and γ the velocity of light, then the light cone represents the propagation of light in the (x_2, x_3)-plane.

2.18. Definition. A regular curve $c\colon I \to I\!R^3_1$ is called

space-like, if	$\langle \dot c, \dot c\rangle_1 > 0$ everywhere,
time-like, if	$\langle \dot c, \dot c\rangle_1 < 0$ everywhere,
light-like or *isotropic* or a *null curve*, if	$\langle \dot c, \dot c\rangle_1 = 0$ everywhere.

EXAMPLE: The hyperbola $x_1^2 = x_2^2 + 1, x_3 = 0$ is space-like. This can be seen using the parametrization $c(t) = (\cosh t, \sinh t, 0)$. Since $\dot c(t) = (\sinh t, \cosh t, 0)$ which implies that $\langle \dot c, \dot c\rangle_1 = 1$, the parameter t is actually the arc length.

Similarly, the hyperbola $x_1^2 = x_2^2 - 1, x_3 = 0$ is time-like with a similar parametrization $c(t) = (\sinh t, \cosh t, 0)$. The line $c(t) = (t, t, 0)$ is isotropic. This line lies (with the exception of the point for $t = 0$) entirely on the light-cone.

2.19. Lemma. A regular curve $c\colon I \to I\!R^3_1$ which is space-like or time-like everywhere can be parametrized by arc length in the sense that $\langle \dot c, \dot c\rangle_1 = \pm 1$ is valid everywhere. For a curve which is everywhere light-like this is not possible in general, but one can parametrize a light-like line in such a way that $\ddot c = 0$. These parametrizations are not unique, but only determined up to a translation $t \mapsto at + b$. The parameter is therefore also referred to as an *affine parameter.*

The proof of this statement for space-like or time-like curves is similar to that given in 2.2. For isotropic lines the statement is quite trivial.

In order to get *derivative equations* of *Frenet type*, we first observe that in $I\!R^3_1$ a (modified) vector product of two vectors A and B can be defined, by requiring the relation

$$\langle A \times B, C\rangle_1 = \mathrm{Det}(A, B, C)$$

for all C. In the same way one can define three-frames as follows. For two vectors e_1 and e_2, for which $\langle e_i, e_i\rangle_1 = \pm 1$ and $\langle e_1, e_2\rangle_1 = 0$, a third is defined by $e_3 := e_1 \times e_2$, and these three vectors form an orthonormal three-frame. If we define $\epsilon, \eta \in \{1, -1\}$ by $\langle e_1, e_1\rangle_1 = \epsilon,\quad \langle e_2, e_2\rangle_1 = \eta$, then it follows that $\langle e_3, e_3\rangle_1 = -\epsilon\eta$. Hence each vector X can be uniquely decomposed into its three components

$$X = \epsilon\langle X, e_1\rangle_1 e_1 + \eta\langle X, e_2\rangle_1 e_2 - \epsilon\eta\langle X, e_3\rangle_1 e_3.$$

2.20. Theorem. (Frenet equations in Minkowski space)
Let c be a space-like or time-like curve, which we assume is parametrized by arc length and satisfies $\langle c'', c'' \rangle_1 \neq 0$. Then this curve induces a Frenet three-frame $e_1 = c', e_2 = c''/\sqrt{|\langle c'', c'' \rangle_1|}, e_3 = e_1 \times e_2$, for which the following *Frenet equations* hold (here ϵ and η are defined as above):

$$\begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}' = \begin{pmatrix} 0 & \kappa\eta & 0 \\ -\kappa\epsilon & 0 & -\tau\epsilon\eta \\ 0 & -\tau\eta & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}.$$

The quantities defined by this relation, namely

$$\kappa = \langle e_1', e_2 \rangle_1 \quad \text{and} \quad \tau = \langle e_2', e_3 \rangle_1,$$

are called the *curvature* and *torsion* of the curve c.

PROOF: As in 2.8, we only need to calculate the components of e_1', e_2', e_3' in the Frenet three-frame, for example,

$$e_1' = c'' = \eta \langle c'', e_2 \rangle_1 e_2 = \eta\kappa e_2,$$
$$\langle e_2', e_1 \rangle_1 = -\langle e_1', e_2 \rangle_1 = -\kappa,$$
$$\langle e_3', e_2 \rangle_1 = -\langle e_2', e_3 \rangle_1 = -\tau.$$

2.21. Example. (Curves with constant curvature and torsion)
The following plane curves have constant curvature:

$$\begin{aligned} c_1(t) &= (0, \cos t, \sin t); \\ c_2(t) &= (\cosh t, \sinh t, 0); \\ c_3(t) &= (\sinh t, \cosh t, 0). \end{aligned}$$

Here, c_1 and c_2 are space-like, while c_3 is time-like. Space curves with constant curvature and constant torsion can be obtained as the trajectories of a particle under a helicoidal motion in Minkowski space. The corresponding rotation matrices are discussed in 3.42. One then can add a translation in the direction of the axis of rotation. Depending on a constant a one gets in this manner the following curves, all with constant curvature and torsion:

$$\begin{aligned} c_4(t) &= (at, \cos t, \sin t), \\ c_5(t) &= (\cosh t, \sinh t, at), \\ c_6(t) &= (\sinh t, \cosh t, at). \end{aligned}$$

Remark: In the n-dimensional *pseudo-Euclidean space* $I\!R^n_k$ with an analogous inner product

$$\langle X, Y\rangle_k = -\sum_{i=1}^{k} x_i y_i + \sum_{j=k+1}^{n} x_j y_j$$

and a curve $c(s)$ (parametrized by the arc length) an analogous Frenet-n-frame $e_1, \ldots, e_n$ with $\epsilon_i := \langle e_i, e_i\rangle_k \in \{1, -1\}$ leads to the following Frenet matrix:

$$K = \begin{pmatrix} 0 & \kappa_1\epsilon_2 & 0 & 0 & \cdots & 0 \\ -\kappa_1\epsilon_1 & 0 & \kappa_2\epsilon_3 & 0 & \ddots & \vdots \\ 0 & -\kappa_2\epsilon_2 & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 & \kappa_{n-1}\epsilon_n \\ 0 & \cdots & \cdots & 0 & -\kappa_{n-1}\epsilon_{n-1} & 0 \end{pmatrix}$$

The proof is essentially the same as the proof of Theorem 2.13 (the Euclidean case). The only change concerns the modified representation of a vector in an orthonormal basis as follows:

$$e_i' = \sum_{j=1}^{n} \epsilon_j \langle e_i', e_j\rangle_k e_j.$$

To see this equation, one can take the inner product of both sides of the equation with a fixed e_m. We will come back to these pseudo-Euclidean spaces $I\!R^n_k$ in Chapter 7.

2F The global theory of curves

2.22. Definition. (Closed curve)
A (regular) curve $c\colon [a, b] \to I\!R^n$ is called *closed*, if there is a (regular) curve $\widetilde{c}\colon I\!R \to I\!R^n$ with $\widetilde{c}|_{[a,b]} = c$ and $\widetilde{c}(t + b - a) = \widetilde{c}(t)$ for all $t \in I\!R$, where in particular $c(a) = c(b)$ and $c'(a) = c'(b)$. The lifted curve $\widetilde{c}$ is also called *periodic*. A closed curve c is said to be *simply closed*, if $c|_{[a,b)}$ is injective, i.e., if there are no *double points* for which $c(t_1) = c(t_2)$ for some $a \le t_1 < t_2 < b$. Alternatively, one can define a closed curve [or simply closed curve] as an immersion [or embedding] of the circle S^1 in $I\!R^n$.

The global theory of curves studies the properties of closed curves, in particular their curvature properties, for example with an eye toward their total (i.e., integrated) curvature or torsion. The *total curvature* of a closed curve is defined as the integral

$$\int_a^b \kappa(t)||\dot{c}(t)||dt = \int_0^L \kappa(s)ds,$$

where L is the total length of the curve. Similarly one has the total torsion of a space curve, in case it is a Frenet curve. Note that for a plane curve the curvature κ has a sign, and hence so does the total curvature, while for a Frenet curve in three-space (or higher dimensions) the total curvature is by definition positive. This difference is exemplified in the results 2.28, 2.32 and 2.34 which follow.

2.23. Lemma. (Curvature in polar coordinates)

Let $c\colon [a,b] \to I\!R^2$ be a regular (C^2-)curve with Frenet two-frame $e_1(t)$, $e_2(t)$, and let $e_1(t) = \big(\cos(\varphi(t)), \sin(\varphi(t))\big)$ be the representation in local polar coordinates. Then we have

$$\kappa = \frac{d\varphi}{ds} = \frac{d\varphi}{dt}\cdot\frac{dt}{ds} = \frac{\dot{\varphi}(t)}{||\dot{c}(t)||}.$$

Proof: From the representation of $e_1(t)$ it follows that $e_2(t) = \big(-\sin(\varphi(t)), \cos(\varphi(t))\big)$, and with this, in virtue of the Frenet equations, that also $\kappa e_2 = \frac{de_1}{ds} = \frac{de_1}{dt}\cdot\frac{dt}{ds} = \dot{\varphi}\big(-\sin\varphi, \cos\varphi\big)\frac{dt}{ds}$. □

Consequence: As long as φ is a differentiable function of the parameter t, one has for the total curvature the relation

$$\int_a^b \kappa(t)||\dot{c}(t)||dt = \int_a^b \dot{\varphi}(t)dt = \varphi(b) - \varphi(a).$$

Because of the potential many-valuedness of the angle φ, one must verify separately that this relation holds in the large. In particular, it does not necessarily follow for curves which satisfy $c(a) = c(b)$ that also $\varphi(b) = \varphi(a)$. The same problem also occurs in the theory of a function of a complex variable, because of the many-valuedness of the natural logarithm. Indeed, the logarithm of $e^{i\theta} = \cos\theta + i\sin\theta$ can be any number $\theta + 2k\pi$ for an arbitrary integral value of k.

2.24. Theorem and Definition. (Polar angle function, winding number)
Let $\gamma\colon [a,b] \to I\!R^2 \setminus \{0\}$ be a continuous curve. Then there is a continuous function $\varphi\colon [a,b] \to I\!R$ with

$$\gamma(t) = ||\gamma(t)|| \left(\cos(\varphi(t)), \sin(\varphi(t))\right).$$

The difference $\varphi(b) - \varphi(a)$ is independent of the choice of the function φ. The function φ is called the *polar angle function.* If γ is differentiable, then φ is also, because outside the origin the transition from cartesian coordinates to polar coordinates is differentiable.

CONSEQUENCE: In case γ is a closed curve the value of $W_\gamma = \frac{1}{2\pi}\big(\varphi(b) - \varphi(a)\big)$ is an integer. This number is called the *winding number* of the curve γ.

PROOF: First of all, it is quite clear that φ is determined for all points of any particular half-plane $H = \{x \in I\!R^2 \mid \langle x, x_0\rangle > 0\}$, once the value has been fixed at a single point. Because γ is uniformly continuous, there is a partition $a = t_0 < t_1 \cdots < t_n = b$ of the interval $[a,b]$, such that for every subinterval the image set of the part $\gamma|_{[t_i,t_{i+1}]}$ is completely contained in such a half-plane. For a given initial value $\varphi(a)$, the function φ is uniquely determined on $[t_0,t_1]$ by the fact that it is continuous. Next, if φ is continuous on the whole interval $[t_0,t_i]$, there is a unique continuous extension to $[t_i,t_{i+1}]$ and hence to $[t_0,t_{i+1}]$. Arguing inductively, we see that φ is uniquely determined by the initial value $\varphi(a)$. Of course, the choice of $\varphi(a)$ is quite arbitrary. If however φ and $\widetilde{\varphi}$ are two such continuous functions, then their difference is an integral function, multiplied by 2π. Finally, an integral continuous function is necessarily constant; hence $\varphi - \widetilde{\varphi}$ is constant, implying that $\widetilde{\varphi}(b) - \widetilde{\varphi}(a) = \varphi(b) - \varphi(a)$. $\square$

2.25. Definition. (Rotation index)
Let $c\colon [a,b] \to I\!R^2$ be a regular closed curve. Then the *rotation index* U_c of c is defined as the winding number $W_{\dot c}$ of the tangent $\dot c$, where we view $\dot c\colon [a,b] \to I\!R^2 \setminus \{0\}$ as a continuous curve by gluing the tangent vector at every point of the curve at the origin of $I\!R^2$.

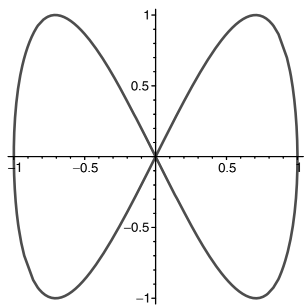


Figure 2.9. Curve with rotation index $U = 0$

2.26. Corollary. The rotation index of a closed and regular plane curve is equal to its total curvature divided by 2π.

PROOF: According to 2.23 and 2.24 we have

$$2\pi U_c = 2\pi W_{\dot c} = \varphi(b) - \varphi(a) = \int_a^b \dot\varphi(t)dt = \int_a^b \kappa(t)||\dot c(t)||dt,$$

where $\dot c = r(\varphi(t))\big(\cos(\varphi(t)), \sin(\varphi(t))\big)$. □

REMARK: The winding number is a homotopy invariant, i.e., closed curves in $\mathbb{R}^2 \setminus \{0\}$ which are homotopic as closed curves have the same winding numbers. Consequently, the rotation index is a regular homotopy invariant, meaning that two regular closed curves which are regularly homotopic have the same rotation index. The notion *regularly homotopic* used here means that the homotopy (i.e., the corresponding one-parameter family of curves) consists exclusively of regular curves. The Whitney–Graustein theorem[3] states that even the converse of this statement holds: *If two regular closed curves have the same rotation index, then they are regularly homotopic to each other.*

2.27. Lemma. Let $e\colon A \to \mathbb{R}^2 \setminus \{0\}$ be continuous, and let $A \subset \mathbb{R}^2$ be star-like with respect to x_0, i.e., for every $x \in A$ the segment $\overline{x_0 x}$ lies completely in A. Then there is a continuous polar angle function $\varphi\colon A \to \mathbb{R}$ with $e(x) = ||e(x)||\big(\cos(\varphi(x)), \sin(\varphi(x))\big)$.

[3] H. Whitney, *On regular closed curves in the plane*, Compositio Math. **4**, 276–284 (1937).

PROOF: We choose $\varphi(x_0)$ to be some fixed value. Then the restriction of e on the segment $x_0 + t(x - x_0)$ is a continuous curve with t as parameter, where $t \in [0, 1]$. According to 2.24 φ is then uniquely defined along the segment $\overline{x_0 x}$ as a continuous polar angle function. From the assumption that A is star-like, we conclude that φ is uniquely defined on all of A. It is sufficient then to verify that φ is continuous on compact subsets which are star-like with respect to x_0. We may assume that $x_0 = 0$ for this. Since e and $e/||e||$ are uniformly continuous, there is a $\delta > 0$ such that – as in the proof of 2.24 – $e(x)$ and $e(y)$ always lie in an open half-plane with respect to 0 (in other words, these two points are never antipodal), provided that $||x - y|| < \delta$. Let $\{x_n\}$ be a convergent sequence, converging to a point $x \in A$. Arguing by contradiction, we assume that $\liminf_{n\to\infty} |\varphi(x_n) - \varphi(x)| \geq 2\pi$. We may also assume here that $||x_n - x|| < \delta$ for all n and that consequently $||tx_n - tx|| < \delta$ for $0 \leq t \leq 1$. For fixed n, we consider the distance $|\varphi(tx_n) - \varphi(tx)|$ as a function of t. This function is certainly continuous in t, has the value 0 at $t = 0$, and is larger than $\frac{3}{2}\pi$ for $t = 1$ and for sufficiently large n. On the other hand, the two points $e(tx_n)$ and $e(tx)$ are never antipodal. Hence this function can never attain the value of π, yielding a contradiction to our assumption. $\square$

2.28. Theorem. (Theorem on turning tangents)
Let $c\colon [a, b] \to \mathbb{R}^2$ be a simply closed regular (C^2-)curve. Then we have

$$\frac{1}{2\pi}\int_a^b \kappa(t)||\dot{c}(t)||dt = U_c = \pm 1.$$

PROOF (following H. Hopf[4]): Surely there is a tangent such that the curve lies completely on one side of this tangent. By choosing an appropriate coordinate system we may further assume that $c(t) = (x(t), y(t))$ with $y(a) = y(b) = 0, y(t) \geq 0$ for all t.

We then define

$$A = \{(s, t) \in \mathbb{R}^2 \mid a \leq s \leq t \leq b\}$$

[4] *Über die Drehung der Tangenten und Sehnen ebener Kurven*, Compositio Math. **2**, 50–62 (1935).

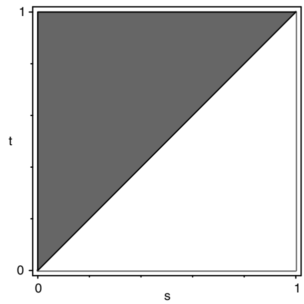


Figure 2.10. The set A (for $a = 0, b = 1$)

as well as $e\colon A \to \mathbb{R}^2 \setminus \{0\}$ by

$$e(s,t) = \begin{cases} \dfrac{c(t) - c(s)}{||c(t) - c(s)||} & \text{in case} \quad s \neq t \text{ and } (s,t) \neq (a,b), \\ \dfrac{\dot{c}(t)}{||\dot{c}(t)||} & \text{in case} \quad s = t, \\ -\dfrac{\dot{c}(a)}{||\dot{c}(a)||} & \text{in case} \quad (s,t) = (a,b). \end{cases}$$

Since the curve is simply closed, one has $c(t) \neq c(s)$ for all $t \neq s$ except $(s,t) = (a,b)$. It follows that e is well-defined. The continuity of e follows from the continuous differentiability of c. This is easily verified by passing to the limit from the secant to the tangent. Clearly $e(t,t)$ is a unit tangent of the curve $c(t)$. According to 2.27 there exists a polar angle function $\varphi\colon A \to \mathbb{R}$ with $e(s,t) = (\cos\varphi(s,t), \sin\varphi(s,t))$ and $\varphi(a,a) = 0$. The function $\varphi(t) := \varphi(t,t)$ is then the polar angle function along the curve φ, hence it is differentiable in t, and by 2.26 we get

$$\frac{1}{2\pi}\int_a^b \kappa(t)||\dot{c}(t)||dt = \frac{1}{2\pi}\int_a^b \dot{\varphi}(t)dt = \frac{1}{2\pi}\big(\varphi(b,b) - \varphi(a,a)\big).$$

On the other hand, $\varphi(a,b) - \varphi(a,a) = \pi$, in case $\dot{x}(a) > 0$ (otherwise $= -\pi$), and also $\varphi(b,b) - \varphi(a,b) = \pi$, in case $\dot{x}(a) > 0$ (otherwise $= -\pi$). This is verified upon consideration of the polar angle for the family of all secants $c(t) - c(a)$ on the one hand, and $c(b) - c(s)$ on the other. In sum, we have that $\varphi(b,b) - \varphi(a,a)$ is either $+2\pi$ or -2π. $\Box$

2.29. Corollary. The *total absolute curvature* $\int |\kappa| ds$ of a simply closed and regular plane curve satisfies the inequality

$$\int_a^b |\kappa(t)| \cdot ||\dot{c}(t)|| dt \geq 2\pi,$$

with equality if and only if the curvature does not change its sign.

This leads to the question of what the condition $\kappa \geq 0$ or $\kappa \leq 0$ means geometrically for a closed curve, and this question in turn leads us to the consideration of convex curves.

2.30. Definition. (Convex)
A simply closed plane curve is called *convex*, if the image set of the boundary is a convex subset $C \subset \mathbb{R}^2$. The convexity of a subset C is defined in the usual way, namely, for any two points contained in C, also the segment joining these two points is completely contained in C.

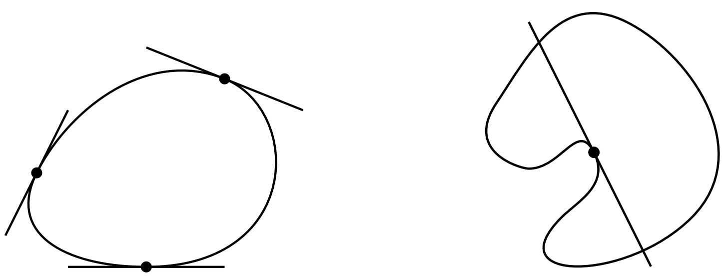


Figure 2.11. convex and non-convex curve

2.31. Theorem. (Characterization of convex curves)

For a simply closed and regular plane curve c whose image is the boundary of a compact connected set $C \subset \mathbb{R}^2$, the following conditions are equivalent:

(1) The curve c is convex (i.e., C is convex).

(2) Every line meets the curve, if at all, either in a segment (which may also degenerate to a single point) or in two points.

(3) For every tangent of the curve, the image of the curve (and also of the set C) always lies on one side of that tangent, compare Figure 2.11.

(4) The curvature of c does not change its sign.

PROOF: (1) $\Rightarrow$ (2): Let g be a line; then $g \cap C$ is a compact subset of C, which is an interval because of the convexity. In case $g \cap C$ does not contain any inner points of C, then $g \cap C$ lies completely in the curve c as the boundary of C. In case $g \cap C$ does contain inner points of C, then $g \cap C$ can only have the two endpoints of the interval in common with c. Indeed, if some subinterval of $g \cap C$ were contained in the boundary of C, then by rotating this line we would have a segment which intersected C in a non-connected set, a contradiction to the convexity of C.

(2) $\Rightarrow$ (3): We assume (2) and argue by contradiction. Suppose that C was not on one side of the tangent T of c at a certain point p. If $\kappa(p) \neq 0$, then locally the curve is contained in one of the half planes determined by T. If there is another point q on the curve in the other half plane, then T meets the curve in at least two other points because the two arcs on the curve near p must be contained with q. This contradicts our assumption (2). The same holds if p is an isolated zero of κ without a change of sign. If p is an isolated zero of κ with a change of sign (that is to say, if p is an inflection point), then by a slight rotation of T around p we obtain a line that meets the curve in at least three isolated points. Again this contradicts (2). If κ vanishes on a (maximal) interval around p, then c is a straight line segment, and by (2) the tangent T contains no other points of c. Hence c (and also C) lies in one of the half planes determined by T. This contradicts our assumption. Any other point p with $\kappa(p) = 0$ is an accumulation point of the types we already discussed. However, our assumption carries over to nearby points. This leads to a contradiction in any case.

(3) $\Rightarrow$ (4): Again we argue by contradiction. We assume that $\kappa(p) = 0$ holds and that κ has a change of sign at that point. Here we explicitly allow κ to vanish on a segment (in which case the curve is a segment at the corresponding points). If we rotate the tangent at p appropriately, then we get a line which contains at least three isolated points of c,

namely p and one point each on either side of the tangent. Hence c cannot lie on one side of the tangent.

(4) $\Rightarrow$ (1): If C is not convex, then there is a line g for which $g \cap C$ has at least two components, which we can describe as intervals $[x_1, x_2]$ and $[x_3, x_4]$ with $x_1 < x_2 < x_3 < x_4$. On the various segments connecting these four points there are points on the curve which have a maximal distance from g, in fact four of these. At these points the unit normal vectors are perpendicular to g, hence there must be two parallel and oriented unit normal vectors at two different points, for which the unit normal in between is not constant. We now assume moreover that $\kappa \geq 0$ and derive a contradiction to this. By 2.23 and 2.24, κ can be viewed as the derivative φ' of a polar angle function φ. Because of our assumption $\kappa \geq 0$, this is a nondecreasing function, ranging from 0 to 2π according to the theorem on turning tangents. We then consider the unit tangent (and similarly the oriented unit normal) as a map from S^1 to S^1. Since φ is (monotonically) increasing, this map has the property that the inverse image of a point is always connected. The condition $\kappa > 0$ would imply under these circumstances that φ is strictly increasing, as otherwise distinct connected arcs would have the same image. But this is a contradiction to what we have said above. Hence c is convex. $\square$

2.32. Corollary. (Total absolute curvature)
The *total absolute curvature* $\int |\kappa| ds$ of a given closed and regular plane curve fulfills the inequality

$$\int_a^b |\kappa(t)| \cdot ||\dot{c}(t)|| dt \geq 2\pi,$$

in which equality holds if and only if the curve is simple and convex.

For (simply closed) convex curves the equality $\int |\kappa| ds = 2\pi$ is clear by the results 2.29 and 2.31. For non-convex curves one sees the validity of the inequality by comparing the curve with the boundary of its convex hull (the convex hull is defined to be the smallest convex set containing a given set). This boundary is then a simply closed convex curve (to be sure, only C^1 and piecewise C^2), whose total absolute curvature cannot exceed that of the given curve. The exceptional points, where the curve is not C^2, can be approximately “smoothed”.

Then one can apply 2.29 and 2.31 to this boundary curve. Equality can only hold if the given curve happens to coincide with the boundary curve, which means precisely that it is convex.

2.33. Theorem. (Four vertex theorem)
A simply closed, regular and convex plane curve which is of class C^3 has at least four local extremal points for its curvature κ (such a point is referred to as a *vertex*).

PROOF: If κ is constant, there is nothing to prove. Hence we may assume that κ is not constant. Local extrema of κ can be recognized as points where $\kappa' = 0$ and κ' changes sign. Here it is possible that κ might be constant on an interval near this extremal point. First, we know that κ takes on an absolute minimum and maximum on the compact interval $[a, b]$ (resp. S^1). At such a point, certainly one has $\kappa' = 0$. Suppose, as we may without restriction of generality, that $\kappa(0)$ is a minimum, and let $\kappa(s_0)$ denote the maximum. Suppose the curve $c\colon [0, L] \to \mathbb{R}^2$ is parametrized by arc length, with $c(0) = c(L)$. The coordinate system (x, y) in the plane may be chosen in such a way that the x-axis contains the two points $c(0)$ and $c(s_0)$, so that we can write $c(s) = (x(s), y(s))$ with $y(0) = y(s_0) = 0$. The curve meets the x-axis at no other point, because according to 2.31 it would then have the entire segment $\overline{c(0)c(s_0)}$ in common with the x-axis, using the convexity. This would imply $\kappa(0) = \kappa(s_0) = 0$, which is a contradiction to κ being non-constant. This in turn implies that $y(s)$ only changes sign in the points $s = 0$ and $s = s_0$.

We now argue by contradiction. Assume that $c(0)$ and $c(s_0)$ are the only vertex points on c. Then κ' changes sign only at $s = 0$ and $s = s_0$, and the function $\kappa'(s)y(s)$ doesn't change its sign at all. The Frenet equations for x tell us that

$$e_1 = (x', y'), \quad e_2 = (-y', x'), \quad (x'', y'') = e_1' = \kappa e_2 = \kappa(-y', x'),$$

from which it in particular follows that $x'' = -\kappa y'$. Applying integration by parts, we get

$$\int_0^L \kappa'(s)y(s)ds = \kappa y\Big|_0^L - \int_0^L \kappa(s)y'(s)ds$$

$$= \int_0^L x''(s)ds = x'(L) - x'(0) = 0.$$

Here we have used the closedness of the curve, in other words the fact that $y(0) = y(L), x'(0) = x'(L)$. The integrand $\kappa' y$ on the left-hand side does not, however, change its sign. If the integral is to vanish anyway, then it must vanish identically, which implies $\kappa' \equiv 0$, which contradicts κ being non-constant.

Thus the assumption led us to a contradiction and must be false. Hence there is a third zero of κ' with a change of sign. Because of the periodicity of κ', the number of changes in sign altogether cannot be an odd number, hence there must also be a fourth such point. This theorem is actually also true for non-convex simply closed plane curves, although the proof has to be modified in that case.[5] □

2.34. Theorem. (Total curvature of space curves, W. Fenchel 1928/29)
For every closed and regular space curve $c\colon [a, b] \to \mathbb{R}^3$ of total length l one has the inequality

$$\int_0^l \kappa(s)ds = \int_a^b \kappa(t)||\dot{c}(t)||dt \geq 2\pi,$$

with equality if and only if the curve is a convex, simple plane curve.

PROOF (following H. Liebmann 1929[6]): Let c be parametrized by arc length. Then for the spherical curve c' one has

$$||(c')'||ds = \kappa ds$$

hence the element of arc length coincides with κds everywhere along the curve c where c' is regular. Note that s is *not* the arc length on c' and that c' need not be regular everywhere. The absolute curvature

[5] See L. Viëtoris, *Ein einfacher Beweis des Vierscheitelsatzes der ebenen Kurven*, Archiv d. Math. **3**, 304–306 (1952) and S. B. Jackson, *Vertices of plane curves*, Bull. Amer. Math. Soc. **50**, 564–578 (1944).

[6] *Elementarer Beweis des Fenchelschen Theoremes über die Krümmung geschlossener Raumkurven*, Sitzungsber. Preußische Akad. Wiss., Physik.-Math. Klasse 1929, 392–393; see also R. A. Horn, *On Fenchel's theorem*, Amer. Math. Monthly **78**, 380–381 (1971).

$\int_0^l \kappa(s)ds$ is therefore nothing more than the total length of c' as a spherical curve. Here one must count those parts of c' which are covered several times with the corresponding multiplicities.

Because of 2.32, which holds for plane curves, it is sufficient to show the following assertion:

The length L of the spherical curve c' is strictly greater than 2π, if c is a closed curve not lying in any plane.

In what follows we use the elementary geometric fact that the length of every curve joining two points on the sphere is greater than or equal to the length of the smaller part of the great circle joining these two points, with equality holding if and only if the curve is that arc itself. We denote by $dist(A, B)$ the (oriented) arc length distance on the curve c', and by $d(A, B)$ the spherical distance, so we have $dist(A, B) \geq d(A, B)$ in any case with equality only for smaller parts of great circles. First, for a coordinate function $x'(s)$ of c' we have the equation

$$\int_0^l x'(s)ds = x(l) - x(0) = 0,$$

from which it follows that the image of c' is at any rate intersected by the great circle given by $x = 0$. By rotating the coordinate system, one sees that the same must hold for every great circle on the sphere. More precisely, it follows that the image of c' is not contained in a closed hemisphere, unless c' is itself a great circle.

Now let A and B be two points on the curve c' which are antipodal on this curve, i.e., the length from A to B is equal to the length from B to A (running along the curve in the same direction):

$$dist(A, B) = dist(B, A) = \tfrac{L}{2}.$$

We assume here that inside the sphere A and B are connected by an arc of a great circle of length $\leq \pi$. If we have $d(A, B) = \pi$, then the length L of c' is greater than or equal to 2π, with equality holding if and only if c consists of two halves of great circles. These arcs must then be the two halves of a single great circle, since otherwise c' would not be intersected by every great circle. Thus c is a plane curve, which contradicts our assumption above.

It remains to consider the case in which $d(A, B) < \pi$. In this case there is a great circle G in a symmetric position in the sense that the plane spanned by G is perpendicular to the line through the origin and the midpoint between A and B on a great circle through them. By our assumption G meets the curve c' at a certain point P. By construction this point P does not lie in the smaller part of the great circle through A and B. Because of the symmetric position taken by G, one has $d(A,P)+d(P,B)=d(A,P)+d(P,-A)=d(A,-A)=\pi$. On the other hand we have $dist(A,P)>d(A,P)$ or $dist(P,B)>d(P,B)$ because by assumption c' is not contained in a great circle. Consequently we have

$$\tfrac{L}{2}=dist(A,P)+dist(P,B)>d(A,P)+d(P,B)=\pi,$$

and the assertion above follows. This proves 2.34. This theorem is in fact more generally true for all closed curves in $I\!R^n$, with essentially the same proof. $\square$

We mention without proof the following relation between various characteristic numbers for closed curves in the plane. Let $c\colon [a,b]\to I\!R^2$ be closed, and let D denote the number of double points, W the number of inflection points (i.e., points with $\kappa=0$). Moreover let N^+ (resp. N^-) be the number of double tangents, so that near the points of contact, the curve lies on the same side (resp. on opposite sides) of the double tangent.

2.35. Theorem. (Fr. Fabricius-Bjerre[7])
For every generic and closed plane curve, one has the equality

$$N^+=N^-+D+\frac{1}{2}W.$$

Here the term "generic" means that (i) the curve has only ordinary double points and double tangents (no three-fold or higher order points or tangents), (ii) the tangents at these double points are linearly independent, (iii) for all points with $\kappa=0$ we have $\kappa'\neq 0$, and (iv) no double tangent of the curve has a point of contact at an inflection point.

For the example depicted in Figure 2.9 we have $W=N^+=2$, $D=1$, and $N^-=0$.

[7] *On the double tangents of plane closed curves*, Mathematica Scandinavica **11**, 113–116 (1962).

Exercises

1. The curvature and the torsion of a Frenet curve $c(t)$ in $\mathbb{R}^3$ are given by the formulas

$$\kappa(t) = \frac{||\dot{c} \times \ddot{c}||}{||\dot{c}||^3} \quad \text{and} \quad \tau(t) = \frac{\mathrm{Det}(\dot{c}, \ddot{c}, \dddot{c})}{||\dot{c} \times \ddot{c}||^2}$$

for an arbitrary parametrization. For a plane curve we have $\kappa(t) = \mathrm{Det}(\dot{c}, \ddot{c})/||\dot{c}||^3$.

2. At every point p of a regular plane curve c with $c''(p) \neq 0$ (or, equivalently, $\kappa(p) \neq 0$) there is a parabola which has a point of third order contact with the curve at p. The point of contact is the vertex of the parabola if and only if $\kappa'(p) = 0$.
Hint: There is a two-parameter family of parabolas which have a given point as a point of contact on a given line. If we choose this line to be the tangent of a given curve at p, then by prescribing $\kappa(p)$ and $\kappa'(p)$, a unique parabola of the two-dimensional family is determined. The curvature of the parabola given by $x \mapsto (x, \frac{a}{2}x^2)$ calculates by Exercise 1 to $\kappa(x) = a(1 + a^2x^2)^{-3/2}$. This implies $\kappa'(x) = \frac{d\kappa}{dx} \cdot \frac{dx}{ds} = -3ax\kappa^2$. Consequently one can express a and x by κ und κ': $a = \kappa\left(1 + \frac{\kappa'^2}{9\kappa^4}\right)^{3/2}$ and $x = -\frac{\kappa'}{3a\kappa^2}$.

3. The evolute $\gamma(t) = c(t) + \frac{1}{\kappa(t)}e_2(t)$ of a curve $c(t)$ is regular precisely where $\kappa' \neq 0$. The tangent to γ at the point $t = t_0$ intersects the curve c at $t = t_0$ perpendicularly.

4. A regular curve between two points p, q in $\mathbb{R}^n$ with minimal length is necessarily the line segment from p to q. Hint: Consider the Schwarz inequality $\langle X, Y \rangle \leq ||X|| \cdot ||Y||$ for the tangent vector and the difference vector $p - q$.

5. If all tangent vectors to the curve $c(t) = (3t, 3t^2, 2t^3)$ are drawn from the origin, then their endpoints are on the surface of a circular cone with axis the line $x - z = y = 0$.

6. If a circle is rolled along a line (without friction), then a fixed point on that circle has as its trajectory the so-called *cycloid*, see Figure 2.12. Find the equation or a parametrization for the cycloid.

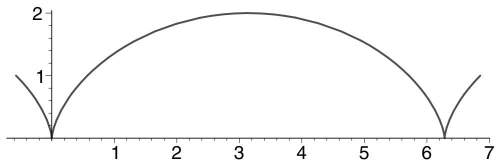


Figure 2.12. Cycloid

7. Calculate explicitly the parametrization of the plane curve which has $\kappa(s) = s^{-1/2}$. Hint: 2.6.
8. The Frenet two-frame of a plane curve with given curvature function $\kappa(s)$ can be described by the exponential series for the matrix
$$\begin{pmatrix} 0 & \int_0^s \kappa(t)dt \\ -\int_0^s \kappa(t)dt & 0 \end{pmatrix}.$$
It follows that
$$\begin{pmatrix} e_1(s) \\ e_2(s) \end{pmatrix} = \sum_{i=0}^{\infty} \frac{1}{i!} \begin{pmatrix} 0 & \int_0^s \kappa \\ -\int_0^s \kappa & 0 \end{pmatrix}^i.$$
9. Let a plane curve be given in polar coordinates (r, φ) by $r = r(\varphi)$. Using the notation $r' = \frac{dr}{d\varphi}$, the arc length in the interval $[\varphi_1, \varphi_2]$ can be calculated as $s = \int_{\varphi_1}^{\varphi_2} \sqrt{r'^2 + r^2}d\varphi$, and the curvature is given by
$$\kappa(\varphi) = \frac{2r'^2 - rr'' + r^2}{(r'^2 + r^2)^{3/2}}.$$
10. Calculate the curvature of the curve given by $r(\varphi) = a\varphi$ (a constant), the so-called *Archimedean spiral*, see Figure 2.13, left.
11. Show the following: (i) The length of the curve given in polar coordinates by $r(t) = \exp(t), \varphi(t) = at$ with a constant a (the *logarithmic spiral*) in the interval $(-\infty, t]$ is proportional to the radius $r(t)$, see Figure 2.13, right. (ii) The position vector of the logarithmic spiral has a constant angle with the tangent vector.
12. In plane polar coordinates (r, φ), let a curve be given by $r = \cos(2\varphi), 0 \leq \varphi \leq 2\pi$. Check whether this curve is regular, and if so, calculate its rotation index and the total curvature.

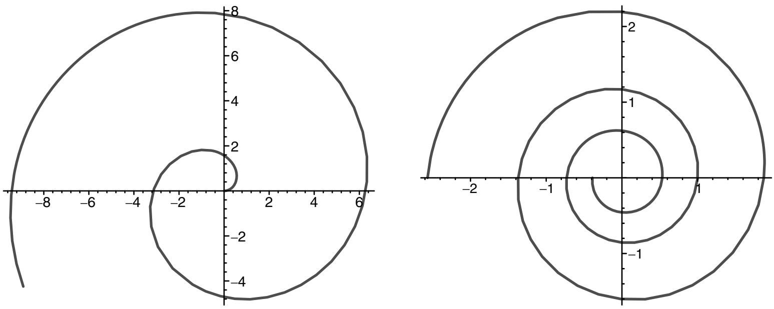


Figure 2.13. Archimedean spiral and logarithmic spiral

13. Show the following: the plane curve $c(t) = (\sin t, \sin(2t))$ is regular and closed, but not simply closed, and the rotation index is equal to 0, cf. Figure 2.9.

14. Show that the *osculating cubic parabola* of a Frenet curve c in $\mathbb{R}^3$, defined by

$$s \mapsto c(o) + se_1(0) + \frac{s^2}{2}\kappa(0)e_2(0) + \frac{s^3}{6}\kappa(0)\tau(0)e_3(0),$$

has at the point $s = 0$ the same curvature $\kappa(0)$ and torsion $\tau(0)$ as c itself. Moreover, both have a contact of third order at that point if $\kappa'(0) = 0$.

15. In spherical coordinates φ, ϑ, let a regular curve be given by the functions $(\varphi(s), \vartheta(s))$ inside the sphere with parametrization $(\cos\varphi\cos\vartheta, \sin\varphi\cos\vartheta, \sin\vartheta)$. For $s = 0$ the tangent to this curve is tangent to the equator $\vartheta = 0$, i.e., $\vartheta'(0) = 0$. Then the geodesic curvature is given by $\vartheta''(0) = \frac{d^2\vartheta}{ds^2}|_{s=0}$, and the curvature is consequently

$$\kappa(0) = \sqrt{1 + (\vartheta''(0))^2}.$$

Hint: 2.10 (iii), where the geodesic curvature is denoted J.

16. Show that a slope line with $\tau \neq 0$ lies on a sphere if and only if an equation $\kappa^2(s) = (-A^2s^2 + Bs + C)^{-1}$ is satisfied for some constants A, B, C, where $A = \frac{\tau}{\kappa}$. Hint: 2.10 (ii)

Prove that a spherical slope line through a point on the equator can never reach the north pole. It ends at a point where it cuts a small circle around the north pole orthogonally.

17. In the orthogonal (but not normal) three-frame $c', c'', c' \times c''$ the Frenet equations of a space curve take the equivalent form

$$\begin{pmatrix} c' \\ c'' \\ c' \times c'' \end{pmatrix}' = \begin{pmatrix} 0 & 1 & 0 \\ -\kappa^2 & \frac{\kappa'}{\kappa} & \tau \\ 0 & -\tau & \frac{\kappa'}{\kappa} \end{pmatrix} \begin{pmatrix} c' \\ c'' \\ c' \times c'' \end{pmatrix}.$$

Here the entries of the matrix depend in some sense rationally (i.e., without roots) on $\kappa^2 = \langle c'', c'' \rangle$ and τ (because of the relation $\kappa'/\kappa = \frac{1}{2}(\log(\kappa^2))')$.

18. Show the that the Frenet equations for a space curve are equivalent to the *Darboux equations* $e_i' = D \times e_i$ for $i = 1, 2, 3$, where $D = \tau e_1 + \kappa e_3$ is the *Darboux rotation vector.*

19. Show that the Darboux rotation vector D is perpendicular to e_1', e_2', e_3', and because of this lies in the kernel of the Frenet matrix. The normal form of the Frenet matrix is

$$\begin{pmatrix} 0 & \sqrt{\kappa^2 + \tau^2} & 0 \\ -\sqrt{\kappa^2 + \tau^2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In this normal form, the Darboux vector points in the direction of the third coordinate axis. Since the Frenet matrix is the derivative of the rotation of the Frenet three-frame, it follows that the Darboux vector points in this direction, and its length is the angular velocity. Similarly, the Darboux vector describes the accompanying screw-motion around that axis.

20. Show the following: c is a helix if and only if D is constant. c is a slope line if and only if $D/||D||$ is constant.

21. The axis of the accompanying screw-motion at a point $c(0)$ is the line in the direction of the Darboux vector $D(0) = \tau(0)e_1(0) + \kappa(0)e_3(0)$ through the point

$$P(0) = c(0) + \frac{\kappa(0)}{\kappa^2(0) + \tau^2(0)} e_2(0).$$

Show that under these circumstances the tangent to the curve which passes through all of these points, namely

$$P(s) = c(s) + \frac{\kappa}{\kappa^2 + \tau^2} e_2(s),$$

is proportional to $D(s)$ if and only if $\kappa/(\kappa^2 + \tau^2)$ is constant.

22. Verify the constancy of the curvature and torsion for the curves c_4, c_5, c_6 in 2.21.

23. Let c be a Frenet curve in $\mathbb{R}^n$. Show that

$$\operatorname{Det}(c', c'', \dots, c^{(n)}) = \prod_{i=1}^{n-1} (\kappa_i)^{n-i}.$$

24. Construct a non-planar C^∞-curve which is a Frenet curve except for a single point, and outside of this point satisfies $\tau \equiv 0$.

25. A Frenet curve in $\mathbb{R}^3$ is called a *Bertrand curve*, if there is a second curve such that the principle normal vectors to these two curves (at corresponding points) are identical, viewed as lines in space. One speaks in this case of a *Bertrand pair of curves.* Show that non-planar Bertrand curves are characterized by the existence of a linear relation $a\kappa + b\tau = 1$ with constants a, b, where $a \neq 0$.

26. Let c_1, c_2 be two plane closed curves with the property that the segment $\overline{c_1(t)c_2(t)}$ connecting them never contains the origin. Show that then $W_{c_1} = W_{c_2}$.

27. Does the equivalence (1) $\Leftrightarrow$ (4) in 2.31 hold also for curves which are not necessarily simply closed?

28. Show that one can integrate the Frenet equations for slope lines in 3-space *explicitly* by the same method as described in Section 2.16. In this case one just has to replace the expression sK by the integral $\int K(s)ds$, compare Exercise 8.

Chapter 3

The Local Theory of Surfaces

By passing from curves to surfaces we in principle just replace the parameter of the curve by two independent parameters, which then describe a two-dimensional object, which is what is called a parametrized surface. For a proper development of the theory we require that the surface is not just given by a differentiable map in two variables, but that moreover it admits a *geometric linearization* in the sense that at every point there is a *linear* surface (i.e., a plane) which touches the surface at least to order one at that point. Hence it is quite natural to demand that the derivative of the parametrization at every point has maximal rank. A map satisfying this condition is called an *immersion*, cf. 1.3.

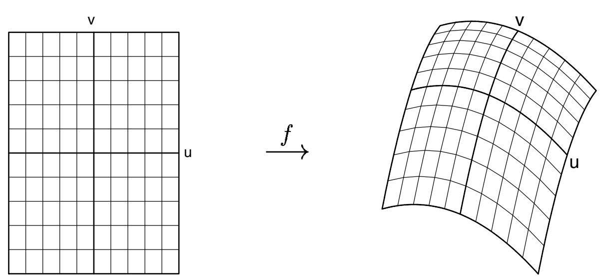


Figure 3.1. Parametrized surface element with a coordinate grid

3A Surface elements and the first fundamental form

3.1. Definition. Let $U \subset \mathbb{R}^2$ be an open set. A *parametrized surface element* is an immersion

$$f\colon U \longrightarrow \mathbb{R}^3.$$

f is also called a *parametrization*, the elements of U are called the *parameters*, and their images under f are called *points*. The cartesian coordinates in U are then mapped by f onto *coordinate lines* in the surface element; see Figure 3.1 for such a grid of coordinate lines.

A (non-parametrized) *surface element* is an equivalence class of parametrized surface elements, where two parametrizations $f\colon U \to \mathbb{R}^3$ and $\widetilde{f}\colon \widetilde{U} \to \mathbb{R}^3$ are viewed as being *equivalent* if there is a diffeomorphism $\varphi\colon \widetilde{U} \to U$ such that $\widetilde{f} = f \circ \varphi$.

Sometimes one also speaks of *regular* surface elements if the rank of the map f is maximal, i.e., if f is an immersion. If there turn out to be points, however, where the rank is not maximal, one speaks of *singular points* or *singularities*.

Similarly, one defines a *hypersurface element* in $\mathbb{R}^{n+1}$ by means of an immersion of an open subset U of $\mathbb{R}^n$ in $\mathbb{R}^{n+1}$ (cf. Section 3F), and even more generally a k-dimensional surface element in $\mathbb{R}^n$.

REMARKS:

1. The classical notion of a parametrization is given by a *triple of functions* x, y, z in Cartesian coordinates

$$f(u,v) = \Big(x(u,v), y(u,v), z(u,v)\Big) \in \mathbb{R}^3.$$

The parameter (u,v) is mapped here to the point (x,y,z). The property of $f = f(u,v)$ of being an immersion is equivalent to the property that the vectors $\frac{\partial f}{\partial u}$ and $\frac{\partial f}{\partial v}$ are linearly independent at every point. These span the *tangent plane*. The orthogonal complement to this plane is the (1-dimensional) *normal space*.

We introduce the following notations for a parametrized surface element $f\colon U \to \mathbb{R}^3, u \in U, p = f(u)$:

$T_u U$	is the *tangent space* of U at u,	$T_u U = \{u\} \times \mathbb{R}^2$,
$T_p \mathbb{R}^3$	is the *tangent space* of $\mathbb{R}^3$ at p,	$T_p \mathbb{R}^3 = \{p\} \times \mathbb{R}^3$,
$T_u f$	is the *tangent plane* of f at p,	$T_u f := Df\|_u(T_u U) \subset T_{f(u)} \mathbb{R}^3$,
$\perp_u f$	is the *normal space* of f at p,	$T_u f \oplus \perp_u f = T_{f(u)} \mathbb{R}^3$.

The elements of $T_u f$ are called *tangent vectors* and the elements of $\perp_u f$ are called *normal vectors*. Similarly, the vectors in the tangent space $T_p M$ of a submanifold $M \subset \mathbb{R}^3$ are called tangent vectors (to M at p) and the elements of the subspace $\perp_p M$ are called normal vectors (cf. 1.7, 1.8). We call a vector $X \in \mathbb{R}^3$ *tangential* (resp. *normal*) at a point p if $(p, X) \in T_u f$ (resp. $(p, X) \in \perp_u f$).

2. A two-dimensional submanifold of $\mathbb{R}^3$ (cf. Def. 1.5) can be locally described as a surface element. The parametrization in this case however is far from being unique. For example, certain parts of the unit sphere $S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ can be parametrized by

$$(u, v) \mapsto (u, v, \pm\sqrt{1 - u^2 - v^2}), \quad u^2 + v^2 < 1,$$

or by the so-called *spherical coordinates* (cf. Figure 3.2)

$$(\varphi, \vartheta) \mapsto (\cos\varphi \cos\vartheta, \sin\varphi \cos\vartheta, \sin\vartheta), \quad 0 < \varphi < 2\pi,\ -\tfrac{\pi}{2} < \vartheta < \tfrac{\pi}{2}.$$

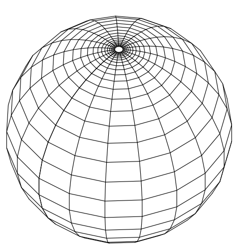


Figure 3.2. Sphere with spherical coordinates

3. The *graph* of an arbitrary real-valued differentiable function $h(u,v)$ can be viewed as the image of the immersion

$$f(u,v) := (u, v, h(u,v)).$$

Here $\frac{\partial f}{\partial u} = (1, 0, h_u)$, $\frac{\partial f}{\partial v} = (0, 1, h_v)$ are always linearly independent. Conversely, by Theorem 1.4 every two-dimensional submanifold (and also every surface element) can locally be described by the graph of a function, if the coordinates are chosen appropriately.

4. As to what exactly is to be understood under a *surface in the large*, there are several different possibilities for how this is precisely defined. A two-dimensional submanifold certainly also can be viewed as a global surface. This excludes self-intersections. This matter can only be completely clarified upon introduction of the notion of an (abstract) two-dimensional manifold, which we postpone until Section 5.1. A *surface in the large* will then be defined as an immersion of a two-dimensional manifold in $\mathbb{R}^3$.

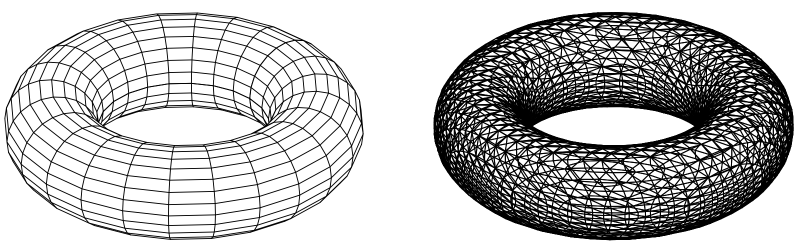


Figure 3.3. Rotational torus

Example: A *rotational torus* or *torus of revolution* is defined as a surface element by

$$f(u,v) = \big((a + b\cos u)\cos v, (a + b\cos u)\sin v, b\sin u\big),$$

$$0 < u, v < 2\pi, 0 < b < a.$$

Because of the periodicity of sine and cosine, this parametrization closes after a period of 2π in every coordinate direction, if one goes beyond the interval $u, v \in (0, 2\pi)$. One then obtains the *(two-dimensional) torus* as a global submanifold, cf. Figure 3.3. The latter is given for example by the equation $(a^2 - b^2 + x^2 + y^2 + z^2)^2 = 4a^2(x^2 + y^2)$.

3.2. Definition. (First fundamental form)
We denote by $\langle\ ,\ \rangle$ the Euclidean inner product in $\mathbb{R}^3$ as well as in every tangent space $T_p\mathbb{R}^3$, i.e., we use the notation $\langle (p,\xi),(p,\eta)\rangle = \langle \xi,\eta\rangle$. The *first fundamental form* I of a surface element (resp. of a two-dimensional submanifold) is just the restriction of $\langle\ ,\ \rangle$ to all tangent planes $T_u f$ (resp. T_pM), i.e.,

$$I(X,Y) := \langle X,Y\rangle$$

for any two tangent vectors $X,Y \in T_u f$ (resp. T_pM) or for vectors $X,Y \in \mathbb{R}^3$ which are tangent to the surface element.

In an explicit parametrization one can view this also as a symmetric bilinear form on T_uU, that is, as a mapping

$$T_uU \times T_uU \ni (V,W) \mapsto \Big\langle Df|_u(V), Df|_u(W)\Big\rangle.$$

This is also often referred to as the first fundamental form, and is denoted by I or $Df \cdot Df$ or $df \cdot df$ or $df \otimes df$. Two surface elements with the same first fundamental form are called *isometric*.

REMARKS:

In coordinates $f(u,v) = (x(u,v), y(u,v), z(u,v))$, the first fundamental form is described by the following symmetric, positive definite matrix:

$$(g_{ij}) = \begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} I(\frac{\partial f}{\partial u},\frac{\partial f}{\partial u}) & I(\frac{\partial f}{\partial u},\frac{\partial f}{\partial v}) \\ I(\frac{\partial f}{\partial v},\frac{\partial f}{\partial u}) & I(\frac{\partial f}{\partial v},\frac{\partial f}{\partial v}) \end{pmatrix}$$
$$= \begin{pmatrix} \langle\frac{\partial f}{\partial u},\frac{\partial f}{\partial u}\rangle & \langle\frac{\partial f}{\partial u},\frac{\partial f}{\partial v}\rangle \\ \langle\frac{\partial f}{\partial v},\frac{\partial f}{\partial u}\rangle & \langle\frac{\partial f}{\partial v},\frac{\partial f}{\partial v}\rangle \end{pmatrix}.$$

In case the parametrization f is k-times continuously differentiable, the matrix (g_{ij}) of the first fundamental form is $(k-1)$-times continuously differentiable. This matrix (g_{ij}) is also called the *measure tensor*, because it can be interpreted as a tensor (cf. section 6A) which determines the metric properties (that is the notion of measure here). More precisely, one can also write

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} E(u,v) & F(u,v) \\ F(u,v) & G(u,v) \end{pmatrix}$$

to indicate that E, F, G are functions of u and v. In terms of these parameters, one often writes the first fundamental form as a quadratic differential:

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2;$$

ds^2 (or ds) is also called the *element of arc length* or the *arc element* or the *line element*. It is in fact true that for a curve $c(t) = f(u(t), v(t))$, the expression

$$\sqrt{E\left(\frac{du}{dt}\right)^2 + 2F\frac{du}{dt} \cdot \frac{dv}{dt} + G\left(\frac{dv}{dt}\right)^2}$$

is equal to the length $||\dot{c}||$ of the tangent vector $\dot{c}(t)$, which is easily seen by applying the chain rule: $\dot{c} = f_u\dot{u} + f_v\dot{v}$ implies $\langle \dot{c}, \dot{c}\rangle = \langle f_u, f_u\rangle \dot{u}^2 + 2\langle f_u, f_v\rangle \dot{u}\dot{v} + \langle f_v, f_v\rangle \dot{v}^2 = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2$. For this reason the first fundamental form is also called *metric* since lengths and angles determine the metrical structure on the surface. Similarly this leads to the notions *isometric* and *isometry*, compare 4.29.

Note that for an injective f every regular curve c whose image is contained in $f(U)$ can be written as $c(t) = f(\gamma(t))$ with a regular curve γ whose image is contained in U, a fact we have used here. For this it is sufficient to set $\gamma(t) = f^{-1}(c(t))$.

The first fundamental form I can be clearly distinguished from the Euclidean inner product on T_uU. In symbolic notation, in which $\frac{\partial}{\partial u}, \frac{\partial}{\partial v}$ denotes the standard basis of the tangent space T_uU, the inner product is always given by the following matrix:

$$\begin{pmatrix} \left\langle \frac{\partial}{\partial u}, \frac{\partial}{\partial u}\right\rangle & \left\langle \frac{\partial}{\partial u}, \frac{\partial}{\partial v}\right\rangle \\ \left\langle \frac{\partial}{\partial v}, \frac{\partial}{\partial u}\right\rangle & \left\langle \frac{\partial}{\partial v}, \frac{\partial}{\partial v}\right\rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Compare this with the spherical coordinates

$$f(\varphi, \vartheta) = (\cos\varphi\cos\vartheta, \sin\varphi\cos\vartheta, \sin\vartheta)$$

on the sphere and the properties of the length function there. The first fundamental form is

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} \cos^2\vartheta & 0 \\ 0 & 1 \end{pmatrix}.$$

In U the length of an interval determined by the parameter values $\vartheta = \vartheta_0, 0 \leq \varphi \leq \pi$ is always equal to π, while the length of the image curve in $f(U)$ is equal to $\pi \cos \vartheta_0$. This factor in which the lengths differ, $\cos \vartheta$, occurs explicitly in the matrix of the first fundamental form.

3.3. Lemma. The matrix of the first fundamental form behaves as follows under a transformation of the parameters $\widetilde{f} = f \circ \varphi$ (here $D\varphi$ denotes the Jacobi matrix of φ):

$$(\widetilde{g}_{ij}) = (D\varphi)^T (g_{ij})(D\varphi).$$

PROOF: The equation $(g_{ij}) = (Df)^T \cdot (Df)$ results easily from the matrix multiplication of the corresponding matrices; compare Exercise 1 at the end of the chapter. With this we can calculate

$$\begin{aligned}(\widetilde{g}_{ij}) &= (D\widetilde{f})^T (D\widetilde{f}) = (Df \cdot D\varphi)^T (Df \cdot D\varphi) \\ &= (D\varphi)^T (Df)^T (Df)(D\varphi) = (D\varphi)^T (g_{ij})(D\varphi).\end{aligned}$$

The determinant of the first fundamental form plays an important role in the integration of functions which are defined on surface elements (so-called *surface integrals*). We provide here the following definition. For more details as well as the rule for substitutions we refer the reader to [**27**], Chapter XX, and [**28**].

3.4. Definition. (Surface integral)
Let $f: U \to \mathbb{R}^3$ be a surface element, and suppose that f is injective, viewed as a map. Let α be a continuous, real-valued function which is defined on all of $f(U)$. For every compact subset $Q \subset U$, the expression

$$\iint_{f(Q)} \alpha \, dA = \iint_Q (\alpha \circ f)(u, v) \sqrt{\operatorname{Det}(g_{ij})} \, du dv$$

is well-defined, and is called a *surface integral.* For $\alpha \equiv 1$, one just gets the *surface area.* The injectivity of f can be weakened to the assumption that no open set is covered more than once. In that case one would have to count the contribution of this set to the integral with a corresponding multiplicity.

REMARKS: One can similarly define an integral for integrable functions on measurable subsets of U, for example the Lebesgue integral.

The surface integral defined by 3.4 is invariant under transformations of the parameter according to Lemma 3.3. More precisely one has for $\widetilde{f} = f \circ \varphi, Q = \varphi(\widetilde{Q}), (u, v) = \varphi(\widetilde{u}, \widetilde{v})$ the substitution rule

$$\begin{aligned}\iint_{\widetilde{f}(\widetilde{Q})} \alpha \, dA &= \iint_{\widetilde{Q}} (\alpha \circ \widetilde{f})(\widetilde{u}, \widetilde{v}) \sqrt{\mathrm{Det}(\widetilde{g}_{ij})} \, d\widetilde{u} d\widetilde{v} \\ &= \iint_{\widetilde{Q}} (\alpha \circ \widetilde{f})(\widetilde{u}, \widetilde{v}) |\mathrm{Det} D\varphi| \sqrt{\mathrm{Det}(g_{ij})} \, d\widetilde{u} d\widetilde{v} \\ &= \iint_{Q} (\alpha \circ f)(u, v) \sqrt{\mathrm{Det}(g_{ij})} \, du dv = \iint_{f(Q)} \alpha \, dA.\end{aligned}$$

$g = \mathrm{Det}(g_{ij})$ is also called the *Gram determinant*, and $\sqrt{g}$ describes an infinitesimal distortion of f, which is made quite explicit by expressions like $dA = \sqrt{g} \, du dv$. The symbol dA is meant to remind one of "area" (*element of surface area*). Moreover, one has

$$g = \Big\| \frac{\partial f}{\partial u} \times \frac{\partial f}{\partial v} \Big\|^2,$$

where $\times$ denotes the *cross product* or *vector product* in $\mathbb{R}^3$. (Note that in the book [**1**], $\frac{\partial f}{\partial u} \wedge \frac{\partial f}{\partial v}$ is written instead of $\frac{\partial f}{\partial u} \times \frac{\partial f}{\partial v}$.) The surfaces with the minimal possible surface area (with some fixed boundary) play an important role in differential geometry and analysis; see Section 3D for more details.

3.5. Definition. (Vector fields along f)
For a surface element $f : U \to \mathbb{R}^3$ we call a map $X : U \to \mathbb{R}^3$ a *vector field along f*. In this definition, we view the vector $X(u)$ for every $u \in U$ as a vector at the point $p = f(u)$. In full mathematical rigor we would have to view X as a map from U to the tangent bundle $T\mathbb{R}^3$, where the parameter u gets mapped to $(f(u), X(u)) \in T_{f(u)}\mathbb{R}^3$. One also refers to this situation by saying that $f(u)$ is the *position vector* and $X(u)$ is the *directional vector*. The idea is that the directional vector $X(u)$ is based at the point $p = f(u)$, and then (viewed quite formally) this vector together with p defines an element $(p, X(u)) \in T_p\mathbb{R}^3 \cong \mathbb{R}^3$, compare Definition 1.6.

Similarly, X is called *tangential* (resp. *normal*), if for every $u \in U$ one has the relation $(f(u), X(u)) \in T_u f$ (resp. $(f(u), X(u)) \in \perp_u f$) (note that $T_u f \oplus \perp_u f = T_{f(u)}\mathbb{R}^3$ as an orthogonal direct sum).

A tangential vector field can always be uniquely written (with $u = (u_1, u_2)$) as

$$X(u) = \alpha(u)\frac{\partial f}{\partial u_1}\Big|_u + \beta(u)\frac{\partial f}{\partial u_2}\Big|_u,$$

while a normal vector field can always be uniquely written in the form

$$X(u) = \gamma(u)\cdot\frac{\partial f}{\partial u_1}\Big|_u \times \frac{\partial f}{\partial u_2}\Big|_u.$$

X is said to be *continuous* (resp. *differentiable*), if α, β and γ are all continuous (resp. differentiable).

EXAMPLES:

1. On the cylinder $f(\varphi, x) = (\cos\varphi, \sin\varphi, x)$ the vector field

   $$X(\varphi, x) := (-\sin\varphi, \cos\varphi, x_0)$$

   with constant x_0 is a tangential vector field and at the same time a tangent vector to the family of lines of inclination $t \mapsto (\cos t, \sin t, x_0 t + c)$ with parameter c (cf. Figure 3.4). Each of these curves is a helix, cf. Figure 2.1.

2. Starting with some (variable) point, the unit vector

   $$\nu = \pm\Big(\frac{\partial f}{\partial u_1} \times \frac{\partial f}{\partial u_2}\Big)\Big/\Big|\Big|\frac{\partial f}{\partial u_1} \times \frac{\partial f}{\partial u_2}\Big|\Big|$$

   is a normal vector field. The *unit normal* ν can also be viewed as a map

   $$\nu \colon U \to S^2 \subset \mathbb{R}^3.$$

   Here the vector is based at the origin. This so-called *Gauss map* is of great importance in the theory of surfaces, because it determines the second fundamental form and through this also the curvature, cf. 3.8–3.10.

3.6. Definition. (Orientability)

A submanifold of $\mathbb{R}^n$ is called *orientable*, if one can cover it by images of parametrized surface elements (*charts in an atlas*, cf. [**29**]) with the following property: all the Jacobi determinants of the local coordinate transformations are positive. The choice of such a cover by means of charts is called an *orientation*. For two-dimensional surfaces one can also view an orientation as a definite choice of rotational direction

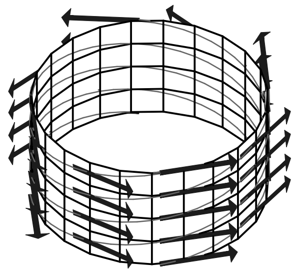


Figure 3.4. A tangential vector field on a cylinder

(mathematically "positive" is usually defined to be counter-clockwise) in each tangent plane, which is not changed inside the individual charts. In this case one has the *element of surface area*

$$dA := \sqrt{g}\, du_1 \wedge du_2$$

as a globally defined differential form (a two-form), cf. [**29**] and [**27**], Chapter XXI. On a single chart there is of course an obvious orientation, defined simply by choice of which coordinate is the first.

For a surface element $f\colon U \to \mathbb{R}^3$ the choice of an orientation can also be expressed as the choice of the order of the two tangent vectors

$$\frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2}$$

as vector fields along f. However, this choice cannot be carried over to the image of f if f is not injective. See the example of a Möbius strip below. A change of parameters with positive Jacobi determinant would preserve the orientation, although in general it would not preserve this particular two-frame. If the orientation of $\mathbb{R}^3$ is considered as being fixed, then the sign of the determinant

$$\mathrm{Det}\left(\frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2}, \nu\right)$$

gives information about the orientation of the surface element in terms of a given unit normal ν. From this we obtain the following lemma:

3.7. Lemma. A two-dimensional submanifold M of $\mathbb{R}^3$ is orientable if and only if there is a continuous unit normal vector field ν on M, i.e., a globally defined, continuous mapping

$$M \ni p \longmapsto (p, \nu(p)) \in \perp_p M.$$

In local coordinates $f(u_1, u_2)$ the vector field ν can be expressed as follows:

$$\nu = \pm\Big(\frac{\partial f}{\partial u_1} \times \frac{\partial f}{\partial u_2}\Big)\Big/\Big|\Big|\frac{\partial f}{\partial u_1} \times \frac{\partial f}{\partial u_2}\Big|\Big|.$$

Example: The Möbius strip is a non-orientable surface. The image of the parametrized surface element $f\colon \mathbb{R} \times (-\epsilon, \epsilon) \to \mathbb{R}^3$ with

$$f(u, v) = \Big(\sin u + v \sin\frac{u}{2}\sin u, \cos u + v\sin\frac{u}{2}\cos u, v\cos\frac{u}{2}\Big)$$

is closed in the u-direction after one revolution $0 \leq u \leq 2\pi$, but this in such a manner that a chosen unit normal vector for $u = 0$ is continuously transformed to the opposite unit normal at $u = 2\pi$. This

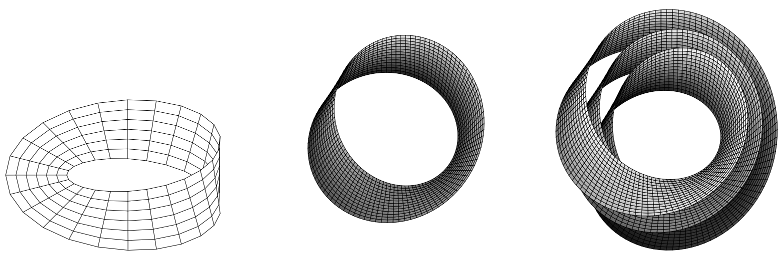


Figure 3.5. Möbius strip (left and middle) with parallel surface (right)

surface is called the *Möbius strip*, named after the German mathematician A. Möbius. From this it follows that the image of f, viewed as a submanifold, is not orientable. We note also that this surface is a ruled surface in the sense of Definition 3.20 below, since the v-curves are segments of straight lines (orthogonal to the circle $v = 0$). This surface has only one side since globally the two sides cannot be distinguished. Consequently the parallel surface in a small distance is connected, see Figure 3.5 right.

3B The Gauss map and the curvature of surfaces

Just as the curvature of curves is described by the changes of the tangents, we would expect that the curvature of surfaces is related to the changes in the tangent planes. Since each plane is essentially determined by just one direction, namely that of its normal vector (compare with the *Hessian normal form* of a plane $\{X \mid \langle X, V\rangle = c\}$, where V is a constant unit normal vector and c is a real constant), we can just as well study the variation of the normal vectors instead. This is what is behind the Gauss mapping, which we introduce now. Let S^2 denote the unit sphere $S^2 = \{(x,y,z) \in I\!R^3 \mid x^2+y^2+z^2 = 1\}$ with a fixed origin which is independent of the surface element f.

3.8. Definition. (Gauss map)
For a surface element $f\colon U \to I\!R^3$, the *Gauss map*

$$\nu\colon U \to S^2$$

is defined by the formula

$$\nu(u_1,u_2) := \frac{\frac{\partial f}{\partial u_1} \times \frac{\partial f}{\partial u_2}}{\left|\left|\frac{\partial f}{\partial u_1} \times \frac{\partial f}{\partial u_2}\right|\right|}.$$

The idea here is that the unit normal vector $\nu(u)$ no longer is thought of as being based at the image point $f(u)$, but rather by means of a parallel translation is based at the origin of space, cf. Figure 3.6. One could replace the ν which appears in the above by $-\nu = -\left(\frac{\partial f}{\partial u_1} \times \frac{\partial f}{\partial u_2}\right)/||\frac{\partial f}{\partial u_1} \times \frac{\partial f}{\partial u_2}||$, as the choice of a sign here is arbitrary and amounts to the choice of a (local) orientation. In fact, there are actually two different Gauss mappings, one for each choice of this sign. Under the assumption of orientability there exists according to 3.6 and 3.7 also a global Gauss map ν. This ν is (locally) continuously differentiable if $f \in C^2$. For this reason, we now make for the remainder of the discussion the following

General assumption: Assume that f is at least twice continuously differentiable.

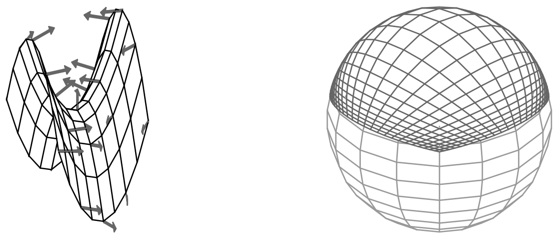


Figure 3.6. Unit normals and their images under the Gauss map

3.9. Lemma and Definition. (Weingarten map, shape operator) Let $f\colon U \to I\!R^3$ be a surface element with Gauss map $\nu\colon U \to S^2 \subset I\!R^3$.

(i) For every $u \in U$ the image plane of the linear map

$$D\nu\big|_u\colon T_uU \to T_{\nu(u)}I\!R^3$$

is parallel to the tangent plane T_uf. By canonically identifying $T_{\nu(u)}I\!R^3 \cong I\!R^3 \cong T_{f(u)}I\!R^3$ we may therefore view $D\nu$ at every point as the map

$$D\nu\big|_u\colon T_uU \to T_uf.$$

Moreover, by restricting to the image, we may view the map $Df\big|_u$ as a linear isomorphism

$$Df\big|_u\colon T_uU \to T_uf.$$

In this sense the inverse mapping $(Df\big|_u)^{-1}$ is well-defined and is also an isomorphism.

(ii) The map $L := -D\nu \circ (Df)^{-1}$ defined pointwise by

$$L_u := -(D\nu\big|_u) \circ (Df\big|_u)^{-1}\colon T_uf \to T_uf$$

is called the *Weingarten map* or the *shape operator* of f. Obviously, for every parameter u this is a linear endomorphism of the tangent plane at the corresponding point $f(u)$.

(iii) L is independent of the parametrization f (up to the choice of the sign of the unit normal vector ν), and it is self-adjoint with respect to the first fundamental form I.

By a slight abuse of notation occasionally we write $L(p, X) = (p, LX)$ for tangent vectors X at a point p.

PROOF: (i) follows simply from the relation $0 = \frac{\partial}{\partial u_i}\langle\nu, \nu\rangle = 2\langle\frac{\partial\nu}{\partial u_i}, \nu\rangle$. Therefore both vectors $\frac{\partial\nu}{\partial u_1}$ and $\frac{\partial\nu}{\partial u_2}$ are perpendicular to the normal vector. That the restriction $Df|_u\colon T_uU \to T_uf$ is a linear isomorphism follows from the assumption that Df has maximal rank.

To prove (iii), let $\widetilde{f} = f\circ\varphi$ be given, so that the corresponding normal is $\widetilde{\nu} = \pm\nu\circ\varphi$ and

$$\begin{aligned}\widetilde{L} &= -(D\widetilde{\nu})\circ(D\widetilde{f})^{-1} = \mp(D\nu)\circ(D\varphi)\circ(D\varphi)^{-1}\circ(Df)^{-1} \\ &= \mp(D\nu)\circ(Df)^{-1} = \pm L.\end{aligned}$$

The property that L is self-adjoint is most easily seen in the basis $\frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2}$ where we have $L\frac{\partial f}{\partial u_i} = -\frac{\partial\nu}{\partial u_i}$:

$$I\Big(L\frac{\partial f}{\partial u_i}, \frac{\partial f}{\partial u_j}\Big) = \Big\langle -\frac{\partial\nu}{\partial u_i}, \frac{\partial f}{\partial u_j}\Big\rangle = -\frac{\partial}{\partial u_i}\underbrace{\Big\langle\nu, \frac{\partial f}{\partial u_j}\Big\rangle}_{=0} + \Big\langle\nu, \frac{\partial^2 f}{\partial u_i\partial u_j}\Big\rangle.$$

The last expression is clearly symmetric in i and j because of the commutativity of the second derivatives. □

3.10. Definition. (Second and third fundamental form)
Let $f\colon U \to \mathbb{R}^3$ and $\nu\colon U \to S^2, L$ be given as in 3.9. Then for tangent vectors X and Y, one defines:

(i) the *second fundamental form II* of f by

$$II(X, Y) := I(LX, Y),$$

(ii) the *third fundamental form III* of f by

$$III(X, Y) := I(L^2X, Y) = I(LX, LY).$$

II and III are symmetric bilinear forms on T_uf for every $u \in U$, as follows from the fact that L is self-adjoint with respect to I.

CONSEQUENCE: The following equation holds between the three fundamental forms I, II, III:

$$III - \mathrm{Tr}(L)II + \mathrm{Det}(L)I = 0.$$

This is most easily verified by inserting a basis consisting of eigenvectors of L. It also follows from the Hamilton-Cayley theorem, see [**31**], Chapter X.

For the fundamental forms we have the following expressions in local coordinates:

$$\begin{aligned}
I\colon \quad g_{ij} &= \Big\langle \frac{\partial f}{\partial u_i}, \frac{\partial f}{\partial u_j} \Big\rangle, && \text{(first fundamental form)} \\
II\colon \quad h_{ij} &= \Big\langle \nu, \frac{\partial^2 f}{\partial u_i \partial u_j} \Big\rangle \\
&= -\Big\langle \frac{\partial \nu}{\partial u_i}, \frac{\partial f}{\partial u_j} \Big\rangle, && \text{(second fundamental form)} \\
III\colon \quad e_{ij} &= \Big\langle \frac{\partial \nu}{\partial u_i}, \frac{\partial \nu}{\partial u_j} \Big\rangle. && \text{(third fundamental form)}
\end{aligned}$$

The matrix h_i^j of the Weingarten map with $L\big(\frac{\partial f}{\partial u_i}\big) = \sum_j h_i^j \frac{\partial f}{\partial u_j}$ satisfies the equation $\big\langle L\big(\frac{\partial f}{\partial u_i}\big), \frac{\partial f}{\partial u_k}\big\rangle = h_{ik} = \sum_j h_i^j g_{jk}$ and consequently $h_i^j = \sum_k h_{ik} g^{kj}$. Here, (g^{ij}) denotes the inverse matrix $(g_{ij})^{-1}$, i.e.,

$$(g^{ij}) = \frac{1}{EG - F^2} \begin{pmatrix} G & -F \\ -F & E \end{pmatrix} = \frac{1}{\mathrm{Det}(g_{ij})} \begin{pmatrix} g_{22} & -g_{12} \\ -g_{12} & g_{11} \end{pmatrix}.$$

Although (h_{ij}) is always a symmetric matrix, the matrix (h_i^j) is *not* always symmetric. This is not in contradiction with the self-adjointness of L. One also often writes

$$II = \begin{pmatrix} L & M \\ M & N \end{pmatrix} \quad \text{or} \quad II = \begin{pmatrix} e & f \\ f & g \end{pmatrix}$$

for the matrix (h_{ij}), just as one writes

$$I = \begin{pmatrix} E & F \\ F & G \end{pmatrix}.$$

Geometrically, the Weingarten map is not very easy to visualize; it turns out to be easier to visualize the matrix h_{ij} as the Hessian matrix of a function h, which represents the surface as a graph over its tangent plane, cf. 3.13 and Figure 3.8. By definition the third fundamental form also may be viewed as the first fundamental form of ν, which in turn is viewed as a surface element (at least if $\mathrm{Rank}(D\nu) = 2$). For this reason, it is also referred to as the "metric of the spherical image", because the Gauss map ν is in a sense a parametrization of the sphere and III is then the first fundamental form of this surface element. I and III are independent of the choice of ν, while the sign of II depends on the sign of ν.

EXAMPLE: For the (unit) sphere S^2 one can set simply

$$\nu = -f$$

for the Gauss map, independent of the special form of the parametrization f. It then follows that $L = -(D\nu) \circ (Df)^{-1} = \text{Identity}$. In this case one then also has $I = II = III$.

3.11. Remark. (Motivation of the different notions of curvature for surfaces)

From Chapter 2 we know what the curvature of a curve in space is. For curves which lie entirely on some surface, it is natural to ask how much of the curvature comes only from that of the surface. We can test this with curves $c = c_X$ on a surface through a fixed point p with an arbitrary unit tangent vector $c'(p) = X$. The curvature κ of c is defined as the length of the vector c''. We decompose c'' into its tangent and its normal parts:

$$c'' = \underbrace{(c'')^{Tang.}}_{\text{tangential component}} + \underbrace{\langle c'', \nu\rangle\nu}_{\text{normal component}} .$$

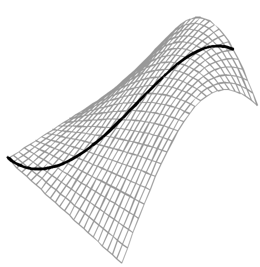


Figure 3.7. A curve on a surface

The normal component of p is then quite simply

$$\langle c'', \nu\rangle\nu = \Big\langle \frac{d^2c}{ds^2}, \nu\Big\rangle\nu = -\Big\langle c', \frac{\partial\nu}{\partial s}\Big\rangle\nu = \langle X, LX\rangle\nu = I\!I(X,X)\nu$$

and thus clearly only depends on the tangent $c' = X$ at the point p, but not on the choice of curve. This state of affairs is referred to as the *Theorem of Meusnier*.

For this reason, one calls $I\!I(X,X)$ the *normal curvature* κ_ν of the curve c_X. One always has $\kappa^2 \geq \kappa_\nu^2$, with equality holding if and only if c'' and ν are linearly dependent, or equivalently in case of a Frenet curve, if the osculating plane of the curve contains ν. This is in particular the case when the curve is obtained as the section of the surface with a plane which is perpendicular to the tangent plane at p which contains X (a so-called *normal section*). The normal curvature is then the (oriented) curvature of the normal section, viewed as a plane curve. In this case the tangent component vanishes. In case this happens to be true for a whole interval, then the curve moves inside the surface without any curvature. One calls such curves *geodesic lines* or *geodesics*; cf. also 4.9 in this respect. In all other cases there is a tangent component, which is referred to as the *geodesic curvature*.

The normal curvature κ_ν certainly does not depend on the curve, but is completely determined by the surface. The directions of the extremal normal curvatures are therefore particularly interesting geometric invariants of the surface, which are given by the extremal values of $I\!I(X,X)$. This motivates the following definition.

3.12. Definition. (Principal curvature)
Let $X \in T_u f$ denote a unit tangent vector, i.e, $I(X,X) = 1$. X is called a *principal curvature direction* for f, if one of the following equivalent conditions is satisfied:

(i) $I\!I(X,X)$ (The normal curvature κ_ν in the direction of X) has a stationary value among all X with $I(X,X) = 1$.

(ii) X is an eigenvector of the Weingarten map L.

The corresponding eigenvalue λ (where $LX = \lambda X$) is called the *principal curvature*.

The eigenvalue λ occurs as a *Lagrange multiplier* for the following extremal value problem: "$II(X,X)$ should become extremal under the constraint that $I(X,X) = 1$". The equivalence of (i) and (ii) is often called the *Theorem of Olinde Rodrigues.* In fact this equivalence follows directly from the Lagrange rule that a point is stationary for one function $II(X,X)$ under the constraint that $I(X,X)$ is constant, if and only if the two gradients are linearly dependent, see [**28**]. Here the gradient of II at X is LX; the gradient of I is just X itself.

For a two-dimensional surface both principal curvatures are simply the minimum and the maximum of the normal curvature. For an n-dimensional hypersurface we have a similar definition with n principal curvatures, among the minimum and the maximum of the normal curvature, as well as $n-2$ saddle points in between. The two principal curvatures of f are denoted κ_1, κ_2. The corresponding principal curvature directions (pcd) X_1, X_2 are perpendicular to one another, if $\kappa_1 \neq \kappa_2$. This follows from the self-adjointness of L together with the relations

$$\kappa_1 \langle X_1, X_2 \rangle = \langle LX_1, X_2 \rangle = \langle X_1, LX_2 \rangle = \kappa_2 \langle X_1, X_2 \rangle.$$

The sign of κ_1, κ_2 depend on the choice of L, hence on the choice of ν and ultimately on the orientation. If both of these are positive (resp. negative), then II is positive (resp. negative) definite. If both signs occur, then II is indefinite. These cases are again independent of the orientation, and hence are of geometric significance.

3.13. Definition.

(i) The determinant $K = \mathrm{Det}(L) = \kappa_1 \cdot \kappa_2$ is called the *Gaussian curvature.*

(ii) The average value $H = \frac{1}{2}\mathrm{Tr}(L) = \frac{1}{2}(\kappa_1 + \kappa_2)$ is called the *mean curvature.*

(iii) A point p on the surface is called

elliptic,	if $K(p) > 0$,
hyperbolic,	if $K(p) < 0$,
parabolic,	if $K(p) = 0$ and $H(p) \neq 0$,
umbilic,	if $\kappa_1(p) = \kappa_2(p)$,
properly umbilic,	if $\kappa_1(p) = \kappa_2(p) \neq 0$,
a level point,	if $\kappa_1(p) = \kappa_2(p) = 0$.

Consequence. One always has $H^2 - K = \frac{1}{4}(\kappa_1 - \kappa_2)^2 \geq 0$, with equality holding precisely for umbilical points. The quantities H and K can be expressed in local coordinates according to 3.10 as follows:

$$K = \frac{\mathrm{Det}(h_{ij})}{\mathrm{Det}(g_{ij})} = \frac{h_{11}h_{22} - h_{12}^2}{g_{11}g_{22} - g_{12}^2},$$

$$H = \frac{1}{2}\sum_i h_i^i = \frac{1}{2}\sum_{i,j} h_{ij}g^{ji} = \frac{1}{2\ \mathrm{Det}(g_{ij})}\big(h_{11}g_{22} - 2h_{12}g_{12} + h_{22}g_{11}\big).$$

Remark: The Gaussian curvature is closely related with the Gauss map: It can be interpreted as the "infinitesimal area distortion" (with sign) of the Gauss map, cf. 4.45.

Examples: An ellipsoid with the equation $a^2x^2 + b^2y^2 + c^2z^2 = 1$ has only elliptic points. The sphere has only proper umbilics because $L = \pm\mathrm{Id}$, while the single-sheeted hyperboloid $x^2 + y^2 - z^2 = 1$ only has hyperbolic points. The circular cylinder only has parabolic points, the plane only level points (because $L = 0$). The paraboloid of rotation $z = x^2 + y^2$ has only elliptic points with just a single isolated umbilic, namely the origin, while the monkey saddle $z = x^3 - 3xy^2$ consists completely of hyperbolic points, with an isolated level point at the origin. For pictures see Figures 3.8 and 3.9.

The different types of points can be seen particularly clearly in a description of the surface as a *graph* of a function h over the tangent plane of a fixed point. The corresponding coordinates are also called *Monge coordinates.* We parametrize the surface element by $f(u_1, u_2) = \big(u_1, u_2, h(u_1, u_2)\big)$ with $h(0,0) = 0$ and $\mathrm{grad} h|_{(0,0)} = 0$. The type of the point $f(0,0)$ can then be read off of the second fundamental form as follows ($\nu = (0,0,1)$):

$$\big(h_{ij}(0,0)\big)_{ij} = II|_{(0,0)} = \left(\left\langle \frac{\partial^2 f}{\partial u_i \partial u_j}, \nu \right\rangle\right)_{ij} = \left(\frac{\partial^2 h}{\partial u_i \partial u_j}\right)_{ij}.$$

The matrix $\big(\frac{\partial^2 h}{\partial u_i \partial u_j}\big)_{ij} = \mathrm{Hess}(h)$ is called the *Hessian matrix* or the *Hessian* of h, [**27**], Chapter XVII, §5.

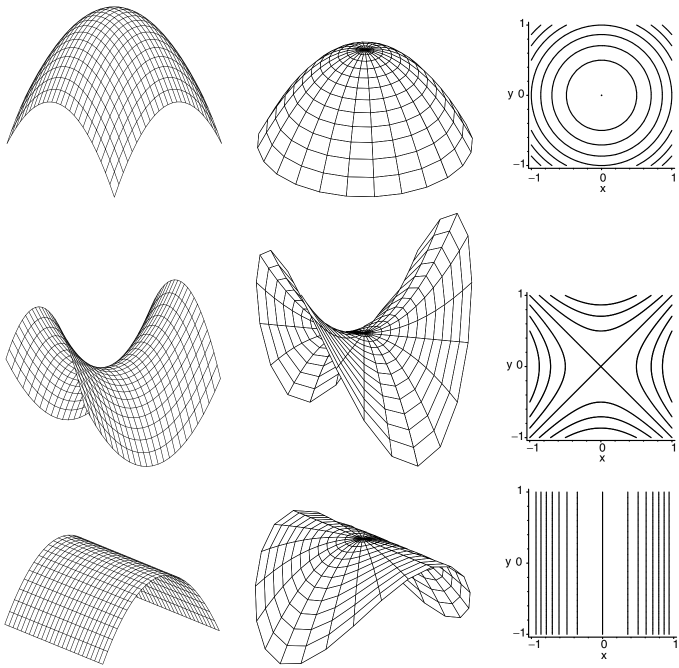


Figure 3.8. Elliptic, hyperbolic and parabolic points with level curves

The point $f(0,0)$ is

elliptic,	in case $\mathrm{Hess}(h)\vert_{(0,0)}$ is positive (or negative) definite,
hyperbolic,	in case $\mathrm{Hess}(h)\vert_{(0,0)}$ is indefinite,
parabolic,	in case $\mathrm{Rank}(\mathrm{Hess}(h))\vert_{(0,0)} = 1$,
umbilic,	in case $\mathrm{Hess}(h)\vert_{(0,0)} = \lambda\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$,
properly umbilic,	in case in addition $\lambda \neq 0$,
a level point,	in case $\mathrm{Hess}(h)\vert_{(0,0)} = \left(\begin{smallmatrix}0 & 0\\ 0 & 0\end{smallmatrix}\right)$.

The notations are derived from the type of the approximating quadratic surface (elliptic, parabolic, hyperbolic), in which h is replaced by the Taylor polynomial of second degree[1] The type of a point is also described by the *Dupin indicatrix*, which is defined by the condition that the corresponding quadratic form takes a constant value. This yields, according to type of point, an ellipse, a hyperbola or a pair of lines, respectively, and in the case of an umbilic a circle.

Examples: An ordinary saddle point is given by $h(x,y) = x^2 - y^2$, while a monkey saddle is given by the equation $h(x,y) = x^3 - 3xy^2$, and a dog saddle by the equation $h(x,y) = xy(x^2 - y^2)$.

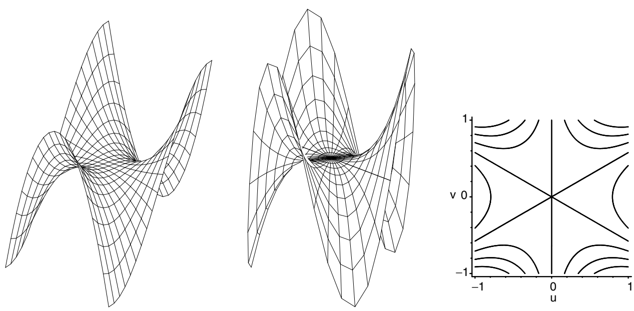


Figure 3.9. A monkey saddle with level curves

Those surfaces which consist solely of umbilics are often called *totally umbilical surfaces.* They are classified in the following theorem:

3.14. Theorem. All points of a connected surface element of class C^2 are umbilics if and only if the surface is contained in either a plane or a sphere.

Proof:[2] First we have $L = 0$ for a plane and $L = \pm\frac{1}{r}\cdot\mathrm{Id}$ for a sphere of radius r. Conversely, we clearly have $\kappa_1 = \kappa_2$ if and only if L is a scalar multiple of the identity. Call this scalar factor κ such that $D\nu(u,v) = -\kappa(u,v)\cdot Df(u,v)$ at any point of the surface element. We have to show that this function κ is constant. We can locally introduce Monge coordinates $f(u,v) = (u,v,h(u,v))$ which implies

[1] Strictly speaking, these quadratic surfaces are called *elliptic paraboloid, parabolic cylinder and hyperbolic paraboloid*, respectively.
[2] following A.Pauly, *Flächen mit lauter Nabelpunkten*, Elemente d. Math. **63**, 141–144 (2008)

$$\nu_u(u,v) = -\kappa(u,v) f_u(u,v) = -\kappa(u,v)(1,0,h_u),$$

$$\nu_v(u,v) = -\kappa(u,v) f_v(u,v) = -\kappa(u,v)(0,1,h_v).$$

From

$$\nu(u,v) = \frac{f_u \times f_v}{||f_u \times f_v||} = \frac{1}{\sqrt{1+h_u^2+h_v^2}}\Big(-h_u, -h_v, 1\Big)$$

we obtain

$$\frac{\partial}{\partial u}\Big(h_u/\sqrt{1+h_u^2+h_v^2}\Big) = \kappa(u,v), \quad \frac{\partial}{\partial v}\Big(h_u/\sqrt{1+h_u^2+h_v^2}\Big) = 0,$$

$$\frac{\partial}{\partial v}\Big(h_v/\sqrt{1+h_u^2+h_v^2}\Big) = \kappa(u,v), \quad \frac{\partial}{\partial u}\Big(h_v/\sqrt{1+h_u^2+h_v^2}\Big) = 0.$$

The equations on the right hand side imply that $h_u/\sqrt{1+h_u^2+h_v^2}$ depends only on u and $h_v/\sqrt{1+h_u^2+h_v^2}$ depends only on v:

$$h_u/\sqrt{1+h_u^2+h_v^2} = a(u), \qquad h_v/\sqrt{1+h_u^2+h_v^2} = b(v).$$

From the two other equations we obtain $a'(u) = \kappa(u,v) = b'(v)$. Therefore κ must be constant since it simultaneously depends only on u and only on v. The case $\kappa = 0$ corresponds to the plane (because $D\nu = 0$, which implies that ν is constant), and $\kappa \neq 0$ corresponds to the case of a sphere of radius $1/|\kappa|$. Here the quantity $\frac{1}{\kappa}\nu + f$ is constant because $D(\frac{1}{\kappa}\nu + f) = 0$, and it defines the center of the sphere. □

3.15. Definition.
A regular curve $c = f \circ \gamma$, $\gamma\colon I \to U$, $f : U \to \mathbb{R}^3$ is called a *line of curvature*, if the unit tangent vector $\dot{c}(t)/\parallel \dot{c}(t) \parallel$ is a principal curvature direction at every point.

One says that a surface is parametrized by *lines of curvature parameters*, if the u_i-lines are lines of curvature everywhere. This is the case if and only if in these parameters one has $g_{12} = h_{12} = 0$, i.e., if

$$I = \begin{pmatrix} g_{11} & 0 \\ 0 & g_{22} \end{pmatrix}, \quad II = \begin{pmatrix} h_{11} & 0 \\ 0 & h_{22} \end{pmatrix}, \quad \kappa_i = \frac{h_{ii}}{g_{ii}}.$$

Every surface element without umbilics can locally be parametrized in such a way that the new parameters are lines of curvature parameters. This follows from the theory of partial differential equations, cf. [**2**], 3.6.

3C Surfaces of rotation and ruled surfaces

In this section we will study two classes of surfaces in more detail, which first of all occur often and secondly allow quite simple computations of all of the relevant geometric quantities. The *surfaces of rotation* or *surfaces of revolution* are formed from circles centered at one of the axes, with variable radii (perpendicular to the axis), and the *ruled surfaces* are formed from lines along some fixed curve, but in variable direction. Similarly one can define a *canal surface* by the condition that the fixed axis in the definition of surfaces of rotation is replaced by a fixed curve, and *scrolls*, in which the lines in the definition of ruled surfaces are replaced by a fixed curve.

3.16. Definition. (Surface of rotation)
A surface is called a *surface of rotation*, if it is obtained by rotating a regular, plane curve (the *meridian curve* or *profile curve*)

$$t \mapsto (r(t), h(t))$$

around the z-axis in $I\!R^3$, in other words, if it admits a parametrization of the following form:

$$f(t, \varphi) = \big(r(t)\cos\varphi, r(t)\sin\varphi, h(t)\big).$$

Surfaces of rotation occur naturally in all technical disciplines in which rotations occur, for example in mechanical engineering. Objects with rotational symmetries occur often in physics. A symmetry of this kind greatly simplifies computations; in fact, in some cases only under this circumstance can one do calculations at all, which is why one often assumes the existence of a symmetry of this kind. By the definition above, a surface of rotation is invariant under all rotations about the z-axis, which are described by the following maps:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \longmapsto \begin{pmatrix} \cos\ \varphi & -\sin\ \varphi & 0 \\ \sin\ \varphi & \cos\ \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

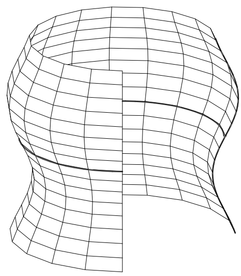


Figure 3.10. Surface of rotation

For surfaces of rotation the most important geometrical quantities can be easily calculated. For example, from the expressions

$$\frac{\partial f}{\partial t} = (\dot{r}\cos\varphi, \dot{r}\sin\varphi, \dot{h}), \quad \frac{\partial f}{\partial \varphi} = (-r\sin\varphi, r\cos\varphi, 0)$$

it immediately follows that the first fundamental form is

$$I = \begin{pmatrix} \dot{r}^2 + \dot{h}^2 & 0 \\ 0 & r^2 \end{pmatrix}.$$

Hence, if a curve is regular (i.e., $\dot{r}^2 + \dot{h}^2 \neq 0$), then for $r \neq 0$ the surface is also regular, meaning that f is an immersion. We choose as normal vector

$$\nu = \frac{\frac{\partial f}{\partial t} \times \frac{\partial f}{\partial \varphi}}{\| \frac{\partial f}{\partial t} \times \frac{\partial f}{\partial \varphi} \|} = \frac{1}{\sqrt{\dot{r}^2 + \dot{h}^2}} \Big(- \dot{h}\cos\varphi, -\dot{h}\sin\varphi, \dot{r} \Big)$$

and calculate the second fundamental form by the second derivatives

$$\begin{aligned} \frac{\partial^2 f}{\partial t^2} &= (\ddot{r}\cos\varphi, \ddot{r}\sin\varphi, \ddot{h}), \\ \frac{\partial^2 f}{\partial t \partial \varphi} &= (-\dot{r}\sin\varphi, \dot{r}\cos\varphi, 0), \\ \frac{\partial^2 f}{\partial \varphi^2} &= (-r\cos\varphi, -r\sin\varphi, 0). \end{aligned}$$

Again it follows immediately from this that

$$II = \frac{1}{\sqrt{\dot{r}^2 + \dot{h}^2}} \begin{pmatrix} -\ddot{r}\dot{h} + \dot{r}\ddot{h} & 0 \\ 0 & r\dot{h} \end{pmatrix}.$$

From this it is clear that t, φ are lines of curvature parameters in the sense of 3.15. The principal curvatures (i.e., the eigenvalues of II with respect to I) are thus

$$\begin{aligned} \kappa_1 &= \frac{1}{(\dot{r}^2+\dot{h}^2)^{3/2}}(-\ddot{r}\dot{h}+\dot{r}\ddot{h}), \\ \kappa_2 &= \frac{1}{(\dot{r}^2+\dot{h}^2)^{1/2}}\cdot\frac{\dot{h}}{r}. \end{aligned}$$

In case t is the arc length parameter, $r'^2+h'^2=1$ holds and

$$I = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}, \qquad II = \begin{pmatrix} -r''h'+r'h'' & 0 \\ 0 & rh' \end{pmatrix},$$

$$\kappa_1 = -r''h'+r'h'', \qquad \kappa_2 = \frac{h'}{r}.$$

The first principal curvature κ_1 is here nothing but the curvature of the plane curve $(r(t),h(t))$, which is easily seen from the Frenet equations $e_1' = \kappa e_2, e_2' = -\kappa e_1$, cf. 2.5. Indeed, one has $e_1 = (r', h')$, $e_1' = (r'', h''), e_2 = (-h', r')$, hence $\kappa = \langle e_1', e_2\rangle = -r''h'+r'h''$.

Other expressions for the same quantities are

$$\kappa_1 = -r''h'+r'h'' = \frac{h'h''}{r'}h'+r'h'' = \frac{h''}{r'}(h'^2+r'^2) = \frac{h''}{r'} = -\frac{r''}{h'}.$$

For the second and the third equality, note that $r'^2+h'^2$ is constant, so $r'r''+h'h''=0$ is also. It follows that

$$K = \kappa_1\kappa_2 = -\frac{r''}{r},$$

$$H = \frac{1}{2}(\kappa_1+\kappa_2) = \frac{1}{2}\Big(\frac{h''}{r'}+\frac{h'}{r}\Big) = \frac{rh''+r'h'}{2rr'} = \frac{(rh')'}{(r^2)'}.$$

From this we in turn recognize among other things the following facts: every condition on K and H (for example that one of the curvatures is constant) leads to an ordinary differential equation for r, if one replaces h' by $\pm\sqrt{1-r'^2}$. In particular, one has

$$\begin{aligned} K = c &\iff r''+cr=0, \\ H = c &\iff (rh')' = c(r^2)', \\ \kappa_1 = \kappa_2 &\iff \tfrac{h''}{h'} = \tfrac{r'}{r} \qquad \iff r'^2+c^2r^2=1, \end{aligned}$$

where in each case c is constant, see 3.17. For obtaining explicit solutions it may be more convenient to use other parameters, compare

3.27 where r is the parameter and h is a function of r. Cases in which a singularity appears also occur, for example when one of the principal curvatures vanishes while the other becomes infinite. Also interesting are the extreme cases in which $r' = 0, h' = 1$ and $r' = 1, h' = 0$.

REMARK: A surface of rotation can be a regular (or even C^∞ or analytic) surface along the axis of rotation $r = 0$ (perhaps in a different parametrization), even though the matrix

$$I = \begin{pmatrix} \dot{r}^2 + \dot{h}^2 & 0 \\ 0 & r^2 \end{pmatrix}$$

is apparently degenerate there. To see this, one has to determine κ_2 by passing to the limit and applying the rule of Bernoulli-l'Hospital: $\kappa_2 = \lim \frac{h'}{r} = \lim \frac{h''}{r'} = \pm \lim h''$, in case $\lim h' = 0$, from which it follows that $\lim r' = \pm 1$. As a simple example of this, consider the sphere with $r(t) = \sin t,\ h(t) = -\cos t$. One has

$$\kappa_1 = -r''h' + r'h'' = \sin^2 t + \cos^2 t = 1,$$

$$\kappa_2 = \frac{h'}{r} = \frac{\sin t}{\sin t} = 1 \text{ (also as } t \to 0).$$

This surface is also regular for $r = 0$. A necessary condition for this is that $h' = 0$ at this point, since κ_2 otherwise cannot have any finite value. If the surface is regular (C^2) in $r = 0$, then there is necessarily an umbilical point there ($\kappa_1 = \kappa_2$).

3.17. Example. (Surfaces of rotation with constant curvature)

In order to determine the surfaces of rotation with constant Gaussian curvature K, we seek according to 3.16 all solutions to the differential equation

$$r'' + Kr = 0.$$

Here the parameter of the curve we are looking for $(r(t), h(t))$ is arc length, and hence we have $h'^2 = 1 - r'^2$.

The general solution is then the following expression, with constants a, b:

$$r(t) = \begin{cases} a\cos(\sqrt{K}t) + b\sin(\sqrt{K}t), & \text{in case } \ K > 0, \\ at + b \text{ with } |a| \leq 1, & \text{in case } \ K = 0, \\ a\cosh(\sqrt{-K}t) + b\sinh(\sqrt{-K}t), & \text{in case } \ K < 0. \end{cases}$$

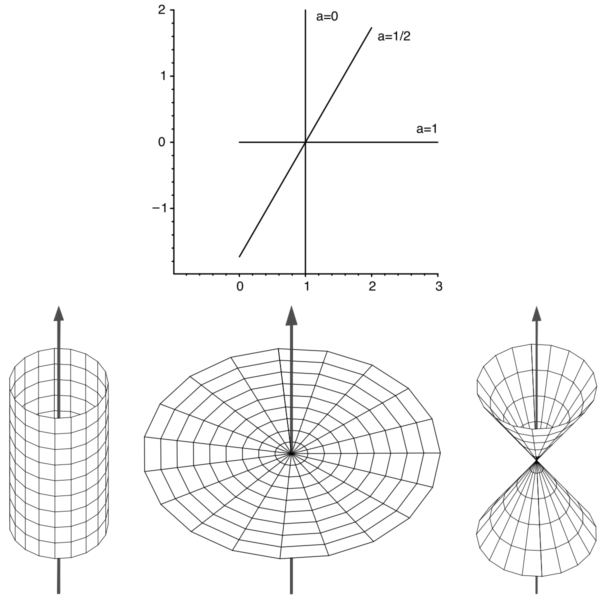


Figure 3.11. Surfaces of rotation with vanishing Gaussian curvature

For $K = 0$ we get a cylinder of radius $r = b$ if $a = 0$, a plane which is orthogonal to the axis of rotation if $|a| = 1$, and a circular cone if $0 < |a| < 1$.

In case $K > 0$ we can achieve, by a translation of parameters if necessary, that $b = 0$. In order that the equation $h'^2 = 1 - r'^2$ has a real solution h, it is necessary that an inequality of the kind $0 \leq a^2 K \sin^2(\sqrt{K}t) \leq 1$ holds and consequently that

$$h(t) = \int_0^t \sqrt{1 - a^2 K \sin^2(\sqrt{K}x)}dx,$$

which is an elliptic integral. The case $a^2K = 1$ corresponds to a *sphere*, in case $0 < a^2K < 1$ one has an *elongated sphere* (Figure 3.12, right side), while for $a^2K > 1$ one has an *oblate sphere* (Figure 3.12, middle).

If $K < 0$, we get for $b^2 > a^2$ the so-called *conic type* (Figure 3.13, right) and for $b^2 < a^2$ the so-called *hyperboloid type* (Figure 3.13,

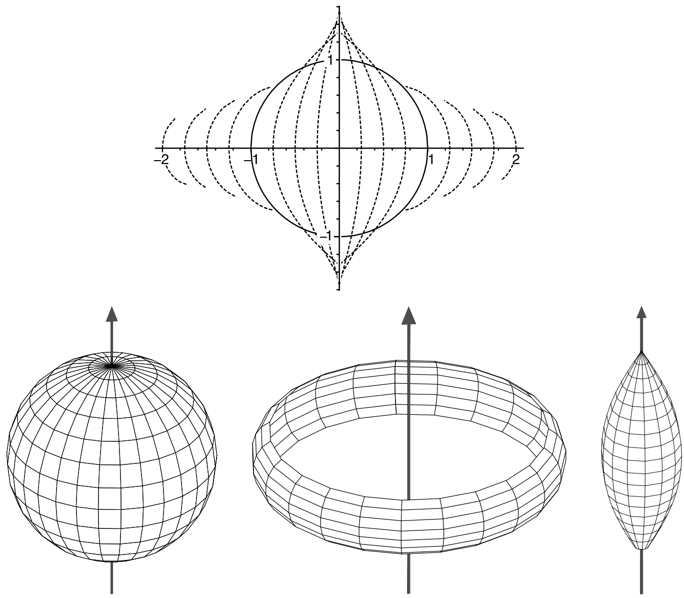


Figure 3.12. Surfaces of rotation with constant positive Gaussian curvature

center). In the special case in which $a = b$ and $K = -1$, one gets the famous *pseudo-sphere* with

$$r(t) = a\exp(t), \quad h(t) = \int_0^t \sqrt{1 - a^2\exp(2x)}\, dx, \quad t \in (-\infty, 0)$$

which is also known as the *Beltrami surface* or the *tractroid* or the *bugle surface*. A meridian of this surface is a *tractrix*, as discussed in Section 2.3 (Figure 3.13, left[3]). The curve ends at a point $t = 0$ with infinitely large curvature (this is the point where the tangent becomes exactly horizontal), hence the surface itself ends there in a singularity. While the product of both principle curvatures is necessarily constant, at the singular point one of the principal curvatures becomes infinite, while the other vanishes. Additional remark: Regarding this surface as a purely intrinsic object without ambient space, this singularity is not really there, compare 3.44 in combination with Theorem 4.16.

[3]While Figure 3-22 in [**1**] provides a nice and correct picture of Beltrami's surface as well, Figure 3-23 in [**1**] is far from being correct.

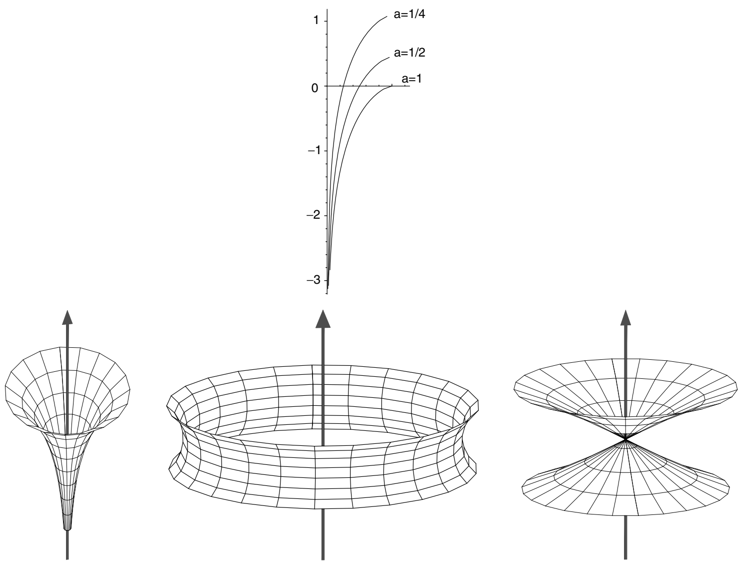


Figure 3.13. Surfaces of rotation with constant negative Gaussian curvature

3.18. Definition. A curve $c = f \circ \gamma$ is called an *asymptotic curve* of f, if $II(\dot{c}, \dot{c}) = 0$ identically.

The name arises from the asymptotic lines of the hyperbola which is defined by the Dupin indicatrix at a hyperbolic point. Obviously, asymptotic curves do not exist on elliptic surface elements. On hyperbolic surface elements one can introduce parameters in such a way that the parameter lines (curves where the parameters are constant) are asymptotic curves, cf. [**2**], 3.6. For example, every straight line which lies on a surface is an asymptotic curve, because $\ddot{c} = 0$. This holds in particular for the lines on the one-sheeted hyperboloid (hyperboloid of rotation) with the equation $x^2 + y^2 - z^2 = 1$ (see Figure 3.14). Surfaces of this kind, which are composed of lines which lie on them, will be studied in more detail in Sections 3.20 – 3.24. All these surfaces have nonpositive Gaussian curvature.

If the curvature of an asymptotic curve does not vanish, then we have the following geometric interpretation for its torsion:

3.19. Theorem. (Beltrami–Enneper)
Every asymptotic curve with curvature $\kappa \neq 0$ and torsion τ satisfies the equation $\tau^2 = -K$.

PROOF: Let $c(s)$ be an asymptotic curve with $II(c', c') = 0$. Then the normal curvature of c vanishes (cf. 3.11). Hence e_2 is tangential to the surface and consequently the vector $e_3 = \nu$ is a unit normal, possibly up to sign. We now calculate $\tau = \langle e_2', e_3 \rangle = \langle e_2', \nu \rangle = II(e_1, e_2)$. From this it follows that

$$K = \mathrm{Det} II / \mathrm{Det} I = II(e_1, e_1) II(e_2, e_2) - (II(e_1, e_2))^2 = 0 - \tau^2. \qquad \square$$

3.20. Definition. (Ruled surface)
A surface is called a *ruled surface*, if it has a C^2-parametrization of the following kind:

$$f(u, v) = c(u) + v \cdot X(u),$$

where c is a (differentiable, but not necessarily regular) curve and X is a vector field along c which vanishes nowhere (cf. 3.5).

Clearly the v-lines (with constant u) are Euclidean lines in space. The intuition we have of the situation is that the surface results from the motion of a line in space, similarly to the way a curve represents the motion of a point (particle), cf. Figure 3.5 or 3.14 for an example. These lines on X are also called *generators* or the *ruling* of the surface, and the curve c is called the *directrix* of the surface. Movements of this kind of surfaces or segments occur often in technology in the description of mechanical processes, like for example the motion of the arm of a robot.

3.21. Lemma. (Standard parameters)
Let $f(t, s) = c(t) + s \cdot X(t)$ be a ruled surface with $||X|| = 1$ and $\frac{dX}{dt} \neq 0$ in an interval $t_1 < t < t_2$. Then f can be reparametrized in a unique way as $f_*(u, v) = c_*(u) + v \cdot X_*(u)$ so that $||X_*|| = ||X_*'|| = 1$ and $\langle c_*', X_*' \rangle = 0$.

The curve c_* is uniquely determined (with the exception of the plane) by this property and is called the *striction line* of the surface. The parameter u is then the arc length on the spherical curve X. In the case of the one-sheeted hyperboloid with the equation $x^2+y^2-z^2=1$, the striction line is nothing but the "waist", cf. Figure 3.14. If $\frac{d}{dt}X \equiv 0$ holds on an interval, then X is constant there (and the surface is a cylinder over the curve c); hence there is no such exceptional curve and parameter of this kind, as the condition $||X'|| = 1$ can no longer be satisfied. If the plane is, however, parametrized with a constant X, then one can vary the vector field X and get other standard parameters, which now depend on the choice of X.

PROOF: Since X is a regular curve, we can choose the parameter u for c and X in a certain interval $u_1 < u < u_2$ in such a way that $X_*(u) := X(t)$ is parametrized by arc length u, i.e., $\langle X_*', X_*' \rangle = 1$. We then follow the Ansatz $c_*(u) = c(u) + v(u)X_*(u)$ with a certain function $v(u)$. Then we have $\langle c_*', X_*' \rangle = \langle c' + v(u)X_*' + v'(u)X_*, X_*' \rangle = \langle c', X_*' \rangle + v(u)$, and this expression vanishes if and only if $v(u) = -\langle c', X_*' \rangle$. The curve c_* is uniquely determined by these data. It is not necessarily regular. The plane is exceptional since it is a ruled surface in infinitely many ways. □

3.22. Theorem. Using standard parameters, a ruled surface is, up to Euclidean motions, uniquely determined by the following quantities:

$$F = \langle c', X \rangle,$$
$$\lambda := \langle c' \times X, X' \rangle = \mathrm{Det}(c', X, X'),$$
$$J := \langle X'', X \times X' \rangle = \mathrm{Det}(X, X', X''),$$

each of which is a function of u. Conversely, every choice of these three quantities uniquely determines a ruled surface.

Consequence: For a ruled surface, given by standard parameters $f(u,v) = c(u) + v \cdot X(u)$, the first fundamental form is given as follows:

$$I = \begin{pmatrix} \langle c', c' \rangle + v^2 & \langle c', X \rangle \\ \langle c', X \rangle & 1 \end{pmatrix} = \begin{pmatrix} F^2 + \lambda^2 + v^2 & F \\ F & 1 \end{pmatrix}$$

with $\mathrm{Det}(I) = \lambda^2 + v^2$.

The quantity $F = g_{12}$ determines the angle φ between the striction line and X by $F = ||c'|| \cos\varphi$, J determines the curvature of the spherical curve X and consequently X itself by 2.10 (iii) and 2.13, and λ is called the *parameter of distribution.*

PROOF OF 3.22: The fact that a given surface uniquely determines these three quantities is clear. Conversely, according to 2.10 (iii) and 2.13, X is uniquely determined by the prescription of J (up to Euclidean motions). To determine the curve we use the orthonormal frame $X, X', X \times X'$ and calculate

$$c' = \langle c', X\rangle X + \langle c', X'\rangle X' + \langle c', X \times X'\rangle X \times X' = FX + \lambda X \times X'.$$

For given X, F, λ, this is a system of linear differential equations with the solution $c(u) = c(u_0) + \int_{u_0}^{u} (FX + \lambda X \times X')dt$. The initial conditions are determined by the choice of a starting point $c(u_0)$ on the curve and the three-frame $X, X', X \times X'$ at that point. □

3.23. Consequence. (Special case: helicoidal ruled surfaces)

(i) The three quantities λ, F, J which determine a ruled surface are constant if and only if the surface is a so-called *helicoidal ruled surface*, which is formed by the screw-motion of a single line (Figure 3.14; compare 2.3). This includes the special case of a rotation. The striction line is then the trajectory of the point on the line which is nearest to the axis of rotation, i.e., either the axis itself or a helix or a circle.

(ii) In addition to the three quantities being constant as in (i), one has moreover $F = J = 0, \lambda \neq 0$ if and only if the surface is a *(right) helicoid* (Figure 3.19) $f(u, v) = (v\cos(\alpha u), v\sin(\alpha u), bu)$ with constant α, b, where $\lambda^2 = \alpha^2 b^2$.

(iii) The only surfaces of rotation which are also ruled surfaces are those with $K = 0$ (see Figure 3.11) and the one-sheeted hyperboloids with the equations $x^2 + y^2 - a^2z^2 = c^2$ (see Figure 3.14).

PROOF: (i) can be seen as follows. Since J is constant, X is a circle by 2.10. The equation which determines c is then $c' = FX + \lambda X \times X'$ by the proof of 3.22. Moreover one has $(X \times X')' = X \times X'' =$

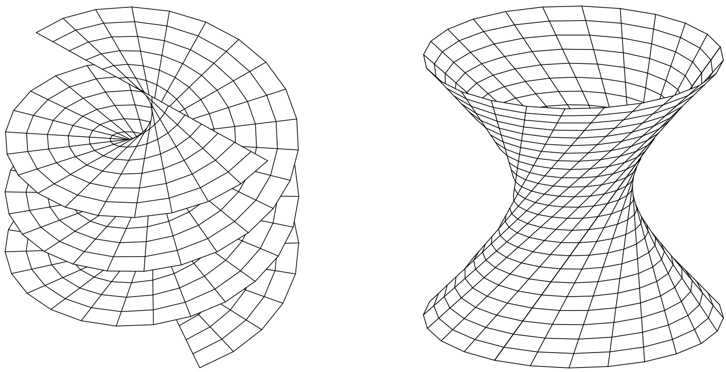


Figure 3.14. Helicoidal ruled surface and a one-sheeted hyperboloid of revolution as ruled surface

$-JX'$. For constant F and λ it follows that $c'' = FX' + \lambda X \times X'' = (F - \lambda J)X'$. Hence c' is a constant multiple of X plus an additive constant Y_0, where Y_0 is perpendicular to the plane which is spanned by the circle X (i.e., the X', X''-plane). To see this one just has to calculate $\langle Y_0, X'\rangle = \langle Y_0, X''\rangle = 0$. Therefore c' coincides with the tangent to a helix, and one further integration determines c as a helix. This in turn determines a screw-motion, and the surface arises as the trajectory of a line under the one-parameter group of all of these screw-motions. Conversely, for a helicoidal ruled surface the three determining quantities λ, F, J must be constant, since they are invariant under the one-parameter group of Euclidean motions. The case of $FJ + \lambda = 0$ reduces to that of pure surfaces of rotation, in which the screw-motion degenerates into a rotation because in this case we have $\mathrm{Det}(c', c'', c''') = 0$.

For the proof of (ii), the fact that $J = 0$ implies that X is a great circle with constant $X \times X'$, and $F = 0$ implies $c' = \lambda X \times X'$. Hence c is a line in the direction of $X \times X'$. If we choose $X \times X'$ as the vector $(0, 0, 1)$ in space, then we get the above parametrization of the right helicoid.

Part (iii) is an easy exercise: If the rotating line meets the axis of rotation, then we get either a plane or a double cone. If it is parallel to the axis of rotation, we get a cylinder, and if they have different slopes in space and do not meet, we get a hyperboloid of rotation, cf. Exercise 11 at the end of this chapter. A special case occurs when the surface degenerates to a plane minus an open disc, which occurs if the line lies on a plane which is perpendicular the axis of rotation. □

EXERCISE: Using standard parameters, calculate the Gaussian curvature and the mean curvature of a ruled surface as follows:

$$K = -\frac{\lambda^2}{(\lambda^2 + v^2)^2}, \quad H = -\frac{1}{2(\lambda^2 + v^2)^{3/2}} \Big(Jv^2 + \lambda' v + \lambda(\lambda J + F) \Big).$$

From this one can determine all ruled surfaces which fulfill $H \equiv 0$ with ease, cf. also Exercise 12. We see also that for a helicoidal ruled surface we have $K = 0 \Leftrightarrow \lambda = 0$ and that consequently either $K \equiv 0$ or $K \neq 0$ everywhere. The interesting case $K \equiv 0$ is the following.

3.24. Definition and Theorem. (Developable surfaces)

A ruled surface is said to be *developable* if it can be mapped locally to the plane, preserving the first fundamental form and the generating lines. Intuitively this means that one puts one of the generators as a line into the plane and then "develops" the strips of the surface on both sides into the plane, preserving both angles and lengths. For a ruled surface the following conditions are equivalent:

(1) The surface is developable.

(2) $K \equiv 0$.

(3) Along every one of the generators all the surface normals are parallel, i.e., the Gauss mapping is constant along each line.

A ruled surface which satisfies one of the conditions (1), (2) or (3) is also called a *torse* or a *developable.* Moreover, one has

(4) An open and dense subset of every torse consists of pieces of planes, cones, cylinders or tangent developable, where *tangent developables* or *tangent surfaces* are ruled surfaces for which the vector X is tangent to the curve c, for an example see Figure 3.15.

(5) Every surface element without level points and with $K \equiv 0$ is a ruled surface.

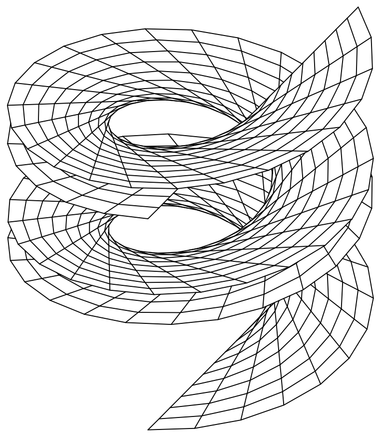


Figure 3.15. Tangent developable of a helix, also called "developable helicoid"

Proof: Without loss of generality we can assume $||X|| = 1$ in any case.

(2) $\Leftrightarrow$ (3): The unit normal $\nu(u, v)$ satisfies the equations $\langle \nu, X\rangle = 0, \langle \nu, c' + vX'\rangle = 0$. The derivative with respect to v of the first of these equations yields $\langle \frac{\partial \nu}{\partial v}, X\rangle = 0$. The derivative with respect to v of the second yields

$$\Big\langle \frac{\partial \nu}{\partial v}, c' + vX'\Big\rangle + \langle \nu, X'\rangle = 0.$$

But the vector $\frac{\partial \nu}{\partial v}$ is tangent to the surface. Hence the vanishing of $\frac{\partial \nu}{\partial v}$ is equivalent to $0 = \langle \nu, X'\rangle = \big\langle \nu, \frac{\partial^2 f}{\partial v \partial u}\big\rangle = h_{12}$. Now, for every ruled surface one has the equality $K = -(h_{12})^2 \big/ \mathrm{Det}(I)$ because of $h_{22} = \big\langle \frac{\partial X}{\partial v}, \nu\big\rangle = 0$; thus the statement.

(4): Here we have to consider the different cases where the three determining quantities of the ruled surface either vanish identically along an interval or are non-vanishing on an interval. We may disregard the endpoints of the interval for these considerations.

1st case: $X' = 0$ on an interval. Then $X(u) = X_0$ is constant, and the surface is a piece of a cylinder. A special case of this is the plane, if for example c is a line.

2nd case: $X' \neq 0$ on an interval. In this case we can introduce standard parameters according to 3.21, and $\frac{\partial \nu}{\partial v} = 0$ implies by the above $\nu = X \times X'$, since ν is perpendicular to X and X'. Then we get the relation $c' = FX$ for the three-frame $X, X', X \times X'$. If $c' = 0$ on an interval, then c is constant and the surface is a piece of a cone. If $c' \neq 0$, then we can conclude from $c' = FX$ that the vector field X is tangent to c, hence we have a tangent surface (with singularities along the curve, since the first fundamental form is degenerate there).

(2) $\Rightarrow$ (1): After what we have already shown, it only remains to show that the four named types of surfaces are developable. For a surface which is composed of elements, developments of the elements can be again composed, since one can always transform the generating lines into other generating lines. It is trivial that the plane is developable. For a cylinder, choose c in such a way that c' is a unit vector and orthogonal to the constant vector X_0. In these parameters the first fundamental form is given by $E = G = 1, F = 0$, which corresponds to the Euclidean metric in Cartesian coordinates. For a cone one gets similarly for the first fundamental form $E = v^2, G = 1, F = 0$. The same values are provided by polar coordinates in the plane. For a tangent surface in standard parameters, the quantities for the first fundamental form are given by $E = F^2 + v^2, F = \langle c', X \rangle, G = 1$ (note that for the determinant we have $EG - F^2 = v^2$). The same fundamental form is obtained locally if $c(u)$ is an arbitrary plane curve and X is a unit tangent to c with $c' = FX$. More precisely, we must consider the cases $v > 0$ and $v < 0$ separately. Anyway, it is now sufficient to develop the directrix in the plane, and as a result, every tangent surface is developable.

(1) $\Rightarrow$ (3): Here, instead of using standard parameters, we assume that the directrix c has been parametrized by arc length and that it is perpendicular to the vector field X of unit length. We can always choose such a directrix locally by applying the same Ansatz as in 3.12. The first fundamental form then becomes $E = 1 + 2v\langle X', c' \rangle + v^2 ||X'||^2, F = 0, G = 1$. By assumption there is a development to the plane. This maps c to a plane curve γ, which is also parametrized by arc length; similarly, X is mapped to a vector field ξ of unit length which is perpendicular to γ. The corresponding Frenet two-frame is $e_1 = \gamma'$ and $e_2 = \pm\xi$. Because

$\gamma' + v\xi' = e_1 + ve_2' = (1 \mp v\kappa)e_1 = (1 \mp v\kappa)\gamma'$ (where κ is the curvature of γ), the corresponding first fundamental form is determined by the relations $E^* = (1 \mp v\kappa)^2, F^* = 0, G^* = 1$. By assumption, the development preserves the first fundamental form and hence we have $E = E^*, F = F^*, G = G^*$; in particular we have $(1 \mp v\kappa)^2 = 1 + 2v\langle X', c'\rangle + v^2||X'||^2$. Comparing coefficients for v yields $\kappa^2 = ||X'||^2$ and $\langle X', c'\rangle = \mp\kappa$. Since c' is a unit vector, this is only possible if c' and X' are linearly dependent. But in this case the unit normal to the surface is simply $\nu = \pm c' \times X$ and therefore depends only on u and not on v. Hence ν is constant along every line.

It only remains to show (5). By assumption there are no umbilics, hence both principal curvature directions are uniquely defined. We now utilize lines of curvature parameters (u, v), such that $L(\frac{\partial f}{\partial v}) = 0$ and $L(\frac{\partial f}{\partial u}) = \mu(u, v)\frac{\partial f}{\partial u}$, so that $\frac{\partial \nu}{\partial v} = 0$ and $\frac{\partial \nu}{\partial u} = -\mu\frac{\partial f}{\partial u}$ as well as $\left\langle\frac{\partial f}{\partial u}, \frac{\partial f}{\partial v}\right\rangle = 0$. In particular we then have $\frac{\partial^2\nu}{\partial u\partial v} = 0$. We claim that the curve $c(v) = f(u_0, v)$ is a Euclidean line for every fixed value of u_0. Setting $\dot{c} = \frac{\partial f}{\partial v}$, $\ddot{c} = \frac{\partial^2 f}{\partial v^2}$, we have

$$\langle\ddot{c}, \nu\rangle = I\!I(\dot{c}, \dot{c}) = 0,$$

$$\left\langle\ddot{c}, \frac{\partial f}{\partial u}\right\rangle = -\left\langle\frac{\partial f}{\partial v}, \frac{\partial^2 f}{\partial u\partial v}\right\rangle = \left\langle\frac{\partial f}{\partial v}, \frac{\partial}{\partial v}\Big(\frac{1}{\mu}\frac{\partial \nu}{\partial u}\Big)\right\rangle = \left\langle\frac{\partial f}{\partial v}, \frac{1}{\mu}\frac{\partial^2 \nu}{\partial u\partial v}\right\rangle = 0.$$

Thus the two vectors $\dot{c}$ and $\ddot{c}$ are linearly dependent, and it follows from this that c is a line (up to parametrization). Indeed, if we choose the arc length as in 2.2 (by means of a reparametrization), then we have on the one hand that c'' and c' are orthogonal, and on the other that they are linearly dependent. □

Remark: The implication (1) ⇒ (2) is particularly important for the intrinsic geometry of the situation. This will be seen more clearly later when we discuss the Theorema Egregium 4.16, which says that the Gaussian curvature is already determined by the first fundamental form. Thus the Gaussian curvature must always vanish when the first fundamental form is Euclidean, cf. the remark in 4.30. In particular we obtain a different (and surely more beautiful) proof of the equivalence of (1) and (2) above.

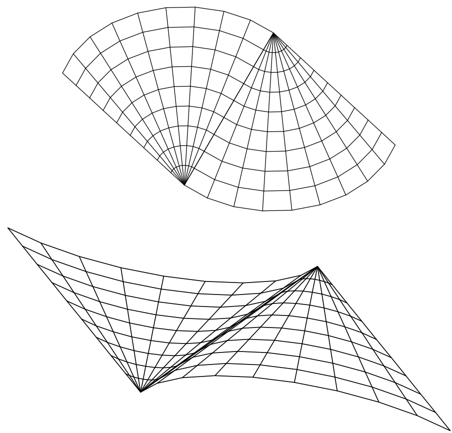


Figure 3.16. Surface with $K = 0$ which is not ruled

3.25. Example. (A surface with $K = 0$ which is not a ruled surface) We consider a cone over a curve $c(x) = (x, 0, z(x))$ in the $(x, 0, z)$-plane whose vertex is at the point $(0, 1, 0)$. We assume that this cone contains the point $(0, -1, 0)$. Similarly we consider a second cone over the same curve with vertex at $(0, -1, 0)$, and assume that this cone contains the point $(0, 1, 0)$. Suppose the curve c is chosen in such a way that it passes through the origin as follows: $(0, 0, 0) = c(0)$ and $c'(0) = (1, 0, 0)$. Moreover assume that all higher derivatives of c at this point (and only at this point) vanish identically. Curves with these properties can be constructed explicitly with the help of the function $\exp(-x^{-2})$. Under these assumptions, we can proceed to connect the part of one cone with $x \geq 0$ with the part of the other cone with $x \leq 0$. In this way we get a surface of class C^∞ with level points along the $(0, y, 0)$-axis. The surface we have constructed is, in a neighborhood of the level point, not a ruled surface in the sense of Definition 3.20, since it cannot be parametrized in the class C^2 by lines because the vector field is of the type $X(x) = (|\sin x|, \cos x, 0)$, up to terms of higher order. Therefore, the lines have a "bend" in the first derivative near where the two cone parts are put together, see the first picture in Figure 3.16. It follows that C^2-parameters on the surface must be chosen differently.

A more precise description is as follows. Set $A = \{(x,y) \in I\!R^2 \mid x^2 + (y+1)^2 < 4, x \geq 0, -1 < y < 1\} \cup \{(x,y) \in I\!R^2 \mid x^2 + (y-1)^2 < 4, x \leq 0, -1 < y < 1\}$. A is the union of two quarter-circles, which are joined along the segment $-1 < y < 1$ on the y-axis. The midpoints, $(0,-1)$ and $(0,1)$, respectively, must be excluded; later they will be used as the vertices of the two cones. On A we define "crossed polar coordinates" (r,φ) by

$$r := \sqrt{(1+y)^2 + x^2} - 1, \quad \text{in case } x \geq 0, \qquad \cos\varphi := \tfrac{1+y}{1+r}, \quad \varphi \geq 0,$$

$$r := 1 - \sqrt{(1-y)^2 + x^2}, \quad \text{in case } x \leq 0, \qquad \cos\varphi := \tfrac{1-y}{1-r}, \quad \varphi \leq 0.$$

Here, in contrast to our usual policy, we have chosen the "radius" r as $-1 < r < 1$. For $x = 0$ we get $\varphi = 0$ and $r = y$, hence the two definitions fit together continuously. We then set

$$f(r,\varphi) := \begin{cases} (0,-1,0) \;+\; (r+1)\big[c(\varphi) + (0,1,0)\big], & \text{in case } \varphi \geq 0, \\ (0,1,0) \;-\; (r-1)\big[c(\varphi) - (0,1,0)\big], & \text{in case } \varphi \leq 0. \end{cases}$$

For $\varphi = 0$ we get $f(r,0) = (0,r,0)$, and for $r = 0$ we have the equality $f(0,\varphi) = c(\varphi)$. Thus the two definitions fit together here also, in a C^∞ manner, since the tangent plane depends only on $c(\varphi)$ and $\dot{c}(t)$. Thus the surface materializes as the graph of a C^∞-function over the (x,y)-plane. The vector field ist $X = \frac{\varphi}{|\varphi|}c(\varphi) + (0,1,0)$. Therefore it is not differentiable for $\varphi = 0$. Another example is implicitly given by assigning the first and second fundamental form, as described in [**2**], 3.9.4 (pp. 68-69).

3.26. Definition and Theorem. (Weingarten surface, W-surface)
A *Weingarten surface* or *W-surface* is a surface for which a nontrivial relation holds between the two principal curvatures (or between H and K), i.e., if there is a function Φ in two variables with $\Phi(\kappa_1,\kappa_2) = 0$ (resp. $\Phi(H,K) = 0$). Then:

1. Every surface of rotation is a Weingarten surface.
2. Among the ruled surfaces, the class of Weingarten surfaces is precisely the set of all developable surfaces and all helicoidal ruled surfaces.[4]

[4] This theorem was proved independently in 1865 by E. Beltrami and U. Dini.

PROOF: 1. For a surface of rotation each curvature depends on only one parameter. Hence the gradients of H and K in the (r, φ)-plane are linearly dependent, which implies that the level curves coincide. More explicitly, if we set $r'' = -rK$ in the equation $2H = h''/r' + h'/r$ and use the fact that $r'^2 + h'^2 = 1$ and $r'r'' + h'h'' = 0$, we get

$$2H = \frac{r}{\sqrt{1 - r'^2}} K + \frac{\sqrt{1 - r'^2}}{r}.$$

On the other hand, r can be interpreted as a function of H or of K unless H and K both are stationary. We again recognize the unit sphere as the special solution for which $r = \sqrt{1 - r'^2}$.

2. A surface with $K = 0$ belongs to the set of Weingarten surfaces, as can be seen by simply setting $\Phi(H, K) := K$. If we consider a ruled surface which is not developable, then for the expressions for H and K above in standard parameters (see p. 88) we get

$$2H = -\frac{J}{(\lambda^2 + v^2)^{1/2}} - \frac{\lambda' v}{(\lambda^2 + v^2)^{3/2}} - \frac{\lambda F}{(\lambda^2 + v^2)^{3/2}}.$$

Here we can replace $\lambda^2 + v^2$ by $\sqrt{-\lambda^2/K}$ throughout, because $K = -\frac{\lambda^2}{(\lambda^2+v^2)^2}$, and v can be replaced by the expression $\sqrt{\sqrt{-\lambda^2/K} - \lambda^2}$. Since v no longer appears explicitly in these expressions, a nontrivial relation between H and K can only hold if all coefficients (which depend only on u) are constant in the expression above for $2H$, which means that J, F, λ are constant. Then the second summand in the expression vanishes, and the equation between H and K is necessarily of the form (at least for $\lambda > 0$)

$$2H = -\frac{J}{\lambda^{1/2}}(-K)^{1/4} - \frac{F}{\lambda^{1/2}}(-K)^{3/4}.$$

The statement then follows from 3.23 (i). □

3.27. Example. (Surfaces of rotation with a linear relation between the principal curvatures according to H. Hopf[5])
We are looking for surfaces of rotation with a constant ratio of the two principal curvatures, for example $\kappa_1 = c\kappa_2$, with a constant $c \neq 0$ (the case $c = 0$ is contained in 3.17). For this, it is convenient to choose the

[5] *Über Flächen mit einer Relation zwischen den Hauptkrümmungen*, Math. Nachr. **4**, 232–249 (1951).

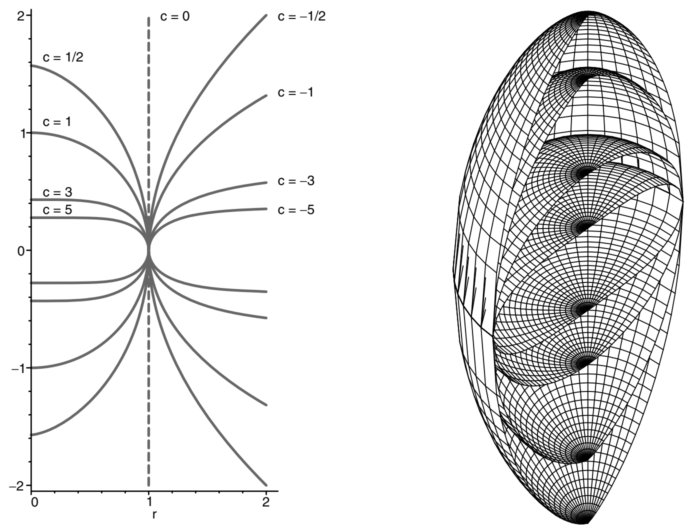


Figure 3.17. Several surfaces of rotation with constant $c = \kappa_1/\kappa_2$

parametrization $f(r, \varphi) = (r\cos\varphi, r\sin\varphi, h(r))$. By 3.16 the principal curvatures are given by the following expressions:

$$\kappa_1 = \frac{h''}{(1 + h'^2)^{3/2}}, \quad \kappa_2 = \frac{h'}{r(1 + h'^2)^{1/2}};$$

in particular one has $\kappa_1 = \frac{d}{dr}(r\kappa_2)$. The equation $\kappa_1 = c\kappa_2$ is then equivalent to the differential equation $c\kappa_2 = (r\kappa_2)'$ or $(c-1)\kappa_2 = r\kappa_2'$ with the solution $\kappa_2 = br^{c-1}$, where b is a constant. If we set $b = 1$ for simplicity, then the surface is described in terms of the variable c (at least for $c \neq 0$) by the relation

$$f_c(r, \varphi) = \Big(r\cos\varphi, r\sin\varphi, \pm\int_1^r t^c(1 - t^{2c})^{-1/2}dt\Big).$$

Here f_1 is the unit sphere, f_{-1} is the catenoid (Figure 3.19, see also 3.37). In General Relativity $f_{-1/2}$ is known as *Flamm's paraboloid*, and f_2 is a mathematical model for the *mylar balloon*. The circular cylinder can be inserted as the particular member f_0 of this family. The surface f_c is real-analytic even at the apex $r = 0$ if c is an odd integer. The point $r = 0$ on the axis of rotation is a level point for $c > 1$, but a singularity for $0 < c < 1$, see Fig. 3.17.

3D Minimal surfaces

A soap film which is spanned across a fixed boundary will, for physical reasons, minimize its surface area.[6] In this way very interesting mathematical phenomena occur (see [**10**]), which we will treat to a certain extent in this section. More precisely, we are concerned with regular surface elements which (at least locally) minimize the surface area and are therefore referred to as *minimal surfaces*. Surprisingly, this leads to unexpected connections with the theory of functions of a complex variable. For this reason, we will be using, only in this section, some basic facts from that theory, like the notion of holomorphic and meromorphic functions, Cauchy-Riemann differential equations and complex contour integrals. For background on this, we recommend the books [**35**] and [**36**].

Problem. (Surfaces with minimal surface area)
For a given boundary curve, find the surface spanned by that boundary with the smallest possible surface area, or find geometric conditions which such surfaces must fulfill.

In order to find geometric conditions, we assume that we have a surface element $f(u_1, u_2)$ which has minimal surface area, and consider a *variation in the normal direction* (also known as a *normal variation*) of the following kind:

$$f_\varepsilon(u_1, u_2) := f(u_1, u_2) + \varepsilon \cdot \varphi(u_1, u_2) \cdot \nu(u_1, u_2),$$

where φ is an arbitrary C^2-function, which vanishes at the boundary. In other words, we consider a suitable 1-parameter family f_ε of surface elements, all with the same boundary, such that our given surface element occurs as the particular member $f = f_0$. Then we are going to calculate the derivative of the surface area of f_ε at $\varepsilon = 0$. For this we need a few prerequisites, as follows.

For sufficiently small $|\varepsilon|$, the surface element f_ε is regular, as follows from the equation

$$\frac{\partial f_\varepsilon}{\partial u_i} = \frac{\partial f}{\partial u_i} + \varepsilon \frac{\partial \varphi}{\partial u_i} \cdot \nu + \varepsilon \cdot \varphi \cdot \frac{\partial \nu}{\partial u_i}$$

[6]For a mathematical video on minimal surfaces with soap film experiments see `http://www.youtube.com/watch?v=X9YrqUxJzSY`.

by calculation of the first fundamental form:

$$
\begin{aligned}
g_{ij}^{(\varepsilon)} &= \Big\langle \frac{\partial f_\varepsilon}{\partial u_i}, \frac{\partial f_\varepsilon}{\partial u_j} \Big\rangle \\
&= g_{ij} + 2\varepsilon\varphi \Big\langle \frac{\partial f}{\partial u_i}, \frac{\partial \nu}{\partial u_j} \Big\rangle + \varepsilon^2 \Big(\varphi^2 \Big\langle \frac{\partial \nu}{\partial u_i}, \frac{\partial \nu}{\partial u_j} \Big\rangle + \frac{\partial \varphi}{\partial u_i} \frac{\partial \varphi}{\partial u_j} \Big) \\
&= \underbrace{g_{ij} - 2\varepsilon\varphi h_{ij}}_{\text{linearization}} + O(\varepsilon^2).
\end{aligned}
$$

If also g_{ij} is positive definite, then g_{ij}^{ε} remains positive definite for sufficiently small values of $|\varepsilon|$, where φ can be chosen arbitrarily. Comparison of the surface area $\int_U dA = \int_U \sqrt{\mathrm{Det} g_{ij}}\, du_1 du_2$ of $f = f_0$ with the surface area $\int_U \sqrt{\mathrm{Det} g_{ij}^{(\varepsilon)}}\, du_1 du_2$ of f_ε for small values of ε yields the following relation (we again set $g := \mathrm{Det}(g_{ij}) = g_{11}g_{22} - g_{12}^2$):

$$
0 = \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \Big(\int_U \sqrt{\mathrm{Det}(g_{ij}^{(\varepsilon)})}\, du_1 du_2 \Big)
$$

$$
\begin{aligned}
&= \int_U \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \sqrt{\mathrm{Det}(g_{ij}^{(\varepsilon)})}\, du_1 du_2 = \int_U \frac{\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \big(\mathrm{Det} g_{ij}^{(\varepsilon)}\big)}{2\sqrt{\mathrm{Det} g_{ij}}} du_1 du_2 \\
&= \int_U \frac{1}{2\sqrt{g}} \Big(\frac{\partial g_{11}^{(\varepsilon)}}{\partial \varepsilon}\Big|_{\varepsilon=0} g_{22} + g_{11} \frac{\partial g_{22}^{(\varepsilon)}}{\partial \varepsilon}\Big|_{\varepsilon=0} - 2g_{12} \frac{\partial g_{12}^{(\varepsilon)}}{\partial \varepsilon} \Big) du_1 du_2 \\
&= \int_U \frac{1}{2\sqrt{g}} \big((-2\varphi h_{11}) g_{22} + g_{11}(-2\varphi h_{22}) - 2g_{12}(-2\varphi h_{12}) \big) du_1 du_2 \\
&= -\int_U \varphi \cdot \underbrace{\frac{1}{g} \big(h_{11}g_{22} + h_{22}g_{11} - 2h_{12}g_{12} \big)}_{=2H} \sqrt{g}\; du_1 du_2 \\
&= -\int_U \varphi \cdot 2H \underbrace{\sqrt{g}\; du_1 du_2}_{=dA},
\end{aligned}
$$

where in the last line we have used the formula for H from 3.13. If we choose $\varphi = H$ in the interior (such that φ decreases in size towards the boundary), we get the following result:

3.28. Theorem and Definition. (Minimal surface)
Let $U \subset \mathbb{R}^2$ be an open set, let $\overline{U}$ be a compact set with boundary ∂U, and let $f \colon \overline{U} \to \mathbb{R}^3$ be a surface element. A necessary condition for the surface area of f to be less than or equal to the surface areas of all normal variations

$$f_\varepsilon \colon \overline{U} \to \mathbb{R}^3 \text{ with } f_\varepsilon\big|_{\partial U} = f\big|_{\partial U}$$

is the vanishing of the mean curvature H in all of U. One calls a surface element with $H \equiv 0$ a *minimal surface.*

Remark: Strictly speaking the equation $H \equiv 0$ only expresses the fact that the surface area is *stationary* (a critical point), so that a change in H must be of higher than linear order. For example, it could also be maximal or "saddle-like", just as saddle points occur in minimization problems in several variables. Still the notion of "minimal surface" is generally introduced as in Definition 3.28. It is in fact the case that minimal surfaces *locally* minimize the surface area, cf. [**11**].

If we are given a surface $f \colon U \longrightarrow \mathbb{R}^3$ of class C^∞, then one can define on the space of all such f with fixed $f\big|_{\partial U}$ the *area functional* $\mathbf{A}$ by the formula

$$\mathbf{A}(f) := \int_U \sqrt{g}\; du_1 du_2.$$

Earlier we computed that the "directional derivative" of $\mathbf{A}$ in the direction of a normal variation φ can be expressed as follows:

$$D_\varphi \mathbf{A}(f) = \frac{\partial \mathbf{A}(f_\varepsilon)}{\partial \varepsilon}\Big|_{\varepsilon=0} = -2 \int_U \varphi \cdot H \cdot \sqrt{g}\; du_1 du_2.$$

The *gradient* of $\mathbf{A}$ is then uniquely determined by the equation

$$\langle\!\langle \mathbf{grad}\ \mathbf{A}(f), \varphi \rangle\!\rangle = D_\varphi \mathbf{A}(f),$$

where the inner product $\langle\!\langle\ ,\ \rangle\!\rangle$ is defined on the space of C^∞-functions on U by

$$\langle\!\langle \psi_1, \psi_2 \rangle\!\rangle := \int_U \psi_1 \psi_2 \sqrt{g}\; du_1 du_2.$$

This inner product is positive definite and therefore in particular non-degenerate. Only the null-function (whose values are identically zero) is "perpendicular" in this sense to all functions φ. Hence

grad $\mathbf{A}(f) = -2H$, since

$$\langle\!\langle -2H, \varphi \rangle\!\rangle = -2\int_U \varphi \cdot H\sqrt{g}\, du_1 du_2 = D_\varphi \mathbf{A}(f).$$

This shows the following fact: if f is not a minimal surface, then the "evolution" $f_\varepsilon = f + \varepsilon H\nu$ leads to a surface whose surface area is strictly smaller.

In the definition of the angle between two tangent vectors X, Y, only the value of the inner product $\langle X, Y\rangle$ in relation to the length of both vectors occurs, cf. 1.1. Therefore the angle is preserved if both are multiplied by the same factor. This is precisely what characterizes an "angle preserving mapping", for an example see Figure 3.18 and Exercise 9 at the end of this chapter.

3.29. Definition. (conformal parametrization)
A parametrization $f\colon U \longrightarrow I\!R^3$ of a surface element is called *conformal* or *angle preserving*, if the first fundamental form, written in these parameters, is a scalar multiple of the unit matrix, i.e., if the equation

$$(g_{ij}) = \lambda(u_1, u_2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

holds for some function $\lambda\colon U \to I\!R$. More generally, two surface elements which are parametrized by the same set $f, \widetilde{f}\colon U \to I\!R^3$ are said to be *conformally equivalent*, if the first fundamental form (g_{ij}) of one of them is a scalar multiple of the other $(\widetilde{g_{ij}})$:

$$(\widetilde{g_{ij}}) = \lambda(u_1, u_2)(g_{ij}), \qquad (u_1, u_2) \in U,$$

with some positive function $\lambda\colon U \to I\!R$. A similar statement holds after a change of parameters: $f\colon U \to I\!R^3, \widetilde{f}\colon \widetilde{U} \to I\!R^3$ are said to be *conformally equivalent* with the *conformal factor* λ, if there is a change of parametrizations $\Phi\colon U \to \widetilde{U}$ such that

$$\left\langle \frac{\partial(\widetilde{f}\circ\Phi)}{\partial u_i}, \frac{\partial(\widetilde{f}\circ\Phi)}{\partial u_j} \right\rangle = \lambda(u_1, u_2) \cdot \left\langle \frac{\partial f}{\partial u_i}, \frac{\partial f}{\partial u_j} \right\rangle \quad \text{for all } i, j.$$

A conformal parametrization is also called *isothermal*, and the parameters are also referred to as *isothermal coordinates*. The notion of angles is then identical with that in the Euclidean (u_1, u_2)-plane.

3.30. Consequence.

(i) The Gauss map ν is conformal for a minimal surface with $K \neq 0$, i.e., ν and f are conformally equivalent, with conformal factor $-K$.

(ii) The following relation holds for a conformal parametrization $f\colon U \longrightarrow I\!R^3$ with $(g_{ij}) = \lambda\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$:
$$\frac{\partial^2 f}{\partial u_1^2} + \frac{\partial^2 f}{\partial u_2^2} = 2H\lambda \cdot \nu.$$
In particular, a conformal parametrization f defines a minimal surface if and only if the three component functions f_1, f_2, f_3 of f are *harmonic*, which means that the relation
$$\Delta f_i = \frac{\partial^2 f_i}{\partial u_1^2} + \frac{\partial^2 f_i}{\partial u_2^2} = 0$$
holds. The vector $\mathbf{H} = H \cdot \nu$ is also called the *mean curvature vector.*

Proof: (i) follows immediately from the equation $I\!I\!I - 2H \cdot I\!I + K \cdot I = 0$ from 3.10, which implies that $I\!I\!I = -K \cdot I$. But $I\!I\!I$ is the first fundamental form of the map ν.

For (ii) we start with the observation
$$\left\langle \frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_1} \right\rangle = \lambda = \left\langle \frac{\partial f}{\partial u_2}, \frac{\partial f}{\partial u_2} \right\rangle, \left\langle \frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2} \right\rangle = 0.$$
By further differentiating it follows from this that
$$\left\langle \frac{\partial^2 f}{\partial u_1^2}, \frac{\partial f}{\partial u_1} \right\rangle = \left\langle \frac{\partial^2 f}{\partial u_1 \partial u_2}, \frac{\partial f}{\partial u_2} \right\rangle = -\left\langle \frac{\partial f}{\partial u_1}, \frac{\partial^2 f}{\partial u_2^2} \right\rangle.$$
We conclude $\left\langle \frac{\partial^2 f}{\partial u_1^2} + \frac{\partial^2 f}{\partial u_2^2}, \frac{\partial f}{\partial u_1} \right\rangle = 0$, and similarly $\left\langle \frac{\partial^2 f}{\partial u_1^2} + \frac{\partial^2 f}{\partial u_2^2}, \frac{\partial f}{\partial u_2} \right\rangle = 0$. Hence the vector $\frac{\partial^2 f}{\partial u_1^2} + \frac{\partial^2 f}{\partial u_2^2}$ is perpendicular to the tangent plane, thus linearly dependent on the unit normal ν. But since $2H = \lambda^{-2}(g_{22}h_{11} + g_{11}h_{22}) = \lambda^{-1}(h_{11} + h_{22})$, we have finally
$$\left\langle \frac{\partial^2 f}{\partial u_1^2} + \frac{\partial^2 f}{\partial u_2^2}, \nu \right\rangle = h_{11} + h_{22} = 2H\lambda. \qquad \square$$

3.31. Corollary. (Complexification)
For a surface element $f\colon U \to I\!R^3$ with components $f = (f_1, f_2, f_3)$ we define the map $\varphi\colon U \to \mathbb{C}^3$ by the relation $\varphi(u+iv) = \frac{\partial f}{\partial u}(u,v) - i\frac{\partial f}{\partial v}(u,v)$, which is, in components,

$$\varphi_1(u+iv) = \frac{\partial f_1}{\partial u}(u,v) - i\frac{\partial f_1}{\partial v}(u,v);$$
$$\varphi_2(u+iv) = \frac{\partial f_2}{\partial u}(u,v) - i\frac{\partial f_2}{\partial v}(u,v);$$
$$\varphi_3(u+iv) = \frac{\partial f_3}{\partial u}(u,v) - i\frac{\partial f_3}{\partial v}(u,v).$$

Then we have:

(i) f is conformal if and only if $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = 0$.

(ii) If f is a conformal parametrization, f is a minimal surface if and only if the functions $\varphi_1, \varphi_2, \varphi_3$ are complex analytic (holomorphic).

(iii) If conversely $\varphi_1, \varphi_2, \varphi_3$ are complex analytic with $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = 0$, then the function f defined by the above equation is regular (hence an immersion) if and only if $\varphi_1\overline{\varphi}_1 + \varphi_2\overline{\varphi}_2 + \varphi_3\overline{\varphi}_3 \neq 0$.

In what follows we use the following basic fact from the theory of functions of a complex variable. A complex valued function $\varphi(u+iv) = x(u,v) + iy(u,v)$ with real quantities u, v, x, y is *complex analytic* or *holomorphic* if and only if the *Cauchy-Riemann differential equations* (abbreviated as CR-equations)

$$\frac{\partial x}{\partial u} = \frac{\partial y}{\partial v}, \quad \frac{\partial x}{\partial v} = -\frac{\partial y}{\partial u}$$

hold. One also writes $x = \text{Re}\,\varphi, y = \text{Im}\,\varphi$ for the real and imaginary parts of the function φ. The CR-equations are equivalent to the fact that the (real) Jacobi matrix of φ is at every point a composition of a rotation and a scalar multiplication (with varying angle of rotation and scalar multiple).

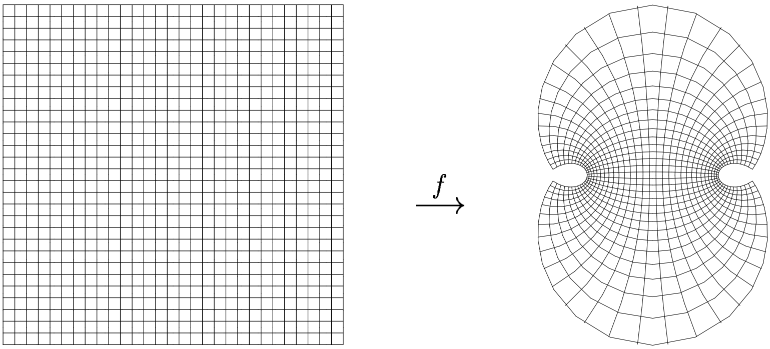


Figure 3.18. Coordinate grid and its conformal image under the complex function $f(z) = \frac{z}{z^2+1}$

Proof (of 3.31): (i): By definition we have

$$\begin{aligned}\varphi_1^2 + \varphi_2^2 + \varphi_3^2 &= \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial u} \right\rangle + i^2 \left\langle \frac{\partial f}{\partial v}, \frac{\partial f}{\partial v} \right\rangle - 2i \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \right\rangle \\ &= g_{11} - g_{22} - 2i g_{12}.\end{aligned}$$

Thus the left hand side vanishes if and only if $g_{11} = g_{22}$ and $g_{12} = 0$ hold.

(ii): We calculate the second derivatives:

$$\frac{\partial^2 f_k}{\partial u^2} = \frac{\partial}{\partial u}(\mathrm{Re}\ \varphi_k), \qquad \frac{\partial^2 f_k}{\partial v^2} = -\frac{\partial}{\partial v}(\mathrm{Im}\ \varphi_k),$$
$$\frac{\partial^2 f_k}{\partial u \partial v} = \frac{\partial}{\partial v}(\mathrm{Re}\ \varphi_k) = -\frac{\partial}{\partial u}(\mathrm{Im}\ \varphi_k).$$

The validity of the equation $\frac{\partial^2 f}{\partial u^2} + \frac{\partial^2 f}{\partial v^2} = 0$ is then equivalent to the validity of the CR-equations for $\varphi_1, \varphi_2, \varphi_3$. By 3.30, this is precisely the case when f is a minimal surface.

(iii): We have $\varphi_1 \overline{\varphi}_1 + \varphi_2 \overline{\varphi}_2 + \varphi_3 \overline{\varphi}_3 = \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial u} \right\rangle + \left\langle \frac{\partial f}{\partial v}, \frac{\partial f}{\partial v} \right\rangle \geq 0$, with equality if and only if $\frac{\partial f}{\partial u} = \frac{\partial f}{\partial v} = 0$. By the proof of (i), either both vectors vanish or both are non-vanishing and linearly independent. This implies (iii). □

The zeros of the complex map φ correspond by (iii) to the points at which f is not regular (so-called singularities of f). The reason for these considerations is that in the theory of functions of a complex variable it does not make sense to exclude zeros, while in differential geometry one usually assumes regular surface elements.

3.32. Corollary. Let $U \subset \mathbb{C}$ be a simply connected domain and $\varphi_k \colon U \to \mathbb{C}$ $(k = 1, 2, 3)$ given holomorphic functions with $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = 0$ and $\varphi_1\bar{\varphi}_1 + \varphi_2\bar{\varphi}_2 + \varphi_3\bar{\varphi}_3 \neq 0$. Then the mapping $f \colon U \to I\!R^3$ defined by

$$f_k(z) = \mathrm{Re} \int_{z_0}^{z} \varphi_k(\zeta) d\zeta, \qquad k = 1, 2, 3,$$

is a regular minimal surface element.

PROOF: First we convince ourselves that the integration of the equations in 3.31 does not require prior knowledge of partial differential equations. Indeed, one can use the well-known fact that locally a holomorphic function possesses a complex primitive, which can be obtained by integration along an appropriately chosen curve (for example in a star-shaped domain). Passing from this local consideration to the global situation requires that the integration is independent of the path of integration, which means that the primitive is globally defined. This is why we must assume that the domain U is simply connected. For details on this argument we refer to [**35**]. Assuming this, 3.32 follows directly from 3.31. Without the assumption $|\varphi| \neq 0$ one gets a real minimal surface element with isolated singularities. □

3.33. Lemma. (Isothermal coordinates)
In lines of curvature parameters (u_1, u_2) for a minimal surface with $K \neq 0$ (this means $g_{12} = h_{12} = 0$), one can construct by a simple integration a conformal parametrization, that is, isothermal coordinates.

PROOF: In lines of curvature parameters we have $\frac{\partial \nu}{\partial u_i} = -\kappa_i \frac{\partial f}{\partial u_i}$, hence

$$\frac{\partial^2 \nu}{\partial u_1 \partial u_2} = -\frac{\partial}{\partial u_1}\Big(\kappa_2 \cdot \frac{\partial f}{\partial u_2}\Big) = -\frac{\partial}{\partial u_2}\Big(\kappa_1 \cdot \frac{\partial f}{\partial u_1}\Big).$$

If we now set $\kappa := \kappa_1 = -\kappa_2 > 0$, it follows that

$$\frac{\partial \kappa}{\partial u_2}\frac{\partial f}{\partial u_1} + 2\kappa \frac{\partial^2 f}{\partial u_1 \partial u_2} + \frac{\partial \kappa}{\partial u_1}\frac{\partial f}{\partial u_2} = 0,$$

and hence

$$\Big\langle \frac{\partial\kappa}{\partial u_2}\frac{\partial f}{\partial u_1} + 2\kappa\frac{\partial^2 f}{\partial u_1\partial u_2} + \frac{\partial\kappa}{\partial u_1}\frac{\partial f}{\partial u_2}, \frac{\partial f}{\partial u_1}\Big\rangle = 0;$$

consequently

$$\frac{\partial\kappa}{\partial u_2}g_{11} + \kappa\frac{\partial g_{11}}{\partial u_2} + \frac{\partial\kappa}{\partial u_1}\cdot 0 = \frac{\partial}{\partial u_2}\Big(k\cdot g_{11}\Big) = 0.$$

It follows that $\kappa \cdot g_{11}$ is constant in the u_2-direction, hence it is a function of u_1 only. Similarly, $\kappa \cdot g_{22}$ is constant in the u_1-direction, hence a function of u_2 only. We now set $\kappa \cdot g_{11} = \Phi_1(u_1) > 0$ and $\kappa \cdot g_{22} = \Phi_2(u_2) > 0$ as well as $v_i := \int \sqrt{\Phi_i(u_i)}du_i, \quad i = 1, 2$. Then the Jacobi matrix of the transformation $(u_1, u_2) \mapsto (v_1, v_2)$, which is

$$\Big(\frac{\partial v_i}{\partial u_j}\Big) = \begin{pmatrix}\sqrt{\Phi_1} & 0\\ 0 & \sqrt{\Phi_2}\end{pmatrix},$$

is of maximal rank. The line element ds^2 transforms under this above transformation as follows:

$$\begin{aligned} ds^2 &= g_{11}du_1^2 + g_{22}du_2^2 \\ &= \frac{\Phi_1}{\kappa}du_1^2 + \frac{\Phi_2}{\kappa}du_2^2 \quad = \quad \frac{\Phi_1}{\kappa}\Big(\frac{1}{\sqrt{\Phi_1}}dv_1\Big)^2 + \frac{\Phi_2}{\kappa}\Big(\frac{1}{\sqrt{\Phi_2}}\cdot dv_2\Big)^2 \\ &= \frac{1}{\kappa}\big(dv_1^2 + dv_2^2\big). \end{aligned}$$

Thus v_1, v_2 are isothermal coordinates, i.e., $I(v_1, v_2) = \frac{1}{\kappa}\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$. This holds whenever f is a minimal surface without level points, i.e., in case $H \equiv 0$ and $K \neq 0$. It is easy to see that near a level point the factor $\frac{1}{\kappa}$ can become arbitrarily large and that this process will not work near such a point. $\square$

3.34. Corollary. (Analyticity)
Let $f: U \to I\!R^3$ be a minimal surface element without level points, with $f \in C^3$. Then there is a parametrization with the property that the three component functions are real-analytic (C^ω), hence locally developable in a Taylor series.

This follows from 3.31 - 3.33 together with the fact that a complex-analytic function is in particular real-analytic.

Summarizing our results up to this point, we have that every minimal surface locally allows a conformal parametrization, provided no level points occur. In this conformal parametrization the surface is analytic, occurring as the real part of a complex-analytic function, cf. 3.32. For a given φ with the constraints $|\varphi| \neq 0$ and $\varphi_1^2+\varphi_2^2+\varphi_3^2=0$ one even gets according to 3.32 a completely explicit minimal surface. A natural question at this point is whether we can freely prescribe the function without the use of a constraint. The answer is provided by the so-called *Weierstrass representation*. This allows a more or less free choice of two functions.

3.35. Lemma. One can associate to three arbitrarily given holomorphic functions $\varphi_1, \varphi_2, \varphi_3 \colon U \to \mathbb{C}$ with $\varphi_1^2+\varphi_2^2+\varphi_3^2=0$ (where we assume that none of the φ_i vanishes identically) a holomorphic function $F \colon U \to \mathbb{C}$ and a meromorphic function $G \colon U \to \mathbb{C} \cup \{\infty\}$ with the property that FG^2 is holomorphic and

$$\varphi_1 = \frac{F}{2}(1-G^2), \quad \varphi_2 = \frac{iF}{2}(1+G^2), \quad \varphi_3 = FG.$$

Conversely, every given pair (F, G) of such functions induces a corresponding φ with $\varphi_1^2+\varphi_2^2+\varphi_3^2=0$.

Proof: For a given φ we set

$$F = \varphi_1 - i\varphi_2, \quad G = \frac{\varphi_3}{\varphi_1 - i\varphi_2}.$$

This is well-defined except in the case that $\varphi_1 = i\varphi_2$, which implies that in addition $\varphi_3 = 0$, which has been excluded by assumption. Thus we get

$$FG^2 = \frac{\varphi_3^2}{\varphi_1 - i\varphi_2} = -\frac{\varphi_1^2+\varphi_2^2}{\varphi_1 - i\varphi_2} = -(\varphi_1 + i\varphi_2),$$

which is a holomorphic function. The equations

$$\varphi_1 = \frac{F}{2}(1-G^2), \quad \varphi_2 = \frac{iF}{2}(1+G^2), \quad \varphi_3 = FG$$

can be easily verified with the aid of the definitions. Conversely, let F and G be given; then the corresponding $\varphi_1, \varphi_2, \varphi_3$ fulfill the equation

$$\varphi_1^2+\varphi_2^2+\varphi_3^2 = \frac{F^2}{4}(1-G^2)^2 - \frac{F^2}{4}(1+G^2)^2 + F^2G^2 = 0.$$

Moreover, φ_1, φ_2 are holomorphic, since FG^2 is so. In any case FG is also holomorphic, hence φ_3 is holomorphic. $\square$

In addition, the relation $|\varphi|^2 = 0$ can hold at a point only if $F = FG = FG^2 = 0$ there. The excluded case $\varphi_1 = i\varphi_2$ and $\varphi_3 = 0$ corresponds geometrically to a plane which is parallel to the (x_1, x_2)-plane. This is clear from the formulas in 3.31. The following Weierstrass representation therefore excludes this case. Apart from this, the two functions F and G can be defined essentially arbitrarily, inducing (at least locally) a corresponding minimal surface, given by a completely explicit formula.

3.36. Corollary. (Weierstrass representation) Every conformal parametrized minimal surface f that is not a plane can locally be represented as follows:

$$\begin{aligned} f_1(z) &= \mathrm{Re}\int_{z_0}^{z} \frac{1}{2}F(\zeta)(1 - G^2(\zeta))d\zeta; \\ f_2(z) &= \mathrm{Re}\int_{z_0}^{z} \frac{i}{2}F(\zeta)(1 + G^2(\zeta))d\zeta; \\ f_3(z) &= \mathrm{Re}\int_{z_0}^{z} F(\zeta)G(\zeta)d\zeta, \end{aligned}$$

where F is holomorphic and G is meromorphic such that FG^2 is holomorphic (just the same conditions as in 3.35). The domain of definition of the parametrization must be chosen in such a way that the occurring integrals are independent of the path of integration (for example, a small disc or a simply connected domain).

Conversely, every pair $(F(z), G(z))$ with holomorphic FG^2 defines a conformal parametrized minimal surface element f. This f is regular if F has zeros only at the poles of G and there it holds that $FG^2 \neq 0$.

With the help of the examples in 3.37 we will see that even very simple looking functions F, G lead to interesting minimal surfaces. If, however, G is constant, then there is a linear relation between f_1, f_2, f_3 and consequently it is a parametrization of the plane. On the other hand, F can be constant, as the example of the Enneper surface shows.

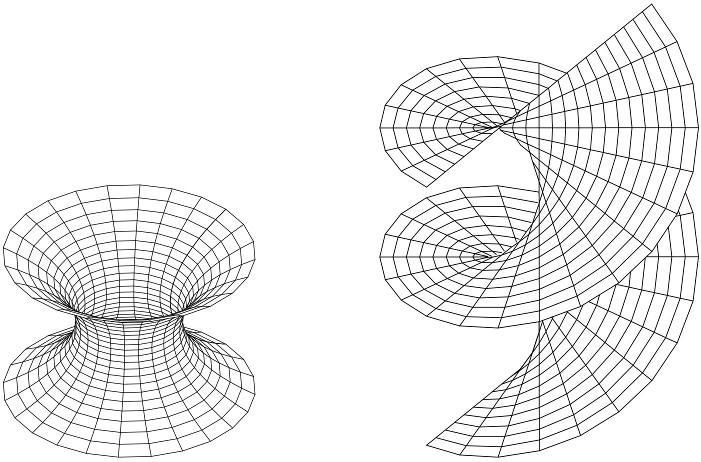


Figure 3.19. The interior part of a catenoid and a right helicoid (staircase surface)

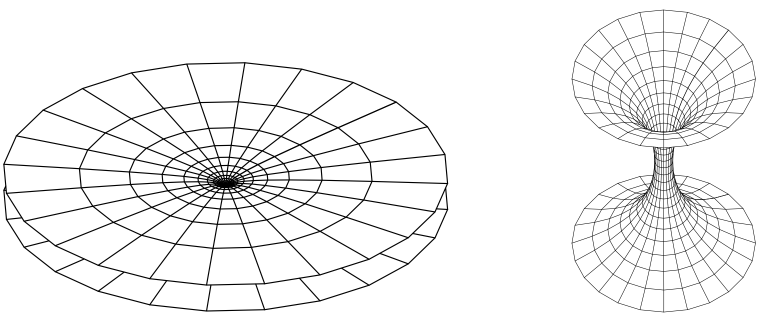


Figure 3.20. A catenoid in the large and a catenoid which is scaled in the vertical direction

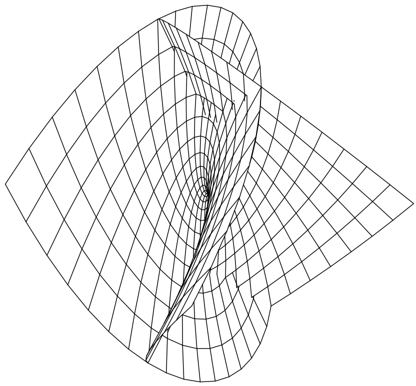


Figure 3.21. Enneper surface

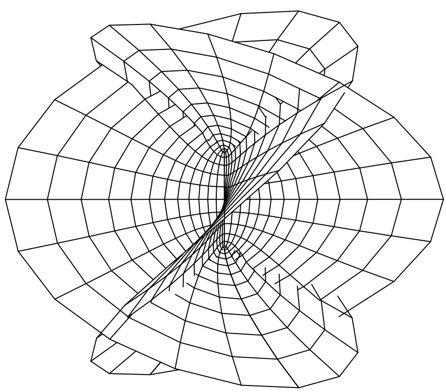


Figure 3.22. Henneberg surface

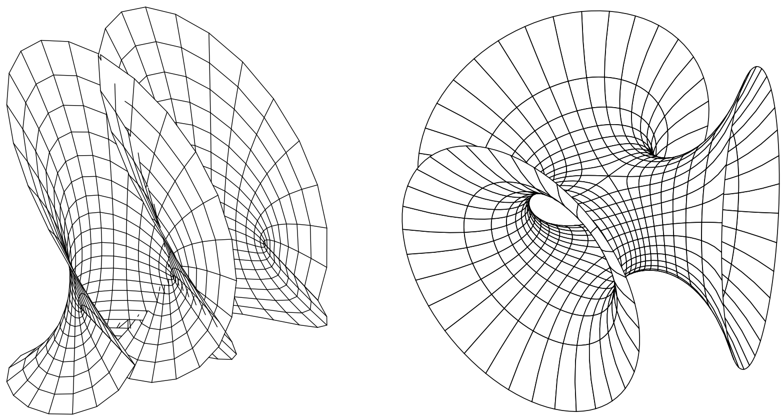


Figure 3.23. Catalan surface and trinoid[7] (catenoid with three ends)

3.37. Examples.

1. The *catenoid* $f(u, v) = (\cosh u \cos v, \cosh u \sin v, u)$ is the surface of revolution generated by a catenary (details on this are found in Section 2.3). Here one has

$$\begin{aligned}
\varphi_1(u+iv) &= \sinh u \cos v + i \cosh u \sin v \\
&= \sinh(u+iv), \\
\varphi_2(u+iv) &= \sinh u \sin v - i \cosh u \cos v \\
&= -i \cosh(u+iv), \\
\varphi_3(u+iv) &= 1.
\end{aligned}$$

Clearly $\varphi_1,\ \varphi_2,\ \varphi_3$ are holomorphic with

$$\sum_i \varphi_i^2(z) = \sinh^2 z + i^2 \cosh^2 z + 1 = 0.$$

According to 3.31, f is a conformally parametrized minimal surface. The Weierstrass representation is the following: $F(z) = -e^{-z}, G(z) = -e^{z}$.

[7]This surface was found by L.P.Jorge and W.H.Meeks: *The topology of complete minimal surfaces of finite Gaussian curvature*, Topology **22**, 203–221 (1983). They also give the Weierstrass representation.

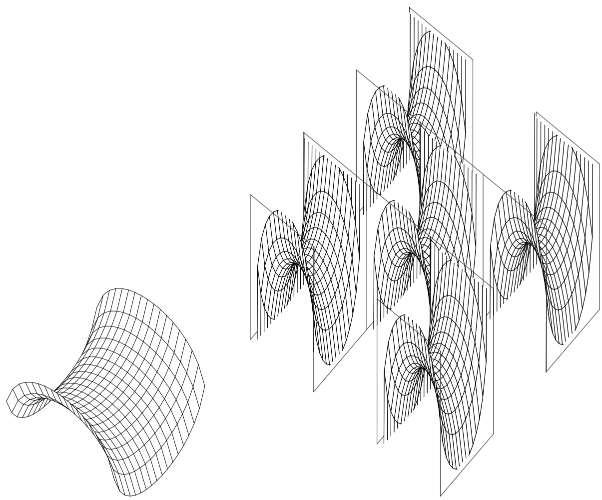


Figure 3.24. Building blocks of the Scherk minimal surface

2. The *helicoid* is at the same time a ruled surface and a minimal surface:

$$h(u,v) = (0,0,-u) + v(-\sin u, \cos u, 0) = (-v\sin u, v\cos u, -u).$$

Here we have standard parameters with $\lambda = 1, F = 0, J = 0$, and this in turn implies $H = 0$, cf. 3.23. However, this parametrization is not conformal. If we reparametrize the surface as

$$h^*(u,v) = (-\sinh u \sin v, \sinh u \cos v, -v),$$

then the complexification which this induces is

$$\varphi_1(z) = i\sinh z, \quad \varphi_2(z) = \cosh z, \quad \varphi_3(z) = i.$$

These are exactly the same functions $\varphi_1, \varphi_2, \varphi_3$ as occurred above in the case of the catenoid, up to a factor of i. In particular, h^* is a conformally parametrized minimal surface. We also see that the catenoid f and the helicoid h^* are isometric, since $g_{12} = 0$ and $g_{11} = g_{22} = \frac{1}{2}(\varphi_1\overline{\varphi}_1 + \varphi_2\overline{\varphi}_2 + \varphi_3\overline{\varphi}_3)$, an expression which is invariant under multiplication by i. For this reason one speaks of *conjugate pairs* of minimal surfaces, if the complexification of one is obtained by multiplication by i of the other. In this sense, the catenoid and the helicoid are conjugate to one another, see Figure 3.19. One can even view both of them as the real and imaginary parts, respectively, of a common complex surface. If one multiplies $\varphi_1, \varphi_2, \varphi_3$ by $e^{i\theta}$ then one obtains a family of conjugate and isometric minimal surfaces. Such a continuous transformation of the catenoid into the helicoid is indicated by Figure 4-4 in [**1**].[8]

3. The Weierstrass representation of the *Enneper surface* is given by the easy formulas $F = 2, G(z) = z$ with $\varphi_1 = 1 - z^2, \varphi_2 = i(1+z^2), \varphi_3 = 2z$ and
$$f(u,v) = \big(u - \tfrac{1}{3}u^3 + uv^2, -v + \tfrac{1}{3}v^3 - vu^2, u^2 - v^2\big).$$
For the choice of $F = 2z^2,\ G = z^{-1}$ one has $\varphi_1 = z^2 - 1,\ \varphi_2 = i(z^2+1),\ \varphi_3 = 2z$ and thus, up to a reflection $f_1 \mapsto -f_1$, the Enneper surface. The Weierstrass representation is thus even in the geometric sense far from being unique, see Figure 3.21.

4. The *Scherk minimal surface* is given by the functions $\varphi_1(z) = \frac{2}{1+z^2},\ \varphi_2(z) = \frac{2i}{1-z^2},\ \varphi_3(z) = \frac{4z}{1-z^4}$. The φ_k are analytic except for the points $z = \pm i, \pm 1$. One has $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = 0$. The Weierstrass representation is $F(z) = 4/(1-z^4), G(z) = z$. A very simple parametrization of the Scherk surface is as a graph over the (u,v)-plane: $f(u,v) = \big(u, v, \log\frac{\cos v}{\cos u}\big)$. However, this parametrization is not conformal, see Figure 3.24.

5. The *Catalan surface* (see Figure 3.23) is
$$f(u,v) = (u - \sin u \cosh v, 1 - \cos u \cosh v, 4\sin\tfrac{u}{2}\sinh\tfrac{v}{2}),$$
with the complexification
$$\varphi_1(z) = 1 - \cosh(-iz), \varphi_2(z) = i\sinh(-iz), \varphi_3(z) = 2\sinh(-\tfrac{iz}{2}).$$
Here $F(z) = 1 - e^{iz}$ and $G(z) = \varphi_3(z)/F(z)$ is the Weierstrass representation.

6. The *Henneberg surface* (see Figure 3.22) has the coordinates
$$\begin{aligned} f_1(u,v) &= 2\sinh u\cos v - \tfrac{2}{3}\sinh(3u)\cos(3v), \\ f_2(u,v) &= 2\sinh u\sin v + \tfrac{2}{3}\sinh(3u)\sin(3v), \\ f_3(u,v) &= 2\cosh(2u)\cos(2v). \end{aligned}$$

[8]For an animated transition from the catenoid to the helicoid and back see `http://www.youtube.com/watch?v=-Pa6FOK3gpM`.

Finally we mention the three-fold periodic minimal surfaces. These arise from compact building blocks through periodic repetition or gluing along common boundaries. Especially famous in this respect is the *Schwarz minimal surface* with cube-like building blocks. One can glue arbitrarily many of these together like cubical bricks in space, see Figure 3.25 which shows a modified version where the block is based on a cuboid rather than a cube. There is an almost unlimited amount of further literature on minimal surfaces, often with beautiful pictures.[9]

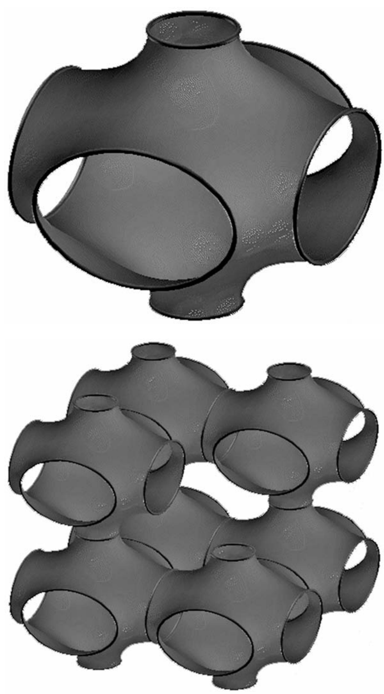


Figure 3.25. Building blocks of the Schwarz minimal surface[10]

[9] Cf. [**12**] as well as the literature which is cited there, see also [**11**], Sections 3.5 and 3.8.

[10] Reproduced with kind permission of K. Polthier, M. Steffens and Ch. Teitzel, see also `http://met.iisc.ernet.in/~lord/webfiles/tpmbs.pdf`.

3E Surfaces in Minkowski space $I\!R^3_1$

We continue now the investigations begun in Section 2E. Just as one can define and study curves in Minkowski space, we can develop a theory of surfaces in Minkowski space. In addition to the motivation considered in 2E, there is an interesting phenomena here, namely the existence of a simple model of *hyperbolic* or *non-Euclidean geometry*, which (at least globally) has no counterpart in Euclidean three-space. From the differential geometric point of view this surface is simply a surface of constant negative Gaussian curvature without singularities at the boundary (which occurred for example in the pseudo-sphere of Section 3.17). A (regular) *surface element* is defined as an immersion $f\colon U \to I\!R^3_1$, exactly as in $I\!R^3$. Because of the different types of vectors in 2.17, there are different kinds of planes, in particular tangent planes. The *first fundamental form* can be formally defined as in 3.2. However, this form is not necessarily positive definite, not even of maximal rank. At least the rank cannot vanish, since there cannot be a two-dimensional plane in $I\!R^3_1$ which consists solely of null vectors. The rank can be 1, see the examples below. This leads to the following classification of surfaces into different types:

3.38. Definition. A surface element $f\colon U \to I\!R^3_1$ is called

space-like, in case the first fundamental form is positive definite,
time-like, in case the first fundamental form is indefinite,
isotropic, in case the first fundamental form has rank 1.

EXAMPLES: The two-sheeted hyperboloid $x_1^2 = x_2^2+x_3^2+1$ is a surface which is everywhere space-like. We will see this in a convincing manner below. Geometrically, the two-sheeted hyperboloid is obtained by rotating the space-like hyperbola from 2.18 about the x_1-axis. Similarly, the one-sheeted hyperboloid $x_1^2 = x_2^2+x_3^2-1$ is a surface which is everywhere time-like. It is obtained by rotating the time-like hyperbola from 2.18 around the x_1-axis.

The *null-cone* or *light-cone* $x_1^2 = x_2^2+x_3^2$ is itself an isotropic surface, except for the origin which must be excluded since this point is already topologically a singularity: no neighborhood of the origin on the light-cone can be parametrized by a (regular) differentiable map to an open disc.

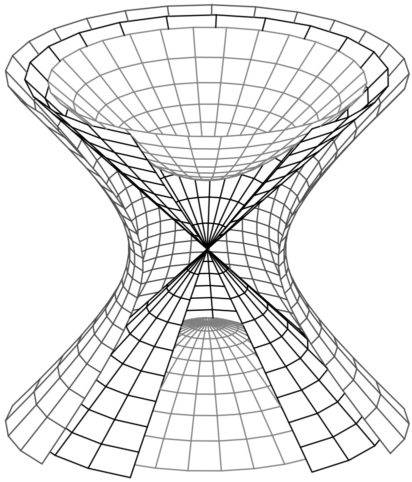


Figure 3.26. Light-cone with a one-sheeted and two-sheeted hyperboloid

3.39. Lemma. A surface element f is

$$\left\{\begin{array}{c}\text{space-like}\\ \text{time-like}\\ \text{isotropic,}\end{array}\right\} \begin{array}{l}\text{if and only if, at every}\\ \text{point } p = f(u) \text{ there is a}\end{array} \left\{\begin{array}{c}\text{time-like}\\ \text{space-like}\\ \text{isotropic}\end{array}\right\}$$

vector $X \neq 0$ which is perpendicular, with respect to the inner product $\langle\ ,\ \rangle_1$ in Minkowski space, to the tangent plane $T_u f$.

PROOF: First we start with a tangent plane and look for the vector X. In the first case of space-like tangent plane we choose an orthonormal basis $\{V_1, V_2\}$ in that plane and complete this to a basis of three-space by adding a vector X. By means of the Schmidt orthogonalization procedure 2.4 we can assume that X is perpendicular to the tangent plane. But then X is necessarily time-like, as otherwise the inner product in $\mathbb{R}^3_1$ would be positive (semi-)definite.

We proceed similarly in the second case. Here we choose V_1, V_2 such that V_1 is space-like and V_2 is time-like. This implies that X is space-like, since otherwise the inner product on $\mathbb{R}^3_1$ would not be positive definite on a two-dimensional plane.

In the last case we could choose the basis of the tangent plane such that V_1 is isotropic and V_2 is either space-like or time-like, but perpendicular to V_1. Then we can set $X = V_1$.

For the converse, we observe that the orthogonal complement of a space-like vector is a time-like plane and that the orthogonal complement of a time-like vector is a space-like plane. For a given isotropic vector X there is no orthogonal complement in the classical sense, as the vector is perpendicular to itself. But if there is plane which is perpendicular to X, then it also contains X and is therefore necessarily isotropic. To see this, we argue indirectly, that there can be no plane which is perpendicular to X and does not contain X at the same time. Indeed, if this were the case then either the inner product would be positive semi-definite (if the plane is space-like), or there would exist a plane which consists only of isotropic vectors, which is impossible (if the plane which is perpendicular to X contains an isotropic vector). $\square$

3.40. Corollary. A space-like surface element has a unique (up to sign) unit normal, which is necessarily time-like, and a time-like surface element has a unique (up to sign) unit normal, which is then necessarily space-like. An isotropic surface element has a unique one-dimensional normal space, but this is contained in a tangent space. (In this regard, the tangent space and the normal space together do not span the ambient space, as one is accustomed to.)

Example: If we return to the examples above, then we see easily that the two-sheeted hyperboloid is space-like, since its unit normal is a time-like vector. Similarly, the one-sheeted hyperboloid is time-like, because its unit normal is a space-like vector. In the case of the light-cone we also see that each position vector at a point on the cone is itself a normal vector, which is clearly contained in the tangent plane. Note the similarity with the situation of the unit sphere S^2, for which also the vectors to the points on the sphere are unit normals.

3.41. Definition. (Weingarten map, curvatures)

For a space-like or time-like surface in $I\!R^3_1$ there is a unit normal, which is unique (up to sign) by 3.40. These unique unit normals can be used to define the *Gauss map* just as in 3.8. More precisely, the Gauss map is a map

$$\nu\colon U \to S^2(1) = \{(x,y,z) \in \mathbb{R}^3_1 \mid -x_1^2 + x_2^2 + x_3^2 = 1\},$$

in case the surface is time-like (that is, the normal vector is space-like), and

$$\nu\colon U \to S^2(-1) = \{(x,y,z) \in \mathbb{R}^3_1 \mid -x_1^2 + x_2^2 + x_3^2 = -1\},$$

in case the surface is space-like (that is, the normal vector is time-like). Then the statements of Lemma 3.9 continue to hold, and we can define the *Weingarten map* as $L = -D\nu \circ (Df)^{-1}$. The *first fundamental form* I of a surface element is given in local coordinates (just like the Euclidean case) by

$$g_{ij} = \Big\langle \frac{\partial f}{\partial u_i}, \frac{\partial f}{\partial u_j} \Big\rangle_1;$$

the second fundamental form is just the normal component of the matrix of the second derivatives (cf. 3.10), and is therefore actually vector-valued. In the Euclidean case this was irrelevant, we simply set there $II(X,Y) = I(LX,Y)$ and viewed the second fundamental form as the scalar factor of this compared with ν. Because of the different types of unit normals, we have to consider here a *vector-valued second fundamental form* and define $II(X,Y)$ as the normal vector which satisfies

$$\big\langle II(X,Y), \nu \big\rangle_1 = \big\langle LX, Y \big\rangle_1,$$

which in the Euclidean case is just the same thing. In coordinates we then have

$$II\Big(\frac{\partial f}{\partial u_i}, \frac{\partial f}{\partial u_j}\Big) = h_{ij}\nu = \epsilon \Big\langle \frac{\partial^2 f}{\partial u_i \partial u_j}, \nu \Big\rangle_1 \nu,$$

where $\epsilon = \langle \nu, \nu \rangle_1$ is the sign which is defined by ν. To get the Gaussian curvature, instead of taking the determinant of h_{ij}, one has to take the determinant of $h_{ij}\nu$ in the following sense. The *Gaussian curvature* is defined as

$$K = \frac{\langle II(X,X), II(Y,Y)\rangle_1 - \langle II(X,Y), II(Y,X)\rangle_1}{I(X,X)\cdot I(Y,Y) - I(X,Y)\cdot I(Y,X)} = \frac{\mathrm{Det}(h_{ij})}{\mathrm{Det}(g_{ij})} \cdot \epsilon.$$

Here X, Y is an arbitrary basis of the tangent plane, for example $X = \frac{\partial f}{\partial u_1}, Y = \frac{\partial f}{\partial u_2}$. If we take an orthonormal basis e_1, e_2 with $\langle e_i, e_i \rangle_1 = \epsilon_i$, then we get

$$K = \epsilon_1 \epsilon_2 \Big(\langle II(e_1, e_1), II(e_2, e_2) \rangle_1 - \langle II(e_1, e_2), II(e_2, e_1) \rangle \Big).$$

Similarly one defines the mean curvature in vectorial form, namely as the trace of II with respect to I, or one defines the *mean curvature vector* $\mathbf{H}$, which we already met in 3.30:

$$\mathbf{H} = H \cdot \nu = \tfrac{1}{2}\big(\epsilon_1 II(e_1, e_1) + \epsilon_2 II(e_2, e_2)\big).$$

Since the mean curvature is only defined up to a sign anyhow, the signs which are involved are not very important. In contrast, for the Gaussian curvature the sign is of fundamental importance, since the determinant does not depend on the sign of ν. This means that we are confronted with the phenomenon that if the Weingarten map is the identity we cannot conclude that $K = 1$, but only that $K = \epsilon = \langle \nu, \nu \rangle_1 = \pm 1$. This leads directly to the definition of hyperbolic space as a surface in Minkowski space, see 3.44.

3.42. Surfaces of rotation in Minkowski space. A surface of rotation in Euclidean space is generated by rotating an arbitrary curve about an arbitrary axis, cf. 3.16. In Minkowski space, however, there are different types of curves (space-like, time-like and isotropic) as well as different type of rotation axes (space-like, time-like and isotropic), so that there are different flavors of surfaces of rotation in this context.

A rotation whose axis is time-like (for example the x_1-axis) is described by a matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix}.$$

Formally, this looks identical to a Euclidean rotation matrix. Thus, the surfaces of rotation obtained through rotations of axes of this kind should “look like” Euclidean surfaces of rotation. The surface itself will be space-like if the curve is space-like (for example a two-sheeted hyperboloid), and will be time-like if the curve is so (for example

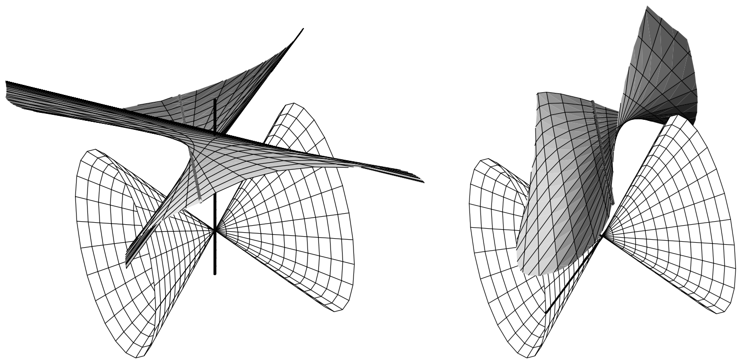


Figure 3.27. Surfaces of rotation in Minkowski space (with axis), simultaneously ruled surfaces

the one-sheeted hyperboloid); an isotropic curve leads to an isotropic surface of rotation.

A rotation whose axis of rotation is space-like (for example, the x_3-axis) is described by a matrix of the form

$$\begin{pmatrix} \cosh\varphi & \sinh\varphi & 0 \\ \sinh\varphi & \cosh\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is easy to see that the linear map defined by this matrix preserves the inner product (so that it is legitimate to speak of a "rotation"). The surface of rotation thus obtained replaces each point of the curve by a hyperbola instead of a circle. Therefore surfaces of this type look quite different from the Euclidean surfaces of rotation. According to the type of curve which is rotated, one again gets different types of surfaces of rotation, for an example see Figure 3.27, left side.

Finally there are rotations whose axis is isotropic (light-like), for example the diagonal in the (x_1, x_2)-plane. The matrix which describes such a rotation is

$$\begin{pmatrix} 1+\frac{\varphi^2}{2} & -\frac{\varphi^2}{2} & \varphi \\ \frac{\varphi^2}{2} & 1-\frac{\varphi^2}{2} & \varphi \\ \varphi & -\varphi & 1 \end{pmatrix}.$$

This matrix no longer resembles a Euclidean rotation matrix in any way, but it does preserve the Minkowski inner product, and it fixes the line spanned by the isotropic vector $(1,1,0)$, for an example see Figure 3.27, right side.

The formulas for the first and second fundamental forms as well as for the curvatures can be derived in the same way as in Section 3.16.

3.43. Ruled surfaces. A ruled surface can be defined in Minkowski space just as in Euclidean space, see 3.20, since a Euclidean line is also a line in Minkowski space. Correspondingly, most of the formulas for ruled surfaces derived in Euclidean space retain their validity in Minkowski space. One point where more caution is required is the situation in which a vector field X or its tangent $\dot{X}$ is isotropic. In this case, for example, one no longer has the standard parameters of 3.21. On the other hand, the developable surfaces in Minkowski space are the same as in Euclidean space; in particular the four standard types plane, cone, cylinder, tangent developable are the same, compare Section 3.24. In contrast with the Euclidean case, however, there are in Minkowski space four different types of ruled surfaces, which are simultaneously also minimal surface; for details on this, see Exercise 22 at the end of the chapter.

3.44. The hyperbolic plane. We return to the two-sheeted hyperboloid given bythe equation $-x_1^2 + x_2^2 + x_3^2 = -1$. For reasons of symmetry it suffices to consider just one of these sheets, say the one with positive x_1. In each point the unit normal $\nu = \pm(x_1, x_2, x_3)$ is the same as the corresponding position vector, up to a sign. Just as in the case of the Euclidean sphere (following 3.10), we choose here $\nu = -(x_1, x_2, x_3)$ with $\epsilon = \langle \nu, \nu \rangle_1 = -1$. The surface itself is then space-like, hence has a positive definite first fundamental form I. According to 3.41, the Weingarten mapping is just the identity map, and one has $I = \langle I\!I, \nu \rangle_1$. Hence $I\!I(X,Y) = -\langle X, Y \rangle_1 \nu$, so that the Gaussian curvature is

$$K = \langle I\!I(e_1,e_1), I\!I(e_2,e_2) \rangle_1 - \langle I\!I(e_1,e_2), I\!I(e_2,e_1) \rangle_1 = \langle \nu, \nu \rangle_1 = -1.$$

The mean curvature vector is

$$H \cdot \nu = \tfrac{1}{2}\big(I\!I(e_1,e_1) + I\!I(e_2,e_2)\big) = \tfrac{1}{2}\big(-I(e_1,e_1) - I(e_2,e_2)\big) \cdot \nu = -\nu;$$

consequently we have the relation $H = -1$. Here, e_1, e_2 is an arbitrary orthonormal basis of the tangent plane. With the normal $-\nu$ of opposite sign, we of course get $H = 1$.

Definition The set

$$H^2 = \left\{(x_1, x_2, x_3) \in I\!R^3_1 \mid -x_1^2 + x_2^2 + x_3^2 = -1,\ x_1 > 0\right\},$$

together with the inner product induced by $\langle\ ,\ \rangle_1$ on each tangent plane, is called the *hyperbolic plane*. The curvature satisfies the equation $K = -1$ as a surface in Minkowski space. It is also called the *Bolyai-Lobachevski plane*.

REMARK: The hyperbolic plane defines a *non-Euclidean geometry* in the sense that all axioms of Euclid are satisfied except for the famous parallel axiom, see [**51**, Sec.7] or [**4**, Sec.4-7]. The *parallel axiom* states that for each line and each point not on this line, there is a unique line through that given point which does not meet the given line. As a *point* in the hyperbolic plane one takes the points of H^2 or, alternatively, the corresponding lines in Minkowski space which pass through the origin (each such line intersects H^2 in exactly one point, if at all), and one defines the set of *lines* to be the set of intersection curves of H^2 with all two-dimensional planes in Minkowski space which pass through the origin and meet H^2 (or, alternatively, these two-dimensional planes themselves), see Figure 3.28.

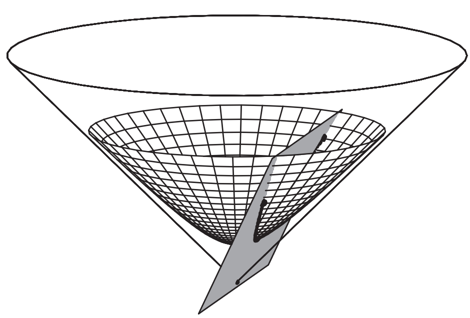


Figure 3.28. Hyperbolic line segment between two points in $H^2 \subset I\!R^3_1$

It follows that any two "points" in H^2 can be joined by exactly one "line". However, there are no unique parallels since it happens quite often that two lines do not meet: The intersection of the two corresponding two-dimensional planes in Minkowski space can be disjoint with H^2 even if it is not empty.

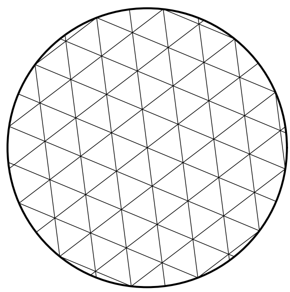


Figure 3.29. Lines in H^2 after projection into the interior of a conic in $\mathbb{R}P^2$

After projection into the real projective plane (defined as the set of all lines in three-space through the origin), the hyperbolic plane H^2 appears as its *projective disc model* where the "lines" appear in ordinary line segments, see Figure 3.29 or Figure 7.2 (right-hand side). The light cone appears as a conic in this model (just the exterior circle in Figure 3.29). In particular one then has the fact that through two arbitrary "points" of H^2 there is exactly one "line" of H^2 which contains them both, but from the figure it is intuitively clear that there need not be unique parallel lines. It is also easy to see that each "line" is infinite in both directions, measured now with the inner product of Minkowski space. We will see later on in Chapter 4 (Exercise 24) that these "lines" have a further geometric meaning as being the *geodesics* of the hyperbolic plane (compare Figure 4.9). Moreover, both the geodesics and the Gaussian curvature actually are independent of the ambient Minkowski space, but are rather "intrinsic" objects of H^2, defined exclusively with the aid of the first fundamental form. But we won't be able to prove this until later, in 4.16. Similarly, it follows that the hyperbolic and the Euclidean plane are diffeomorphic to one another (there is a bijective, differentiable in both directions, map between them, for example by means of the orthogonal projection $F(x_1, x_2, x_3) := (x_2, x_3)$,) but such a diffeomorphism can never preserve the first fundamental form. Otherwise it would have to map the geodesics in one geometry to the geodesics of the other (that is, the hyperbolic "lines" into the Euclidean lines), which we can already see is impossible from the non-validity of the parallel axiom in hyperbolic geometry. A comprehensive source in the literature on this topic is the classical [**49**], see also [**51**].

3F Hypersurfaces in $\mathbb{R}^{n+1}$

The entire theory of Chapter 3 can be extended to higher dimensions, with few exceptions, which we describe here. We replace the two-dimensional parameter domain by a n-dimensional one and the three-dimensional ambient Euclidean space by the $(n+1)$-dimensional one. In this case one speaks of *hypersurfaces*, in analogy with hyperplanes.

3.45. Definition. (Hypersurface element)
$f\colon U \to \mathbb{R}^{n+1}$ is called a *regular hypersurface element*, if $U \subset \mathbb{R}^n$ is open and f is a (C^2-) immersion. The parameter $u = (u_1, \ldots, u_n)$ is associated with the point $f(u)$ with $n+1$ coordinates $f(u) = (f_1(u), \ldots, f_{n+1}(u))$. The *tangent hyperplane* $T_u f$ is then is defined to be the image of $T_u U$ under the map $Df\big|_u$. Similarly, one defines

- the *Gauss map* $\nu\colon U \to S^n$ by the unit normal vector $\nu(u)$, which is perpendicular to $T_u f$ (but note: in $\mathbb{R}^{n+1}$ for $n \geq 3$ there is no bilinear vector product of tangent vectors; still one can formally define ν as an n-linear vector product),
- the *Weingarten map* $L = -D\nu \circ (Df)^{-1}$,
- the *first, second and third fundamental forms* (cf. 3.10)

$$\begin{aligned}
I &= \big(g_{ij}\big)_{i,j=1,\ldots,n} &= \Big(\langle \frac{\partial f}{\partial u_i}, \frac{\partial f}{\partial u_j}\rangle\Big)_{ij},\\
II &= \big(h_{ij}\big)_{i,j=1,\ldots,n} &= \Big(\langle \frac{\partial^2 f}{\partial u_i, \partial u_j}, \nu\rangle\Big)_{ij},\\
III &= \big(e_{ij}\big)_{i,j=1,\ldots,n} &= \Big(\langle \frac{\partial \nu}{\partial u_i}, \frac{\partial \nu}{\partial u_j}\rangle\Big)_{ij}.
\end{aligned}$$

Example: The three-sphere in $\mathbb{R}^4$ can be described (with a certain exceptional set) by coordinates as follows, where we have set $\nu = \pm f$ and $L = \pm\mathrm{Id}$:

$$f(\phi,\psi,\theta) = (\cos\phi\cos\psi\cos\theta, \sin\phi\cos\psi\cos\theta, \sin\psi\cos\theta, \sin\theta).$$

3.46. Definition. (Curvature of a hypersurface)

The considerations of 3.12 remain valid, as long as the normal curvature is still given by $I\!I(X,X)$ with unit vectors X. Hence, one looks for the stationary values with constraints, using Lagrange multipliers. Thus one defines the *principal curvatures* $\kappa_1,\dots,\kappa_n$ as the eigenvalues of L, the *mean curvature* by $H = \frac{1}{n}(\kappa_1+\ldots+\kappa_n) = \frac{1}{n}\,\mathrm{Tr}(L)$, the *Gauss-Kronecker curvature* by $K = \kappa_1\cdot\ldots\cdot\kappa_n = \mathrm{Det}(L)$, and finally, the *i-th mean curvature* K_i as the coefficient of the characteristic polynomial

$$\mathrm{Det}(L-\lambda\cdot Id) = \sum_{i=0}^{n}(-1)^{n-i}\binom{n}{i}K_i\lambda^{n-i},$$

$$K_i := \binom{n}{i}^{-1}\sum_{j_1<\cdots<j_i}\kappa_{j_1}\cdot\ldots\cdot\kappa_{j_i}.$$

In particular, one gets $H = K_1$, $K = K_n$, $K_0 = 1$. For $n > 2$, the formulas analogous to 3.13 for H in terms of coordinates require subdeterminants (cofactors) of the matrix g_{ij}, as the inverse matrix g^{ij} occurs. In low dimensions, we have

$$\begin{aligned} n=3:\quad K_1 &= \tfrac{1}{3}(\kappa_1+\kappa_2+\kappa_3),\\ K_2 &= \tfrac{1}{3}(\kappa_1\kappa_2+\kappa_1\kappa_3+\kappa_2\kappa_3),\\ K_3 &= \kappa_1\kappa_2\kappa_3. \end{aligned}$$

$$\begin{aligned} n=4:\quad K_1 &= \tfrac{1}{4}(\kappa_1+\kappa_2+\kappa_3+\kappa_4),\\ K_2 &= \tfrac{1}{6}(\kappa_1\kappa_2+\kappa_1\kappa_3+\kappa_1\kappa_4+\kappa_2\kappa_3+\kappa_2\kappa_4+\kappa_3\kappa_4),\\ K_3 &= \tfrac{1}{4}(\kappa_1\kappa_2\kappa_3+\kappa_1\kappa_2\kappa_4+\kappa_1\kappa_3\kappa_4+\kappa_2\kappa_3\kappa_4),\\ K_4 &= \kappa_1\kappa_2\kappa_3\kappa_4. \end{aligned}$$

The corollary following 3.10 is true for arbitrary n, where it is the so-called *Hamilton-Cayley theorem:* A self-adjoint linear endomorphism L satisfies its characteristic polynomial, see [**31**], Ch. X. In our case one would introduce the k-th fundamental form by $I(L^{k-1}X, Y)$, hence, e.g., a fourth fundamental form $IV(X, Y) := I(L^3 X, Y)$ and so on. All of these together are derived from the single equation

$$\sum_{i=0}^{n}(-1)^i\binom{n}{i}K_i \cdot I\Big(L^{n-i}X, Y\Big) = 0.$$

In particular, in the three-dimensional case one gets $IV - 3K_1 III + 3K_2 II - KI = 0$, and in four dimensions, $V - 4K_1 IV + 6K_2 III - 4K_3 II + KI = 0$. For $n > 2$ there are of course more "types" of points than just "elliptic", "hyperbolic" and "parabolic" leading to more types of Dupin indicatrices. The Gauss-Kronecker curvature no longer determines this type by itself. Rather, the type is dependent on the distribution of signs of $\kappa_1, \ldots, \kappa_n$. In algebra the number of negative eigenvalues is called the *index* of L.

Theorem 3.14 can be extended to higher dimensions word for word. One only needs to replace the "planes" by "hyperplanes" and the "spheres" by "hyperspheres", the set $S^n(r) = \{x \in \mathbb{R}^{n+1} \mid ||x|| = r\}$. An *umbilic* is defined as a point in which the Weingarten map is a multiple of the identity. The n-dimensional Monge coordinates are given by $f(u_1, \ldots, u_n) = \big(u_1, \ldots, u_n, h(u_1, \ldots, u_n)\big)$ with the unit normal $\nu(u_1, \ldots, u_n) = \frac{1}{\sqrt{1+h_{u_1}^2+\cdots+h_{u_n}^2}}\Big(-h_{u_1}, \ldots, -h_{u_n}, 1\Big)$.

3.47. Theorem. A connected surface element of the class C^2 consists only of umbilics if and only if it is contained in a hyperplane or a hypersphere $S^n(r)$. It is said to be *totally umbilical.*

In contrast, the existence of special parameters (for example, isothermal parameters) does *not* easily generalize to higher dimensions. It is not to be expected that in higher dimensions the first fundamental form can be described by a single scalar function. This only holds in very special cases, the so-called *conformally flat metrics*, for which we refer to Section 8E. Moreover, one has the following fact: I already determines II completely, if the rank of L or of II is at least 3, cf. 4.31. The "intrinsic geometry" in this case already determines the "extrinsic geometry". The step from two to three dimensions is thus in this respect a quite important one.

There are also higher-dimensional analogs of Minkowski space. Instead of the Euclidean inner product on $\mathbb{R}^n$ as a vector space, one can introduce different bilinear forms (also referred to as "inner products"), which are not positive definite, but are non-degenerate. This leads to the *pseudo-Euclidean spaces* $\mathbb{R}^n_k$, $k = 1, \ldots, n-1$. We will come back to this in Chapter 7 for the construction of higher-dimensional spaces of constant curvature.

Exercises

1. Verify that the matrix g_{ij} of the first fundamental form of $f\colon U \to \mathbb{R}^{n+1}$ can be written as a matrix product $(Df)^T \cdot (Df)$.

2. Show that for a curve c inside a given surface element, the two following statements are equivalent:
   (i) c is a line of curvature.
   (ii) The ruled surface f defined by the surface normal ν along c is developable (i.e., satisfies the equation $K = 0$). More precisely we have $f(u,v) = c(u) + v\nu(c(u))$.

3. Let c be a curve parametrized by arc length, and suppose that its image is contained in a surface element $f\colon U \to \mathbb{R}^3$. The *Darboux three-frame* E_1, E_2, E_3 is then defined by the relations $E_1(s) = c'(s), E_3(s) = \nu(c(s)), E_2(s) = E_3(s) \times E_1(s)$. Here, as usual, ν denotes the unit normal on the surface f.

   Derive the following derivative equations for this three-frame, which correspond to the Frenet equations:

$$\begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix}' = \begin{pmatrix} 0 & \kappa_g & \kappa_\nu \\ -\kappa_g & 0 & \tau_g \\ -\kappa_\nu & -\tau_g & 0 \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix}.$$

   The notations are as follows. The *geodesic curvature* is $\kappa_g = \langle c'', E_2 \rangle$, the *normal curvature* is $\kappa_\nu = II(c', c')$, and τ_g denotes a certain *geodesic torsion.*

4. Show that at a fixed point p on a surface element, the mean curvature is equal to the integral mean of all normal curvatures,

i.e.,

$$H(p) = \frac{1}{2\pi}\int_0^{2\pi} \kappa_\nu(\varphi)d\varphi.$$

Here we view κ_ν as a function of the angle φ, which parametrizes the set of unit vectors at this point (for example in some fixed orthonormal basis).

5. A surface of rotation can always be locally parametrized in such a way that the new parametrization is angle preserving. Hint: find $\psi = \psi(u)$ with the property that $(u,v) \mapsto f(\psi(u), v)$ is angle preserving.

6. Let $f\colon [0,A]\times[0,B] \to I\!R^3$ be a parametrized surface element. Show that the following conditions (i) and (ii) are equivalent:
   (i) For each rectangle $R = [u_1, u_1 + a] \times [u_2, u_2 + b] \subset U$, the opposite sides of $f(R)$ are of equal length.
   (ii) One has $\frac{\partial g_{11}}{\partial u_2} = \frac{\partial g_{22}}{\partial u_1} = 0$ in all of U.
   The coordinate grid (or two-parameter family of curves) formed by the u_1 and the u_2 lines is called a *Tchebychev grid*. Show that under these conditions there is a parameter transformation $\varphi\colon U \to \widetilde{U}$ such that for $\widetilde{f} = f\circ\varphi^{-1}$ the first fundamental form can be written as

$$(\widetilde{g}_{ij}) = \begin{pmatrix} 1 & \cos\vartheta \\ \cos\vartheta & 1 \end{pmatrix},$$

   where ϑ is the angle between the coordinate lines.
   Hint: Set $\varphi(u_1,u_2) = \big(\int\sqrt{g_{11}}du_1), \int\sqrt{g_{22}}du_2\big)$.

7. Suppose we are given a surface element with $K < 0$. Show that this surface is a minimal surface if and only if the asymptotic curves at each point are perpendicular to one another.

8. More generally, show the following analog of 3.19: If an asymptotic curve of a surface element with $K < 0$ is a Frenet curve with torsion τ, then for the mean curvature we have the formula $H = \pm\tau\cot\varphi$, where φ is the angle between the two asymptotic curves.

9. The *Mercator projection* (see Figure 3.30)

$$f(u,\varphi) = \tfrac{1}{\cosh u}\big(\cos\varphi, \sin\varphi, \sinh u\big)$$

is a parametrization of the surface of the sphere without the north and the south pole. Show that this parametrization is angle preserving, i.e., that u, φ are isothermal parameters. In the science of cartography, a map with this property is referred to as *angle preserving* or *conformal*. Think of it as a circular cylinder touching the sphere along the equator. For more information concerning mathematical cartography, compare [**5**], §§66,67, or [**8**].

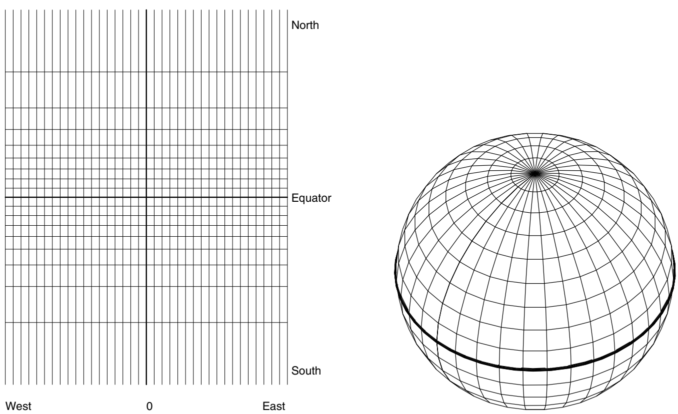


Figure 3.30. Coordinate grid of the Mercator projection

10. Investigate for which parameters the 3-sphere

$$f(\phi,\psi,\theta) = (\cos\phi\cos\psi\cos\theta, \sin\phi\cos\psi\cos\theta, \sin\psi\cos\theta, \sin\theta)$$

is an immersion. Compare your results with the case of the two-dimensional sphere.

11. Show that the one-sheeted hyperboloid with the equation $x^2+y^2-z^2=1$ (cf. Figure 3.14) as well as the hyperbolic paraboloid with the equation $x^2-y^2-4z=0$ can be parametrized as ruled surfaces. Which quantities λ, J, F in standard parameters occur? Compare this with 3.23 (iii).

12. Verify the formulas obtained in 3.23 for the Gaussian and the mean curvature of a ruled surface in standard parameters, and prove the Theorem of Catalan, which states that among all ruled surfaces the right helicoid is characterized by the condition $H \equiv 0, K \not\equiv 0$. Find all ruled surfaces for which $H = (-K)^{1/4}$ or $H = (-K)^{3/4}$.

13. Let c be a Frenet curve in $\mathbb{R}^3$ and let $D = \tau e_1 + \kappa e_3$ be the Darboux vector. Show that the ruled surface this defines, given by
$$f(u, v) = c(u) + vD(u),$$
is a developable surface (cf. 3.24), the so-called *rectifying developable*. The name comes from the fact that developing this surface in the plane, the curve c maps to a straight line (hence is rectified). Moreover, for $v = 0$ the tangent plane coincides with the rectifying plane.
Hint: In the parameters u, v above (not standard parameters), show that $\mathrm{Det}(II) = 0$ by calculating $\left\langle \frac{\partial^2 f}{\partial u \partial v}, \frac{\partial f}{\partial u} \times \frac{\partial f}{\partial v} \right\rangle$.

14. Show that the rectifying developable is a cylinder exactly for the slope lines (cf. 2.11). It is a cone if and only if $\frac{\tau}{\kappa} = as + b$ with constant a, b, where $a \neq 0$.

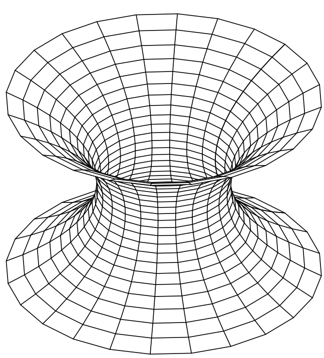


Figure 3.31. catenoid

15. Show that the catenoid (depicted in Figure 3.31) is the only surface of rotation for which $H \equiv 0$ and $K \not\equiv 0$.

Hint: Use a parametrization $f(z,\varphi) = (r(z)\cos\varphi, r(z)\sin\varphi, z)$ leading to a differential equation for the function $r(z)$ as in Example 3.27. Here the $(0,0,z)$-axis is the axis of rotation.

16. The *rotational torus* is given by

$$f(u,v) = \big((a+b\cos u)\cos v, (a+b\cos u)\sin v, b\sin u\big),$$

$0 \le u, v \le 2\pi$, cf. Figure 3.3. Here $a > b > 0$ are arbitrary (but fixed) parameters. Calculate the *total mean curvature* of this torus as the surface integral of the function $(H(u,v))^2$, $0 \le u, v \le 2\pi$, explicitly as a function of a and b. What is the smallest possible value of the total mean curvature?
Hint: The minimum occurs at $a = \sqrt{2}b$. Note that the integral is invariant under the homotheties $x \mapsto \lambda x$ of space with a fixed number λ. This relation $a = \sqrt{2}b$ is attained for the stereographic projection of the Clifford torus from the north pole, see the end of Remark 2.16.
Remark: The *Willmore conjecture* states that there is no immersed torus in $\mathbb{R}^3$ which has a smaller total mean curvature that the above rotational torus (also called *Willmore torus*), no matter what is looks like geometrically. This conjecture had been verified in many cases (see [**17**], 5.1–5.3, 6.5), but in general it was open until recently.[11]

17. For a surface element $f\colon U \to \mathbb{R}^3$ we define the *parallel surface* at distance ε by

$$f_\varepsilon(u_1,u_2) := f(u_1,u_2) + \varepsilon \cdot \nu(u_1,u_2),$$

cf. Section 3D and the Möbius strip in Figure 3.5. As usual ν denotes the unit normal of the surface f. Decide for which ε this defines a regular surface, and show the following.
   (a) The principal curvatures of f_ε and f have a ratio of $\kappa_i^{(\varepsilon)} = \kappa_i/(1-\varepsilon\kappa_i)$.
   (b) In case f has constant mean curvature $H \neq 0$, f_ε has constant Gaussian curvature for $\varepsilon = \frac{1}{2H}$.

[11] For the solution see F.C.Marques and A.Neves, *Min-Max theory and the Willmore conjecture*, Annals of Mathematics (2) **179**, 683–782 (2014), E-print: arXiv:1202.6036v2 [math.DG]. For an introduction into the complicated material see `http://www.math.uni-augsburg.de/~eschenbu/willmore.pdf`.

18. Prove that the ellipsoid of rotation defined by the equation
$$\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{c^2} = 1$$
is a Weingarten surface satisfying the equation $\kappa_1 = \frac{a^4}{c^2}(\kappa_2)^3$. Conversely, every compact surface of rotation with a constant ratio between κ_1 and κ_2^3 is an ellipsoid, cf. Figure 3.32.

    Hint for the converse direction: The same approach as in 3.27.

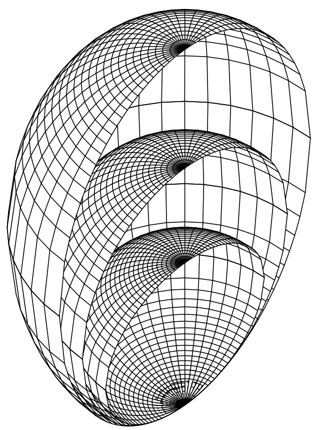


Figure 3.32. Several ellipsoids of rotation

19. Calculate the functions $\varphi_1, \varphi_2, \varphi_3$ for the Henneberg surface and verify the relation $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = 0$.
20. Calculate the Gaussian curvature for the hyperboloid $-x_1^2+x_2^2+x_3^2 = 1$ in Minkowski space, that is, with the induced geometry.

    Hint: The unit normal coincides with the position vector (up to sign), as in 3.44. However, here the position vector is space-like, compare Lemma 3.39.
21. Viewing the hyperbolic plane H^2 as a subset of $\mathbb{R}^3_1$, calculate the first fundamental form in polar coordinates around the point $(1, 0, 0)$ by using the set of all "lines" through this point, where on these lines the arc length should be used as a parameter.

Hint: The "lines" through this point appear in $\mathbb{R}^3$ as usual hyperbolas. In the Euclidean plane the first fundamental form in polar coordinates (r, φ) is given by the arc length element $ds^2 = dr^2 + r^2 d\varphi^2$. A similar formula for the hyperbolic plane is what you must derive, compare Figure 3.33.

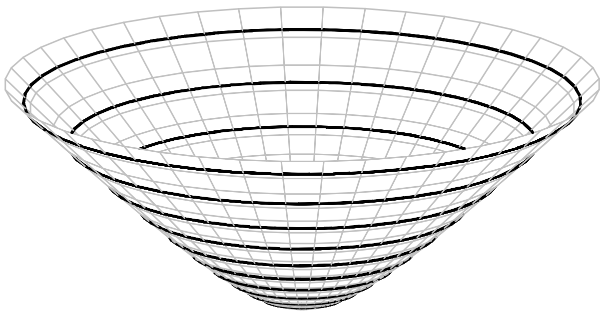


Figure 3.33. Equidistant circles in the hyperbolic plane $H^2 \subset \mathbb{R}^3_1$

22. Show that for an arbitrary choice of constant $a \neq 0$, each of the following four surfaces in Minkowski space is a ruled surface which is simultaneously a minimal surface, i.e., satisfies $H = 0$:

$$\begin{aligned} f_1(u,v) &= (au, v\cos u, v\sin u); \\ f_2(u,v) &= (v\sinh u, v\cosh u, au); \\ f_3(u,v) &= (v\cosh u, v\sinh u, au); \\ f_4(u,v) &= \big(a(\tfrac{u^3}{3}+u)+uv, a(\tfrac{u^3}{3}-u)+uv, au^2+v\big). \end{aligned}$$

More precisely, this holds everywhere where the surface is either space-like or time-like, since for isotropic surfaces mean curvature is not a defined concept. However, the four types of surfaces are isotropic only for special values of the parameter v.

23. Show that the four types of surfaces of the previous exercise are *helicoidal ruled surfaces* in the sense that each can be obtained from a helicoidal motion from a line. A *helicoidal motion* is a one-parameter group of linear transformations which preserve the inner product of $\mathbb{R}^3_1$, similar to the Euclidean screw-motions of Section 2.3.

Hint: Consider the matrices in 3.42. The surfaces f_1, f_2, f_3 in Exercise 22 are called *helicoid of the first, second, and third kind.* The surface f_4 is *Cayley's cubic ruled surface*, depicted in Figure 3.34.

24. A ruled surface with rulings, all lines of which are isotropic, is also called a *Monge surface*. Show that every Monge surface satisfies the equation $K = H^2$. Compare this with the case of umbilic points on surfaces in Euclidean space, which satisfy $H^2 - K = \frac{1}{4}(\kappa_1 - \kappa_2)^2 \geq 0$, with equality holding if and only if the point at which the curvatures are taken is an umbilic.

25. Show that any (regular) tubular surface around any given C^2-curve in constant distance $r > 0$ is a *linear Weingarten surface* in the sense that there is a linear relation of the type $aH + bK = c$ with constants a, b, c. Hint: One of the principal curvatures is constant and equal to $1/r$.

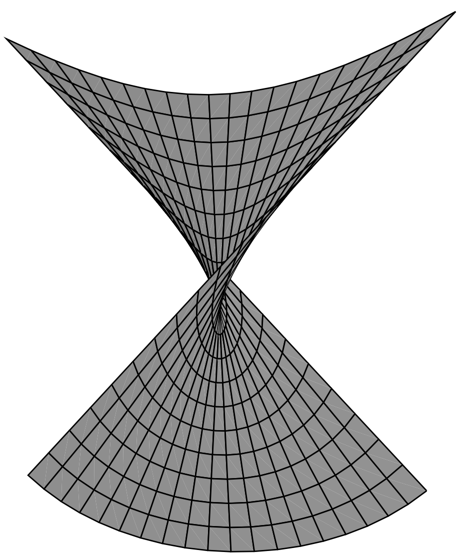


Figure 3.34. The cubic ruled surface of Cayley[12]

[12]For animated pictures of this remarkable surface see `http://www.mathcurve.com/surfaces/cubic/cubique_reglee.shtml`.

Chapter 4

The Intrinsic Geometry of Surfaces

The "intrinsic geometry" of surfaces refers to all those properties of a surface which only depend on the first fundamental form. Expressed more figuratively, the intrinsic geometry is the geometry which pure two-dimensional beings (the inhabitants of "flatland"[1]) can recognize, without any knowledge of the third dimension. Surely angles and lengths are among these properties. The question naturally arises as to what geometric quantities are intrinsic, in particular, which of the curvature quantities are of this kind. On the one hand it is intuitively clear that a change in lengths and angles can lead to a change in the curvature. On the other hand it is not at all clear whether the first fundamental form alone is sufficient to determine the curvature.

A further problem in this connection is as follows. How can one form derivatives using only the properties of the surface itself, without reference to the ambient space? The directional derivative of scalar functions is defined in terms of difference quotients. This is no longer so for the directional derivative of vector fields. For vector fields living in Euclidean space it is sufficient to take the derivatives of the coordinate functions, since one has a constant basis. This is no longer true on an arbitrary surface. Instead, one has to form the derivative of the basis itself, a process which is not *a priori* well-defined. To alleviate these problems, one first reduces the process of

[1] E. A. Abbott, *Flatland – a romance of many dimensions – by a square*, 1884, reprint by Dover 1953, new edition by Princeton University Press, 1991.

taking derivatives to differentiation in the ambient space, and then studies whether or not the notion thus defined only depends on the first fundamental form.

In what follows, U will denote an open set in $\mathbb{R}^n$ and $f\colon U \to \mathbb{R}^{n+1}$ will denote a hypersurface element. We will often just speak of a "surface element" in this context. The notions of "tangent" and "normal" will always be taken with respect to this f if nothing to the contrary is stated. If the general dimension n is too abstract to the reader, he or she may safely just think of n as being 2, and the results are then the classical theory of the "intrinsic geometry of surfaces". However, since most of the formulas which we will give are just the same in dimension $n = 2$ as in general dimensions (in particular, the indices on the objects like g_{ij}, h_{ij}, etc.), we formulate everything in Chapter 4 for hypersurfaces in higher dimensions whenever this makes sense. This is also preparation for Chapters 5 to 8 which follow, in which higher dimensions occur out of necessity. In the discussion of special parameters in Section 4E, in the Gauss-Bonnet theorem in Section 4F, as well as in the global surface theory in Section 4G, we will return to consider the special case of dimension two in more detail.

There are good reasons to write the indices on coordinate functions as superscripts, and we shall do this throughout, writing for example $u = (u^1, \ldots, u^n)$ and $x = (x^1, \ldots, x^{n+1})$. The reason for this is the so-called Ricci calculus, as well as the Einstein summation convention. In the latter convention, summation symbols are omitted for sums over indices which are superscripts in one place and subscripts in another. This will be explained in more detail in Chapters 5 and 6, while we will explicitly write all summation signs in this chapter.

4A The covariant derivative

The analysis of the ambient space $\mathbb{R}^{n+1}$ leads to the notion of directional derivatives of functions and vector fields, as is well-known (cf. 4.1). For the theory of surfaces this has the disadvantage that even the derivative of *tangential* vector fields in the *tangent* direction may very well have a normal component (this is just a directional derivative in space). This would leave the realm of "intrinsic geometry of the surface". There is a way out of this, by considering only the component of that directional derivative which is tangent to the surface (cf. 4.3). The so-called *covariant derivative* obtained in this manner has in addition a series of very pleasant properties. It is, for example, a property of the intrinsic geometry, cf. 4.6.

4.1. Definition and Lemma. (Directional derivative)
Let Y be a differentiable vector field, defined on an open set of $\mathbb{R}^{n+1}$, and let X be a fixed directional vector at some fixed point p of this open set. (In other words, assume $(p, X) \in T_p\mathbb{R}^{n+1}$). Then the expression

$$D_X Y\big|_p := DY\big|_p(X) = \lim_{t\to 0} \frac{1}{t}\big(Y(p+tX) - Y(p)\big)$$

is called the *directional derivative* of Y in the direction of X (cf. [**27**], Chapter XVII, §2). Here DY denotes the Jacobi matrix.
Furthermore, $D_X Y\big|_p$ is already uniquely defined by the value of Y along an arbitrary differentiable curve $c\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^{n+1}$ with $c(0) = p$ and $\dot{c}(0) = X$. More precisely, one has

$$D_X Y\big|_p = \lim_{t\to 0} \frac{1}{t}\big(Y(c(t)) - Y(p)\big).$$

The (vector-valued) *partial derivatives of* Y correspond to the case $X = e_i$ with the standard basis $e_1, \dots, e_n$, meaning that we have the equation $D_{e_i} Y = \frac{\partial Y}{\partial x^i}$. Consequently we have with $X = \sum_i X^i e_i$

$$D_X Y\big|_p = \sum_i X^i D_{e_i} Y\big|_p = \sum_i X^i \lim_{t\to 0} \frac{1}{t}\big(Y(p+te_i) - Y(p)\big).$$

Proof of this claim: By the chain rule we have
$\lim_{t\to 0} \frac{1}{t}\big(Y(c(t)) - Y(c(0))\big) = \frac{d}{dt}\big|_{t=0} Y(c(t)) = DY\big|_p(X) = D_X Y\big|_p$.

4.2. Consequence. For a (hyper-)surface element $f\colon U \to \mathbb{R}^{n+1}$ let Y denote a differentiable vector field along f, and let X be some fixed tangent vector to f at the point $p = f(u)$ (see Definition 3.5). Then, according to 4.1, the directional derivative $D_X Y\big|_p$ is well-defined as a vector $D_X Y\big|_p \in T_p\mathbb{R}^{n+1}$. More precisely, the following relation always holds at the point $p = f(u)$:

$$D_X Y\big|_p = DY\big|_u\big((Df)^{-1}(X)\big) = \lim_{t\to 0} \frac{1}{t}\Big(Y\big(u + t(Df)^{-1}(X)\big) - Y(u)\Big).$$

Here, $c(t) = f\big(u+t(Df)^{-1}(X)\big)$ is a particular curve for which $\dot{c}(0) = X$. Hence we apply 4.1 to this. Note that Y is not defined at points of the surface element, but rather on the set of parameters. Thus $Y\big(u + t(Df)^{-1}(X)\big)$ is a well-defined vector field along this curve.

The derivative in the direction of the ith coordinate u^i is nothing but the case $X = \frac{\partial f}{\partial u^i}$. It follows that

$$D_{\frac{\partial f}{\partial u^i}} Y\Big|_p = \lim_{t\to 0} \frac{1}{t}\big(Y(u^1, \ldots, u^i + t, \ldots, u^n) - Y(u^1, \ldots, u^i, \ldots, u^n)\big)$$

and in particular

$$D_{\frac{\partial f}{\partial u^i}} \frac{\partial f}{\partial u^j} = \frac{\partial^2 f}{\partial u^i \partial u^j}.$$

4.3. Definition. (Covariant derivative)
If X, Y are tangent to a hypersurface element f, then the expression

$$\nabla_X Y := (D_X Y)^{\text{Tang.}} = D_X Y - \langle D_X Y, \nu\rangle \nu$$

is called the *covariant derivative* of Y in the direction of X. If X, Y are tangent vector fields, then the covariant derivative $\nabla_X Y$ is again a tangent vector field. The normal component of $D_X Y$ is nothing but the second fundamental form of f, since the equality

$$\langle D_X Y, \nu\rangle = I\!I(X, Y)$$

holds because of the relation $\langle Y, \nu\rangle = 0$, and consequently $\langle D_X Y, \nu\rangle = -\langle Y, D_X \nu\rangle$. Hence we can also write

$$D_X Y = \nabla_X Y + I\!I(X, Y)\nu.$$

Remark: In [**1**], D is written instead of ∇ for the covariant derivative. It is at any rate important to differentiate between the two differential operators (directional derivative and covariant derivative):

The directional derivative D is defined for vector fields on the ambient Euclidean space.
The covariant derivative ∇ is defined only for tangent vector fields on the hypersurface element.

For a scalar function φ along f there is only one kind of directional derivative in the direction of X, defined as the limit of a difference quotient and written $D_X\varphi = \nabla_X\varphi$. In addition, we can multiply such scalar functions pointwise with vector fields, with the notation φX for the vector field $p \mapsto (\varphi X)(p) := \varphi(p) \cdot X(p)$.

4.4. Lemma. (Properties of D and ∇)

(i) $D_{\varphi_1 X_1 + \varphi_2 X_2} Y = \varphi_1 D_{X_1} Y + \varphi_2 D_{X_2} Y$,
$\nabla_{\varphi_1 X_1 + \varphi_2 X_2} Y = \varphi_1 \nabla_{X_1} Y + \varphi_2 \nabla_{X_2} Y$. (*linearity*)

(ii) $D_X(Y_1 + Y_2) = D_X Y_1 + D_X Y_2$,
$\nabla_X(Y_1 + Y_2) = \nabla_X Y_1 + \nabla_X Y_2$. (*additivity*)

(iii) $D_X(\varphi Y) = \varphi D_X Y + D_X\varphi \cdot Y$,
$\nabla_X(\varphi Y) = \varphi \nabla_X Y + \nabla_X\varphi \cdot Y$. (*product rule*)

(iv) $D_X\langle Y_1, Y_2\rangle = \langle D_X Y_1, Y_2\rangle + \langle Y_1, D_X Y_2\rangle$,
$\nabla_X\langle Y_1, Y_2\rangle = \langle \nabla_X Y_1, Y_2\rangle + \langle Y_1, \nabla_X Y_2\rangle$. (*compatibility with*
. *the inner product*)

WARNING: For the directional derivative and the covariant derivative, commutativity fails, i.e., in general $D_X Y \neq D_Y X$ and $\nabla_X Y \neq \nabla_Y X$.

An example of this:

Let e_1, e_2 be the standard basis of $\mathbb{R}^2$ with coordinates (x^1, x^2). Then one has $D_{e_i} e_j = 0$ for all i, j. Choosing $X := x^1 \cdot e_2$, $Y := e_1$, we get

$$D_X Y = D_{x^1 e_2} e_1 = x^1 D_{e_2} e_1 = 0,$$

$$\text{but} \quad D_Y X = D_{e_1}(x^1 e_2) = x^1 \underbrace{D_{e_1} e_2}_{=0} + \underbrace{D_{e_1} x^1}_{=1} \cdot e_2 = e_2 \neq 0.$$

4.5. Definition. For two vector fields X, Y in $\mathbb{R}^{n+1}$ or two vector fields along f, the expression

$$[X, Y] := D_X Y - D_Y X$$

is called the *Lie bracket* of X and Y. One has $[X, Y] = \nabla_X Y - \nabla_Y X$, if X and Y are tangent vector fields. Moreover,

$$\left[\frac{\partial f}{\partial u^i}, \frac{\partial f}{\partial u^j}\right] = 0 \qquad \text{because} \qquad \frac{\partial^2 f}{\partial u^i \partial u^j} = \frac{\partial^2 f}{\partial u^j \partial u^i}.$$

In addition, in arbitrary coordinates, one has

$$[X,Y] = \sum_{i,j}\Big(\xi^i\frac{\partial\eta^j}{\partial u^i} - \eta^i\frac{\partial\xi^j}{\partial u^i}\Big)\frac{\partial f}{\partial u^j},$$

if $X = \sum_i \xi^i \frac{\partial f}{\partial u^i}, Y = \sum_j \eta^j \frac{\partial f}{\partial u^j}$. An abbreviated notation for this is

$$[X,Y]^j = X(Y^j) - Y(X^j).$$

Here the index j denotes the jth coordinate.

Consequence: For given vector fields X, Y the vanishing of the Lie bracket is a necessary condition for X and Y to be basis vector fields $X = \frac{\partial f}{\partial u^i}, Y = \frac{\partial f}{\partial u^j}$ for certain coordinates $u^1, \ldots, u^n$.

4.6. Theorem. The covariant derivative ∇ depends only on the first fundamental form, and as such is a quantity of the intrinsic geometry of the surface.

Proof: We set $X = \sum_i \xi^i \frac{\partial f}{\partial u^i}$ and $Y = \sum_j \eta^j \frac{\partial f}{\partial u^j}$. In order to determine $\nabla_X Y$, it is sufficient to know the quantities $\big\langle \nabla_X Y, \frac{\partial f}{\partial u^k}\big\rangle$ for all k. From the calculus rules 4.4 we get the equation

$$\nabla_X Y = \sum_i \xi^i \nabla_{\frac{\partial f}{\partial u^i}} Y = \sum_i \xi^i \sum_j \nabla_{\frac{\partial f}{\partial u^i}}\Big(\eta^j\frac{\partial f}{\partial u^j}\Big) =$$

$$= \sum_{ij}\xi^i\Big(\frac{\partial\eta^j}{\partial u^i}\frac{\partial f}{\partial u^j} + \eta^j\nabla_{\frac{\partial f}{\partial u^i}}\frac{\partial f}{\partial u^j}\Big),$$

and consequently

$$\Big\langle\nabla_X Y, \frac{\partial f}{\partial u^k}\Big\rangle = \sum_{ij}\xi^i\Big(\frac{\partial\eta^j}{\partial u^i}g_{jk} + \eta^j\Big\langle\nabla_{\frac{\partial f}{\partial u^i}}\frac{\partial f}{\partial u^j}, \frac{\partial f}{\partial u^k}\Big\rangle\Big).$$

Here we use the notation

$$\Gamma_{ij,k} := \big\langle\nabla_{\frac{\partial f}{\partial u^i}}\tfrac{\partial f}{\partial u^j}, \tfrac{\partial f}{\partial u^k}\big\rangle,$$

and these quantities are symmetric in the indices i and j, as we know by 4.5 that the Lie brackets of basis fields vanish. On the other hand, we also have

$$\frac{\partial}{\partial u^k}g_{ij} = \frac{\partial}{\partial u^k}\Big\langle\frac{\partial f}{\partial u^i}, \frac{\partial f}{\partial u^j}\Big\rangle = \Gamma_{ik,j} + \Gamma_{jk,i}.$$

By cyclically permuting the indices one gets

$$\frac{\partial}{\partial u^i} g_{jk} = \Gamma_{ji,k} + \Gamma_{ki,j}, \qquad \frac{\partial}{\partial u^j} g_{ki} = \Gamma_{kj,i} + \Gamma_{ij,k}.$$

From this we get, by adding or subtracting these equations,

$$2\Gamma_{ij,k} = -\frac{\partial}{\partial u^k} g_{ij} + \frac{\partial}{\partial u^i} g_{jk} + \frac{\partial}{\partial u^j} g_{ki},$$

which is an expression which clearly depends only on the first fundamental form. □

4.7. Definition. (Christoffel symbols)

(i) The quantities $\Gamma_{ij,k}$ defined by the expressions

$$\Gamma_{ij,k} := I\left(\nabla_{\frac{\partial f}{\partial u^i}} \frac{\partial f}{\partial u^j}, \frac{\partial f}{\partial u^k}\right)$$

are called the *Christoffel symbols of the first kind.*

(ii) The quantities Γ_{ij}^k defined by

$$\nabla_{\frac{\partial f}{\partial u^i}} \frac{\partial f}{\partial u^j} = \sum_k \Gamma_{ij}^k \frac{\partial f}{\partial u^k}$$

are called the *Christoffel symbols of the second kind.*

(iii) By definition one has $\Gamma_{ij,k} = \Gamma_{ji,k}, \Gamma_{ij}^k = \Gamma_{ji}^k$ as well as $\Gamma_{ij,k} = \sum_m \Gamma_{ij}^m g_{mk}$.

Consequence: The first fundamental form (g_{ij}) uniquely determines the Christoffel symbols and thus also the covariant derivative of $X = \sum_i \xi^i \frac{\partial f}{\partial u^i}$ and $Y = \sum_j \eta^j \frac{\partial f}{\partial u^j}$ through the equation

$$\nabla_X Y = \sum_{i,k} \xi^i \Big(\frac{\partial \eta^k}{\partial u^i} + \sum_j \eta^j \Gamma_{ij}^k\Big) \frac{\partial f}{\partial u^k}.$$

4.8. Corollary. (Equations of Gauss and Weingarten)
For every (hyper-)surface element f of class C^2, the following equations hold:

(i) The *Gauss formula*

$$\frac{\partial^2 f}{\partial u^i \partial u^j} = \sum_k \Gamma_{ij}^k \cdot \frac{\partial f}{\partial u^k} + h_{ij} \cdot \nu.$$

(ii) The *Weingarten equation*

$$\frac{\partial \nu}{\partial u^i} = -\sum_{j,k} h_{ij} g^{jk} \cdot \frac{\partial f}{\partial u^k} = -\sum_k h_i^k \cdot \frac{\partial f}{\partial u^k}.$$

These equations are also called the *partial differential equations of surface theory.* The proof is more or less contained in the above definitions (for the Gauss formula) and in 3.9 (for the Weingarten equation). In fact, compare the equations

$$D_X Y = \nabla_X Y + I\!I(X,Y)\nu \quad \text{and} \quad D\nu = -L \circ Df.$$

Like the Frenet equations in the theory of curves, 4.8 can also be written as a motion of the *Gaussian three-frame* $\frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^2}, \nu$ using the following matrix:

$$\frac{\partial}{\partial u^i} \begin{pmatrix} \frac{\partial f}{\partial u^1} \\ \frac{\partial f}{\partial u^2} \\ \nu \end{pmatrix} = \begin{pmatrix} \Gamma_{i1}^1 & \Gamma_{i1}^2 & h_{i1} \\ \Gamma_{i2}^1 & \Gamma_{i2}^2 & h_{i2} \\ -h_i^1 & -h_i^2 & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial u^1} \\ \frac{\partial f}{\partial u^2} \\ \nu \end{pmatrix},$$

and similarly in higher dimensions.

4B Parallel displacement and geodesics

That a vector field Y in Euclidean space is constant means just that the directional derivatives $D_X Y$ vanish in all directions X. Since one has to think of the different vectors in space as being based at different points, a constant vector field is characterized by the property that all these vectors are *parallel* to one another (and have the same length). However, the naive parallel transport of (p, X) to (q, X) would not work for surfaces since, if X is tangent at p, it is not necessarily tangent at q. Instead, this notion of being parallel has been transferred to the notion of covariant derivative in differential geometry as follows:

4.9. Definition. (Parallel vector field, geodesic)
If Y is a tangent vector field along a surface element f, then Y is called *parallel*, if $\nabla_X Y \equiv 0$ for every tangent vector X.

If Y is a vector field (tangent to f) along a regular curve $c = f \circ \gamma$, then Y is said to be *parallel along* c, if $\nabla_X Y = 0$ for every X which is tangent to c or, equivalently, if $\nabla_{\dot{c}} Y = 0$.

A non-constant curve c on a surface is called a *geodesic* (or *autoparallel*), if $\nabla_{\dot{c}}\dot{c} \equiv 0$ holds along the curve c or, equivalently, if $\nabla_{\dot{c}}\dot{c}$ and $\dot{c}$ are always linearly dependent. The equation $\nabla_{\dot{c}}\dot{c} \equiv 0$ requires that the parameter is proportional to the arc length.

PHYSICAL INTERPRETATION: If we view $c(t)$ as the motion of a mass particle, then the expression $D_{\dot{c}}\dot{c} = \ddot{c}$ is just the acceleration vector in Euclidean space. The motions free of acceleration (the lines) are characterized by the vanishing of this expression. Similarly, on the surface the expression $\nabla_{\dot{c}}\dot{c}$ is the vector of acceleration on the surface, i.e., the tangential component of the acceleration. In this sense the geodesics are the motions on the surface which are free of acceleration (meaning without consideration of the forces which act perpendicular to the surface). On the surface of the sphere, the geodesics are precisely the great circles. The consideration of $\nabla_{\dot{c}}\dot{c}$ requires by 4.6 only the knowledge of the first fundamental form. From this it is clear that geodesics are quantities of the intrinsic geometry of a surface.

WARNING: In general there is no non-trivial parallel vector field on open sets of surfaces, however there are always parallel vector fields along given curves (4.10), and locally one always has geodesics (4.12).

REMARK: If one (in special cases) wishes to consider *non-regular* curves $c(t)$, then one can still define a covariant derivative of a vector field Y along that curve by the relation $\nabla_{\dot{c}} Y = \left(\frac{dY}{dt}\right)^{\text{Tang.}}$. The calculus rules in 4.4 hold in this case also. For a constant curve c, a vector field Y is parallel along c if and only if Y is constant.

A further notation for this is

$$D_{\dot{c}} Y = DY\Big(\frac{d}{dt}\Big) = \frac{dY(t)}{dt} = Y'(t), \quad \nabla_{\dot{c}} Y = \frac{\nabla Y(t)}{dt}.$$

4.10. Theorem. (Parallel displacement)
For every continuously differentiable curve $\gamma\colon I \to U$ on a hypersurface element $f\colon U \to I\!R^{n+1}$ there exists for every given vector Y_0 which is tangent to the surface at a point $f(\gamma(t_0))$ a unique vector field Y along $c := f \circ \gamma$ which is parallel to c and whose value at $c(t_0)$ is Y_0. It is called the *parallel displacement* of Y_0 along c.

PROOF: As above we denote by $\eta^j(t)$ the coefficients of the vector $Y(t) = \sum_j \eta^j(t)\frac{\partial f}{\partial u^j}$ and by $u^i(t)$ the coordinates of the curve $\gamma(t)$. In the formulas for $\nabla_X Y$ following 4.7 above we get the relation $\xi^i(t) = \dot{u}^i(t)$, and consequently (using the chain rule)

$$\begin{aligned}\nabla_{\dot{c}} Y &= \sum_{i,k} \dot{u}^i(t)\Big(\frac{\partial \eta^k(t)}{\partial u^i} + \sum_j \eta^j(t)\Gamma^k_{ij}(c(t))\Big)\frac{\partial f}{\partial u^k}\\ &= \sum_k \Big(\frac{d\eta^k}{dt} + \sum_{i,j} \dot{u}^i\eta^j\Gamma^k_{ij}\Big)\frac{\partial f}{\partial u^k}.\end{aligned}$$

The requirement that Y should be parallel is thus equivalent to the system of ordinary differential equations

$$\dot{\eta}^k(t) + \sum_{i,j} \dot{u}^i(t)\eta^j(t)\Gamma^k_{ij}(c(t)) = 0$$

for the function $\eta^k(t), k = 1, \ldots, n$. This system is linear, hence there exists for given initial conditions $\eta^1(t_0), \ldots, \eta^n(t_0)$ exactly one solution for every t in the given interval ([**27**], Chapter XIX). $\square$

4.11. Corollary. A parallel vector field along a curve has a constant length. In particular, for a geodesic with $\nabla_{\dot{c}}\dot{c} = 0$, the length $||\dot{c}||$ of the tangent is necessarily constant.

This follows easily from the equation (properties 4.4) $\nabla_{\dot{c}}\langle X, X\rangle = 2\langle \nabla_{\dot{c}} X, X\rangle = 0$, provided that $\nabla_{\dot{c}} X = 0$. Similarly, the inner product (and hence also the angle) of two parallel vector fields along the same curve is constant.

REMARK: An arbitrary regular curve $c(t)$ can be transformed into a geodesic by a reparametrization if and only if $\nabla_{\dot{c}}\dot{c}$ and $\dot{c}$ are always linearly dependent. This expresses the fact that an acceleration can only be forward or backward, not to the side.

4.12. Theorem. (Geodesics)
For every point $p_0 = f(u_0)$ on a surface element f, and for every vector Y_0 with $\|Y_0\| = 1$ which is tangent to the surface at p_0, there exist an $\varepsilon > 0$ and a unique geodesic $c = f \circ \gamma$, parametrized by arc length, for which $c(0) = p_0$ and $\dot{c}(0) = Y_0$, where $\gamma\colon (-\varepsilon, \varepsilon) \to U$ is differentiable with $\gamma(0) = u_0$.

Proof: To set up the differential equation for a geodesic $c(t)$, we must set $\eta^i = \dot{u}^i$ in the corresponding equation in the proof of 4.10. We get (again utilizing the chain rule)

$$\begin{aligned} 0 = \nabla_{\dot{c}}\dot{c} &= \sum_{i,k} \dot{u}^i(t)\Big(\frac{\partial \dot{u}^k(t)}{\partial u^i} + \sum_j \dot{u}^j(t)\Gamma^k_{ij}(c(t))\Big)\frac{\partial f}{\partial u^k} \\ &= \sum_k \Big(\ddot{u}^k(t) + \sum_{i,j} \dot{u}^i(t)\dot{u}^j(t)\Gamma^k_{ij}(c(t))\Big)\frac{\partial f}{\partial u^k}, \end{aligned}$$

which leads to the system of equations

$$\ddot{u}^k(t) + \sum_{i,j} \dot{u}^i(t)\dot{u}^j(t)\Gamma^k_{ij}(c(t)) = 0$$

for $k = 1, \ldots, n$. Since $c(t)$ is determined by the functions $u^i(t)$ according to the relation $c(t) = f(\gamma(t)) = f(u^1(t), \ldots, u^n(t))$, this is a system of ordinary differential equations of the second order for $u^i(t)$ (not of the first order for $\dot{u}^i(t)$). The local existence of solutions for given initial conditions $u^i(0), \dot{u}^i(0)$ then follows from general results ([**27**], Chapter XIX). $\square$

Remark: By Definition 4.3 the expression $\nabla_{\dot{c}}\dot{c}$ is nothing but the tangent component of $D_{\dot{c}}\dot{c} = \ddot{c}$. The (oriented) length of the normal component of this is the *normal curvature* κ_ν, while the (oriented) length of the tangent component is also called the *geodesic curvature*, cf. 3.11 and 4.37. Thus geodesics are also characterized as the curves with vanishing geodesic curvature.

Examples: It is clear that in Euclidean space the geodesics are precisely the straight lines since all Γ^k_{ij} vanish identically. On surfaces $(t, \varphi) \longmapsto \big(r(t)\cos\varphi, r(t)\sin\varphi, h(t)\big)$ of rotation, the curves $\varphi = \text{const}$ are always geodesics (up to reparametrization), while the curves $t = \text{const}$ are geodesics only for values t_0 with $\dot{r}(t_0) = 0$.

4.13. Theorem. ("Shortest paths are geodesics")
Let p, q be fixed points on a surface element $f\colon U \to I\!R^{n+1}$, joined by a regular C^∞-curve $c = f \circ \gamma$. If c is a "shortest path" (i.e., if any other C^∞-curve with the same endpoints is at least as long), then c is a geodesic (up to reparametrizations).

WARNING: This does not answer the question as to whether shortest paths exist, or whether they are differentiable and regular if they exist.

PROOF: This is based on the same variational principle that we already applied in 3.27 in the investigation of minimal surfaces. We start with a given curve and compare its length with the length of curves which are "near" to it. For this it is convenient to parametrize the given curve $c(s)$ by arc length in the interval $[0, L]$ and then embed this in a family of curves with parameter t, fixing the endpoints. Thus, let $C(s, t)$ be a differentiable map with

$$C(s,0) = c(s), \quad C(0,t) = c(0), \quad C(L,t) = c(L),$$

for all $s \in [0, L]$ and $t \in [-\varepsilon, \varepsilon]$. The curves we compare $c(s)$ with are then $c_t(s) = C(s, t)$. They have the same starting and ending points as c, but they are in general not parametrized by arc length. The length of c_t is simply given by

$$L(c_t) = \int_0^L \Big\langle \frac{\partial c_t}{\partial s}, \frac{\partial c_t}{\partial s} \Big\rangle^{1/2} ds.$$

We take the derivative of this expression with respect to t at the point $t = 0$:

$$\begin{aligned}
\frac{\partial}{\partial t}\Big|_{t=0} L(c_t) &= \frac{\partial}{\partial t}\Big|_{t=0} \int_0^L \Big\langle \frac{\partial c_t}{\partial s}, \frac{\partial c_t}{\partial s} \Big\rangle^{1/2} ds \\
&= \int_0^L \frac{\partial}{\partial t}\Big|_{t=0} \Big\langle \frac{\partial c_t}{\partial s}, \frac{\partial c_t}{\partial s} \Big\rangle^{1/2} ds \\
&= \int_0^L \frac{\langle \frac{\partial^2 C}{\partial t \partial s}, \frac{\partial C}{\partial s} \rangle}{\langle c', c' \rangle^{1/2}} ds = \int_0^L \Big\langle \frac{\partial^2 C}{\partial s \partial t}, \frac{\partial C}{\partial s} \Big\rangle ds \\
= \int_0^L \Big\langle \nabla_{c'} \frac{\partial C}{\partial t}, c' \Big\rangle ds &= \int_0^L \Big(\frac{\partial}{\partial s} \langle \frac{\partial C}{\partial t}, c' \rangle - \langle \frac{\partial c}{\partial t}, \nabla_{c'} c' \rangle \Big) ds
\end{aligned}$$

$$= \left\langle \frac{\partial C}{\partial t}, c' \right\rangle \Big|_{s=L} - \left\langle \frac{\partial C}{\partial t}, c' \right\rangle \Big|_{s=0} - \int_0^L \left\langle \frac{\partial C}{\partial t}, \nabla_{c'} c' \right\rangle ds$$
$$= - \int_0^L \left\langle \frac{\partial C}{\partial t}, \nabla_{c'} c' \right\rangle ds$$

because $\frac{\partial C}{\partial t}|_{s=0} = \frac{\partial C}{\partial t}|_{s=L} = 0$.

Now, by assumption we have $\frac{\partial}{\partial t}|_{t=0} L(c_t) = 0$ for every mapping of this kind. This is only possible if $\nabla_{c'} c' = 0$ holds along the whole curve c. Otherwise we could construct a C in such a way that the integral is non-vanishing. Suppose for example $\nabla_{c'} c' \neq 0$ for some parameter s_0. Since $\nabla_{c'} c'$ is always perpendicular to c', we can construct a parameter map $C(s,t)$ with $\frac{\partial C}{\partial t}|_{s=s_0} = \nabla_{c'} c'$ (for example, over the parameter domain). By applying an appropriate cutoff function, one can also ensure that the endpoints are the given points. Moreover, one can ensure that $C(s,t) = c(s)$ outside of a small neighborhood of s_0, where the integrand $\left\langle \frac{\partial C}{\partial t}, \nabla_{c'} c' \right\rangle$ does not change sign. This is a contradiction to the above calculation under the assumption that the curve c is the shortest path. Thus we see the role played by the map C: it is arbitrary, and there are certainly sufficiently many such maps to carry out the argument. For this it is irrelevant whether one can actually realize an *arbitrary* one-parameter family c_t in this manner.

□

4C The Gauss equation and the Theorema Egregium

In this section we want to investigate the equations of Gauss and Weingarten in 4.8 with a view to the situation in which the surface f itself is not given, but rather its (hypothetical) first and second fundamental forms. In the process, we naturally run across more equations, which describe the integrability conditions for the derivative equations (these are conditions which guarantee that the derivative equations can be solved, yielding a surface f with the given fundamental forms), and this has various implications (some of them quite surprising), for example the *Theorema Egregium* of Gauss 4.16, which is anything but obvious.

4.14. Remark. (Strategy for determining f)

We want to integrate the equations of Gauss and Weingarten in 4.8 as a system of partial differential equations (for $n \geq 2$). Note that, even in the simplest cases, for the existence of a solution f of a system of equations

$$\frac{\partial f}{\partial u^1} = b_1(u^1, \ldots, u^n), \ldots, \frac{\partial f}{\partial u^n} = b_n(u^1, \ldots, u^n)$$

for given functions $b_1, \ldots, b_n$, necessary conditions must be fulfilled, the so-called *integrability conditions*

$$\frac{\partial b_j}{\partial u^i} = \frac{\partial b_i}{\partial u^j} \quad \text{for all } i, j,$$

since the second partial derivatives of a solution f (if it were to exist) with respect to u_i and u_j are symmetric in i and j. One can for example integrate from the origin $(0, \ldots, 0)$:

$$f(0, \ldots, 0, u^n) = f(0, \ldots, 0) + \int_0^{u^n} b_n(0, \ldots, 0, x) dx;$$

$$f(0, \ldots, 0, u^{n-1}, u^n) = f(0, \ldots, 0, u^n) + \int_0^{u^{n-1}} b_{n-1}(0, \ldots, 0, x, u^n) dx;$$

$$\vdots$$

$$f(u^1, \ldots, u^n) = f(0, u^2, \ldots, u^n) + \int_0^{u^1} b_1(x, u^2 \ldots, u^n) dx.$$

Because of the integrability conditions these integrals are independent of the path of integration, at least inside an n-dimensional cube with the corners $(0, \ldots, 0)$ and $(u^1, \ldots, u^n)$. In particular, changing the order of integration does not change the result. We must expect similar integrability conditions in 4.8:

$$\frac{\partial^3 f}{\partial u^i \partial u^j \partial u^k} = \frac{\partial^3 f}{\partial u^i \partial u^k \partial u^j},$$

$$\frac{\partial^2 \nu}{\partial u^i \partial u^j} = \frac{\partial^2 \nu}{\partial u^j \partial u^i},$$

and the corresponding expressions for the right-hand sides of 4.8 (i) and (ii). Obviously this will require third derivatives of f.

4.15. Theorem. For a (hyper-)surface element f of class C^3, the integrability conditions for the equations of Gauss and Weingarten (4.8) are the two following equations:

(i) The *Gauss equation*

$$\frac{\partial}{\partial u^k}\Gamma_{ij}^s - \frac{\partial}{\partial u^j}\Gamma_{ik}^s + \sum_r \left(\Gamma_{ij}^r \Gamma_{rk}^s - \Gamma_{ik}^r \Gamma_{rj}^s\right)$$
$$= \sum_m \left(h_{ij}h_{km} - h_{ik}h_{jm}\right)g^{ms} \quad \text{for all } i,j,k,s$$

(ii) The *Codazzi-Mainardi equation*

$$\frac{\partial}{\partial u^k}h_{ij} - \frac{\partial}{\partial u^j}h_{ik} + \sum_r \left(\Gamma_{ij}^r h_{rk} - \Gamma_{ik}^r h_{rj}\right) = 0 \ \text{ for all } i,j,k.$$

Proof: We take the derivative of the Gauss formula from 4.8 and insert both equations from 4.8 in the right place in the result:

$$0 = \frac{\partial}{\partial u^k}\frac{\partial^2 f}{\partial u^i \partial u^j} - \frac{\partial}{\partial u^j}\frac{\partial^2 f}{\partial u^i \partial u^k}$$

$$= \sum_s \left(\frac{\partial}{\partial u^k}\Gamma_{ij}^s\right)\frac{\partial f}{\partial u^s} - \sum_s \left(\frac{\partial}{\partial u^j}\Gamma_{ik}^s\right)\frac{\partial f}{\partial u^s} + \sum_r \Gamma_{ij}^r \frac{\partial^2 f}{\partial u^k \partial u^r}$$

$$-\sum_r \Gamma_{ik}^r \frac{\partial^2 f}{\partial u^j \partial u^r} + \left(\frac{\partial}{\partial u^k}h_{ij}\right)\nu - \left(\frac{\partial}{\partial u^j}h_{ik}\right)\nu + h_{ij}\frac{\partial \nu}{\partial u^k} - h_{ik}\frac{\partial \nu}{\partial u^j}$$

$$= \sum_s \left(\frac{\partial}{\partial u^k}\Gamma_{ij}^s - \frac{\partial}{\partial u^j}\Gamma_{ik}^s\right)\frac{\partial f}{\partial u^s} + \sum_r \Gamma_{ij}^r \Big(\sum_s \Gamma_{kr}^s \frac{\partial f}{\partial u^s} + h_{kr}\nu\Big)$$

$$-\sum_r \Gamma_{ik}^r \Big(\sum_s \Gamma_{jr}^s \frac{\partial f}{\partial u^s} + h_{jr}\nu\Big)$$

$$+\Big(\frac{\partial}{\partial u^k}h_{ij} - \frac{\partial}{\partial u^j}h_{ik}\Big)\nu - h_{ij}\sum_{m,s} h_{km}g^{ms}\frac{\partial f}{\partial u^s} + h_{ik}\sum_{m,s} h_{jm}g^{ms}\frac{\partial f}{\partial u^s}.$$

The Gauss equation simply expresses the vanishing of the coefficients of $\frac{\partial f}{\partial u^s}$, while the Codazzi-Mainardi equation expresses the vanishing of the coefficients of ν. The analogous integrability condition for the Weingarten equation from 4.8,

$$0 = \frac{\partial^2 \nu}{\partial u^i \partial u^j} - \frac{\partial^2 \nu}{\partial u^j \partial u^i},$$

yields no new equations, but rather is just a reformulation of the Codazzi-Mainardi equation. This makes sense, since the motion of the normal is always coupled to the motion of the tangent (hyper-) plane. □

REMARK: The left-hand side of the Gauss equation is called the *curvature tensor* and is in general expressed in the form

$$R^s_{ikj} := \frac{\partial}{\partial u^k}\Gamma^s_{ij} - \frac{\partial}{\partial u^j}\Gamma^s_{ik} + \sum_r \left(\Gamma^r_{ij}\Gamma^s_{rk} - \Gamma^r_{ik}\Gamma^s_{rj}\right).$$

For the significance of the curvature tensor see 4.19 below.

> **4.16. Corollary.** (*Theorema Egregium* of C. F. Gauss)
> The Gaussian curvature K of a two-dimensional surface element $f\colon U \to I\!R^3$ of class C^3 depends only on the first fundamental form[2] (and is consequently an intrinsic quantity of the surface).

Recall that the Gaussian curvature K was defined by $K = \mathrm{Det}(L) = \mathrm{Det}(I\!I)/\mathrm{Det}(I)$, cf. 3.13.

PROOF: We set $i = j = 1, k = 2$ in the Gauss equation and multiply through by g_{s2}. The right-hand side results from a summation over s in the expression

$$\sum_{m,s}\left(h_{11}h_{2m} - h_{12}h_{1m}\right)g^{ms}g_{s2} = h_{11}h_{22} - h_{12}h_{12} = \mathrm{Det}(I\!I).$$

This expression depends only on the first fundamental form, as follows from the Gauss equation (left-hand side of the equation). Thus this also holds for $K = \mathrm{Det}(I\!I)/\mathrm{Det}(I) = \sum_s g_{2s}R^s_{121}/(g_{11}g_{22} - g_{12}^2)$. □

REMARK: The mean curvature H does not depend only on the first fundamental form. For example, on the one hand we have the plane with $H = 0$ and the cylinder with $H \neq 0$, cf. 4.25. However, both surfaces have the same first fundamental form.

[2] This is also true if f is only of class C^2, see Ph.Hartman & A.Wintner, *On the fundamental equations of differential geometry*, American Journal of Math. **72** (1950), 757–774.

4.17. Lemma. Let X, Y, Z be vector fields, defined on an open set in $\mathbb{R}^n$. Then one has

$$D_X(D_Y Z) - D_Y(D_X Z) = D_{[X,Y]}Z.$$

PROOF: First note that the statement is trivial if X, Y are the standard basis elements e_i, e_j of $\mathbb{R}^n$, since $D_{e_i}D_{e_j}Z = D_{e_j}D_{e_i}Z$ and $[e_i, e_j] = 0$.

Next, let $X = \sum_i \xi^i e_i$ and $Y = \sum_j \eta^j e_j$. The calculus rules in 4.4 imply the relations

$$D_X D_Y Z - D_Y D_X Z$$

$$= \sum_i \xi^i D_{e_i}\Big(\sum_j \eta^j D_{e_j} Z\Big) - \sum_j \eta^j D_{e_j}\Big(\sum_i \xi^i D_{e_i} Z\Big)$$

$$= \sum_{i,j} \xi^i \eta^j (D_{e_i}D_{e_j}Z - D_{e_j}D_{e_i}Z) + \sum_{i,j} \xi^i \frac{\partial \eta^j}{\partial x^i} D_{e_j}Z - \sum_{i,j} \eta^j \frac{\partial \xi^i}{\partial x^j} D_{e_i}Z$$

$$= \sum_{i,j}\Big(\xi^i \frac{\partial \eta^j}{\partial x^i} - \eta^i \frac{\partial \xi^j}{\partial x^i}\Big) D_{e_j}Z = D_{[X,Y]}Z.$$

The last equality is legitimate by 4.5 (Lie brackets in local coordinates). □

4.18. Theorem. (Variants of the integrability conditions)
Let X, Y, Z be tangent vector fields along a surface element $f\colon U \to \mathbb{R}^{n+1}$. Then the following equations hold:

(i) The Gauss equation

$$\begin{aligned}\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z &= \langle LY, Z\rangle LX - \langle LX, Z\rangle LY\\ &= I\!I(Y,Z)LX - I\!I(X,Z)LY\end{aligned}$$

(ii) The Codazzi-Mainardi equation

$$\nabla_X(LY) - \nabla_Y(LX) - L([X,Y]) = 0.$$

PROOF: We simply decompose the equation of 4.17 into the tangent and normal components with the aid of the formula $D_X Y = \nabla_X Y + \langle LX, Y\rangle\nu$ and then apply the calculus rules in 4.4:

$$
\begin{aligned}
0 &= D_X D_Y Z - D_Y D_X Z - D_{[X,Y]} Z = D_X \big(\nabla_Y Z + \langle LY, Z\rangle \nu\big) \\
&\quad - D_Y \big(\nabla_X Z + \langle LX, Z\rangle \nu\big) - \nabla_{[X,Y]} Z - \langle L([X,Y]), Z\rangle \nu \\
&= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z - \langle LY, Z\rangle LX + \langle LX, Z\rangle LY \\
&\quad + \Big(\langle \nabla_X (LY), Z\rangle - \langle \nabla_Y (LX), Z\rangle - \langle L([X,Y]), Z\rangle\Big)\nu.
\end{aligned}
$$

4.19. Corollary and Definition. (Curvature tensor)

The value of $\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$ at a point p depends only on the value of X, Y, Z at p, because the right hand side of the Gauss equation in 4.18 does. One also refers to this state of affairs by saying

$$R(X,Y)Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

is a tensor field, which is called the *curvature tensor* of the surface, cf. Chapter 6. This tensor field only depends on the first fundamental form. The Gauss equation can be written

$$R(X,Y)Z = \langle LY, Z\rangle LX - \langle LX, Z\rangle LY.$$

The notation $R(X,Y)Z$ comes from the fact that for fixed vectors X, Y, the so-called *curvature transformation* $R(X,Y)$ can be viewed as an endomorphism of the tangent space. For the two-dimensional unit sphere, this transformation is just a rotation by $\pi/2$, as L is the identity. Here we are assuming that X and Y are orthonormal. In the parameters $u^1, \ldots, u^n$ we have the equation

$$R\Big(\frac{\partial f}{\partial u^k}, \frac{\partial f}{\partial u^j}\Big)\frac{\partial f}{\partial u^i} = \sum_s R^s_{ikj} \frac{\partial f}{\partial u^s}$$

with the quantities R^s_{ikj} which also occurred in 4.15. The curvature tensor of Euclidean space is according to 4.17 simply $R(X,Y)Z = D_X D_Y Z - D_Y D_X Z - D_{[X,Y]} Z = 0$.

Remark: If one introduces the notation $\nabla_X L$ by means of the "product rule"

$$\nabla_X (LY) = (\nabla_X L)(Y) + L(\nabla_X Y),$$

then the Codazzi-Mainardi equation takes on the following simple form of a *symmetry of* ∇L for the shape operator L:

$$0 = \nabla_X (LY) - \nabla_Y (LX) - L(\nabla_X Y - \nabla_Y X) = (\nabla_X L)(Y) - (\nabla_Y L)(X).$$

For this reason, an endomorphism field A of the tangent space which satisfies $(\nabla_X A)(Y) - (\nabla_Y A)(X) = 0$ for all X, Y is called a *Codazzi tensor*.

4.20. Corollary. (Variant of 4.16, *Theorema Egregium*)
Let X, Y be orthonormal vector fields along $f\colon U \longrightarrow \mathbb{R}^3$. Then one has $\big\langle R(X,Y)Y, X\big\rangle = \mathrm{Det}(L) = K$.

PROOF: In the orthonormal basis X, Y, the shape operator L is represented by the matrix

$$\begin{pmatrix} \langle LX, X\rangle & \langle LX, Y\rangle \\ \langle LY, X\rangle & \langle LY, Y\rangle \end{pmatrix}.$$

The Gauss equation 4.19 now implies (with $Z = Y$)
$\big\langle R(X,Y)Y, X\big\rangle = \langle LY, Y\rangle\langle LX, X\rangle - \langle LX, Y\rangle\langle LY, X\rangle = \mathrm{Det}(L) = K.$

4.21. Corollary. Let $f\colon U \to \mathbb{R}^{n+1}$ be a hypersurface element with orthonormal principle curvature directions $X_1, \ldots, X_n$ and principle curvatures $\kappa_1, \ldots, \kappa_n$. Then one has $\big\langle R(X_i, X_j)X_j, X_i\big\rangle = \kappa_i\kappa_j$ for all $i \neq j$.

PROOF: For $LX_i = \kappa_i X_i$ and $\langle X_i, X_i\rangle = 1, \langle X_i, X_j\rangle = 0$, one has by virtue of the Gauss equation

$$\begin{gathered}\big\langle R(X_i, X_j)X_j, X_i\big\rangle \\ = \langle LX_j, X_j\rangle\langle LX_i, X_i\rangle - \langle LX_i, X_j\rangle\langle LX_j, X_i\rangle = \kappa_j\kappa_i - 0.\end{gathered}$$

The expression $\big\langle R(X_i, X_j)X_j, X_i\big\rangle$ can be viewed as an analog of the Gaussian curvature, a kind of curvature in the i,j-plane. In Riemannian geometry, this quantity is referred to as the *sectional curvature* in this plane (cf. Section 6B).

4.22. Corollary. The second mean curvature (cf. 3.46)

$$K_2 = \frac{1}{\binom{n}{2}} \sum_{i<j} \kappa_i\kappa_j = \frac{1}{n(n-1)} \sum_{i\neq j} \kappa_i\kappa_j$$

is a quantity of the intrinsic geometry, since it is independent of the choice of X_i. Indeed, the Gauss equation shows that K_2 is a trace-quantity, which can be calculated purely in terms of the curvature tensor. The sum of the $\big\langle R(X_i, X_j)X_j, X_i\big\rangle$ over all $i \neq j$ always has the same value, independent of the choice of the orthonormal basis X_i and independent of the Weingarten map. One calls the quantity $\sum_{i\neq j}\kappa_i\kappa_j$ the *scalar curvature*, because it is a scalar curvature function on the surface and for its determination no choice of tangent vectors on the surface is necessary. This last property is to be seen in contrast with, for example, the Ricci curvature, which depends on a direction, cf. 6.10.

4D The fundamental theorem of the local theory of surfaces

In this section we want to carry through the integration of the equations of Gauss and Weingarten in 4.8, since we have already discussed the corresponding integrability condition in detail in 4.15. The initial value problem for this integration has a very intuitively clear geometric interpretation, namely as the parameters of the Euclidean motions (translations and rotations) which one may apply to the surface. Thus we first formulate the invariance of the Gauss and Weingarten equations under Euclidean motions. Recall that the orientation of a surface element is essentially determined by the choice of the unit normal vectors, cf. 3.7.

4.23. Lemma. (Invariance under motions and uniqueness)
Let $f\colon U \to \mathbb{R}^{n+1}$ be a given surface element and let $B\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be a Euclidean motion, i.e., $B(x) = A(x) + b$ with an orthogonal map $A \in \mathbf{SO}(n+1)$. Set $\widetilde{f} := B \circ f$. Then, after an appropriate choice of unit normal vectors, for the two fundamental forms $g, \widetilde{g}, h, \widetilde{h}$ one has the equations

$$g_{ij} = \widetilde{g}_{ij},\ h_{ij} = \widetilde{h}_{ij}.$$

Conversely, if for two surface elements $f, \widetilde{f}\colon U \to \mathbb{R}^{n+1}$ which are oriented in the same way, the equations $g_{ij} = \widetilde{g}_{ij},\ h_{ij} = \widetilde{h}_{ij}$ hold and if U is connected, then the two surface elements are the same up to a Euclidean motion, i.e., one has

$$\widetilde{f} := B \circ f$$

for some Euclidean motion B.

Proof: If A and b are constant, then $\frac{\partial \widetilde{f}}{\partial u^i} = A\Big(\frac{\partial f}{\partial u^i}\Big)$, and, for an appropriate choice of unit normals ν, one has $\widetilde{\nu} = A\nu$. Since A is orthogonal, the claim of the first part follows immediately.

For the second part we define for every $u \in U$ a map $A(u)\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ by $A(u)\Big(\frac{\partial f}{\partial u^i}\Big|_u\Big) = \frac{\partial \widetilde{f}}{\partial u^i}\Big|_u$ and $A(u)(\nu(u)) = \widetilde{\nu}(u)$. The map $A(u)$ is then an orthogonal map $A(u) \in \mathbf{SO}(n+1)$ for all u. We will now show that $A(u)$ does not depend on u. By taking one more derivative, we get on the one hand

$$\frac{\partial^2 \widetilde{f}}{\partial u^i \partial u^j} = \frac{\partial}{\partial u^i}\Big(A\frac{\partial f}{\partial u^j}\Big) = \frac{\partial A}{\partial u^i}\Big(\frac{\partial f}{\partial u^j}\Big) + A\Big(\frac{\partial^2 f}{\partial u^i \partial u^j}\Big),$$

$$\frac{\partial \widetilde{\nu}}{\partial u^i} = \frac{\partial (A\nu)}{\partial u^i} = \frac{\partial A}{\partial u^i}(\nu) + A\Big(\frac{\partial \nu}{\partial u^i}\Big).$$

On the other hand, $\widetilde{g}^{ij} = g^{ij}$ and $\widetilde{\Gamma}_{ij}^k = \Gamma_{ij}^k$. From this and the equations of Gauss and Weingarten it follows that

$$\frac{\partial^2 \widetilde{f}}{\partial u^i \partial u^j} = A\Big(\frac{\partial^2 f}{\partial u^i \partial u^j}\Big),$$

$$\frac{\partial \widetilde{\nu}}{\partial u^i} = A\Big(\frac{\partial \nu}{\partial u^i}\Big).$$

This implies $\frac{\partial A}{\partial u^i} = 0$ for all i. Hence A is constant and $\widetilde{f} - A(f)$ is also because of $\frac{\partial \widetilde{f}}{\partial u^i} = A(\frac{\partial f}{\partial u^i})$. $\square$

4.24. Theorem. (The fundamental theorem of the local theory of surfaces, O. Bonnet)
On an open set $U \subset \mathbb{R}^n$ let symmetric matrix functions

$$g_{ij} = g_{ij}(u^1, \ldots, u^n), \quad h_{ij} = h_{ij}(u^1, \ldots, u^n)$$

of class C^2 resp. C^1 be given, so that (g_{ij}) is everywhere positive definite and so that g_{ij} and h_{ij} fulfill the Gauss and Codazzi-Mainardi equations (4.15).

Then, for a given initial condition

$$u^{(0)} \in U,\ p_0 \in I\!R^{n+1}, X_1^{(0)}, X_2^{(0)}, \ldots, X_n^{(0)} \in R^{n+1} \cong T_{p_0} I\!R^{n+1}$$

with $\langle X_i^{(0)}, X_j^{(0)} \rangle = g_{ij}(u^{(0)})$ and given unit normal $\nu^{(0)}$ in p_0 (i.e., a unit vector which is perpendicular to all $X_i^{(0)}$), there is an open connected subset $V \subset U$, $u^{(0)} \in V$ and a unique (hyper-)surface element $f : V \to I\!R^{n+1}$ of class C^3 whose Gauss map is ν and which has the properties

1. $f(u^{(0)}) = p_0$;
2. $\frac{\partial f}{\partial u^i}\big(u^{(0)}\big) = X_i^{(0)}$ for $i = 1, \ldots, n$;
3. $\nu(u^{(0)}) = \nu^{(0)}$;
4. g_{ij} and h_{ij} are the first and second fundamental form of f (with respect to ν).

Proof: The initial condition $X_i^{(0)}$ uniquely determines the unit normal $\nu^{(0)}$ at the point p_0 by requiring that $X_1^{(0)}, X_2^{(0)}, \ldots, X_n^{(0)}, \nu^{(0)}$ is a positively oriented basis of R^{n+1}. Then, in particular, $\langle \nu^{(0)}, \nu^{(0)} \rangle = 1$, $\langle \nu^{(0)}, X_i^{(0)} \rangle = 0$ for $i = 1, \ldots, n$. Thus, after choosing an orientation we may assume without restriction of generality that ν has been fixed. However, formally the same surface can have both h_{ij} and $-h_{ij}$ as second fundamental form, depending on the choice of the normal. Without the choice of normal in 4.24, the surface would only be unique up to a reflection at the tangent plane $T_{u^{(o)}} f$.

We write down the equations of Gauss and Weingarten 4.8 in two steps as systems of linear partial differential equations of the first order,

$$\frac{\partial X_j}{\partial u^i} = \sum_k \Gamma_{ij}^k X_k + h_{ij}\nu, \quad \frac{\partial \nu}{\partial u^i} = -\sum_{j,k} h_{ij} g^{jk} X_k$$

on the one hand, and

$$\frac{\partial f}{\partial u^j} = X_j$$

on the other.

Step 1: As a first step we seek a solution $X_1, \ldots, X_n, \nu$ of the first system. The integrability conditions

$$\frac{\partial^2 X_j}{\partial u^l \partial u^m} = \frac{\partial^2 X_j}{\partial u^m \partial u^l} \quad \text{und} \quad \frac{\partial^2 \nu}{\partial u^l \partial u^m} = \frac{\partial^2 \nu}{\partial u^m \partial u^l}$$

are exactly the equations of Gauss and Codazzi–Mainardi in 4.15 (just replace there $\frac{\partial f}{\partial u^j}$ by X_j) which are satisfied by assumption. Therefore there exists locally a unique solution to the given initial condition $X_1^{(0)}, X_2^{(0)}, \ldots, X_n^{(0)}, \nu^{(0)}$. This existence and uniqueness result can be found for example in [**6**], Appendix B. It goes back to Frobenius (1877). The method is to reduce the system of partial differential equations to a linear system of ordinary differential equations. The integration is carried out in the same way as we saw at the beginning of Section 4C: first a curve integral from the fixed initial point to a second point is calculated, which is independent of the path of integration by the integrability conditions (at least on some simply connected open set, for example an ε-neighborhood V of $u^{(0)}$, which lies inside of U). If U itself is simply connected, then one can set $V = U$.

It remains to show that

$$\langle \nu, \nu \rangle = 1, \quad \langle \nu, X_i \rangle = 0, \quad \langle X_i, X_j \rangle = g_{ij}.$$

This holds certainly at the point p_0 because of the assumed initial conditions. The three equations then also hold in a neighborhood if both sides turn out to be solutions of one and the same differential equation. To see this, we take the derivative of the three left-hand sides, using the first system of equations above:

$$\frac{\partial}{\partial u^i} \langle \nu, \nu \rangle = 2 \left\langle \frac{\partial \nu}{\partial u^i}, \nu \right\rangle = -2 \sum_{k,l} h_{ik} g^{kl} \langle X_l, \nu \rangle;$$

$$\frac{\partial}{\partial u^i} \langle \nu, X_j \rangle = \left\langle \frac{\partial \nu}{\partial u^i}, X_j \right\rangle + \left\langle \nu, \frac{\partial X_j}{\partial u^i} \right\rangle$$

$$= -\sum_{k,l} h_{ik} g^{kl} \langle X_l, X_j \rangle + \sum_k \Gamma_{ij}^k \langle \nu, X_k \rangle + h_{ij} \langle \nu, \nu \rangle;$$

$$\frac{\partial}{\partial u^k} \langle X_i, X_j \rangle = \sum_r \Gamma_{ik}^r \langle X_r, X_j \rangle + \sum_s \Gamma_{jk}^s \langle X_i, X_s \rangle + h_{ik} \langle \nu, X_j \rangle + h_{jk} \langle X_i, \nu \rangle.$$

Now one can see that not only the three left-hand sides $\langle \nu, \nu \rangle, \langle \nu, X_i \rangle$, and $\langle X_i, X_j \rangle$, but also the three right-hand sides $1, 0, g_{ij}$ satisfy this system of equations. By the uniqueness of solutions, these two solutions must coincide.

Step 2: As a second step we look for a solution f of the second system, now for given X_i. The integrability conditions

$$\frac{\partial X_i}{\partial u^j} = \frac{\partial X_j}{\partial u^i}$$

are satisfied here because of the symmetries

$$h_{ij} = h_{ji} \quad \text{and} \quad \Gamma^k_{ij} = \Gamma^k_{ji}.$$

Thus, by the existence result mentioned above, we get a unique solution f with the initial value $f(u^{(0)}) = p_0$.

It remains to show that g_{ij} and h_{ij} are in fact the first and second fundamental form of f. The first was already shown above:

$$g_{ij} = \langle X_i, X_j \rangle = \Big\langle \frac{\partial f}{\partial u^i}, \frac{\partial f}{\partial u^j} \Big\rangle.$$

ν is the unit normal of f because of the relation $0 = \big\langle \nu, X_i \big\rangle = \big\langle \nu, \frac{\partial f}{\partial u^i} \big\rangle$. Hence we get

$$\Big\langle \frac{\partial^2 f}{\partial u^i \partial u^j}, \nu \Big\rangle = \Big\langle \sum_k \Gamma^k_{ij} \frac{\partial f}{\partial u^k} + h_{ij}\nu, \nu \Big\rangle = h_{ij}. \qquad \square$$

4.25. Remark. For a two-dimensional surface element the first fundamental form does *not* determine the surface up to Euclidean motions. Here are some examples.

(i) The plane $(x, y) \mapsto (x, y, 0)$ and the circular cylinder $(x, y) \mapsto (\cos x, \sin x, y)$ have locally the same first fundamental form.

(ii) A compact convex surface with a flat part can be modified by adding a “hump”, with reflection symmetry with respect to the normal direction. The first fundamental form does not notice this at all, since a reflection preserves the first fundamental form.

(iii) The helicoid f and the catenoid $\overline{f}$ have the same first fundamental form in appropriately chosen parameters, see 3.37. Consequently, one has $K = \overline{K}$ by the Theorema Egregium, and in addition $H = \overline{H} = 0$, so that we even have $\kappa_1 = \overline{\kappa}_1$ and $\kappa_2 = \overline{\kappa}_2$.

But clearly these surfaces are not equivalent under a Euclidean motion, not even locally, since this would contradict 3.23 (iii).

In all of these cases there are different possibilities of (h_{ij}) for a given (g_{ij}) which satisfy the Gauss and Codazzi-Mainardi equations. If the quantities (g_{ij}) and (h_{ij}) are arbitrarily prescribed, without the equation 4.15 being satisfied, then there is no surface element at all with (g_{ij}) as the first fundamental form and (h_{ij}) as the second fundamental form.

4E The Gaussian curvature in special parameters

For two-dimensional surfaces in $\mathbb{R}^3$, the Gauss equation yields in particular an explicit expression for the Gaussian curvature, depending only on the first fundamental form, that is, on the three quantities E, F, G. This follows from the Theorema Egregium. However, in general this expression can be quite complicated. One gets, for example,

$$K = \frac{1}{4(EG - F^2)^2}(\mathbf{D}_1 - \mathbf{D}_2),$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are the following determinants:

$$\mathbf{D}_1 = \mathrm{Det}\begin{pmatrix} -2E_{vv} + 4F_{uv} - 2G_{uu} & E_u & 2F_u - E_v \\ 2F_v - G_u & E & F \\ G_v & F & G \end{pmatrix};$$

$$\mathbf{D}_2 = \mathrm{Det}\begin{pmatrix} 0 & E_v & G_u \\ E_v & E & F \\ G_u & F & G \end{pmatrix}.$$

For a proof of this, see for example [**4**], 3-3. Without going into this proof in more detail, we will give quite independently of this a much simpler formula in special parameters which holds for arbitrary surfaces. This allows rather easy calculations with the Gaussian curvature using only the first fundamental form.

4.26. Special case. (Orthogonal parameters, lines of curvature parameters[3])
If $f: U \to \mathbb{R}^3$ has no umbilical points, then one can introduce local parameters (u, v) such that $F = g_{12} = h_{12} = M = 0$ (cf. 3.15), hence

$$I = \begin{pmatrix} E & 0 \\ 0 & G \end{pmatrix}, \quad II = \begin{pmatrix} L & 0 \\ 0 & N \end{pmatrix},$$

with the principal curvatures $\kappa_1 = \frac{L}{E}, \kappa_2 = \frac{N}{G}$. The equations of Codazzi-Mainardi (i) and Gauss (ii) hold in the following form:

(i)

$$L_v = \frac{E_v}{2}\Big(\frac{L}{E} + \frac{N}{G}\Big) = E_v \cdot H,$$
$$N_u = \frac{G_u}{2}\Big(\frac{L}{E} + \frac{N}{G}\Big) = G_u \cdot H.$$

(ii)

$$K = -\frac{1}{2\sqrt{EG}}\bigg(\Big(\frac{E_v}{\sqrt{EG}}\Big)_v + \Big(\frac{G_u}{\sqrt{EG}}\Big)_u\bigg).$$

This equation for K holds already purely intrinsically in orthogonal parameters, i.e., if $I = \left(\begin{smallmatrix} E & 0 \\ 0 & G \end{smallmatrix}\right)$. In isothermal parameters with $E = G = \lambda$ one has

$$K = -\frac{1}{2\lambda}\bigg(\Big(\frac{\lambda_v}{\lambda}\Big)_v + \Big(\frac{\lambda_u}{\lambda}\Big)_u\bigg) = -\frac{1}{2\lambda}\Delta(\log\lambda).$$

PROOF: The proof is carried out by specializing the formulas of 4.15 and 4.16. First, we calculate the Christoffel symbols, for example $\Gamma_{11,1} = \frac{1}{2}E_u$ and $\Gamma_{11}^1 = \frac{1}{2}E_u/E$, and similarly for the other indices. Then, applying the Codazzi-Mainardi equation 4.15 for $i = j = 1, k = 2$ on the one hand and for $i = j = 2, k = 1$ on the other, we get

$$0 = L_v - 0 + \Gamma_{11}^2 h_{22} - \Gamma_{12}^1 h_{11} = L_v - \frac{1}{2}\frac{E_v}{G}N - \frac{1}{2}\frac{E_v}{E}L,$$

$$0 = N_u - 0 + \Gamma_{22}^1 h_{11} - \Gamma_{12}^2 h_{22} = N_u - \frac{1}{2}\frac{G_u}{E}L - \frac{1}{2}\frac{G_u}{G}N.$$

[3]This means these parameters are such that the parameter lines are lines of curvature, cf. 3.15.

For the Gauss equation we do what was done in 4.16 and calculate

$$\mathrm{Det}(II) = \sum_s \Big((\Gamma^s_{11})_v - (\Gamma^s_{12})_u + \sum_r \big(\Gamma^r_{11}\Gamma^s_{r2} - \Gamma^r_{12}\Gamma^s_{r1}\big)\Big) g_{s2}$$

$$= \Big(-\frac{1}{2}\Big(\frac{E_v}{G}\Big)_v - \frac{1}{2}\Big(\frac{G_u}{G}\Big)_u + \frac{1}{2}\frac{E_u}{E}\frac{1}{2}\frac{G_u}{G} - \frac{1}{2}\frac{E_v}{G}\frac{1}{2}\frac{G_v}{G}$$
$$-\frac{1}{2}\frac{E_v}{E}\Big(-\frac{1}{2}\frac{E_v}{G}\Big) - \frac{1}{2}\frac{G_u}{G}\frac{1}{2}\frac{G_u}{G}\Big) G.$$

It then follows that

$$K = \frac{\mathrm{Det}(II)}{\mathrm{Det}(I)}$$

$$= -\frac{1}{2EG}\Big(E_{vv} + G_{uu} - \frac{E_v G_v}{G} - \frac{G_u^2}{G} - \frac{E_u G_u}{2E} + \frac{E_v G_v}{2G} - \frac{E_v^2}{2E} + \frac{G_u^2}{2G}\Big)$$

$$= -\frac{1}{2EG\sqrt{EG}}\Big(\sqrt{EG}\cdot E_{vv} - E_v\big(\sqrt{EG}\big)_v + \sqrt{EG}\cdot G_{uu} - G_u\big(\sqrt{EG}\big)_u\Big)$$

$$= -\frac{1}{2\sqrt{EG}}\Big(\Big(\frac{E_v}{\sqrt{EG}}\Big)_v + \Big(\frac{G_u}{\sqrt{EG}}\Big)_u\Big). \qquad \square$$

4.27. Definition. (Geodesic parallel coordinates)
The coordinates of a surface element $f\colon U \to \mathbb{R}^3$ are called *geodesic parallel coordinates*, if the u-curves (i.e., the curves defined by $v =$ constant) are geodesics parametrized by arc length, which intersect each of the v-curves orthogonally. By construction, two v-curves cut segments of equal length in the u-curves.

Remark: Such geodesic parallel coordinates occur if and only if the first fundamental form can be written as

$$I = \begin{pmatrix} 1 & 0 \\ 0 & G \end{pmatrix}$$

with a positive function $G = G(u, v)$. The necessity of this form of the first fundamental form is clear by definition. Conversely, this form shows that the u-curves are also parametrized by arc length and are always perpendicular to the v-curves. The equation for the geodesics, $\nabla_{f_u} f_u = \Gamma_{11}^1 f_u + \Gamma_{11}^2 f_v = 0$, can be verified by calculating the Christoffel symbols Γ_{11}^1 and Γ_{11}^2. For the calculation, those Christoffel symbols which contain derivatives of G are irrelevant.

Locally geodesic parallel coordinates always exist on surface elements. They can be constructed from a given fixed curve $u = u^{(0)}$, where $u^{(0)}$ is a constant, by constructing all the geodesics which are orthogonal to it. But one still must show that these are in fact coordinates, a verification of which can be found for example in [**4**, Sec.4-3].

In the particular case in which the starting curve $u = u^{(0)}$ is itself a geodesic which is parametrized by arc length, one speaks of *Fermi coordinates.*

Such coordinates are often used in geodesy. In this case one has

$$G(u^{(0)}, v) = 1, \quad \frac{\partial}{\partial u} G(u^{(0)}, v) = 0, \quad \Gamma_{ij}^k(u^{(0)}, v) = 0$$

for all v and all i, j, k.

4.28. Corollary. In geodesic parallel coordinates $\left(\begin{smallmatrix} 1 & 0 \\ 0 & G \end{smallmatrix}\right)$, the following simple equation for the Gaussian curvature holds:

$$K(u, v) = -\frac{(\sqrt{G})_{uu}}{\sqrt{G}}.$$

Alternatively, we have for $I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & G^2 \end{smallmatrix}\right)$ the expression $K = -G_{uu}/G$.

This is a special case of 4.26 (ii) with $E = 1$. Note that the equation $(\sqrt{G})_{uu} = \frac{1}{2}\left(\frac{G_u}{\sqrt{G}}\right)_u$ holds.

Example: A surface of rotation $f(u, \varphi) = (r(u)\cos\varphi, r(u)\sin\varphi, h(u))$ with $r'^2 + h'^2 = 1$ is always parametrized by geodesic parallel coordinates because of the relation

$$I = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix},$$

cf. 3.16. These are Fermi coordinates in a neighborhood of a circle with $r' = 0$, cf. 4.12. The formula $K = -r_{uu}/r$ in 4.28 specializes to the equation $K = -r''/r$, which is already familiar to us from 3.16. Compare this with the surfaces of rotation of constant Gaussian curvature K_0 as the solution of the differential equation $r'' + K_0 r = 0$ in 3.17. Here it is not obvious that an elongated sphere and an oblate sphere have the same first fundamental form as the standard sphere. However, in Fermi coordinates $f(t, \varphi) = \left(a \cos t \cos \frac{\varphi}{a}, a \cos t \sin \frac{\varphi}{a}, \int_0^t \sqrt{1 - a^2 \sin^2 x}\, dx\right)$ around the geodesic $t = 0$ this is indeed the case: $I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \cos^2 t \end{smallmatrix}\right)$ is independent of a. As a further application of 4.28, we prove Theorem 4.30 below, which is more generally concerned with the isometry problem for pairs of surfaces.

4.29. Definition. (Isometric)

Two surface elements $f, \widetilde{f}$ are said to be *isometric* (to one another), if in appropriately chosen coordinates they have the same first fundamental form, i.e., if for given $f\colon U \to \mathbb{R}^{n+1}$, $\widetilde{f}\colon \widetilde{U} \to \mathbb{R}^{n+1}$ there is a parameter transformation $\Phi\colon U \to \widetilde{U}$ such that

$$\left\langle \frac{\partial f}{\partial u^i}, \frac{\partial f}{\partial u^j} \right\rangle = \left\langle \frac{\partial(\widetilde{f} \circ \Phi)}{\partial u^i}, \frac{\partial(\widetilde{f} \circ \Phi)}{\partial u^j} \right\rangle$$

for all i, j. Two isometric surface elements can thus be mapped in a length preserving manner, i.e., the mapping $\widetilde{f} \circ \Phi \circ f^{-1}\colon f(U) \to \widetilde{f}(\widetilde{U})$ is length preserving. Compare this with Definition 3.29.

4.30. Theorem. ("Surfaces with the same constant Gaussian curvature are isometric")
Let $f\colon U \to \mathbb{R}^3, \widetilde{f}\colon \widetilde{U} \to \mathbb{R}^3$ be surface elements with the same constant Gaussian curvature. Then locally f and $\widetilde{f}$ are isometric.

Proof: Let K be the constant Gaussian curvature. We fix two points in $U, \widetilde{U}$ and introduce in appropriately chosen neighborhoods geodesic parallel coordinates, starting with a given geodesic $u = 0$ (i.e., Fermi coordinates). We denote the parameter for both surface elements by the same symbol (u, v). The first fundamental forms of the surfaces, $I, \widetilde{I}$ then have by 4.27 the form

$$I = \begin{pmatrix} 1 & 0 \\ 0 & G(u,v) \end{pmatrix}, \quad \widetilde{I} = \begin{pmatrix} 1 & 0 \\ 0 & \widetilde{G}(u,v) \end{pmatrix}$$

with $G(0,v) = 1 = \widetilde{G}(0,v)$ for every v. The quantities G and $\widetilde{G}$ are uniquely determined by the differential equations

$$\frac{\partial^2}{\partial u^2}\sqrt{G} = -K\sqrt{G}, \quad \frac{\partial^2}{\partial u^2}\sqrt{\widetilde{G}} = -K\sqrt{\widetilde{G}},$$

which hold by 4.28. The uniqueness follows once we are given initial conditions

$$\frac{\partial}{\partial u}\sqrt{G(u,v)}\big|_{u=0} = 0 = \frac{\partial}{\partial u}\sqrt{\widetilde{G}(u,v)}\big|_{u=0}.$$

Note that for fixed but arbitrary v, this is an ordinary differential equation of second order in the parameter u, and the solution is independent of v. Putting everything together, we have in these parameters $G = \widetilde{G}$, hence $I = \widetilde{I}$ and thus the (local) isometry of f and $\widetilde{f}$. □

WARNING: In general 4.30 no longer holds without the assumption that the curvature is constant, i.e., there are non-isometric surfaces with the same Gaussian curvature in certain parameters. However, this can change when one passes to Fermi coordinates or other parameters. Therefore, the differential equation which we utilized above can no longer be used. See Exercise 7 for an example of this kind.

RETROSPECTIVE: Theorem 4.30 shows in particular that locally there is only one first fundamental form for which $K \equiv 0$, namely the Euclidean one. This puts the results of 3.24 on developable surfaces in a new perspective. At the same time, 4.30 leads to a simpler proof for the relation (1) $\Leftrightarrow$ (2) in 3.24, without any details on the geometry of ruled surfaces. Note, however, that a straight line on a surface is always a geodesic. Hence it must remain a geodesic after developing a ruled surface isometrically into the plane. Compare the definition of "developable" in Section 3.24. Moreover, it is clear from 4.30 that all surfaces of rotation with the same constant curvature in 3.17 are locally isometric to one another. In particular any of these surfaces with $K = -1$ is locally isometric to the hyperbolic plane defined in 3.44. However, by a theorem of Hilbert no surface in Euclidean 3-space can be globally isometric with the hyperbolic plane.

For two-dimensional surfaces there are many examples of pairs of surfaces for which the first fundamental forms coincide, but the second do not. What is the analogous statement in dimensions $n \geq 3$? The following theorem gives a surprising answer that in higher dimensions nothing of the kind happens, at least not if the rank of the Weingarten mapping is at least three. The possibility of different second fundamental forms for a fixed first fundamental form is hence something particular to dimension two.

4.31. Theorem. ("The first fundamental form determines the second")
Let $f, \widetilde{f}\colon U \to I\!R^{n+1}$ be two hypersurface elements with the same first fundamental form $(g_{ij}) = (\widetilde{g}_{ij})$. Suppose that the rank of the Weingarten mapping L is at least three at a point $p = f(u)$. Then one has for the second fundamental form at this point

$$(h_{ij}(u)) = \pm(\widetilde{h}_{ij}(u)),$$

where the sign corresponds to a choice of orientation.

PROOF: Up to a Euclidean motion we may assume we are in the following situation: $p = f(u) = \widetilde{f}(u), T_u f = T_u \widetilde{f}$. Let $R, \widetilde{R}, L, \widetilde{L}$ be the curvature tensors and the Weingarten mapping of f. The Gauss equation then tells us that

$$\begin{aligned}\langle LY, Z\rangle LX - \langle LX, Z\rangle LY &= R(X,Y)Z \\ = \widetilde{R}(X,Y)Z &= \langle \widetilde{L}Y, Z\rangle \widetilde{L}X - \langle \widetilde{L}X, Z\rangle \widetilde{L}Y\end{aligned}$$

for all tangent vectors $X, Y, Z \in T_u f$. By assumption there are X, Y such that the resulting vectors LX, LY are linearly independent. Since there is also a Z which is linearly independent of LX and LY, it follows from the Gauss equation that then also $\widetilde{L}X, \widetilde{L}Y$ are linearly independent. Thus we always have $\widetilde{L} \not\equiv 0$, and by repeated application of this argument it can be seen that the rank of $\widetilde{L}$ is the same as the rank of L. We then choose X after the fact in such a way that $\widetilde{L}X \neq 0$. The Y which belongs to this is initially not fixed.

We now wish to show that $LX, \widetilde{L}X$ are linearly dependent, which we do by contradiction. Thus, we make the *assumption: Assume that $LX, \widetilde{L}X$ are linearly independent.* Because $\mathrm{Rank}(L) \geq 3$, there exists a Y such that the three vectors $LX, \widetilde{L}X, LY$ are linearly independent. Without restriction of generality, we may assume that $\langle \widetilde{L}X, LY\rangle = 0$. Applying the Gauss equation with $Z = LY$, we get

$$\langle LY, LY\rangle LX - \langle LX, LY\rangle LY = \langle \widetilde{L}Y, LY\rangle \widetilde{L}X - \langle \widetilde{L}X, LY\rangle \widetilde{L}Y.$$

Since the last coefficient vanishes, either $LX, \widetilde{L}X$ are linearly dependent (if $\langle LX, LY\rangle = 0$), or the three vectors $LX, \widetilde{L}X, LY$ are linearly dependent. But both conclusions contradict our assumption, proving the claim.

This conclusion can be made for every X for which $LX \neq 0$, with the result that $\widetilde{L}X = c_X LX$ holds for every X and appropriately chosen $c_X \in \mathbb{R}$, which may depend on X. If $LX = 0$, we set $c_X = 0$. In a basis of eigenvectors of L we then have $LX_i = \lambda_i X_i$ and $\widetilde{L}X_i = c_i \lambda_i X_i$, which implies that the latter is also a basis consisting of eigenvectors for L. But since $c_i\lambda_i X_i + c_j \lambda_j X_j = \widetilde{L}(X_i + X_j) = c_{ij} L(X_i + X_j) = c_{ij}(\lambda_i X_i + \lambda_j X_j)$, we get $c_i = c_{ij} = c_j$ for all i, j with $\lambda_i, \lambda_j \neq 0$. By the remark we made above, we have on the other hand the equality $\mathrm{Ker}(\widetilde{L}) = \mathrm{Ker}(L)$. Since at least one of the eigenvalues must be non-vanishing, it follows that $\widetilde{L}X = cLX$ for every X with a constant $c \neq 0$ which is independent of X. Note that in the case that $LX = 0$, the equation is trivial for every c. Applying the Gauss equation once again, we get $c^2 = 1$, hence $\widetilde{L} = \pm L$. All of the preceding considerations were made at a point p. $\square$

If the mentioned condition on the rank of L is satisfied at a point, then it is also satisfied in a whole neighborhood of the point. Thus, the uniqueness result 4.23 yields the following statement.

4.32. Corollary. If U is connected and if for two functions, f and $\widetilde{f}\colon U \to \mathbb{R}^{n+1}$ defined on U, the condition $(g_{ij}) = (\widetilde{g}_{ij})$ holds everywhere, and if in addition we have everywhere $\mathrm{Rank}(L) \geq 3$, then $f(U)$ and $\widetilde{f}(U)$ coincide everywhere up to a Euclidean motion (including reflections).

In other words: *under the assumption that Rank*$(L) \geq 3$*, the first fundamental form alone completely determines the geometry of the hypersurface.*

REMARK: If for the rank of L we have $\mathrm{Rank}(L) \leq 2$, then the analogous statement is no longer true, as can be shown in simple examples. If $\mathrm{Rank}(L) = \mathrm{Rank}(\widetilde{L}) = 1$, take two cylinders over distinct plane curves $c, \widetilde{c}$: take for example $f(t,x) = (c(t), x)$, $\widetilde{f}(t,x) = (\widetilde{c}(t), x), t \in I \subset \mathbb{R}, x \in \mathbb{R}^{n-1}$. For examples with $\mathrm{Rank}(L) = 2$, take similar cylinders over two isometric surfaces like the helicoid and the catenoid.

4F The Gauss-Bonnet Theorem

The Gauss-Bonnet theorem is one of the most important theorems in all of differential geometry. It expresses what appears at first sight to be a surprising invariance of the integrated Gaussian curvature (or its mean value). This can be visualized as follows. Consider a two-dimensional surface element with a boundary curve and make a change on this surface, for example like the ones we met in the consideration of minimal surfaces in 3.28. We now require that this change preserves not only the boundary curve, but also a neighborhood strip of the boundary. In other words, only changes are allowed which vanish near the boundary curve. Then the integrated Gaussian curvature is also unchanged! The change accounts for just as much additional positive as negative curvature. In particular, there is no non-trivial condition for the corresponding variational problem as in 3.28. The functional in question, which associates to a surface the total (integrated) Gaussian curvature, is just a constant. This automatically then holds for compact surfaces without boundary (compact two-dimensional submanifolds). Since by the Theorema Egregium 4.16 the Gaussian curvature is intrinsically defined, the value of the curvature integral here, which does not depend on the embedding but only on the surface itself, yields a topological invariant, the so-called *Euler characteristic*.

In order to prove this theorem (locally as well as globally), we reduce it to the theorem on turning tangents 2.28 and the theorem of Stokes, which will be restated in 4.36. The theorem of Stokes allows an elegant formulation using differential forms, the calculus of which goes back to É. Cartan. That is why we consider differential forms here and formulate the main results of the local theory of surfaces in this language. This is also useful in itself, as the calculus of differential forms is an important tool in mathematics. For background we refer to [**27**], Chapter XXI.

In what follows let V^* denote the *dual space* of V, if V is an $\mathbb{R}$-vector space. More precisely, we have the description $V^* = \{\omega\colon V \to \mathbb{R} \mid \omega \text{ is an } \mathbb{R}\text{-linear map}\}$. The positions of the indices which occur in this section are based on the usual practice of writing the indices corresponding to an ON-basis for the space V as subscripts, and those for the dual space as superscripts.

4.33. Definition. (Differential forms, exterior derivative)

A *Pfaffian form* (or *one-form*) in $I\!R^{n+1}$ (resp. on a hypersurface element) is given by associating to each point a linear form on the tangent space,

$$p \longmapsto \omega_p \in (T_p I\!R^{n+1})^*$$
$$\big[\text{ resp. } u \longmapsto \omega_u \in (T_u f)^*\big].$$

ω is said to be *continuous* or *continuously differentiable*, if the coefficients with respect to the standard basis $e_1, \ldots, e_{n+1}$ (resp. with respect to $\frac{\partial f}{\partial u^1}, \ldots, \frac{\partial f}{\partial u^n}$) have the corresponding property, i.e., if all $\omega(e_i)$ $\big[$resp. $\omega(\frac{\partial f}{\partial u^i})\big]$ are continuous or continuously differentiable functions. Here and in the sequel we use the notation $\omega(X)$ instead of the more formal $\omega_p((p, X))$ for each p.

For an orthonormal basis $X_1, \ldots, X_{n+1}$ of $I\!R^{n+1}$ let $\omega^1, \ldots, \omega^{n+1}$ be the corresponding *dual basis*, i.e., such that

$$\omega^i(X_j) = \delta^i_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

In particular, the dual basis of the standard basis $e_1, \ldots, e_{n+1}$ in $I\!R^{n+1}$ is denoted by $dx^1, \ldots, dx^{n+1}$.

To express the covariant derivative in terms of differential forms, we begin with the directional derivative in the ambient space and define one-forms ω^i_j by the equations $\omega^i_j(Y) = \omega^i(D_Y X_j)$ for every Y. We then get

$$D_Y X_j = \sum_i \omega^i_j(Y) X_i.$$

These ω^i_j satisfiy the equation $\omega^i_j = -\omega^j_i$, because $\omega^i_j(Y) + \omega^j_i(Y) = D_Y \langle X_i, X_j \rangle = 0$. In what follows the vectors $X_1, \ldots, X_n$ will be taken to be *tangential* and X_{n+1} taken to be *normal* to a hypersurface element. Then X_{n+1} is nothing but the familiar unit normal ν, and one has

$$\omega^i_j(Y) = \omega^i(D_Y X_j) = \omega^i(\nabla_Y X_j) \quad \text{for } i, j \leq n$$

and for tangential Y. One also calls the ω^i_j *connection forms*, since they determine the covariant derivative uniquely, and the covariant derivative is also referred to as a *connection* (cf. 5.15). The connection forms play here the same role as the Christoffel symbols. They determine and are determined by one another.

With the Gauss-Bonnet formula in the back of our minds, we consider in particular the case $n = 2$, in which we require only one-forms and two-forms, but not k-forms for $k \geq 3$. The two-forms occur as the derivatives of one-forms, more precisely by the skew-symmetrization of this derivative, as follows.

For a one-form $\omega = \sum_i \omega(X_i)\omega^i$, the *exterior derivative* $d\omega$ is defined by

$$d\omega(X, Y) = D_X(\omega(Y)) - D_Y(\omega(X)) - \omega([X, Y]).$$

This derivative is skew-symmetric: $d\omega(X, Y) = -d\omega(Y, X)$, and the value of $d\omega(X, Y)$ at a point p depends only on the values of X and Y at the point p (exercise). Thus, $d\omega$ is pointwise a skew-symmetric form on the tangent space. In the two-dimensional case, it follows that $d\omega$ is a scalar multiple of the surface element (with a function as multiplier), and in general such a two-form is a linear combination of the $\omega^i \wedge \omega^j, i < j$. For the "wedge product" $\wedge$ one sets $\omega^i \wedge \omega^j = -\omega^j \wedge \omega^i$ and views this as a bilinear operation, which associates to two given one-forms a two-form (and similarly for k-forms with higher k).

Examples:

1. Every vector field X uniquely induces a one-form by means of the equation $\omega(Y) = \langle X, Y \rangle$.
2. The *curve integral* of a one-form along a curve $c\colon [a, b] \to I\!R^n$ is defined as
$$\int_c \omega = \int_a^b \omega(\dot{c}(t))dt.$$
3. One has
$$d(dx^i)(X, Y) = D_X(dx^i(Y)) - D_Y(dx^i(X)) - dx^i([X, Y])$$
$$= X(Y^i) - Y(X^i) - [X, Y]^i,$$
where the upper index i on vectors simply denotes the ith component. The last expression vanishes if we use the formula of 4.5 for the Lie brackets. Hence $d(dx^i) = 0$.
4. In the same way, the equation $d(\alpha \cdot dx^i) = d\alpha \wedge dx^i$ is verified for every scalar function α.

5. The *differential* df of a differentiable scalar function f on $\mathbb{R}^{n+1}$ is the one-form
$$df = \sum_i \frac{\partial f}{\partial x^i} dx^i.$$
It is easy to see that the condition $d(df) = 0$ is equivalent to the symmetry of the second partial derivatives of f:
$$\begin{aligned} d(df) &= \sum_i d\Big(\frac{\partial f}{\partial x^i}\Big) \wedge dx^i = \sum_{i,j} \frac{\partial^2 f}{\partial x^j \partial x^i} dx^j \wedge dx^i \\ &= \sum_{i<j} \Big(\frac{\partial^2 f}{\partial x^i \partial x^j} - \frac{\partial^2 f}{\partial x^j \partial x^i}\Big) dx^i \wedge dx^j. \end{aligned}$$
6. More generaly, the condition $d\omega = 0$ is necessary for it to be possible to write ω as a differential form of a function (integrability condition): $\omega = df$.

The equations of Gauss and Weingarten of the theory of surfaces from 4.8 correspond here to the decomposition into ω_j^i for $i, j \le n$ on the one hand and into ω_{n+1}^i on the other. Note that $X_{n+1} = \nu$ (unit normal):
$$D_Y X_j = \sum_{i=1}^{n+1} \omega_j^i(Y) X_i, \qquad \nabla_Y X_j = \sum_{i=1}^{n} \omega_j^i(Y) X_i,$$
$$\begin{aligned} \omega_j^{n+1}(Y) = \langle D_Y X_j, X_{n+1}\rangle &= -\langle X_j, D_Y X_{n+1}\rangle \\ &= \langle X_j, LY\rangle = II(X_j, Y). \end{aligned}$$

4.34. Theorem. (Maurer-Cartan structural equations)
The following equations express the integrability conditions of the derivatives in the theory of surfaces, in which the first equation is similar to the Gauss equation and the second is similar to the Codazzi-Mainardi equation:
$$\text{(i)} \quad d\omega_j^i + \sum_{k=1}^{n+1} \omega_k^i \wedge \omega_j^k = 0 \quad \text{for } i, j = 1, \ldots, n.$$
$$\text{(ii)} \quad d\omega_{n+1}^i + \sum_{k=1}^{n} \omega_k^i \wedge \omega_{n+1}^k = 0 \quad \text{for } i = 1, \ldots, n.$$

Proof: The proof is, like that of 4.18, based on the decomposition of higher derivatives into tangential and normal parts, where for each occurrence we have to replace the covariant derivative by the connection forms

$$D_Y X_j = \sum_k \omega_j^k(Y) X_k.$$

Then a straightforward calculation leads to

$$\begin{aligned}
0 &= \big\langle X_i, D_X D_Y X_j - D_Y D_X X_j - D_{[X,Y]} X_j \big\rangle \\
&= \omega^i \Big(D_X \big(\sum_k \omega_j^k(Y) X_k \big) - D_Y \big(\sum_k \omega_j^k(X) X_k \big) \\
&\qquad\qquad - \sum_k \omega_j^k([X,Y]) X_k \Big) \\
&= \sum_k \omega_j^k(Y) \omega_k^i(X) - \sum_k \omega_j^k(X) \omega_k^i(Y) \\
&\quad + \sum_k D_X(\omega_j^k(Y)) \omega^i(X_k) - \sum_k D_Y(\omega_j^k(X)) \omega^i(X_k) \\
&\quad - \sum_k \omega_j^k([X,Y]) \omega^i(X_k) \\
&= \sum_k \omega_k^i \wedge \omega_j^k(X,Y) + D_X(\omega_j^i(Y)) - D_Y(\omega_j^i(X)) - \omega_j^i([X,Y]) \\
&= \Big(\sum_k \omega_k^i \wedge \omega_j^k + d\omega_j^i \Big)(X,Y).
\end{aligned}$$

(i) thus corresponds to the case $i, j \leq n$, and (ii) corresponds to the case $i \leq n, j = n+1$. Note here that $\omega_{n+1}^{n+1} = 0$ because of the skew-symmetry $\omega_j^i = -\omega_i^j$. □

In 4.19 and 4.20 we have seen that the curvature of a surface is already completely determined by the curvature tensor $R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$. Since this expression is clearly skew-symmetric in X and Y, one can naturally define two-forms, the so-called curvature forms. These contain the same information as the curvature tensor.

4.35. Definition and Theorem. (Curvature forms)
The *curvature forms* Ω^i_j are defined by the relation $\Omega^i_j(X,Y) = \langle R(X,Y)X_j, X_i\rangle$. Then one has the equation

$$\Omega^i_j = d\omega^i_j + \sum_{k=1}^{n} \omega^i_k \wedge \omega^k_j.$$

In connection with 4.34 (i) this corresponds to the Gauss equation.

For two-dimensional surface elements one has by 4.20

$$\Omega^1_2(X_1, X_2) = \langle R(X_1, X_2)X_2, X_1\rangle = \mathrm{Det}(L) = K$$

and consequently the elegant relation

$$K \cdot \omega^1 \wedge \omega^2 = \Omega^1_2 = d\omega^1_2 + \sum_{k=1}^{2} \omega^1_k \wedge \omega^k_2 = d\omega^1_2.$$

PROOF: We simply rewrite the Gauss equation in the following way:

$$\big\langle R(X,Y)X_j, X_i\big\rangle = \langle LY, X_j\rangle\langle LX, X_i\rangle - \langle LX, X_j\rangle\langle LY, X_i\rangle$$
$$= \omega^{n+1}_j(Y)\,\omega^{n+1}_i(X) - \omega^{n+1}_j(X)\,\omega^{n+1}_i(Y) = -\omega^i_{n+1} \wedge \omega^{n+1}_j(X,Y).$$

From this it follows, using 4.34 (i), that

$$\Omega^i_j = -\omega^i_{n+1} \wedge \omega^{n+1}_j = \sum_{k=1}^{n} \omega^i_k \wedge \omega^k_j - \sum_{k=1}^{n+1} \omega^i_k \wedge \omega^k_j = \sum_{k=1}^{n} \omega^i_k \wedge \omega^k_j + d\omega^i_j.$$

□

4.36. Theorem of Stokes. Let B be a compact set with smooth boundary ∂B in $\mathbb{R}^k$, and let ω be a differentiable $(k-1)$-form, which is defined in a neighborhood of B. Then the following *Stokes integral theorem* holds:

$$\int_B d\omega = \int_{\partial B} \omega.$$

∂B represents the oriented boundary of B. In addition to this, the same relation holds for the image of B under an immersion f. In the special case $k = 2$, the integral on the left-hand side is an ordinary surface integral (in the sense of 3.4), while the integral on the right-hand side is a contour integral along the (oriented) boundary curve.

This general theorem contains several well-known integral theorems as special cases, for example the Gauss integral theorem in the plane. A proof and further information can be found in [**27**], Chapter XIX, and [**28**], Chapter XIV.

4.37. Definition. (Geodesic curvature)
To motivate the notion of curvature on surfaces in 3.11 and 3.12 we have referred to the normal component κ_ν of the curvature κ of a curve as the *normal curvature*. The tangential component of this curvature is an intrinsic quantity which is called the *geodesic curvature* κ_g of the curve. Setting $\kappa = ||D_{e_1}e_1|| = ||e_1'||$ and $\kappa_\nu = \langle D_{e_1}e_1, \nu\rangle = \langle e_1', \nu\rangle = II(e_1, e_1)$, one necessarily has

$$\kappa_g = \langle \nabla_{e_1}e_1, e_2\rangle = \langle e_1', e_2\rangle,$$

where e_1 is the unit tangent vector of the curve and e_2 is the (oriented and normed) normal vector to the curve on the surface (i.e., e_1, e_2 are a ON-basis of the tangent plane). Our definitions are such that $\kappa^2 = \kappa_g^2 + \kappa_\nu^2$, cf. 3.11. The geodesic curvature is an important quantity in the Gauss-Bonnet formula 4.38.

In particular, equations for the derivatives of the frame

$$\nabla_{e_1}e_1 = \kappa_g e_2, \quad \nabla_{e_1}e_2 = -\kappa_g e_1$$

hold then in analogy to the Frenet equations $e_1' = \kappa e_2$, $e_2' = -\kappa e_1$ for the Frenet frame of plane curves (2.5). Note that $\langle \nabla_{e_1}e_1, e_1\rangle = 0$, which is similar to the relation $\langle e_1', e_1\rangle = 0$ in the Frenet theory we met earlier. Once the orientation has been fixed, the sign of κ_g shows for plane curves whether the tangent of the curve moves to the right or to the left when one runs through the curve, see Figure 4.1 or Figure 4.10. Geodesics are characterized by the relation $\kappa_g = 0$, just as the lines are characterized among all plane curves by the relation $\kappa = 0$, cf. 4.9:

$$\nabla_{c'}c' = 0 \iff \nabla_{e_1}e_1 = 0 \iff \langle \nabla_{e_1}e_1, e_2\rangle = 0 \iff \kappa_g = 0.$$

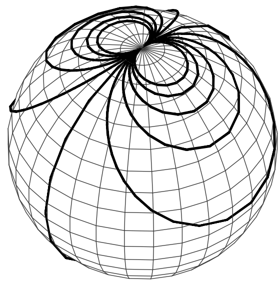


Figure 4.1. Curves of constant geodesic curvature in the 2-sphere

4.38. Theorem. (Gauss-Bonnet formula, first local version)
Let $U \subset \mathbb{R}^2$ be an open subset, and let $B \subset U$ be diffeomorphic to a closed disc (in the terminology of [**28**], Chapter XIV: a region B which is the interior of a closed C^2-path γ which is parametrized counterclockwise). Let $f: U \to \mathbb{R}^3$ be a surface element such that f is injective. We assume that the boundary of B is parametrized by $\gamma: I \to U$ in such a way that the interior of B is to the left of γ, and we set $c = f \circ \gamma$. Then

$$\int_{f(B)} K dA + \int_c \kappa_g ds = 2\pi,$$

where K denotes the Gaussian curvature of f and κ_g is the geodesic curvature of c. This equation belongs to the intrinsic geometry.

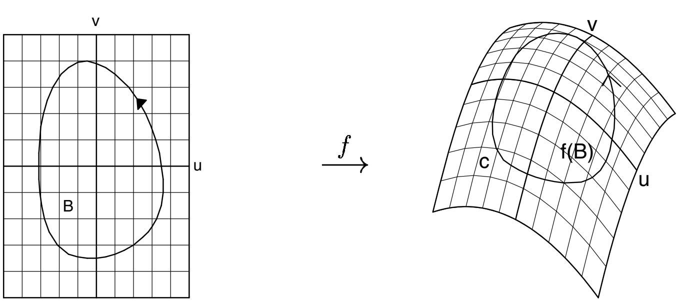


Figure 4.2. Gauss-Bonnet formula

As far as terminology is concerned, the statement that B has a smooth boundary means that every point of B has a neighborhood which is either diffeomorphic to an open disc $D^2 = \{(x, y) \in I\!R^2 \mid x^2+y^2 < 1\}$ or diffeomorphic to a half-disc $D^2_+ = B^2 \cap \{y \geq 0\}$, in which the linear boundary belongs to the set. The points in the former case are called *interior points*, and those of the latter case are called *boundary points* of B. By assumption the boundary of B is then a simply closed curve in the sense of Section 2F. In particular, the theorem on turning tangents 2.28 is applicable. The injectivity of f is only necessary for the integral, guaranteeing that no part of the surface is counted twice, cf. Definition 3.4. From the standpoint of the intrinsic geometry, this can be ignored, provided the integral of K is interpreted as the integral $\int_B K\sqrt{g}\ du \wedge dv$ over the parameter domain, cf. 3.6.

EXAMPLES:

1. A *Euclidean disc* of radius r in $I\!R^2 \subset I\!R^3$. Here one has $K = 0$ and $\kappa_g = \frac{1}{r}$, and this implies $\int KdA + \int \kappa_g ds = 2\pi$.
2. The *upper hemisphere* $\{(x, y, z) \in I\!R^3 \mid x^2 + y^2 + z^2 = 1, z \geq 0\}$. Here one has $K = 1$ and $\kappa_g = 0$ (since the equator is a geodesic) and hence $\int KdA + \int \kappa_g ds = 2\pi$.
3. An intrinsic example: A *hyperbolic disc* of radius r with the abstract arc element $ds^2 = dr^2 + \sinh^2 r\ d\varphi^2$ has Gaussian curvature $K = -1$ and $\int KdA = 2\pi(1 - \cosh r)$. The boundary c is a circle of radius r with $c' = \frac{1}{\sinh r}\frac{\partial}{\partial\varphi}$. Its arc length $s(\varphi)$ satisfies $\frac{ds}{d\varphi} = \sinh r$, and its geodesic curvature is $\kappa_g = \frac{\cosh r}{\sinh r}$. This follows from $\nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi} = \Gamma^1_{22}\frac{\partial}{\partial r} = -\cosh r \sinh r \frac{\partial}{\partial r}$ leading to $\nabla_{c'}c' = -\frac{\cosh r}{\sinh r}\frac{\partial}{\partial r}$. Hence one has $\int \kappa_g ds = \int \kappa_g \sinh r d\varphi = 2\pi \cosh r$ and $\int KdA + \int \kappa_g ds = 2\pi$.

PROOF OF 4.38: We think of $c\colon [a, b] \to I\!R^3$ as being parametrized by arc length with $e_1 = c'$ and an accompanying oriented two-frame e_1, e_2 in the tangent plane of the surface (i.e., e_2 is a unit vector which is perpendicular to the curve, that is in a sense the unit normal *in* the surface). Moreover, we choose orthonormal vector fields X_1, X_2, ν along f such that $X_1 = \frac{\partial f}{\partial u^1}/||\frac{\partial f}{\partial u^1}||$, requiring in addition that ν is a unit normal to the surface. The important thing is that X_1, X_2 and e_1, e_2 have the same orientation. Then we can (as in 2.23 and 2.24) introduce a polar angle φ by

$$e_1 = \cos\varphi X_1 + \sin\varphi X_2,\ e_2 = -\sin\varphi X_1 + \cos\varphi X_2.$$

More precisely, $\varphi\colon [a,b] \to \mathbb{R}$ can be defined as a continuous polar angle function. By the theorem on turning tangents 2.28 we then have $\varphi(b) - \varphi(a) = 2\pi$ (note that B is to the left of the boundary). The theorem on turning tangents holds literally just for the curve γ in U and the polar angle γ' makes with the u^1-axis. The measurement of the angle in U can be continuously deformed into the measurement of the angle in $f(U)$ by the one-parameter family $(1-t)\delta_{ij} + tg_{ij}$ of "inner products". In the process, $\varphi(b) - \varphi(a)$ remains integral, so that in this way the theorem on turning tangents is also valid for the curve c in $f(U)$. However, the equation $\varphi' = \kappa$ from 2.23 does not remain valid in the form $\varphi' = \kappa_g$. Instead, φ' leads to κ_g plus a second term as follows: From $\langle e_1, X_1\rangle = \cos\varphi$ we conclude that $\frac{d}{ds}\langle e_1, X_1\rangle = \frac{d\varphi}{ds}(-\sin\varphi)$. The theorem on turning tangents then implies

$$
\begin{aligned}
2\pi &= \int_a^b \frac{d\varphi}{ds}ds = -\int_a^b \frac{1}{\sin\varphi}\frac{d}{ds}\Big\langle e_1, X_1\Big\rangle ds \\
&= -\int_a^b \frac{1}{\sin\varphi}\Big(\langle \nabla_{e_1} e_1, X_1\rangle + \langle e_1, \nabla_{e_1} X_1\rangle\Big)ds \\
&= -\int_a^b \frac{1}{\sin\varphi}\Big(\cos\varphi \underbrace{\langle \nabla_{e_1} e_1, e_1\rangle}_{=0} - \sin\varphi \underbrace{\langle \nabla_{e_1} e_1, e_2\rangle}_{=\kappa_g} \\
&\qquad + \cos\varphi \underbrace{\langle X_1, \nabla_{e_1} X_1\rangle}_{=0} + \sin\varphi \underbrace{\langle X_2, \nabla_{e_1} X_1\rangle}_{=\omega_1^2(e_1)}\Big)ds \\
&= \int_a^b \big(\kappa_g + \omega_2^1(e_1)\big)ds = \int_c \kappa_g ds + \int_{f(\partial B)} \omega_2^1 = \int_c \kappa_g ds + \int_{f(B)} \Omega_2^1 \\
&= \int_c \kappa_g ds + \int_{f(B)} K \cdot \omega^1 \wedge \omega^2.
\end{aligned}
$$

But $\omega^1 \wedge \omega^2$ is precisely the surface area element ($= dA$) of the surface. The theorem of Stokes 4.36 has been used in the next to last equation, together with the equation $d\omega_2^1 = \Omega_2^1 = K \cdot \omega^1 \wedge \omega^2$ in 4.35. Note that this calculation is purely intrinsic and makes no use whatsoever of the second fundamental form. In fact, 4.38 is a purely intrinsic result. □

4.39. Theorem. (Gauss-Bonnet formula, second local version) Let B be as in 4.38, now not diffeomorphic to a closed disc, but rather homeomorphic to one, with piecewise smooth and connected boundary (that is to say, every point has a neighborhood which is either diffeomorphic to D^2 or D^2_+ or $D^2_{++} = \{(x, y) \in I\!R^2 \mid x^2 + y^2 < 1,\ x \geq 0 \text{ and } y \geq 0\}$ or diffeomorphic to the closure of $D^2 \setminus D^2_{++}$ in D^2, i.e., to $\{(x, y) \in I\!R^2 \mid x^2 + y^2 < 1,\ x \leq 0 \text{ or } y \leq 0\}$).
In the image $f(B)$ let $\alpha_1, \ldots, \alpha_n$ be (oriented) exterior angles at the finitely many places where the boundary is not smooth (the so-called *corners*), where we always assume that $-\pi < \alpha_i < \pi$ holds. Then

$$\int_{f(B)} K dA + \int_{\partial f(B)} \kappa_g ds + \sum_i \alpha_i = 2\pi.$$

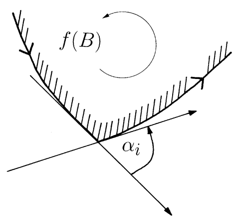


Figure 4.3. Exterior angle in the Gauss-Bonnet formula

The proof of 4.39 can be carried out in one of two different ways: either one generalizes the theorem on turning tangents to the case of a piecewise smooth boundary curve, or one reduces it to 4.38 by smoothing each of the finitely many corners. For this it is sufficient to smooth the corner in the above defined set D^2_{++} by an appropriately chosen convex curve and then to pull back the result to the boundary of B by means of the diffeomorphism. This can be thought of as follows: the exterior angle α_i corresponds to the "skip" of the tangent at the ith corner. We will not go into these technical details here. The smooth pieces of the boundary between the corners are also called *edges*, and $f(B)$ can be viewed as an abstract polygon. This will be important in the combinatorial considerations of Section 4.43.

4.40. Corollary. (Geodesic n-gon)
Let B be as in 4.39, but the boundary is now assumed to consist of finitely many segments of geodesics (this forms what is referred to as a *geodesic n-gon*) with exterior angles $\alpha_1, \ldots, \alpha_n$. Then

$$\int_{f(B)} K dA = 2\pi - \sum_{i=1}^{n} \alpha_i.$$

In the special case $n = 3$ (a *geodesic triangle*) we obtain the *Theorema Elegantissimum* of Gauss

$$\int_{f(B)} K dA = 2\pi - \alpha_1 - \alpha_2 - \alpha_3 = \beta_1 + \beta_2 + \beta_3 - \pi,$$

in which $\beta_i := \pi - \alpha_i$ denotes the interior angles $(0 < \beta_i < 2\pi)$. From this we get the following consequence.

4.41. Corollary. The sum of the interior angles of a geodesic triangle

$$\text{is} \left\{ \begin{array}{c} > \pi, \\ = \pi, \\ < \pi, \end{array} \right\} \text{ in case } \left\{ \begin{array}{c} K > 0 \\ K = 0 \\ K < 0 \end{array} \right\} \begin{array}{l} \text{holds in the interior} \\ \text{of the triangle.} \end{array}$$

What is important in this simple formulation is that the interior of the triangle is really contained in the surface. Thus, it is not sufficient to glue three segments of geodesics together such that they close to a triangle.

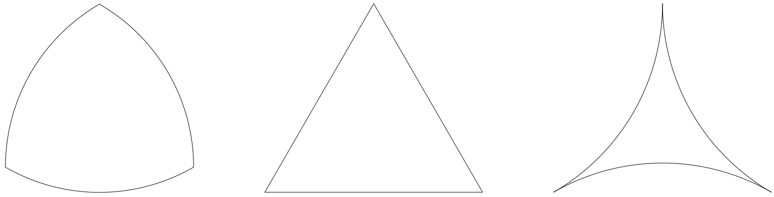


Figure 4.4. Sum of the interior angles in a geodesic triangle in the three cases $K > 0,\ K = 0,\ K < 0$

For a geodesic n-gon, the π in the above formulas is replaced by $(n-2)\pi$, with equality again holding in the Euclidean case.

EXAMPLE: An *octant* of the 2-sphere is the set given by the condition

$$\{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1, x,y,z \geq 0\}.$$

Here one has $K = 1$ and $\kappa_g = 0$ (the boundary consists of segments of great circles) and $\alpha_i = \frac{\pi}{2}$, $i = 1,2,3$. This implies the equation

$$\underbrace{\int K dA}_{=\frac{1}{8}4\pi} + \underbrace{\int \kappa_g}_{=0} + \underbrace{\sum_i \alpha_i}_{=\frac{3}{2}\pi} = 2\pi.$$

The fact that the sum of the interior angles in a Euclidean triangle is precisely $180° = \pi$ is one of the basic insights of Euclidean geometry. In spherical trigonometry it has been known since ancient times that the sum of the interior angles is larger than π and smaller than 5π. The most interesting case is the hyperbolic case ($K = -1$), which is closely related to the parallel axiom and non-Euclidean geometry, cf. 3.44. In the hyperbolic plane the sum of the interior angles always lies between 0 and π. From the Theorema Elegantissimum in 4.40 it follows that the surface area of a geodesic triangle also lies between 0 and π, even though the surface area of the entire hyperbolic plane is infinite.

4.42. Corollary. In case one has $K < 0$ everywhere on a surface element, there is no geodesic two-gon in the sense of 4.40. In other words, in a simply connected domain, it is not possible that two geodesics intersect in two points if $K < 0$.

4.43. Theorem. (Gauss-Bonnet formula, global version)
Let $M \subset \mathbb{R}^3$ be a compact two-dimensional (orientable) submanifold (without boundary). Then

$$\int_M K dA = 2\pi\chi(M),$$

where $\chi(M) \in \mathbb{Z}$ denotes the Euler characteristic of M, which is invariant under homeomorphisms and in particular is independent of the embedding of the submanifold.

Sketch of proof: Since M is compact, it can be covered by finitely many subsets, each of which can be described as an image of a surface element as in Definition 3.1. Therefore, M can be decomposed into finitely many parts $M_1, \ldots, M_m$ such that

1. $M = \bigcup_{i=1}^{m} M_i$,
2. $M_i \cap M_j$ contains no interior points for $i \neq j$, but at most boundary points of M_i resp. M_j,
3. each M_i is a compact set with piecewise smooth and connected boundary as in 4.39 (with finitely many corners and edges and with exterior angles α_{ij}). Corners with an angle $\alpha_{ij} = 0$ can be added or deleted if necessary.

A compact submanifold without boundary is always oriented in the sense of 3.6 and 3.7. Thus, for every $i = 1, \ldots, m$ with orientation as in 4.39, we have

$$\int_{M_i} K dA + \int_{\partial M_i} \kappa_g ds = 2\pi - \sum_j \alpha_{ij}.$$

When we take the sum over all i, the boundary components cancel, i.e., $\sum_i \int_{\partial M_i} \kappa_g ds = 0$, and we get

$$\begin{aligned} \int_M K dA &= 2\pi m - \sum_{i,j} \alpha_{ij} \\ &= 2\pi m - \sum_{i,j} (\pi - \beta_{ij}) \quad \text{(where } \beta_{ij} \text{ denote the interior angles)} \\ &= 2\pi \big(\text{number of corners } - \text{ number of edges } + m\big) =: 2\pi\chi(M). \end{aligned}$$

The next to last equality is verified by realizing that the sum of the interior angles at each corner is 2π and therefore $\sum_{i,j} \beta_{ij}$ is equal to the number of corners, multiplied by 2π. Moreover, the number of summands in the sum is equal to twice the number of edges (every edge occurs in precisely two of the M_i), and thus $\sum_{i,j} \pi$ is equal to the number of edges, multiplied by 2π. □

The last equality is nothing but the definition of the *Euler characteristic* $\chi(M)$. In fact, $\chi(M)$, being a purely combinatorially defined quantity, does not depend on the decomposition of M into the M_i. See [**38**] for its significance and for more information on triangulations of surfaces. The Gauss-Bonnet formula shows that this quantity is independent of combinatorial considerations, since the integral over K has nothing to do with combinatorics.

The *classification theorem for surfaces* in topology tells us that two (abstract) compact surfaces without boundary are homeomorphic (in fact diffeomorphic) to one another if and only if their Euler characteristics coincide and both surfaces are either orientable or non-orientable. For a proof of this see for example [**38**], Chapter 7. The standard models for compact surfaces (more precisely: two-dimensional compact manifolds) are either the sphere with g handles (cylinders connected on both ends to the sphere) or a sphere with g Möbius strips connected to it. To connect a cylinder, one cuts two disjoint holes into the sphere and glues the ends of the cylinder along the boundaries, which are all just copies of the one-sphere. This is done in such a way as to preserve the orientation. To connect a Möbius strip to the sphere, one cuts one hole in the sphere and glues the Möbius strip along this boundary (which makes sense since the boundary of the Möbius strip is connected, hence a copy of the one-sphere). One speaks of *oriented surfaces of genus g* in the former case, of *non-orientable surfaces of genus g* in the latter case. The Euler characteristic is $\chi = 2 - 2g$ in the former case and $\chi = 2 - g$ in the latter. For oriented surfaces the principle of construction is illustrated in Figure 4.5.

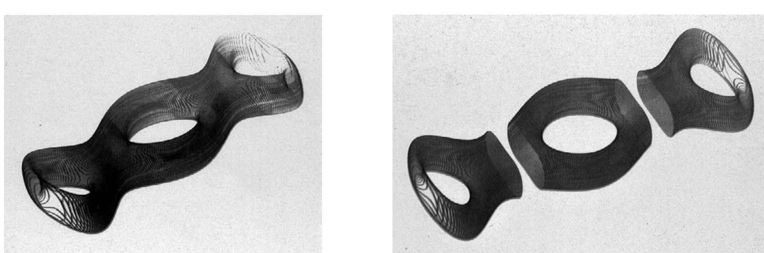


Figure 4.5. Orientable surfaces of genus 3 and the Euler characteristic $\chi = -4$, decomposed in irreducible pieces[5]

[5]Following F. Apéry, "Models of the real projective plane", Vieweg 1987, p. 130.

Orientable examples are the sphere with $g = 0$ and $\chi = 2$, the *torus* with $g = 1$ and $\chi = 0$, and the *pretzel surface* with $g = 2$ and $\chi = -2$. Non-orientable examples are the *real projective plane* with $g = 1$ and $\chi = 1$, and the *Klein bottle* with $g = 2$ and $\chi = 0$.

Two final remarks: (i) The fact that the quantity $\int_M K dA$ is constant for a fixed Euler characteristic (fixed topological type) can also be shown without the local Gauss-Bonnet formula, using the calculus of variations. In fact, it can be shown that the purely intrinsic variation of this integral under variations of the metric (i.e., under variations of the first fundamental form) vanishes identically. See 8.6 and 8.8 for more details on this.

(ii) An anolog (due to H. Hopf) of 4.43 for compact hypersurfaces in $\mathbb{R}^{n+1}$ with even n states that $\int_M K dV = c_n \chi(M)$, where c_n is half the volume of the unit sphere S^n and K denotes the Gauss-Kronecker curvature, which is the determinant of the Weingarten mapping.[6] For odd n such a theorem does not hold, cf. Exercise 21.

4G Selected topics in the global theory of surfaces

The *theory of sufaces in the large* or *global theory of surfaces* is concerned with the properties of compact surfaces or surfaces which are *complete* in some sense. Think of this as meaning that the surface "extends to infinity". Compact surfaces (without self-intersection) are best thought of as two-dimensional compact submanifolds of $\mathbb{R}^3$ (cf. 1.5, and for the orientability 3.7). The Gauss-Bonnet theorem for two-dimensional submanifolds of $\mathbb{R}^3$ is, according to 4.43, the equation

$$\int_M K dA = 2\pi\chi(M).$$

Following this fundamental result, we turn first to the total absolute curvature, and then we will discuss some classical results about surfaces with constant curvature. First we relate the Gaussian curvature with the surface area of the *Gaussian normal image*, i.e., the image

[6] H. Hopf, *Über die curvatura integra geschlossener Hyperflächen*, Math. Annalen **95**, 340–367 (1926); see also D. H. Gottlieb, *All the way with Gauss-Bonnet and the sociology of mathematics*, Amer. Math. Monthly **104**, 457–469 (1996).

of the surface under the Gauss map ν. The symbol Vol_{S^2} denotes the surface area (i.e., two-dimensional volume) of a subset of the sphere.

4.44. Lemma. (Gaussian normal image)
Let $f\colon U \to \mathbb{R}^3$ be a surface element and let $B \subset U$ be compact. Suppose the Gauss map $\nu\colon U \to S^2$ is injective and has maximal rank. Then

$$\int_{f(B)} |K| dA = \mathrm{Vol}_{S^2}\big(\nu(B)\big).$$

PROOF: We orient the surface in such a way that $\int dA$ becomes positive. Then we have, on the one hand,

$$\int_{f(B)} |K| dA = \int_{f(B)} |\mathrm{Det}(L)| dA = \int_B |\mathrm{Det}(L)| \sqrt{\mathrm{Det}(g_{ij})} du^1 du^2,$$

and on the other

$$\begin{aligned}\mathrm{Vol}_{S^2}\big(\nu(B)\big) &= \int_B \sqrt{\mathrm{Det}(e_{ij})} du^1 du^2 \\ &= \int_B \sqrt{(\mathrm{Det}(L))^2 \mathrm{Det}(g_{ij})} du^1 du^2.\end{aligned}$$

Here, e_{ij} denotes the third fundamental form, which is nothing but the first fundamental form of the spherical image, cf. 3.10. The last equality follows from the transformation rule for the Gram determinant. This is most easily seen in a basis of eigenvectors of the Weingarten mapping: letting $LX = \lambda X, LY = \mu Y$, we have $\mathrm{Det}(L) = \lambda\mu$ and

$$\begin{aligned}\mathrm{Det}\begin{pmatrix} \langle LX, LX\rangle & \langle LX, LY\rangle \\ \langle LY, LX\rangle & \langle LY, LY\rangle \end{pmatrix} &= \mathrm{Det}\begin{pmatrix} \lambda^2\langle X, X\rangle & \lambda\mu\langle X, Y\rangle \\ \lambda\mu\langle Y, X\rangle & \mu^2\langle Y, Y\rangle \end{pmatrix} \\ &= \lambda^2\mu^2 \mathrm{Det}\begin{pmatrix} \langle X, X\rangle & \langle X, Y\rangle \\ \langle Y, X\rangle & \langle Y, Y\rangle \end{pmatrix}.\end{aligned}$$

4.45. Corollary. For a surface element $f\colon U \to \mathbb{R}^3$, let $(U_n)_{n\in\mathbb{N}}$ be a sequence of open sets in the domain of definition U of the parametrization. Suppose that $U_{n+1} \subset U_n$ for all n and that $\bigcap_n f(U_n) = \{p\}$. Then we have

$$K(p) = \lim_{n\to\infty} \frac{\mathrm{Vol}_{S^2}(\nu(U_n))}{\mathrm{Vol}_U(U_n)}.$$

4.45 follows immediately from 4.44 and the mean value theorem of calculus. To conclude in this way, however, it is necessary to give the "volume" of the Gauss normal image a definite sign, depending on whether ν is orientation preserving or reversing. Once this is done, the Gaussian curvature is viewed as the infinitesimal change of the volume element (with sign) of the Gauss map ν. The points for which $K = 0$ are characterized as the points at which the Gauss map has rank less than two.

A *convex* two-dimensional submanifold is defined similarly as in 2.30, namely as the smooth boundary of a convex three-dimensional set in $\mathbb{R}^3$. The *convex hull* of a given set A is defined to be the smallest convex set containing A.

4.46. Theorem. (Total absolute curvature)

(i) Let M_0 denote a strictly convex and compact two-dimensional submanifold in $\mathbb{R}^3$, i.e., a surface for which $K > 0$ holds everywhere, which bounds a convex set (one also speaks in this case of an *ovaloid*). Then the Gauss map of this surface is globally bijective, and one has

$$\int_{M_0} K dA = 4\pi.$$

(ii) Now let M an arbitrary two-dimensional submanifold of $\mathbb{R}^3$ with $M_+ = \{x \in M \mid K(x) > 0\}, M_- = \{x \in M \mid K(x) < 0\}$. Then

$$\int_{M_+} K dA \geq 4\pi,$$

with equality if and only if M_+ is contained in the boundary of the convex hull of M.

(iii) Let M be as in (ii). Then

$$\int_M |K| dA \geq 2\pi\big(4 - \chi(M)\big),$$

with equality holding if and only if $\int_{M_+} |K| dA = 4\pi$ and $\int_{M_-} |K| dA = 2\pi\big(2 - \chi(M)\big)$.

Proof: (i) First we must agree that a continuous Gauss map $\nu : M_0 \to S^2$, which is defined everywhere, must be globally bijective (and hence is a diffeomorphism because of the assumption on differentiability). If we chose ν in such a way that $\nu(x)$ at every point $x \in M_0$ points in the outward direction (away from the surface), then this clearly defines a continuous Gauss map. Moreover, ν is surjective: for a given direction $e \in S^2$ we can, by applying parallel translation, find a plane which touches the surface and is perpendicular to e, with the additional property that e points outward; it follows that e coincides with $\nu(x)$. The map ν is also injective: if $\nu(x) = \nu(y)$ for two different points x, y of the surface, then the tangent planes at x and y are parallel to each other. Because $K > 0$, the surface lies in each case strictly in a half-space, which is bounded by these tangent planes (except for the points x and y themselves). In fact, this half-space is the one which is opposite the vector e, i.e., the half-space with the same exterior normal vector. This is a contradiction to the convexity, thus establishing the injectivity, since the segment joining x with y must lie in the interior of the surface, but in the situation at hand it partially lies in the exterior. From the bijectivity of the Gauss map ν, it follows from 4.44 that $\int_{M_0} KdA$ coincides with the surface area of the entire two-sphere (which is 4π).

For the proof of (ii), note that M has a convex hull with boundary $\widetilde{M}$. This $\widetilde{M}$ is a (maybe not strictly) convex surface. At any rate M contains all points of $\widetilde{M}$ with $K > 0$, since otherwise the convex hull would necessarily be smaller. Every point of $\widetilde{M}$ satisfies $K \geq 0$; hence $\int_{\widetilde{M} \setminus M_+} KdA = 0$. Here we would like to apply the equation of (i) to $\widetilde{M}$ with the result

$$\int_{M_+} KdA \geq \int_{M_+ \cap \widetilde{M}} KdA = \int_{\widetilde{M}} KdA = 4\pi.$$

The equation in (i), however, holds *a priori* only for strictly convex surfaces with $K > 0$. Moreover, $\widetilde{M}$ is only C^1 at the points where M leaves the boundary of the convex hull. But one can show that in this case the Gauss map, while not being bijective when restricted to $\widetilde{M} \setminus \{x \in \widetilde{M} \mid K(x) = 0\}$, still maps dominantly on the surfae area of the sphere S^2, since the points where $K = 0$ don't contribute to the integral. This is because on an open connected piece of the surface where $K = 0$, the Jacobi determinant of ν vanishes and thus also the distortion in volume. On the other hand, every element of S^2 is in the image of ν, and a neighborhood of every point with $K > 0$ carries a positive contribution to the integral. In conclusion, the equality $\int K dA = 4\pi$ holds also for arbitrary convex surfaces. Also, the inequality in (ii) holds, and in the case of equality $\int_{M_+} = 4\pi$ there can be no points of positive curvature which are not in $\widetilde{M}$.

(iii) follows now quite easily from (ii) and 4.43 by decomposing the integral

$$\int_M |K| dA = \int_{M_+} |K| dA + \int_{M_-} |K| dA = 2\int_{M_+} K dA - \int_M K dA$$

$$= 2\int_{M_+} K dA - 2\pi\chi(M) \geq 8\pi - 2\pi\chi(M). \qquad \square$$

4.47. Definition. (Tightness)
If a compact two-dimensional submanifold of $\mathbb{R}^3$ satisfies the equality in 4.46 (iii), then it clearly has only as much positive curvature as is absolutely necessary, and the integral of the absolute value $|K|$ of the curvature is a small as possible for a given topological type of the surface. One calls the surface *tight*, if its total absolute curvature is minimal, i.e., if one has

$$\int_M |K| dA = 2\pi(4 - \chi(M)).$$

REMARKS:

1. Definition 4.47 can also be applied to compact surfaces with self-intersections. In fact, both 4.46 and 4.48 hold in this case, and with appropriate modifications also in the non-orientable case. However, the proof of 4.46 has to be modified for these generalizations; in particular, the definitions which are used must first be adapted. For this it is necessary to know something about globally defined immersions of abstract manifolds. In any case, we will not define the notion of an abstract manifold until later on in Chapter 5, and cannot use that here.

2. Just like the total curvature $\int_M K dA$, the total absolute curvature $\int_M |K| dA$ is a quantity of the intrinsic geometry. In this sense, the notion of tightness is determined intrinsically. Note, however, that without the ambient Euclidean space the inequality in 4.46 (iii) is wrong. There are abstractly defined metrics on compact surfaces with vanishing (flat torus, cf. 7.24) or purely negative Gaussian curvature, and thus equality in the weaker inequality
$$\int_M |K| dA \geq 2\pi |\chi(M)|$$
becomes possible. The case of equality in the last inequality simply means that the sign of K is equal to the sign of $\chi(M)$.

3. Figure 4.6 shows a tight surface of genus 2, which is in fact a connected component of the algebraic surface[7]
$$2y(y^2 - 3x^2)(1 - z^2) + (x^2 + y^2)^2 = (9z^2 - 1)(1 - z^2).$$

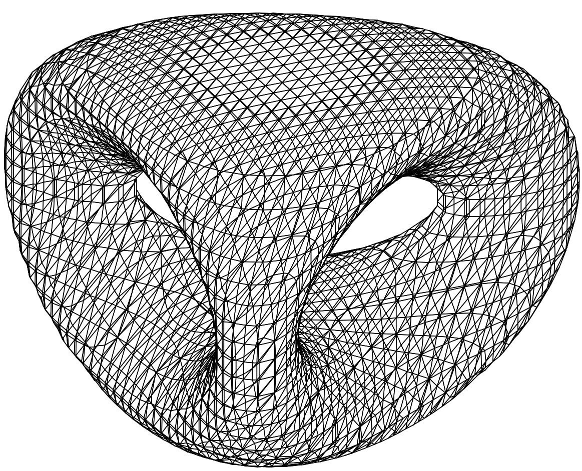


Figure 4.6. A tight surface of genus 2

[7]According to T. Banchoff & N. H. Kuiper, *Geometrical class and degree for surfaces in three-space*, Journal of Differential Geometry **16**, 559-576 (1981), Section 5.

4.48. Corollary. The following conditions on a compact surface $M \subset I\!R^3$ are equivalent:

(i) M is tight.
(ii) $\int_M |K| dA = 2\pi(4 - \chi(M))$.
(iii) $\int_{M_+} K dA = 4\pi$.
(iv) Every plane $\mathcal{E} \subset I\!R^3$ decomposes M into at most two connected components, i.e., $M \setminus \mathcal{E}$ has at most two connected components, each of which is on one side of the plane.

Proof: The equivalence of (i), (ii), (iii) are clear by 4.46 and 4.47.

For the implication (iii) $\Rightarrow$ (iv), we make the assumption that $M \setminus \mathcal{E}$ has at least three components and take this assumption to a contradiction. First, it is easy to see that each component of $M \setminus \mathcal{E}$ has at least one point with positive Gaussian curvature: the plane which is parallel to $\mathcal{E}$ and which contains that point of the component of the surface which is farthest away from $\mathcal{E}$ touches the surface at a point with $K \geq 0$ such that a unit normal ν is perpendicular to $\mathcal{E}$. If we slightly perturb $\mathcal{E}$, we get a (variable) plane $\mathcal{E}'$ and can arrange that the unit normal vectors ν' to these planes cover an open set of the Gaussian normal image, and that in addition the number of components of $M \setminus \mathcal{E}'$ is always at least three, since these components are open in M. Then, by 4.44, at least one of these normal vectors ν' must have three points of tangency with the three components, at points with positive Gaussian curvature. On the other hand, by 4.46 (ii), all points with positive Gaussian curvature lie on the boundary of the convex hull. But this is an impossibility for the third point, since in the boundary of the convex hull there can be at most two such points. In other words, it is impossible that there are three parallel planes, all of which touch one and the same convex surface (in the case at hand the boundary of the convex hull of M).

For (iv) $\Rightarrow$ (iii) we assume conversely that $\int_{M_+} K dA > 4\pi$. Then by 4.46 there is a point x with positive Gaussian curvature, which does not lie in the boundary $\widetilde{M}$ of the convex hull but rather in the interior of the convex hull. By moving the tangent plane at x slightly, we can find a parallel plane $\mathcal{E}$, such that $M \setminus \mathcal{E}$ contains a small neighborhood of x as a separate component. On the other hand, $M \setminus \mathcal{E}$ must have at least two further components, since $\mathcal{E}$ passes through the interior of the convex hull. Thus there are at least three components. $\square$

4.49. Corollary. The tightness of a compact surface is invariant under projective transformations of the ambient space. More precisely, let M be a tight two-dimensional submanifold of $\mathbb{R}^3$, and let F be a projective transformation of the projective closure $\mathbb{R}P^3$, which maps every point of M to a point in the finite (affine) part[8]. Then $F(M)$ is again compact and tight.

This follows simply from the fact that projective transformations map planes to planes. Thus the two-piece property of 4.48 (iv) is preserved, since $F(M) \setminus F(\mathcal{E})$ has as many components as $M \setminus \mathcal{E}$.

REMARKS:[9]

1. Since the property (iv) in 4.48 makes no assumption on differentiability, one considers also the more general condition on a compact subset of $\mathbb{R}^3$ that it is homeomorphic to a surface, and defines the surface to be *tight* if the property (iv) is satisfied (*two-piece property*, TPP). The notion defined in this manner is a generalization of the notion of convexity, since every convex set and the boundary of such a set has the property (iv).
2. The differential topological interpretation of the notion of tightness is the following. In almost all directions z, the function $M \ni p \mapsto \langle p, z\rangle$ has only finitely many (non-degenerate) critical points (i.e., points with vanishing gradient) on the submanifold M. The number of these critical points is always larger than or equal to $4-\chi(M)$. Equality for almost all z is satisfied precisely for the tight surfaces.
3. There are orientable and tight surfaces of arbitrary genus. The sphere and the torus of rotation are obviously tight, and the example in Figure 4.6 indicates how to construct tight surfaces of higher genus. It is in principle sufficient to glue to a given

[8] This part is given by the image of the embedding $(x^1, x^2, x^3) \mapsto [1, x^1, x^2, x^3]$ of $\mathbb{R}^3$ into $\mathbb{R}P^3$. Here $\mathbb{R}P^3$ is the set of all equivalence classes $[x^0, x^1, x^2, x^3]$ (not all entries zero) where $[x^0, x^1, x^2, x^3] = [\lambda x^0, \lambda x^1, \lambda x^2, \lambda x^3]$ for any $\lambda \neq 0$. The affine part is just given by $x^0 \neq 0$. A *projective transformation* is nothing but a linear transformation of 4-space, regarded on this set of equivalence classes.

[9] Cf. also the survey article T. Banchoff & W. Kühnel, *Tight submanifolds, smooth and polyhedral*, in: "Tight and Taut Submanifolds", MSRI Publ. **32**, 51–118, Cambridge Univ. Press 1997.

tight surface of genus g a handle with non-positive curvature, such that the result is a tight surface of genus $g+1$. It is an interesting fact that there are also non-orientable surfaces in $\mathbb{R}^3$ which have the same tightness property. To accommodate this case one must modify the definition appropriately (see the remark following 4.47), since there is no globally defined Gauss map and globally defined surface element in this case. Still, the condition in 4.48 (iv) can be carried over literally. Of course, closed, non-orientable surfaces in $\mathbb{R}^3$ always have some self-intersections, but they do exist as globally defined immersions in abstract non-orientable surfaces. Non-orientable tight surfaces exist for every value of the Euler characteristic $\chi \leq -2$. The only exceptions to this are thus the projective plane ($\chi = 1$), the Klein bottle ($\chi = 0$) and the non-orientable surface of genus 3 ($\chi = -1$).

We also mention without proof the following result on the total curvature on non-compact surfaces:

4.50. Theorem. (S. Cohn-Vossen 1935[10])
Let M be a non-compact but complete two-dimensional submanifold of $\mathbb{R}^3$ (completeness here means that every Cauchy sequence in M converges to a point in M). Then one has the following inequality for the total curvature:

$$\int_M K dA \leq 2\pi\chi(M),$$

with equality at least if the total surface area $\int_M dA$ is finite. More precisely, this holds under the assumption that $\int_M K dA$ either converges as an improper integral or diverges to $-\infty$. As in 4.43, $\chi(M)$ denotes the Euler characteristic, which in the case of non-compact surfaces either is finite or is defined (formally) as $-\infty$.

This theorem is actually true in a more general context, being an intrinsic result for abstract two-dimensional manifolds with an (abstract) first fundamental form, which is complete in the sense that every geodesic can be continued infinitely in both directions. This

[10] *Kürzeste Wege and Totalkrümmung auf Flächen*, Compositio Math. **2**, 69-133 (1935); for a short proof see: S. Rosenberg, *Gauss-Bonnet theorems for noncompact surfaces*, Proc. Amer. Math. Soc. **86**, 184–185 (1982).

notion of completeness is explained in more detail in Section 7C. One can also study minimal surfaces "in the large", looking for minimal surfaces in space, in particular for those without self-intersections and those with finite total curvature, see [**12**].

Following these considerations of the total curvature, we now pass to constancy conditions on the curvature of a surface. This is a quite natural question, comparable with the conditions for the Frenet curvature to be constant in our discussion of curves. The (round) sphere for example has constant Gaussian curvature and constant mean curvature as well as constant principal curvatures. As far back as the nineteenth century global results were discovered which characterize the (round) sphere by curvature conditions of this kind.

4.51. Theorem. (H. Liebmann 1899[11])
Let M be a compact, two-dimensional submanifold in $I\!R^3$ of the class C^4 with constant Gaussian curvature K. Then K is positive, and M is a sphere of radius $r = \frac{1}{\sqrt{K}}$.

PROOF: Every compact surface in $I\!R^3$ has at least one point p for which $K(p) > 0$. This follows for example from 4.46 (cf. also Exercise 10 at the end of the chapter). Thus the constant K is positive. Let $\kappa \geq \lambda > 0$ denote the two principal curvatures. If one always has $\kappa = \lambda$, then the surface is locally already a piece of a sphere by 3.14. All the more, then, this holds globally. Otherwise there would be a point p for which $\kappa(p) > \lambda(p)$, where κ has a local maximum and thus λ a local minimum (from the constancy $K = \kappa \cdot \lambda$). But this is impossible, which we will now show by contradiction.

For this we use, in a neighborhood of p, curvature line parameters (u, v) and the Gauss and Codazzi-Mainardi equations from 4.26. Setting $\kappa = L/E, \lambda = N/G$, we have

$$L_v = \frac{E_v}{2}\Big(\kappa + \lambda\Big), \quad N_u = \frac{G_u}{2}\Big(\kappa + \lambda\Big).$$

[11] *Eine neue Eigenschaft der Kugel*, Nachr. Akad. Göttingen, Math.-Phys. Klasse, 44–55 (1899).

If we differentiate $L = \kappa E$ and $N = \lambda G$ and insert the above expression, we get the relation

$$E_v = \frac{2\kappa_v E}{\lambda - \kappa}, \quad G_u = \frac{2\lambda_u G}{\kappa - \lambda}.$$

At the point p the principal curvatures are stationary; hence we have $\kappa_u(p) = \kappa_v(p) = \lambda_u(p) = \lambda_v(p) = 0$ and consequently $E_v(p) = G_u(p) = 0$. By differentiating one more time we get

$$E_{vv} = -\frac{2\kappa_{vv} E}{\kappa - \lambda} + \kappa_v(\cdots) + \lambda_v(\cdots),$$

$$G_{uu} = \frac{2\lambda_{uu} G}{\kappa - \lambda} + \kappa_u(\cdots) + \lambda_u(\cdots).$$

The expressions $(\cdots)$ here denote some continuous (hence bounded) functions of E, G and their derivatives. We know that κ has a local maximum at p and λ has a local minimum there. Thus $\kappa_{vv}(p) \leq 0, \lambda_{uu}(p) \geq 0$, from which it follows that

$$E_{vv}(p) \geq 0, \quad G_{uu}(p) \geq 0.$$

We now evaluate the Gauss equation 4.26 (ii) at the point p:

$$K = -\frac{1}{2\sqrt{EG}}\Big(\Big(\frac{E_v}{\sqrt{EG}}\Big)_v + \Big(\frac{G_u}{\sqrt{EG}}\Big)_u\Big).$$

Using $E_v(p) = G_u(p) = 0$, we get from this

$$K(p) = -\frac{1}{2EG}\Big(E_{vv}(p) + G_{uu}(p)\Big) \leq 0,$$

which contradicts the assumption that $K(p) > 0$. □

The proof of this theorem implies the following purely local lemma, which we formulate separately. It immediately leads to Theorem 4.53 as well.

4.52. Lemma. (D. Hilbert)

If at a point on a two-dimensional surface element of the class C^4, which is assumed to be non-umbilic, the larger of the two principal curvatures has a local maximum and the other has at the same time a local minimum, then at this point $K \leq 0$.

4.53. Theorem. (H. Liebmann 1900)
Let M be a compact, two-dimensional, C^4-submanifold of $\mathbb{R}^3$ with $K > 0$ everywhere and with constant mean curvature H. Then M is a sphere of radius $\frac{1}{|H|}$.

PROOF: If all points on the surface are umbilics, then M is a sphere by Theorem 3.14. Otherwise there is a point p with $\kappa(p) > \lambda(p)$. From the assumption that H is constant, we have const. $= 2H = \kappa + \lambda$, and κ has a local maximum precisely where λ has a local minimum. But this contradicts the previous lemma 4.52. □

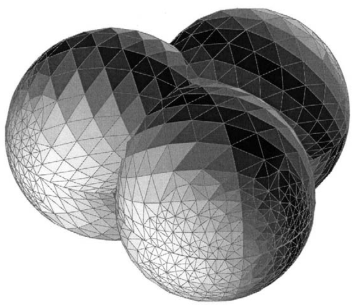


Figure 4.7. Wente-torus with constant mean curvature[12]

REMARK: It was conjectured for some time that the sphere is the only compact surface that has constant mean curvature (attributed to H. Hopf). In fact, there are other surfaces with constant mean curvature. By a theorem of Alexandrov, all of other examples must have self-intersections. The first example of this type was the so-called Wente-torus, named after its discoverer H. C. Wente 1984[13], pictured in Figure 4.7 and Figure 4.8.

[12]Reproduced with kind permission of K. Große-Brauckmann and K. Polthier, for more information see the essay *Numerical examples of compact constant mean curvature surfaces*, "Elliptic and parabolic methods in geometry" (B.Chow et al., eds.), Proceedings Minneapolis, MN 1994, 23–46, A.K.Peters 1996, cf. also `http://www.math.uni-tuebingen.de/user/nick/gallery/WenteTorus.html`.

[13]Cf. U. Abresch, *Constant mean curvature tori in terms of elliptic functions*, J. Reine und Angew. Math. **374**, 169–192 (1987) as well as R. Walter, *Explicit examples to the H-problem of Heinz Hopf*, Geometriae Dedicata **23**, 187–213 (1987); both articles contain computer images and explain the rather complicated interior of the Wente torus.

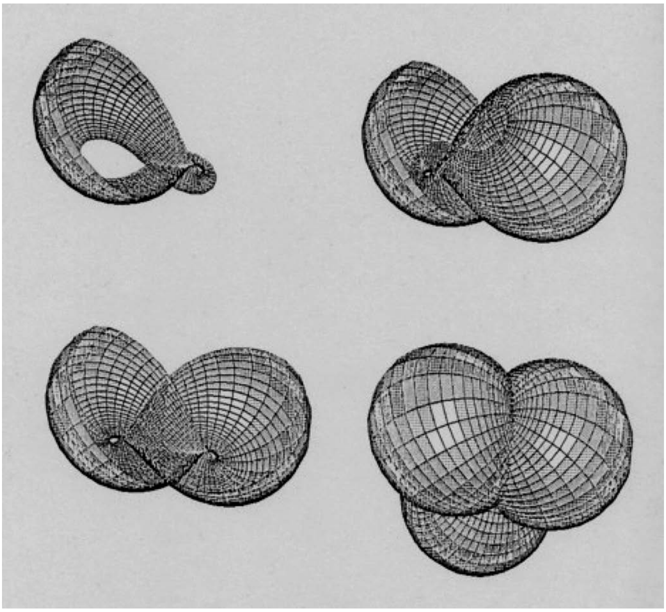


Figure 4.8. Building blocks and the interior of the Wente torus[14]

Exercises

1. Show that all geodesics on a circular cylinder

$$f(u, v) = (\cos u, \sin u, v)$$

are either Euclidean lines, circles, or helices. What do the geodesics on a circular cone look like?

2. Show that the geodesics on the surface of the sphere are precisely the great circles.

3. Suppose we are given a curve c on a surface element, which passes through a fixed point p. Show that the geodesic curvature $\kappa_g(p)$ of c coincides with the curvature $\kappa(p)$ of the plane curve

[14]Reproduced with kind permission of I. Sterling und U. Pinkall. Figures from "Willmore surfaces", U. Pinkall and I. Sterling, *The Mathematical Intelligencer*, vol. 9, no. 2, 1987, pp. 38–43. With kind permission of Springer Science and Business Media.

which is obtained as the orthogonal projection of c in the tangent plane at p.

4. Show that (locally) a curve on a surface element is uniquely determined by the geodesic curvature as a function of the arc length, if one prescribes a point $c(0)$ and the direction $c'(0)$. Compare this with the plane case discussed in Section 2B as well as the case $\kappa_g = 0$ in 4.12.

5. Show that a Frenet curve on a surface element is a geodesic if and only if the unit normal to the surface coincides with the principal normal of the curve (at least up to sign).

6. Do there exist local coordinates u^1, u^2 on an arbitrary surface element with the property that the u^1-curves are perpendicular to the u^2-curves, and all u^i-curves are geodesics parametrized by arc length? Hint: 4.28.

7. Show that the surface elements $f_1(u, v) = (u \sin v, u \cos v, \log u)$ and $f_2(u, v) = (u \sin v, u \cos v, v)$ have the same Gaussian curvature in the parameters u, v. The former surface is a surface of rotation; the latter is the helicoid (cf. 3.37). Are these surfaces (locally) isometric? Hint: Consider the curves for which the Gaussian curvature is constant, as well as the curves which are perpendicular to these, compare [**5**, §91].

8. Show that for a Tchebychev grid (cf. Exercise 6 in Chapter 3) the curvature is given by $K = -\frac{\partial^2 \vartheta}{\partial u_1 \partial u_2} \big/ \sin \vartheta$.

9. The *four-dimensional catenoid* is defined as the hypersurface in $I\!R^5$ which arises through rotation of a (plane) catenary around an axis which lies in this plane. The topological type is $I\!R \times S^3$, and this hypersurface contains the usual catenoid as a section with any three-dimensional subspace which contains this axis of rotation. Show that the hypersurface has vanishing scalar curvature, i.e., the second mean curvature vanishes identically, cf. Definition 4.22. Hint: The principal curvatures are κ_1 and $\kappa_2 = \kappa_3 = \kappa_4 = -\kappa_1$.

10. Show the following without using the results of Section 4G: a two-dimensional compact submanifold of $I\!R^3$ always has an elliptic point.

Hint: Consider a ball of as small a radius as possible, which contains the submanifold, and apply the Taylor-expansion in the neighborhood of a point of contact. Why isn't the Gauss-Bonnet formula 4.43 alone a sufficient argument?

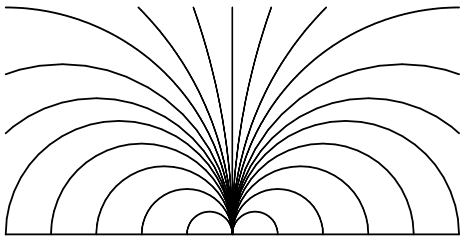


Figure 4.9. Geodesics in the Poincaré upper half-plane

11. The *Poincaré upper half-plane* is defined as the set $\{(x, y) \in \mathbb{R}^2 \mid y > 0\}$ endowed with an abstractly given first fundamental form (or metric) $(g_{ij}) = \frac{1}{y^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Although this metric is not induced by a surface f in $\mathbb{R}^3$, one can nevertheless calculate the Christoffel symbols and the geodesics[15] as quantities of the intrinsic geometry, see Figure 4.9. Hint: The geodesics are the curves with constant x as well as the half-circles whose centers lie on the x-axis. Introduce appropriate polar coordinates.

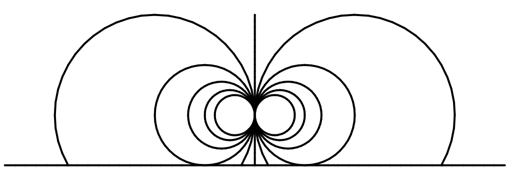


Figure 4.10. Curves of constant geodesic curvature in the Poincaré upper half-plane

12. Calculate the Gaussian curvature of the Poincaré upper half plane along the lines of 4.26 (ii).

[15]These play the role of the "lines" in this non-Euclidean geometry, cf. 3.44.

13. Show that for $z = x + iy \in \mathbb{C}$ all transformations

$$z \mapsto \frac{az+b}{cz+d}, \qquad a,b,c,d \in \mathbb{R},\ ad-bc>0,$$

are *isometries* of the Poincaré upper half-plane, i.e., preserve the abstract first fundamental form g_{ij} above.

14. Let $\lambda(x)$ be a positive differentiable function. For an abstract surface of rotation with metric $ds^2 = dx^2 + \lambda^2(x)dy^2$ ("warped product metric"), calculate the Christoffel symbols and show that the x-lines are geodesics parametrized by arc length. What do the rest of the geodesics look like?

15. Determine all functions λ in Exercise 14 such that the Gaussian curvature of this abstract surface of rotation is -1. Hint: Look at 4.28.

16. Is there a surface element in $\mathbb{R}^3$ with $(g_{ij}(u,v)) = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$ and $(h_{ij}(u,v)) = \left(\begin{smallmatrix}0 & 0\\ 0 & u\end{smallmatrix}\right)$?

17. Is there a surface element in $\mathbb{R}^3$ with $(g_{ij}(u,v)) = \left(\begin{smallmatrix}1 & 0\\ 0 & \cos^2 u\end{smallmatrix}\right)$ and $(h_{ij}(u,v)) = \left(\begin{smallmatrix}1 & 0\\ 0 & \sin^2 u\end{smallmatrix}\right)$?

18. Calculate explicitly the total absolute curvature of a torus of rotation, and compare this with 4.46.

19. Compare the total curvature of a closed space curve with the total absolute curvature of the parallel surface generated by this curve for sufficiently small distance ε. From 4.46, derive an alternative proof of Theorem 2.34 from this[16].

20. Prove the following: a compact 2-dimensional C^4-submanifold of $\mathbb{R}^3$ is necessarily a standard sphere if the equation $\alpha H + \beta K = 0$ is satisfied with two constants $\alpha, \beta \neq 0$.
Hint: By the result of Exercise 10 there is an elliptic point, that is, a point with $\kappa_1 > 0, \kappa_2 > 0$. Exclude points with $\kappa_1 = 0$ or $\kappa_2 = 0$ by considering limits of κ_1/κ_2 respectively κ_2/κ_1 in the part where $\kappa_1 > 0, \kappa_2 > 0$. This implies that we have $\kappa_1 > 0$ and $\kappa_2 > 0$ everywhere. Now use Lemma 4.52.

[16] Cf. K. Voss, *Eine Bemerkung über die Totalkrümmung geschlossener Raumkurven*, Archiv d. Math. **6**, 259–263 (1955).

21. Calculate the total curvature (i.e., the integrated determinant of the Weingarten mapping) for the following hypersurfaces in $\mathbb{R}^4$: consider the parallel set at a distance of ε from a plane circle, as well as the parallel set of a two-dimensional unit sphere. Show in particular that both of these are homeomorphic to the product space $S^1 \times S^2$, but their total curvatures are different. This shows that a direct analog of 4.43 cannot exist in this case: the total curvature is *not* independent of the embedding.

22. For a surface of rotation, 3.16 tells us that $(g_{ij}) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & r^2 \end{smallmatrix}\right)$ and $H = (rh')'/(r^2)'$ if the profile curve is parametrized by arc length. This implies $H = (r\sqrt{1-r'^2})'/(r^2)'$ which is an expression depending only on r and thus only on the coefficients of the first fundamental form. However, this does *not* prove that for surfaces of revolution H is a quantity of the intrinsic geometry because this is not true. Why not? See the remark after 4.16.

23. Prove that the equations of of Gauss and Codazzi-Mainardi in 4.15 are equivalent to the following two equations:
    (a) $R_{ijkl} := \sum_s g_{is} R^s_{jkl} = h_{ik}h_{jl} - h_{il}h_{jk}$,
    (b) $\nabla_i h^j_k = \nabla_k h^j_i$.
    Here $\nabla_i h^j_k$ denotes the jth component of the tangential vector
    $$\left(\nabla_{\frac{\partial f}{\partial u^i}} L\right)\left(\tfrac{\partial f}{\partial u^k}\right) := \nabla_{\frac{\partial f}{\partial u^i}}\left(L\left(\tfrac{\partial f}{\partial u^k}\right)\right) - L\left(\nabla_{\frac{\partial f}{\partial u^i}} \tfrac{\partial f}{\partial u^k}\right)$$
    in local coordinates $u^1, \ldots, u^n$. (Compare the remark in 4.19.) As a consequence we obtain once again the Theorema Egregium in the form $K = \mathrm{Det}(h_{ij})/\mathrm{Det}(g_{ij}) = R_{1212}/\mathrm{Det}(g_{ij})$. (Compare 4.16.)

24. Using the model of the hyperbolic plane H^2 given in 3.44, prove the following by analogy with Exercise 2: The intersection with any ordinary plane in Minkowski 3-space that passes through the origin is a geodesic in H^2. Briefly: *The "lines" in the hyperbolic plane are geodesics.* For the same result in the Poincaré upper half-plane compare Exercise 11.
    Hint: The result of Exercise 5 is also valid in Minkowski space. Therefore, one can apply it to the planar intersection curves in the hyperbolic plane.

Chapter 5

Riemannian Manifolds

In this chapter we want to introduce the notion of an "intrinsic geometry" without making reference to an ambient space $\mathbb{R}^{n+1}$, not only locally, but also as a global notion. This continues the considerations of Chapter 4. The most important tools for this are on the one hand, from a local point of view, a notion of "first fundamental form" independent of an ambient space $\mathbb{R}^{n+1}$ (similar to the notion of intrinsic geometry in the previous chapter), and on the other hand, from a global point of view, the notion of a "manifold". The local notion goes back essentially to the famous lecture of Riemann[1], which explains the modern notions *Riemannian geometry*, *Riemannian manifold* and *Riemannian space*. From the point of view of the development in the book up to now, this is motivated on the one hand by the intrinsic geometry of surfaces, including the Gauss-Bonnet theorem, and on the other hand by the natural occurrence of such spaces which can *not* in any meaningful way be embedded as hypersurfaces in some $\mathbb{R}^n$, as for example the Poincaré upper half-plane as a model of non-Euclidean geometry. Furthermore, the space-times of $3+1$ dimensions which are considered in general relativity do not admit an ambient space in a natural way. This motivates the intention of explaining all geometric quantities in a purely intrinsic manner.

[1]B. Riemann, *Über die Hypothesen, welche der Geometrie zu Grunde liegen*, edited by H. Weyl, Springer, 1921; see also [**7**], Vol. II, Chapter 4. For an English version see also `http://larouchepac.com/node/12479`.

In the previous Chapters 3 and 4 we have basically been considering surface elements $f\colon U \to \mathbb{R}^{n+1}$, where $U \subset \mathbb{R}^n$ was a given open set. From a geometric point of view, we are really more interested in the image set $f(U)$ than we are in the map f itself. Nonetheless, for a description and for local calculations we do use the parameter set U and the parametrization f:

$$U \ni u \overset{f}{\longmapsto} p = f(u) \in f(U).$$

If we decide that the basic object we are considering is the image $f(U)$, then we come to view the inverse mapping

$$f(U) \ni p \overset{f^{-1}}{\longmapsto} u \in U$$

as an image which is "thrown" from $f(U)$, in order to carry out calculations in U. This map is called a "chart" in what follows, which should be thought of as creating a "map" (but the word "map" has a fixed, different meaning in mathematics, so that one uses "chart" instead), and a set of charts which cover the object of interest forms an "atlas", just as a world atlas contains a map containing an arbitrary location on the earth. For the mathematical notion this means that every point has a neighborhood which is contained in one of the charts, in which local computations near that point can be carried out in the corresponding set U. What we have to be able to guarantee is that all defined notions are *independent* of the choice of charts used, just as the Gaussian curvature in the theory of surfaces was independent of the parametrization. In particular, we need to carefully consider the transformations which map us from one chart into a different, nearby one.

5A The notion of a manifold

We have already met submanifolds of $\mathbb{R}^n$ in the form of zero sets of differentiable maps, cf. Chapter 1. If there is no ambient space to begin with, this definition no longer makes any sense. Instead, one uses a description in terms of local coordinates in the form of parametrizations or *charts*, just as one considers maps of the earth to approximate that round object by flat pictures. Note that the chart maps go in the opposite direction from the usual parametrization we have been using up to now.

5.1. Definition. (Abstract differentiable manifold)
A *k-dimensional differentiable manifold* (briefly: a *k-manifold*) is a set M together with a family $(M_i)_{i\epsilon I}$ of subsets such that

1. $M = \bigcup_{i\in I} M_i$ (union),
2. for every $i \in I$ there is an injective map $\varphi_i \colon M_i \to \mathbb{R}^k$ so that $\varphi_i(M_i)$ is open in $\mathbb{R}^k$, and
3. for $M_i \cap M_j \neq \emptyset$, $\varphi_i(M_i \cap M_j)$ is open in $\mathbb{R}^k$, and the composition
$$\varphi_j \circ \varphi_i^{-1} \colon \varphi_i(M_i \cap M_j) \to \varphi_j(M_i \cap M_j)$$
is differentiable for arbitrary i, j.

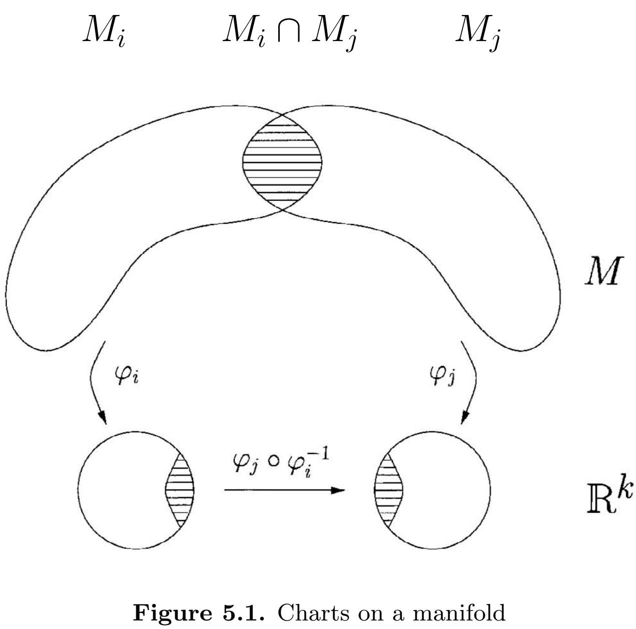


Figure 5.1. Charts on a manifold

Each φ_i is called a *chart*, φ_i^{-1} is referred to as the *parametrization*, the set $\varphi_i(M_i)$ is called the *parameter domain*, and $(M_i, \varphi_i)_{i\in I}$ is called an *atlas*. The maps $\varphi_j \circ \varphi_i^{-1} \colon \varphi_i(M_i \cap M_j) \to \varphi_j(M_i \cap M_j)$, defined on the intersections of two such charts, are called *coordinate transformations* or *transition functions*. Without restriction of generality, we may assume that the atlas is *maximal* with respect to adding more charts satisfying the conditions 2 and 3 above. A maximal atlas in this sense is then referred to as a *differentiable structure.*

EXAMPLES:

1. Every open subset U of $\mathbb{R}^k$ is a k-manifold, where a single chart is sufficient for the entire manifold, namely the inclusion map $\varphi\colon U \to \mathbb{R}^k$. Condition 3 is trivially satisfied in this case.

2. Every k-dimensional submanifold M of $\mathbb{R}^n$ (cf. Chapter 1) is also a k-dimensional manifold in the sense of the above definition. If M is given locally by $M = \{x \in \mathbb{R}^n \mid F(x) = 0\}$, where $F : \mathbb{R}^n \to \mathbb{R}^{n-k}$ is a continuously differentiable submersion (i.e., the differential DF is surjective, or in other words $\mathrm{Rank}(DF) = n - k$), then according to the implicit functions theorem one can locally solve the equation
$$F(x^1, \ldots, x^n) = 0$$
(perhaps after a renumbering) in the explicit form
$$\begin{aligned} x^{k+1} &= x^{k+1}(x^1, \ldots, x^k), \\ &\vdots \\ x^n &= x^n(x^1, \ldots, x^k). \end{aligned}$$
By making the association
$$(x^1, \ldots, x^k) \longmapsto (x^1, \ldots, x^k, x^{k+1}, \ldots, x^n),$$
we get a parametrization, while the association $(x^1, \ldots, x^n) \longmapsto (x^1, \ldots, x^k)$ gives us a chart.

3. The (abstract) torus $\mathbb{R}^2/\mathbb{Z}^2$ is defined as the quotient (group) of these two Abelian groups. To give it a differentiable structure, one defines charts by starting with arbitrary open sets M_i in $\mathbb{R}^2$ (more precisely, take their images in the quotient) which are contained in the open square $(x_0 - \frac{1}{2}, x_0 + \frac{1}{2}) \times (y_0 - \frac{1}{2}, y_0 + \frac{1}{2})$ for an arbitrary point $(x_0, y_0) \in \mathbb{R}^2$. Then set $\varphi(x, y) := (x - x_0, y - y_0)$ to obtain one chart (depending on the choice of (x_0, y_0)). It follows that the coordinate transformations are just translations in $\mathbb{R}^2$. One sees without difficulty that three

of these charts suffice to cover the image, namely the just mentioned squares centered at the points $(0,0)$, $(\frac{1}{3},\frac{1}{3})$, $(\frac{2}{3},\frac{2}{3})$. Two such connected and simply connected sets (each homeomorphic to an open disc) do *not* suffice. For orientable surfaces of higher genus $g \geq 2$ one needs at least four such charts.

Similar results, with appropriate modifications, hold also for the n-dimensional torus $\mathbb{R}^n/\mathbb{Z}^n$. Here one needs $n+1$ charts, each homeomorphic to an open n-ball.

4. The (abstract) *Klein bottle* is a quotient of the two-dimensional torus by the involution $(x,y) \mapsto (x+\frac{1}{2},-y)$. We may take any square in the (x,y)-plane whose length in the x-direction is at most $\frac{1}{2}$ and whose length in the y-direction is at most 1, as charts. As in the case of the torus, one needs three such charts.

5. The real projective plane is the quotient of the two-sphere

$$\mathbb{R}P^2 := S^2/\sim,$$

where the equivalence relation is given by $x \sim -x$. We may take any open set in S^2 as M_i, provided it is contained in a hemisphere (by which we mean half a sphere), and in particular contains no antipodal points. φ can be defined as a projection to a hemisphere, followed by a projection of this onto a disc.
A model of this is the closed disc modulo the identification of the antipodal pairs of points on the boundary. On the other hand, the "classical" model of projective geometry is all of $\mathbb{R}^2$ with an added "line at infinity".
An atlas of the projective plane containing three charts can be constructed as the charts induced by the centrally symmetric atlas on S^2, which consists of the six hemispheres in the three directions (x^1,x^2,x^3).

6. The *rotation group* $\mathbf{SO}(3)$ is defined as the set of all (real) orthogonal (3×3)-matrices with determinant equal to 1. We show that it is a 3-manifold by defining the *Cayley map*

$$CAY\colon \mathbb{R}^3 \to \mathbf{SO}(3), \quad CAY(A) := (\mathbf{1}+A)(\mathbf{1}-A)^{-1}.$$

Here $\mathbf{1}$ denotes the unit matrix, and A denotes the skew-symmetric matrix

$$A = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}$$

with real parameters a, b, c, which can also be viewed as an element of $\mathbb{R}^3$. The Cayley map is injective, and the inverse map can be used as a chart of $\mathbf{SO}(3)$ and determined as follows:

$$CAY(A) = B \Longleftrightarrow B(\mathbf{1} - A) = \mathbf{1} + A$$

$$\Longleftrightarrow (B + \mathbf{1})A = B - \mathbf{1} \Longleftrightarrow A = (B + \mathbf{1})^{-1}(B - \mathbf{1}).$$

Note that $B + \mathbf{1}$ is always invertible, except when -1 is an eigenvalue of B. The matrices B for which this last condition holds are precisely the rotations by π. In fact, the image of the Cayley map is all of $\mathbf{SO}(3)$ with the exception of the set of rotation matrices by a rotational angle of π.
The set of all such rotations by π is naturally bijective to the set of all possible axes of rotation, hence bijective to a projective plane $\mathbb{R}P^2$. To get charts covering this exceptional set of $\mathbf{SO}(3)$, we require three more charts, just as in the above example of an atlas for the projective plane. If we define $\mathbf{E_i}$ as the rotation matrix by an angle of π around the ith axis, and if we formally set $\mathbf{E_0} = \mathbf{1}$, then the following four maps (resp. their inverses) define an atlas of $\mathbf{SO}(3)$:[2]

$$A \mapsto \mathbf{E_i} \cdot CAY(A), \quad i = 0, 1, 2, 3.$$

The four parametrizations of the atlas thus consist of the Cayley maps "centered at" $\mathbf{1}, \mathbf{E_1}, \mathbf{E_2}, \mathbf{E_3}$. The transformations from one chart to another are given by matrix multiplication and are therefore differentiable.

5.2. Definition. (Structures on a manifold)

Given a k-dimensional manifold, one gets *additional structure* by placing additional requirements on the transformation functions $\varphi_j \circ \varphi_i^{-1}$,

[2] I am indebted to Prof. E. Grafarend for a question giving rise to this, which arose from applications in geodesy. Traditionally one considers in geodesy only a single chart for the rotation group, yielding the *Euler angles* or *Cardan angles*.

which belong to the atlas of the manifold; if all $\varphi_j \circ \varphi_i^{-1}$ are (left-hand side), then one speaks of (right-hand side) as follows:

continuous	$\leftrightarrow$	topological manifold
differentiable	$\leftrightarrow$	differentiable manifold
C^1-differentiable	$\leftrightarrow$	C^1-manifold
C^r-differentiable	$\leftrightarrow$	C^r-manifold
C^∞-differentiable	$\leftrightarrow$	C^∞-manifold
real analytic	$\leftrightarrow$	real analytic manifold
complex analytic	$\leftrightarrow$	complex manifold of dimension $\frac{k}{2}$
affine	$\leftrightarrow$	affine manifold
projective	$\leftrightarrow$	projective manifold
conformal	$\leftrightarrow$	manifold with a conformal structure
orientation-preserving	$\leftrightarrow$	orientable manifold

Convention: In what follows we shall understand by the term "manifold" a C^∞-manifold, and "differentiable" will always mean C^∞. One can show that a C^k-atlas always contains a C^∞ one, so that this convention is not a real restriction.

5.3. Definition. (Topology)

A subset $O \subseteq M$ is called *open*, if $\varphi_i(O \cap M_i)$ is open in $\mathbb{R}^k$ for every i. This defines a *topology* on M as the set of all open sets. Then all φ_i are continuous, since the inverse images under them of open sets are again open. M is said to be *compact*, if every open covering contains a finite sub-covering (Heine-Borel covering property). In particular, every compact manifold can be covered with finitely many charts.

Running assumption: In what follows we will always assume that the manifolds which occur satisfy the *Hausdorff separation axiom* (T_2-axiom), formulated as follows. Every two distinct points p, q have disjoint open neighborhoods U_p, U_q. Note that this property does not follow from Definition 5.1.

The important point here is that *locally* (or *in the small*) the topology of a manifold is the same as that of an $I\!R^k$. In particular this means that the inverse images of open ε-balls in $I\!R^k$ are again open in M, although one cannot necessarily make sense of the notion of ε-balls there, as there is no distance function (metric) defined. But this suffices to define the notion of convergence of sequences just as in $I\!R^k$. In addition, the topology of every manifold is locally compact, which means that every point has a compact neighborhood, for example the inverse image of a closed ε-ball in $I\!R^k$.

5.4. Definition. (Differentiable map)
Let M be an m-dimensional differentiable manifold, and let N be an n-dimensional differentiable manifold; furthermore, let $F\colon M \to N$ be a given map. F is said to be *differentiable*, if for all charts $\varphi : U \to I\!R^m, \psi : V \to I\!R^n$ with $F(U) \subset V$ the composition $\psi \circ F \circ \varphi^{-1}$ is also differentiable.

In particular this defines the concept of a *differentiable function* $f\colon M \to I\!R$, where in this case $I\!R$ carries the (identity) standard chart.

This definition is independent of the choice of φ and ψ. A *diffeomorphism* $F : M \to N$ is defined to be a bijective map which is differentiable in both directions. One then calls the two manifolds M and N *diffeomorphic*. Two diffeomorphic manifolds necessarily have the same dimension. This is because for $I\!R^m$ and $I\!R^n$ with $n \neq m$, there is no bijective mapping which is differentiable in both directions, since the corresponding Jacobi matrix is not square and hence cannot have non-vanishing determinant (i.e., cannot be invertible).

Remark: With respect to additional structures on our manifold, one can similarly define when a map is analytic or complex analytic or affine, etc. For example, let us consider here the *Riemann sphere* $\widehat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$. By means of the inclusion $\mathbb{C} \to \widehat{\mathbb{C}}$ one has a chart, and a second is given by $z \mapsto \frac{1}{z}$. These two charts define a *complex structure* on the Riemann sphere, if one adds all compatible charts. Then all meromorphic maps of the Riemann sphere to itself are differentiable maps in the sense of the above definition, for example, also the map $z \mapsto z^{-k}$. Furthermore, this defines a conformal structure on S^2 since every complex-analytic function $f(z)$ with $f' \neq 0$ in one variable z is conformal, cf. Section 3D.

Convention: For a chart φ we will denote by $(u^1, \ldots, u^k)$ the standard coordinates of $\mathbb{R}^k$, and by $(x^1, \ldots, x^k)$ the corresponding coordinates in M. Thus, $x^i(p)$ is the function given by the ith coordinate of $\varphi(p)$, $x^i(p) = u^i(\varphi(p))$. The functions $(u^1, \ldots, u^k)$ as well as $(x^1, \ldots, x^k)$ are thus on the one hand the coordinates of the points considered, while on the other hand $(u^1, \ldots, u^k)$ and $(x^1, \ldots, x^k)$ are also viewed as variables, with respect to which we can form derivatives. For a real-valued function $f : M \to \mathbb{R}$ we set

$$\left.\frac{\partial f}{\partial x^i}\right|_p := \left.\frac{\partial(f \circ \varphi^{-1})}{\partial u^i}\right|_{\varphi(p)}$$

and emphasize this notation by thinking of the partial derivatives as infinitesimal changes of a function in the directions x^i or u^i.

5B The tangent space

Let M be an n-dimensional differentiable manifold and $p \in M$ a fixed point. The tangent space of M at the point p is going to be thought of as the n-dimensional set of “directional vectors”, which – starting at p – point in all directions of M, cf. for example [**27**]. Since there is no ambient space, this notion has to be intrinsically defined and constructed. For this, there are three possible definitions, all of which we describe here.

5.5. Definition. (Tangent vector, tangent space)
Geometric Definition:
A *tangent vector* at p is an equivalence class of differentiable curves $c : (-\varepsilon, \varepsilon) \to M$ with $c(0) = p$, where $c \sim c^* \Leftrightarrow (\varphi \circ c)^{\cdot}(0) = (\varphi \circ c^*)^{\cdot}(0)$ for every chart φ containing p.

Briefly: *tangent vectors are tangents to curves lying on the manifold.*

Unfortunately there is no privileged representative of such an equivalence class, and such a representative would depend on the choice of chart (for example, a line in the parameter domain).

Algebraic Definition:
A *tangent vector* X at p is a derivation (derivative operator) defined on the set of *germs of functions*

$$\mathcal{F}_p(M) := \{f : M \to \mathbb{R} \mid f \text{ differentiable}\}/\sim \ ,$$

where the equivalence relation $\sim$ is defined by declaring $f \sim f^*$ if and only if f and f^* coincide in a neighborhood of p. The value $X(f)$ is also referred to as the *directional derivative* of f in the direction X.

This definition means more precisely the following. X is a map $X : \mathcal{F}_p(M) \to \mathbb{R}$ with the two following properties:

1. $X(\alpha f + \beta g) = \alpha X(f) + \beta X(g), \ \alpha, \beta \in \mathbb{R}, \ f, g \in \mathcal{F}_p(M)$ ($\mathbb{R}$-*linearity*);
2. $X(f \cdot g) = X(f) \cdot g(p) + f(p) \cdot X(g)$ for $f, g \in \mathcal{F}_p(M)$ (*product rule*).
   (For this to make sense, both f and g have to be defined in a neighborhood of p.)

Briefly: *tangent vectors are derivations acting on scalar functions.*

Physical Definition:
A *tangent vector* at the point p is defined as an n-tuple of real numbers $(\xi^i)_{i=1,\dots,n}$ in a coordinate system $x^1, \dots, x^n$ (that is, in a chart), in such a way that in any other coordinate system $\widetilde{x}^1, \dots, \widetilde{x}^n$ (i.e., in any other chart) the same vector is given by a corresponding n-tuple $(\widetilde{\xi^i})_{i=1,\dots,n}$, where

$$\widetilde{\xi^i} = \sum_j \frac{\partial \widetilde{x}^i}{\partial x^j}\bigg|_p \xi^j.$$

Briefly: *tangent vectors are elements of* $\mathbb{R}^n$ *with a particular transformation behavior.*

The *tangent space* T_pM of M at p is defined in all cases as the set of all tangent vectors at the point p. By definition T_pM and T_qM are disjoint if $p \neq q$.

For the special case of an open subset $U \subset \mathbb{R}^n$, the tangent space can be identified with $T_pU := \{p\} \times \mathbb{R}^n$ endowed with the standard basis $(p, e_1), \ldots, (p, e_n)$. The vector e_i corresponds to the curve $c_i(t) := p + t \cdot e_i$ (geometric definition) and to the derivation given by the partial derivative $f \longmapsto \frac{\partial f}{\partial u^i}\big|_p$ (algebraic definition). Therefore 5.5 is compatible with the previous definitions given in 1.7 and 3.1. The directional derivative coincides in $\mathbb{R}^n$ with the directional derivative which was already defined in 4.1.

Special (geometric) tangent vectors are those given by the parameter lines (lines along which parameter values are constant), formally meaning the equivalence classes of them. The corresponding special tangent vectors in the algebraic definition are the partial derivatives $\frac{\partial}{\partial x^i}\big|_p$ defined by

$$\frac{\partial}{\partial x^i}\Big|_p(f) := \frac{\partial f}{\partial x^i}\Big|_p = \frac{\partial(f \circ \varphi^{-1})}{\partial u^i}\Big|_{\varphi(p)}$$

in a chart φ which contains p. As a notational convenience one also writes $\partial_i\big|_p$ instead of $\frac{\partial}{\partial x^i}\big|_p$. The special tangent vectors in the sense of the physical definition are in this case simply the tuples which consist of zeros except in the ith place.

The geometric definition is probably the most intuitive (*a tangent vector is a tangent to a curve*), but not easy to work with. In this definition it is not even clear that the tangent space is a real vector space. The algebraic definition is most convenient for doing computations, and by its very definition it is independent of any chart. The physical definition will be further clarified below. The art of doing computations with the geometric quantities of the physical definition goes back to G. Ricci and is called the *Ricci calculus*, cf. [**16**]. A vector is simply written as ξ^i, and the very fact that the notion involves a superscript indicates the transformation behavior, in this case, for example, as a vector (or 1-contravariant tensor), cf. Section 6.1. This aspect will be of importance in what follows, but for all definitions we will give a coordinate-independent formulation as far as this is feasible. The equivalence of these three definitions is explained for example in [**39**], Chapter 2. In what follows we base our analysis on the algebraic definition and will therefore not require this equivalence.

5.6. Theorem. The (algebraic) tangent space at p on an n-dimensional differentiable manifold is an n-dimensional $\mathbb{R}$-vector space and is spanned in any coordinate system $x^1, \ldots, x^n$ in a given chart by

$$\frac{\partial}{\partial x^1}\Big|_p, \ldots, \frac{\partial}{\partial x^n}\Big|_p.$$

For every tangent vector X at p one has

$$X = \sum_{i=1}^{n} X(x^i) \frac{\partial}{\partial x^i}\Big|_p.$$

Looking at the last equation, we see that the components ξ^i of a tangent vector X in the Ricci calculus are nothing but the $X(x^i)$, that is, the directional derivatives of the coordinate functions x^i in the direction X. To prove the statement of the theorem we require the following lemma.

5.7. Lemma. If X is a tangent vector and f is a constant function, then $X(f) = 0$.

Proof: First suppose $f = 1$ everywhere. Then by the product rule 5.5.2 we have

$$X(1) = X(1 \cdot 1) = X(1) \cdot 1 + 1 \cdot X(1) = 2 \cdot X(1),$$

hence $X(1) = 0$. Now suppose that f has the constant value $f = c$. Then by the linearity 5.5.1 we have

$$X(c) = X(c \cdot 1) = c \cdot X(1) = c \cdot 0 = 0. \qquad \square$$

Proof of 5.6: The proof utilizes an adapted representation of the transition functions in local coordinates. We calculate in a chart $\varphi : U \longrightarrow V$, where without restricting generality we may assume V is an open ε-ball with $\varphi(p) = 0$, hence $x^1(p) = \cdots = x^n(p) = 0$. Let $h : V \to \mathbb{R}$ be a differentiable function, and $f := h \circ \varphi$. We set

$$h_i(y) := \int_0^1 \frac{\partial h}{\partial u^i}(t \cdot y)dt \quad (\text{note: } h \in C^\infty \Rightarrow h_i \in C^\infty)$$

and perform the following computation:

$$\sum_{i=1}^{n} \frac{\partial h}{\partial u^i}(t \cdot y) \cdot \underbrace{\frac{d(tu^i)}{dt}}_{=u^i} = \frac{\partial h}{\partial t}(t \cdot y),$$

which implies

$$\sum_{i=1}^{n} h_i(y) \cdot u^i = \int_0^1 \frac{\partial h}{\partial t}(t \cdot y)dt = h(y) - h(0).$$

From this we get, using the identities $f = h \circ \varphi$, $f_i = h_i \circ \varphi$, $x^i = u^i \circ \varphi$, the equation

$$f(q) - f(p) = \sum_{i=1}^{n} f_i(q) \cdot x^i(q)$$

for a variable point q. Taking derivatives, we get

$$\frac{\partial f}{\partial x^i}\Big|_p = f_i(p).$$

Now if we are given a tangent vector X at p, then it follows from properties 1 and 2 in 5.5 that

$$X(f) = X\Big(f(p) + \sum_{i=1}^{n} f_i x^i\Big) = 0 + \sum_{i=1}^{n} X(f_i) \cdot \underbrace{x^i(p)}_{=0} + \sum_{i=1}^{n} f_i(p) \cdot X(x^i)$$

$$= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_p \cdot X(x^i) = \Big(\sum_{i=1}^{n} X(x^i) \cdot \frac{\partial}{\partial x^i}\Big|_p\Big)(f)$$

for every f. It remains to show that the vectors $\frac{\partial}{\partial x^i}\big|_p$ are linearly independent. But this is easy to see, since $\frac{\partial}{\partial x^i}\big|_p(x^j) = \frac{\partial x^j}{\partial x^i} = \delta_i^j$.

Note that this proof only works for C^∞-manifolds, as otherwise the degree of differentiability of h_i is one less that of h. In fact, the algebraic tangent space on a C^k-manifold is infinite-dimensional. But there are no difficulties in simply passing to the subspace spanned by $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$ and performing the same calculations there. □

5.8. Definition and Lemma. (Derivative, chain rule)
Let $F : M \to N$ be a differentiable map, and let p, q be two fixed points with $F(p) = q$. Then the *derivative* or the *differential* of F at p is defined as the map

$$DF|_p : T_pM \longrightarrow T_qN$$

whose value at $X \in T_pM$ is given by $(DF|_p(X))(f) := X(f \circ F)$ for every $f \in \mathcal{F}_q(N)$ (which automatically implies the relation $f \circ F \in \mathcal{F}_p(M)$). For the derivative as defined in this manner, one has the *chain rule* in the form

$$D(G \circ F)|_p = DG|_{F(p)} \circ DF|_p$$

for every composition $M \xrightarrow{F} N \xrightarrow{G} Q$ of maps, or, more briefly, $D(G \circ F) = DG \circ DF$.

PROOF: By definition we have

$$D(G \circ F)|_p(X)(f) = X(f \circ G \circ F)$$

$$= (DF|_p(X))(f \circ G) = \Big(DG|_q\big(DF|_p(X)\big)\Big)(f). \qquad \square$$

REMARK: One can view $DF|_p$ as a linear approximation of F at p, just as in vector analysis on $\mathbb{R}^n$. In coordinates $x^1, \ldots, x^m$ on M and $y^1, \ldots, y^n$ on N, $DF|_p$ is represented by the Jacobi matrix, for which we have the more precise relation

$$DF|_p\left(\frac{\partial}{\partial x^j}\Big|_p\right) = \sum_i \frac{\partial(y^i \circ F)}{\partial x^j}\Big|_p \frac{\partial}{\partial y^i}\Big|_q.$$

In the physical definition of tangent spaces, the chain rule consists essentially of the product of the Jacobi matrices, applied to the tangent vector. In the geometric definition of the tangent space (i.e., for equivalence classes of curves through the point p), the differential is simply described by the transport of curves, as follows:

$$DF|_p([\ c\ ]) := [\ F \circ c\ ],$$

and the chain rule $DG\big(DF([c])\big) = [G \circ F \circ c]$ is then quite obvious. Note the action on the tangent of a curve:

$$\dot{c}(0) \mapsto (F \circ c)^{\cdot}(0) = DF|_p(\dot{c}(0)).$$

EXAMPLES:

(i) In case $F : U \to \mathbb{R}^{n+1}$ $(U \subset \mathbb{R}^n)$ is a surface element in the sense of Chapter 3 with $u \mapsto F(u) = p$, then the differential of F acts in the following way on the basis $\frac{\partial}{\partial u^1}\big|_u, \ldots, \frac{\partial}{\partial u^n}\big|_u$ of $T_u U$ resp. $\frac{\partial}{\partial x^1}\big|_p, \ldots, \frac{\partial}{\partial x^{n+1}}\big|_p$ of $T_p\mathbb{R}^{n+1}$:

$$DF\big|_u\Big(\frac{\partial}{\partial u^j}\Big|_u\Big) = \sum_i \frac{\partial x^i}{\partial u^j}\Big|_u \cdot \frac{\partial}{\partial x^i}\Big|_p,$$

where the matrix $\frac{\partial x^i}{\partial u^j}$ is the familiar *Jacobi matrix* of the mapping F. Here, x^i is the ith component of $F(u^1, \ldots, u^n)$, also written as the function $x^i(u^1, \ldots, u^n)$.

(ii) If $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$ are two coordinate systems on a single manifold, then one has similarly, for F equal to the identity,

$$\frac{\partial}{\partial x^j} = \sum_i \frac{\partial y^i}{\partial x^j}\frac{\partial}{\partial y^i}.$$

(iii) For the components ξ^i and η^j, respectively, of a tangent vector $X = \sum_j \xi^j \frac{\partial}{\partial x^j} = \sum_i \eta^i \frac{\partial}{\partial y^i}$, one has similarly $X = \sum_j \xi^j \frac{\partial}{\partial x^j} = \sum_{i,j} \xi^j \frac{\partial y^i}{\partial x^j}\frac{\partial}{\partial y^i}$; hence $\eta^i = \sum_j \xi^j \frac{\partial y^i}{\partial x^j}$. This is precisely the transformation behavior of tangent vectors in Ricci calculus (Definition 5.5).

The following *summation convention* is used in Ricci calculus, and is usually referred to as the Einstein summation convention: sums are formed over indices which occur in formulas as both an upper (in the numerator) and a lower (in the denominator) subscript, without the explicit summation symbol, for example

$$h_{ik} = h_i^j g_{jk} \quad \text{instead of} \quad h_{ik} = \sum_j h_i^j g_{jk} \text{ and}$$

$$\eta^i = \xi^j \frac{\partial y^i}{\partial x^j} \quad \text{instead of} \quad \eta^i = \sum_j \xi^j \frac{\partial y^i}{\partial x^j}.$$

5.9. Definition. (Vector field)
A differentiable *vector field* X on a differentiable manifold is an association $M \ni p \longmapsto X_p \in T_pM$ such that in every chart $\varphi\colon U \to V$ with coordinates $x^1, \dots, x^n$, the coefficients $\xi^i\colon U \to \mathbb{R}$ in the representation (valid at a point)

$$X_p = \sum_{i=1}^{n} \xi^i(p) \frac{\partial}{\partial x^i}\bigg|_p$$

are differentiable functions.

Another common notation for this is $X = \sum_i \xi^i \frac{\partial}{\partial x^i}$ or, in Ricci calculus, $X = \xi^i$. Note that in the physical definition, a vector field is identified with the n-tuple $(\xi^1, \dots, \xi^n)$ of functions of the coordinates $x^1, \dots, x^n$.

As to the notations used in conjunction with vector fields, for a scalar function $f : M \to \mathbb{R}$, the symbol fX denotes the vector field $(fX)_p := f(p) \cdot X_p$ (one can say that *the set of vector fields is a module over the ring of functions f on M*), while the symbol $Xf = X(f)$ denotes the function $(Xf)(p) := X_p(f)$ (in other words, Xf is the derivative of f in the direction of X).

5C Riemannian metrics

The first fundamental form of a surface element is a scalar product, which is defined by restricting the Euclidean scalar product to each tangent space, as we have explained in Chapter 3. In our present endeavor, we have to find a way to do this without the ambient space, that is, defining (intrinsically) a scalar product on each tangent space. Recall the following fact from linear algebra, which we will require in this regard.

The space $L^2(T_pM; \mathbb{R}) = \{\alpha\colon T_pM \times T_pM \to \mathbb{R} \mid \alpha \text{ bilinear}\}$ has the basis

$$\{dx^i|_p \otimes dx^j|_p \mid i, j = 1, \dots, n\},$$

where the dx^i form the *dual basis* in the dual space

$$(T_pM)^* = L(T_pM; \mathbb{R}),$$

defined as follows:

$$dx^i\big|_p\Big(\frac{\partial}{\partial x^j}\Big|_p\Big) = \delta^i_j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The bilinear forms $dx^i|_p \otimes dx^j|_p$ are defined in terms of their action on the basis (this action being then extended by linearity):

$$(dx^i|_p \otimes dx^j|_p)\Big(\frac{\partial}{\partial x^k}\Big|_p, \frac{\partial}{\partial x^l}\Big|_p\Big) := \delta^i_k \delta^j_l = \begin{cases} 1 & \text{if } i = k \text{ and } j = l, \\ 0 & \text{otherwise.} \end{cases}$$

By inserting the basis, for the coefficients of the representation

$$\alpha = \sum_{i,j} \alpha_{ij} \cdot dx^i \otimes dx^j$$

one obtains the expression

$$\alpha_{ij} = \alpha\Big(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\Big).$$

In Ricci calculus, the form α is just represented by the symbol α_{ij}; one also refers to this as a *tensor of degree two*, cf. 6.1.

5.10. Definition. (Riemannian metric)
A *Riemannian metric* g on M is an association $p \longmapsto g_p \in L^2(T_pM; I\!R)$ such that the following conditions are satisfied:

1. $g_p(X, Y) = g_p(Y, X)$ for all X, Y, (*symmetry*)
2. $g_p(X, X) > 0$ for all $X \neq 0$, (*positive definiteness*)
3. The coefficients g_{ij} in every local representation (i.e., in every chart)
$$g_p = \sum_{i,j} g_{ij}(p) \cdot dx^i|_p \otimes dx^j|_p$$
are differentiable functions. (*differentiability*)

The pair (M, g) is then called a *Riemannian manifold.* One also refers to the Riemannian metric as the *metric tensor.* In local coordinates the metric tensor is given by the matrix (g_{ij}) of functions. In Ricci calculus this is simply written as g_{ij}.

REMARKS:

1. A Riemannian metric g defines at every point p an *inner product* g_p on the tangent space T_pM, and therefore the notation $\langle X, Y\rangle$ instead of $g_p(X,Y)$ is also used. The notions of angles and lengths are determined by this inner product, just as these notions are determined by the first fundamental form on surface elements. The length or norm of a vector X is given by $||X|| := \sqrt{g(X,X)}$, and the angle β between two tangent vectors X and Y can be defined by the validity of the equation $\cos\beta \cdot ||X|| \cdot ||Y|| = g(X,Y)$, cf. Chapter 1.

2. If the condition that g is positive definite is replaced by the weaker condition that it is *non-degenerate* (meaning that $g(X,Y) = 0$ for all Y implies $X = 0$), then one arrives at the notion of a *pseudo-Riemannian metric* or *semi-Riemannian metric*, in which all notions are defined in exactly the same way as for a Riemannian metric. In particular, a so-called *Lorentzian metric* is defined as one for which the signature of g is $(-,+,+,+)$; such metrics are basic to the general theory of relativity. In this case the tangent spaces are modeled after Minkowski space $\mathbb{R}^4_1$ instead of Euclidean space (cf. Section 3E) with the metric

$$(g_{ij}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The difference compared with Euclidean space is that there are vectors $X \neq 0$ with $g(X,X) = 0$, so-called *null vectors.* We have already studied the three-dimensional Minkowski space in connection with curves and surfaces (compare sections 2E and 3E). The tensor g_{ij} is referred to in the theory of relativity as the *gravitational potential* or *gravitational field*, cf. [**26**], Section 1.3. It gives a metric form to the manifold (four-dimensional space-time) according to the gravity coming from the matter which is contained in the space.

Examples:

(i) The first fundamental form g of a hypersurface element in $\mathbb{R}^{n+1}$ is an example of a Riemannian metric.

(ii) The standard example is $(M,g) = (\mathbb{R}^n, g_0)$, where the metric $(g_0)_{ij} = \delta_{ij}$ (identity matrix) is the Euclidean metric in the standard chart of $\mathbb{R}^n$ (given by Cartesian coordinates). This space is also referred to as *Euclidean space* and denoted by $\mathbb{E}^n$. The metric is

$$(g_0)_{ij} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

so that $g_0(.,.) = \langle \cdot, \cdot \rangle$ is, not unexpectedly, nothing but the *Euclidean inner product.*

(iii) A different Riemannian metric on $\mathbb{R}^n$ is given for example by $g_{ij}(x_1, \dots, x_n) := \delta_{ij}(1 + x_i x_j)$:

$$(g_{ij}) = \begin{pmatrix} 1+x_1^2 & 0 & \dots & 0 \\ 0 & 1+x_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1+x_n^2 \end{pmatrix}$$

Similarly, one can define numerous Riemannian metrics simply by choosing the coefficients g_{ij} arbitrarily, provided only that one has positive definiteness or non-degeneracy of the metric.

(iv) After choosing constants $0 < b < a$, on $(0, 2\pi) \times (0, 2\pi) \subset \mathbb{R}^2$, $0 < r < 1$, one can define a Riemannian metric by

$$(g_{ij}(u, v)) = \begin{pmatrix} b^2 & 0 \\ 0 & (a + b \ \cos u)^2 \end{pmatrix}.$$

This coincides with the first fundamental form on an open subset of the torus of revolution (cf. Chapter 3).

(v) We can give the abstract torus $\mathbb{R}^2/\mathbb{Z}^2$ a uniquely defined Riemannian metric g with the property that the projection

$$(\mathbb{R}^2, g_0) \longrightarrow (\mathbb{R}^2/\mathbb{Z}^2, g)$$

is a local isometry in the sense of 5.11. This is called the *flat torus.* In the chart $(0, 1) \times (0, 1)$ the metric is $(g_{ij}) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, as in the Euclidean plane.

(vi) Similarly, the real projective plane $\mathbb{R}P^2 = S^2/\pm$ can be given a unique Riemannian metric g such that the projection $(S^2, g_1) \to (\mathbb{R}P^2, g)$ is a local isometry in the sense of 5.11, where g_1 is the standard metric on the unit sphere.

(vii) The Poincaré upper half-plane $\{(x,y) \in \mathbb{R}^2 \mid y > 0\}$ with the metric

$$(g_{ij}(x,y)) := \frac{1}{y^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is a Riemannian manifold. In this metric, length is given by $||\frac{\partial}{\partial y}|| = \frac{1}{y}$; thus the half-lines in the y-direction have infinite length: $\int_\eta^1 \frac{1}{t} dt = -\log(\eta) \longrightarrow \infty$ for $\eta \to 0$ and $\int_1^\eta \frac{1}{t} dt = \log(\eta) \longrightarrow \infty$ for $\eta \to \infty$. In fact, every geodesic is of infinite length in both directions. We refer also to the Exercises at the end of Chapter 4 as well as Section 7A for more details.

5.11. Definition. (Maps which are compatible with the metric) A differentiable map $F\colon M \longrightarrow \widetilde{M}$ between two Riemannian manifolds $(M,g), (\widetilde{M}, \widetilde{g})$ is called a *(local) isometry*, if for all points p and tangent vectors X, Y we have

$$\widetilde{g}_{F(p)}(DF|_p(X), DF|_p(Y)) = g_p(X,Y);$$

(M,g) and $(\widetilde{M}, \widetilde{g})$ are called *(locally) isometric* in this case. More generally, F is called a *conformal mapping*, if there is a function $\lambda\colon M \to \mathbb{R}$ without zeros, such that for all p, X, Y, one has

$$\widetilde{g}_{F(p)}(DF_p(X), DF_p(Y)) = \lambda^2(p) g_p(X,Y).$$

See also Definitions 3.29 and 4.29.

By definition a local isometry preserves the length of a vector, angles, and areas and volumes, whereas a conformal mapping preserves angles but rescales the length of any vector by the factor λ.

Examples: The map $(x,y) \mapsto (\cos x, \sin x, y)$ is a local isometry of the plane onto a cylinder. Stereographic projection defines a conformal map between the plane and the punctured sphere.

QUESTION: Does there exist a Riemannian metric on an arbitrary manifold M? Locally there is no problem in constructing one, as we choose any (g_{ij}) which is both positive definite and symmetric. To make this construction global, one can use the method of a *partition of unity*. To introduce this notion, we define the following

NOTATION: For a given function $f\colon M \to \mathbb{R}$, the topological closure

$$supp(f) := \overline{\{x \in M \mid f(x) \neq 0\}}$$

is called the *support of f*.

5.12. Definition and Lemma. (Partition of unity)

A differentiable *partition of unity* on a differentiable manifold M is a family $(f_i)_{i\in I}$ of differentiable functions $f_i\colon M \to \mathbb{R}$ such that the following conditions are satisfied:

1. $0 \leq f_i \leq 1$ for all $i \in I$,
2. every point $p \in M$ has a neighborhood which intersects only finitely many of the $supp(f_i)$, and
3. $\sum_{i\in I} f_i \equiv 1$ (locally this is always to be a finite sum).

If there is a partition of unity on M such that the support $supp(f_i)$ of each function is contained in a coordinate neighborhood, then there exists a Riemannian metric on M.

PROOF: For each $i \in I$ choose $g_{kl}^{(i)}$ as a symmetric, positive definite matrix-valued function (in the chart associated with $supp(f_i)$). This locally defines a Riemannian metric $g^{(i)}$, and $f_i \cdot g^{(i)}$ is differentiable and well-defined on all of M, namely, it vanishes identically outside of $supp(f_i)$. Then we set

$$g := \sum_{i\in I} f_i \cdot g^{(i)}.$$

It follows that g is symmetric and positive semi-definite because $f_i \geq 0$ and $g^{(i)} > 0$, and from $\sum_i f_i \equiv 1$ we see that g is even positive definite at every point. $\square$

WARNING: The same method does *not* show the existence of an indefinite metric $\widetilde{g}$ on M, because in this case $\tilde{g}$ can degenerate, even if all $\widetilde{g}^{(i)}$ are non-degenerate. In fact, there are topological obstructions to the existence of indefinite metrics. For example there is a Lorentz metric of type $(- + + \cdots +)$ on a compact manifold if and only if the Euler characteristic satisfies $\chi = 0$. This is because precisely in this case, a line element field exists[3]. Among the compact surfaces, only the torus and the Klein bottle satisfy this condition.

We mention the following result without proof.

Theorem: If the topology of M (i.e., the system of open sets, cf. 5.3) is locally compact (which always holds for manifolds) and the second countability axiom is satisfied (there exists a countable basis for the topology), then there exists in every open covering an associated partition of unity, in the sense that $supp(f_i)$ is always contained in one of the given open sets.

For a proof, see for example [**40**]. In fact it is sufficient to make the (weaker) assumption that the space is paracompact.

Under the same assumptions there exists a Riemannian metric. In particular, the compactness of M implies the topological assumptions required. Thus, on an arbitrary compact manifold there exists a Riemannian metric.

5D The Riemannian connection

Just as at the beginning of Chapter 4, we have here the problem of defining the derivative on an abstract differentiable manifold or abstract Riemannian manifold not only for scalar functions (this is sufficiently done in the algebraic Definition 5.5), but also for vector fields. What we have to define is the notion of the derivative of a (tangent) vector field with respect to a tangent vector, with a result which is again a tangent vector. This will be defined in 5.13 in such a way that a Riemannian metric is not necessary and both arguments X and Y are treated equally. The so-called *Riemannian connection*, defined in 5.15, is nearer to the notion of covariant derivative of Chapter 4; in fact, it is just a generalization. Here we also require a compatibility with the Riemannian metric. The fundamental lemma

[3] L. Markus, *Line element fields and Lorentz structures on differentiable manifolds*, Annals of Mathematics (2) **62**, 411–417 (1955).

of Riemannian geometry, presented in 5.16, shows the existence of a unique Riemannian connection for an arbitrary Riemannian metric.

5.13. Definition. (The Lie bracket[4])
Let X, Y be (differentiable) vector fields on M, and let $f\colon M \to I\!R$ be a differentiable function. Through the relation

$$[X, Y](f) := X(Y(f)) - Y(X(f))$$

we define a vector field $[X, Y]$, which is referred to as the *Lie bracket* of X, Y (also called the *Lie derivative* $\mathcal{L}_X Y$ of Y in the direction X). At a point $p \in M$ we have $[X, Y]_p(f) = X_p(Yf) - Y_p(Xf)$.

The Lie bracket measures the degree of non-commutativity of the derivatives. In Section 4.5 above we made a similar definition, namely $[X, Y] := D_X Y - D_Y X$, which in $I\!R^n$ is equivalent to the above definition. For the definition of the Lie bracket, no Riemannian metric is required; the differentiable structure is sufficient. The exercises at the end of the chapter help give a geometric interpretation and intuition of the Lie bracket. For scalar functions φ one simply sets $\mathcal{L}_X \varphi = X(\varphi)$ and declares in this way a Lie derivative for scalar functions and for vectors. There is also a Lie derivative in the direction of a vector field defined for one-forms given by the formula $\mathcal{L}_X \omega(Y) := X(\omega(Y)) - \omega(\mathcal{L}_X Y)$. On can similarly define a Lie derivative for tensor fields in general, see [**44**, 2.24]. If the Lie derivative vanishes in the direction of a vector field, this leads naturally to a corresponding notion of "constancy". An example is an isometric vector field X (also called a *Killing field*) on (M, g) characterized by the equation $\mathcal{L}_X g = 0$. Here we have $\mathcal{L}_X g(Y, Z) = g(\nabla_Y X, Z) + g(Y, \nabla_Z X)$.

5.14. Lemma. (Properties of the Lie bracket)

Let X, Y, Z be vector fields, let α, β be real constants, and let $f, h\colon M \to I\!R$ be differentiable functions. Then the Lie bracket has the following properties:

(i) $[\alpha X + \beta Y, Z] = \alpha[X, Z] + \beta[Y, Z]$;

(ii) $[X, Y] = -[Y, X]$;

(iii) $[fX, hY] = f \cdot h \cdot [X, Y] + f \cdot (Xh) \cdot Y - h \cdot (Yf) \cdot X$;

[4]Named after Sophus Lie, the founder of the theory of transformation groups.

(iv) $\big[X,[Y,Z]\big] + \big[Y,[Z,X]\big] + \big[Z,[X,Y]\big] = 0;$ (*Jacobi identity*)

(v) $\left[\dfrac{\partial}{\partial x^i}, \dfrac{\partial}{\partial x^j}\right] = 0$ for every chart with coordinates $(x^1, \ldots, x^n)$;

(vi) $\left[\sum_i \xi^i \dfrac{\partial}{\partial x^i}, \sum_j \eta^j \dfrac{\partial}{\partial x^j}\right] = \sum_{i,j} \left(\xi^i \dfrac{\partial \eta^j}{\partial x^i} - \eta^i \dfrac{\partial \xi^j}{\partial x^i}\right) \dfrac{\partial}{\partial x^j}$ (*representation in local coordinates*).

Proof: The properties (i) and (ii) are obvious. (iii) follows from the product rule 5.5:

$$[fX, hY](\phi) = fX((hY)\phi) - hY((fX)\phi)$$
$$= f(Xh)(Y\phi) + fhX(Y\phi) - h(Yf)(X\phi) - hfY(X\phi)$$
$$= \Big(fh[X,Y] + f(Xh)Y - h(Yf)X\Big)(\phi)$$

for every function ϕ in a neighborhood of the point under consideration.

(v) is nothing but the well-known Schwarz' law

$$\frac{\partial}{\partial x^i}\Big(\frac{\partial}{\partial x^j}(f)\Big) = \frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial}{\partial x^j}\Big(\frac{\partial}{\partial x^i}(f)\Big)$$

for the commutativity of the second derivatives.

The representation in (vi) has already occurred in Section 4.5, and is proved here in an entirely similar manner.

The Jacobi identity (iv) is easily checked as follows, where we symbolically write $[X,Y] = XY - YX$:

$$\big[X,[Y,Z]\big] + \big[Y,[Z,X]\big] + \big[Z,[X,Y]\big]$$
$$= XYZ - XZY - YZX + ZYX + YZX - YXZ$$
$$-ZXY + XZY + ZXY - ZYX - XYZ + YXZ = 0. \qquad \square$$

5.15. Definition. (Riemannian connection)
A *Riemannian connection* ∇ (pronounced "nabla") on a Riemannian manifold (M, g) is a map

$$(X, Y) \longmapsto \nabla_X Y,$$

which associates to two given differentiable vector fields X, Y a third differentiable vector field $\nabla_X Y$, such that the following conditions are satisfied: ($f\colon M \to I\!R$ denotes a differentiable function):

(i) $\nabla_{X_1+X_2} Y = \nabla_{X_1} Y + \nabla_{X_2} Y$; (*additivity in the subscript*)
(ii) $\nabla_{fX} Y = f \cdot \nabla_X Y$; (*linearity in the subscript*)
(iii) $\nabla_X (Y_1 + Y_2) = \nabla_X Y_1 + \nabla_X Y_2$; (*additivity in the argument*)
(iv) $\nabla_X (fY) = f \cdot \nabla_X Y + (X(f)) \cdot Y$; (*product rule in the argument*)
(v) $X\big(g(Y, Z)\big) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$; (*compatibility with the metric*)
(vi) $\nabla_X Y - \nabla_Y X - [X, Y] = 0$. (*symmetry* or *torsion-freeness*)

Remark: For simplicity one often uses the notation $\nabla_X f = X(f)$ for the directional derivative of f in the direction X. Dropping the conditions (v) and (vi) defines a plain "connection", and if the condition (vi) is not satisfied, the difference $T(X, Y) := \nabla_X Y - \nabla_Y X - [X, Y]$ is called the *torsion tensor* of ∇. Instead of "connection" one also speaks of a *covariant derivative* (cf. 4.3), and instead of "Riemannian connection", the term *Levi–Civita connection.*

The meaning of the term lies in a kind of "connection" between the different tangent spaces, which are disjoint by definition. This will occur again in sections 5.17 and 5.18, where the notion of parallel displacement (or parallel transport) of vectors in introduced. In this way it is possible to relate tangent vectors which are based at different points of the manifold. The properties for calculations with the Riemannian connection are identical to those of the covariant derivative in Section 4.4.

EXAMPLES:

1. In Euclidean space $(\mathbb{R}^n, g_o)$ with the standard metric g_0, we can set $\nabla = D$, which means that the directional derivative is a Riemannian connection, cf. the properties mentioned in Chapter 4.
2. On a hypersurface $M^n \to \mathbb{R}^{n+1}$, the covariant derivative in the sense of Definition 4.3 defines a Riemannian connection for the first fundamental form in the above sense.
3. In $\mathbb{R}^3$ set $\nabla_X Y := D_X Y + \frac{1}{2}(X \times Y)$, where $X \times Y$ is the usual cross product of vectors. This ∇ satisfies (i) - (v), but not (vi):

$$\nabla_X Y - \nabla_Y X = D_X Y - D_Y X + X \times Y = [X, Y] + \underbrace{X \times Y}_{\text{torsion}}$$

5.16. Theorem. On every Riemannian manifold (M, g) there is a uniquely determined Riemannian connection ∇.

PROOF: First we prove the *uniqueness*. From properties (i) - (vi) we get, for vector fields X, Y, Z, a relation as the sum of three equalities:

$$\left.\begin{array}{rcl} X\langle Y, Z\rangle & = & \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle \\ Y\langle X, Z\rangle & = & \langle \nabla_Y X, Z\rangle + \langle X, \nabla_Y Z\rangle \\ -Z\langle X, Y\rangle & = & -\langle \nabla_Z X, Y\rangle - \langle X, \nabla_Z Y\rangle \end{array}\right\} +$$

$$X\langle Y, Z\rangle + Y\langle X, Z\rangle - Z\langle X, Y\rangle = \langle Y, \underbrace{\nabla_X Z - \nabla_Z X}_{[X,Z]}\rangle + \langle X, \underbrace{\nabla_Y Z - \nabla_Z Y}_{[Y,Z]}\rangle$$
$$+\langle Z, \underbrace{\nabla_X Y + \nabla_Y X}_{2\nabla_X Y + [Y,X]}\rangle$$

From this we get the *Koszul formula*

$$(*) \qquad 2\langle \nabla_X Y, Z\rangle = X\langle Y, Z\rangle + Y\langle X, Z\rangle - Z\langle X, Y\rangle - \langle Y, [X, Z]\rangle - \langle X, [Y, Z]\rangle - \langle Z, [Y, X]\rangle.$$

The right-hand side is uniquely determined, given Z; hence also $\nabla_X Y$ is uniquely determined.

To show the *existence* of ∇ we define ∇ by the requirement that $(*)$ holds for all X, Y, Z.

It remains to show that $(\nabla_X Y)\big|_p$ is defined (without using Z as a vector field), in other words, the expression $\langle \nabla_X Y\big|_p, Z_p\rangle$ depends only on Z_p, or equivalently,

$$\langle \nabla_X Y, f\cdot Z\rangle = f\cdot\langle \nabla_X Y, Z\rangle$$

for every scalar function f. This is easily verified by applying the properties of the Lie bracket and the product rule

$$X(fh) = f\cdot(Xh) + (Xf)\cdot h.$$

The validity of (i) - (vi) for the ∇ defined in this manner has to be established.

(i) and (iii) are obvious.

(ii) By applying the formula $(*)$ we get

$$\begin{aligned}&2\langle \nabla_{fX} Y, Z\rangle - 2\langle f\nabla_X Y, Z\rangle\\ &\quad = (Yf)\langle X, Z\rangle - (Zf)\langle X, Y\rangle - \langle Y, -(Zf)X\rangle - \langle Z, (Yf)X\rangle = 0.\end{aligned}$$

The proof of (iv) is similar.

(v) We have $2\langle \nabla_X Y, Z\rangle + 2\langle Y, \nabla_X Z\rangle$

$$\begin{aligned}&= X\langle Y,Z\rangle + Y\langle X,Z\rangle - Z\langle X,Y\rangle - \langle Y,[X,Z]\rangle - \langle X,[Y,Z]\rangle - \langle Z,[Y,X]\rangle\\ &\quad + X\langle Z,Y\rangle + Z\langle X,Y\rangle - Y\langle X,Z\rangle - \langle Z,[X,Y]\rangle - \langle X,[Z,Y]\rangle - \langle Y,[Z,X]\rangle\\ &= X\langle Y,Z\rangle + X\langle Z,Y\rangle = 2X\langle Y,Z\rangle\end{aligned}$$

(vi) We have $2\langle \nabla_X Y - \nabla_Y X, Z\rangle$

$$\begin{aligned}&= X\langle Y,Z\rangle + Y\langle X,Z\rangle - Z\langle X,Y\rangle - \langle Y,[X,Z]\rangle - \langle X,[Y,Z]\rangle - \langle Z,[Y,X]\rangle\\ &\quad - Y\langle X,Z\rangle - X\langle Y,Z\rangle + Z\langle Y,X\rangle + \langle X,[Y,Z]\rangle + \langle Y,[X,Z]\rangle + \langle Z,[X,Y]\rangle\\ &= 2\langle [X,Y], Z\rangle.\end{aligned}$$

□

In *local coordinates* we get with the same formula the expression which we already met in Section 4.6 for the *Christoffel symbols*

$$\Gamma_{ij,k} = \frac{1}{2}\Big(-\frac{\partial}{\partial_k} g_{ij} + \frac{\partial}{\partial_j} g_{ik} + \frac{\partial}{\partial_i} g_{jk}\Big),\quad \Gamma_{ij}^m = \sum_k \Gamma_{ij,k} g^{km},$$

where $(g^{km}) := (g_{rs})^{-1}$, and

$$\Big\langle \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}\, ,\, \frac{\partial}{\partial x^k}\Big\rangle = \Gamma_{ij,k},\quad \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \sum_k \Gamma_{ij}^k \frac{\partial}{\partial x^k}.$$

From this we get the following expression for $\nabla_X Y$, in local coordinates, provided $X = \sum_i \xi^i \frac{\partial}{\partial x^i}$ and $Y = \sum_j \eta^j \frac{\partial}{\partial x^j}$:

$$\nabla_X Y = \nabla_{\sum_i \xi^i \frac{\partial}{\partial x^i}} \Big(\sum_j \eta^j \frac{\partial}{\partial x^j} \Big)$$

$$= \sum_k \Big(\sum_i \xi^i \frac{\partial \eta^k}{\partial x^i} + \sum_{i,j} \Gamma^k_{ij} \xi^i \eta^j \Big) \frac{\partial}{\partial x^k}.$$

Especially for $X = \frac{\partial}{\partial x^i}$ we obtain

$$\nabla_X Y = \nabla_{\frac{\partial}{\partial x^i}} \Big(\sum_j \eta^j \frac{\partial}{\partial x^j} \Big) = \sum_k \Big(\frac{\partial \eta^k}{\partial x^i} + \sum_j \Gamma^k_{ij} \eta^j \Big) \frac{\partial}{\partial x^k}.$$

Consequently, in Ricci calculus the notation for this formula is

$$\nabla_i \eta^k = \frac{\partial \eta^k}{\partial x^i} + \Gamma^k_{ij} \eta^j.$$

In this expression, the left-hand side is not to be interpreted as the derivative of a scalar function η^k, but as the kth component of the derivative of the vector $(\eta^1, \dots, \eta^n)$ with respect to the ith variable.

If we consider, instead of vector fields on the manifold itself, vector fields along a curve c, then the coordinate functions η^i are not to be viewed as functions of $x^1, \dots, x^n$, but rather as functions of the curve parameter t. In this case, the following equation may be taken as a definition, where $c^1(t), \dots, c^n(t)$ are the coordinates of c:

$$\nabla_{\dot c} Y = \sum_k \Big(\frac{d\eta^k(t)}{dt} + \sum_{i,j} \dot c^i(t) \eta^j(t) \Gamma^k_{ij}(c(t)) \frac{\partial}{\partial x^k} \Big)$$

$$= \sum_k \Big(\sum_i \dot c^i(t) \frac{\partial \eta^k(t)}{\partial x^i} + \sum_{i,j} \dot c^i(t) \eta^j(t) \Gamma^k_{ij}(c(t)) \Big) \frac{\partial}{\partial x^k}.$$

The Riemannian metric thus determines the Riemannian connection, and this in turn determines the notion of parallelness in the same way that the covariant derivative in the ambient Euclidean space did in Section 4.9.

5.17. Definition. (Parallel, geodesic, cf. also 4.9)

1. A vector field Y is said to be *parallel*, if $\nabla_X Y = 0$ for every X.
2. A vector field Y along a (regular) curve c is said to be *parallel along the curve* c, if $\nabla_{\dot{c}} Y = 0$ (this is independent of the parametrization).
3. A regular curve c is called a *geodesic*, if $\nabla_{\dot{c}}\dot{c} = \lambda \dot{c}$ for some scalar function λ. This is equivalent to the equation $\nabla_{c'} c' = 0$, provided c is parametrized by arc length.

The same remarks made in 4.9 for non-regular curves hold here also.

5.18. Corollary. (Parallel displacement, geodesics)

(i) Along an arbitrary regular curve c there is for each $Y_0 \in T_{c(t_0)}M$ a vector field Y (along c) which is parallel along c and whose value at $c(t_0)$ is Y_0. This vector field Y is called the *parallel displacement* of Y_0 along c.

(ii) Parallel displacement preserves the Riemannian metric, i.e., $\langle Y_1, Y_2 \rangle$ is constant for any two parallel vector fields Y_1, Y_2 along c.

(iii) At every point p and for each $X \in T_pM$ with $g(X,X) = 1$, there is an $\epsilon > 0$ and a uniquely determined geodesic $c : (-\epsilon, \epsilon) \to M$ which is parametrized by arc length and for which $c(0) = p$, $\dot{c}(0) = X$.

The proof is literally the same as in 4.10, 4.11 and 4.12. It is sufficient to consider the parts of the curve which are contained in local charts. The equation which $Y(t) = \sum_j \eta^j(t) \frac{\partial}{\partial x^j}$ satisfies if and only if it is parallel along c: $\nabla_{\dot{c}} Y = 0$ (where $x^i(t)$ are the coordinates of c) is

$$\frac{d\eta^k}{dt} + \sum_{i,j} \dot{x}^i(t) \cdot \eta^j(t) \cdot \Gamma_{ij}^k(c(t)) = 0, \quad k = 1, \ldots, n.$$

The system of equations which c satisfies precisely if it is a geodesic is

$$\frac{d^2x^k}{dt^2} + \sum_{i,j} \dot{x}^i(t) \cdot \dot{x}^j(t) \cdot \Gamma_{ij}^k(c(t)) = 0, \quad k = 1, \ldots, n.$$

5.19. Definition. (Exponential mapping)
For a fixed point $p \in M$ let $c_V^{(p)}$ denote the uniquely determined geodesic through p which is parametrized by arc length in the direction of a unit vector V. In some neighborhood U of $0 \in T_pM$, the following mapping is well-defined:

$$T_pM \supseteq U \ni (p, tV) \longmapsto c_V^{(p)}(t).$$

Here the parameters are chosen is such a way that $(p, 0) \mapsto p$. This map is called the *exponential mapping* at the point p, and it is denoted by $\exp_p : U \longrightarrow M$. For variable points p one can define a mapping exp in a similar manner by the formula $\exp(q, tV) = \exp_q(tV) = c_V^{(q)}(t)$. This can be defined on an open set of the tangent bundle TM, compare exercise 3.

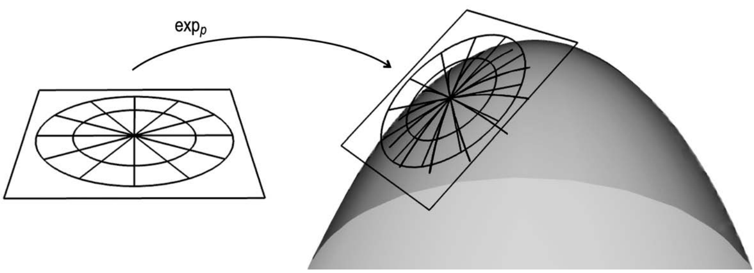


Figure 5.2. Exponential mapping at a point p

Remark: $\exp_p$ maps the lines through the origin of T_pM to geodesics, and this mapping is isometric because the arc length is preserved, see Figure 5.2. In all directions perpendicular to the geodesics through p the map $\exp_p$ is in general not isometric, i.e., it is not length-preserving. This question will be addressed again later in Section 7B, where a more precise study of the transformation of lengths is made.

EXAMPLES:

1. In $\mathbb{R}^n$ the exponential mapping is nothing but the canonical identification of the tangent space $T_p\mathbb{R}^n$ with $\mathbb{R}^n$ itself, where the origin of the tangent space is mapped to the point p. More precisely, $\exp_p(tV) = p + tV$.
2. For the unit sphere S^2 with south pole $p = (0, 0, -1)$, the exponential mapping can be expressed in the following manner using polar coordinates, where we write a tangent vector as $r\cos\phi\frac{\partial}{\partial x} + r\sin\phi\frac{\partial}{\partial y}$, thus viewing it as a function of r and ϕ:

$$\exp_p(r,\phi) = \Big(\cos\phi\cos\big(r - \frac{\pi}{2}\big), \sin\phi\cos\big(r - \frac{\pi}{2}\big), \sin\big(r - \frac{\pi}{2}\big)\Big).$$

   The circle $r = \frac{\pi}{2}$ in the tangent plane gets mapped to the equator, while the circle $r = \pi$ maps to the north pole. At this point the exponential mapping degenerates.
3. In the group $\mathbf{SO}(n, \mathbb{R})$ with the unit element E and with the (bi-invariant) standard metric, $\exp_E$ is given by the exponential series

$$A \longmapsto \sum_{k\geq 0} \frac{A^k}{k!}$$

   evaluated for an arbitrary skew-symmetric real $(n \times n)$-matrix A (cf. the proof of 2.15). This is the origin of the name *exponential mapping.* The exponential rule

$$\exp\big((t+s)A\big) = \exp(tA)\cdot\exp(sA)$$

   expresses the fact that the line $\{tA \mid t \in \mathbb{R}\}$ is mapped by $\exp_E$ onto a 1-parameter subgroup of matrices. A very similar state of affairs holds for other matrix groups such as $\mathbf{GL}(n, \mathbb{R})$, $\mathbf{SL}(n, \mathbb{R})$, $\mathbf{U}(n)$, $\mathbf{SU}(n)$. This mapping is of fundamental importance in the theory of Lie groups. The tangent space at the unit element is the corresponding Lie algebra. In the case of the rotation group $\mathbf{SO}(n, \mathbb{R})$, the Lie algebra is the set of skew-symmetric $(n \times n)$-matrices, together with the multiplication given by the commutator $[X, Y] = XY - YX$ (Compare with 5.13. For more details see [**42**] and [**45**], Chapter 1.

5.20. Definition. (Holonomy group)

Let $P^c : T_pM \longrightarrow T_pM$ denote the parallel translation along a closed curve c with $c(0) = c(1) = p$. For this it suffices that c is continuous and piecewise regular, since the parallel translation is the composition of the corresponding smooth parts and one may then apply 5.18 (i).

For c_1 and c_2 let $c_2 * c_1$ denote the composition of the curves, and let $c^{-1}(t) := c(L - t)$ for $c\colon [0, L] \to M$ (run through in the opposite direction). Then one has

$$P^{c_2 * c_1} = P^{c_2} \circ P^{c_1},$$

$$P^{c^{-1}} = \left(P^c\right)^{-1},$$

and the set of all parallel translations from p to p along piecewise regular curves thus has the structure of a group. It is called the *holonomy group* of the manifold (M, g) at the point p. If M is path connected, then all holonomy groups are isomorphic to each other and one just speaks of *the* holonomy group of (M, g). The holonomy group is always a subgroup of the orthogonal group $\mathbf{O}(n)$, which operates on $T_pM \cong \mathbb{R}^n$. This follows from 5.18 (ii).

Examples:

1. The holonomy group is trivial for $\mathbb{R}^n$ and for the flat torus $\mathbb{R}^n/\mathbb{Z}^n$. The reason for this is that the parallel translation in the sense of the Riemannian metric coincides with the usual parallel translation. For every closed path the result under parallel translation is the vector one starts with.
2. On the standard sphere S^2 the holonomy group contains all rotations (exercise).
3. On a flat cone with a non-trivial opening angle (this is a ruled surface with $K = 0$, cf. 3.24), the holonomy group is not trivial. This is seen by cutting the cone open and developing it on the plane (cf. 3.24). The identification at the boundary leads to non-trivial elements of the holonomy group.
4. On a flat Möbius strip the holonomy group also contains a reflection. This can again be seen most easily by developing the surface in the plane.

Exercises.

1. Show that the open hemispheres $\{(x_1, x_2, x_3) \in S^2 \mid x_i \neq 0\}$ for $i = 1, 2, 3$ define an atlas of the two-dimensional sphere with six (connected) charts $U_1, \ldots, U_6$. Here x_1, x_2, x_3 denote Cartesian coordinates. Determine explicitly the transformation functions between the charts. A picture of the six hemispheres can be found on the cover of the book [**14**].

2. Show that the Cartesian product $M_1 \times M_2$ of two differentiable manifolds is again a differentiable manifold.

3. Show that for a given differentiable n-manifold M the set of all pairs (p, X) with $X \in T_pM$ is again (in a natural way) a differentiable manifold; it is called the *tangent bundle* TM of M. For this, construct for every chart φ in M an *associated bundle chart* by means of

$$\Phi(p, X) := (\varphi(p), \xi^1(p), \cdots, \xi^n(p)) \in I\!R^n \times I\!R^n,$$

where $\xi^1, \ldots, \xi^n$ are the components of X in the corresponding basis, i.e., $X_p = \sum_{i=1}^n \xi^i(p) \frac{\partial}{\partial x^i}\Big|_p$. Check the properties of Definition 5.1. (Note: Formally the definition of the tangent bundle includes the projection from TM to M given by $(p, X) \mapsto p$.)

4. Determine whether this definition of the tangent bundle coincides in the case of $M = I\!R^n$ with Definition 1.6.

5. Show the following. The tangent bundle of the unit circle S^1 is diffeomorphic to the cylinder $S^1 \times I\!R$. The analogous statement does not hold for the two-sphere S^2, but surprisingly it *does* hold for the three-sphere S^3: the tangent bundle of S^3 is diffeomorphic to the product $S^3 \times I\!R^3$, cf. Exercise 12 at the end of Chapter 7.

6. The metrics on two Riemannian manifolds (M_1, g_1) and (M_2, g_2) induce in a canonical manner a Riemannian metric $g_1 \times g_2$ on the Cartesian product $M_1 \times M_2$, the so-called *product metric.* What is the form of this metric in local coordinates?

7. Let a submanifold M of $I\!R^4$ be given by the equation

$$M = \{(x_1, x_2, x_3, x_4) \in I\!R^4 \mid x_1^2 + x_2^2 = x_3^2 + x_4^2 = 1\}.$$

Prove that M is a two-dimensional manifold by displaying an explicit atlas.

8. Construct an explicit Lorentzian metric, i.e., a metric tensor of type $(-+)$, on the (abstract) Klein bottle (cf. the examples following 5.1 in the text).

9. Let (M, g) be a two-dimensional Riemannian manifold, and let $\Delta \subset M$ be a geodesic triangle which is the boundary of a simply connected domain. Show that the parallel translation along this boundary (traced through once) is a rotation in the tangent plane. Calculate the angle of rotation in terms of quantities which only depend on the interior of Δ. Hint: Gauss-Bonnet formula.

10. Show that the holonomy group of the standard two-sphere S^2 really contains all the rotations. Hint: Consider curves which are constructed piecewise from great circles.

11. Determine the holonomy group of the hyperbolic plane as a surface in three-dimensional Minkowski space (cf. 3.44). Here the covariant derivative is to be taken as in Euclidean space, that is, with tangent components which are directional derivatives.

12. Let (M_*, g_*) be an n-dimensional Riemannian manifold and let $f\colon \mathbb{R} \to \mathbb{R}$ be a smooth function without zeros. Then $\mathbb{R} \times M$ endowed with the metric

$$g(t, x^1, \ldots, x^n) = dt^2 + (f(t))^2 \cdot g_*(x^1, \ldots, x^n)$$

is again a Riemannian manifold, the so-called *warped product* with the *warping function* f. Show that the t-lines are always geodesics. What are sufficient conditions in order that geodesics on M_* are also geodesics on M?

13. Let X be a vector field on the manifold M. Show the following.
   (a) At every point $p \in M$ there is a uniquely determined curve $c_p\colon I_p \to M$ with $c_p(0) = p, c_p'(t) = X_{c(t)}$, where I_p is the maximal interval around $t = 0$ with this property.
   (b) For every open neighborhood U of p there is a set, open in $\mathbb{R} \times M$, such that the map ψ which is defined by $\psi(t, q) := \psi_t(q) := c_q(t)$ is differentiable. ψ is called the *local flow* of X at the point p.

(c) In case ψ_t is defined for every $t \in I\!R$, one calls the vector field (or also the flow) *complete*. In this case one has $\psi_{t+s} = \psi_t \circ \psi_s$ for all $t, s \in I\!R$. This property defines a one-parameter group of diffeomorphisms, since $t \mapsto \psi_t$ is a group homomorphism. Why are all ψ_t diffeomorphisms?

14. Let X be a vector field on an n-dimensional manifold M with $X_p \neq 0$ at a point $p \in M$. Using the previous exercise, show that there is a coordinate system $x^1, \ldots, x^n$ near p with $X = \frac{\partial}{\partial x^1}$.

15. Let X, Y be vector fields on M, and let ψ denote the local flow of X at a point $p \in M$. Again using the previous exercise, verify the following equation:

$$[X, Y] = \lim_{t \to 0} \frac{1}{t} \Big(D\psi_{-t}(Y_{\psi_t(p)}) - Y_p \Big).$$

16. Show that the tangent space of the rotation group $\mathbf{SO}(3)$ at the "point" corresponding to the identity matrix can be identified in a natural manner with the set of all skew-symmetric (3×3)-matrices (cf. also the proof of 2.15). Calculate the differential of the Cayley map $CAY \colon I\!R^3 \to \mathbf{SO}(3)$. For the definition of this map see the examples following 5.1.

17. Give an explicit atlas for the manifold $I\!RP^3$ (*real projective space*), which is defined as the quotient of the three-sphere by the antipodal mapping.

18. Show that the exponential series

$$A \longmapsto \sum_{k \geq 0} \frac{A^k}{k!}$$

is actually an orthogonal matrix for an arbitrary skew-symmetric matrix A.

19. Find a formula for the inverse mapping of the exponential mapping (a kind of *logarithm*) for the case of the group $\mathbf{SO}(n)$. Hint: Take a power series and determine the coefficients.

20. The *Schwarzschild half-plane* is defined as the half-plane $E = \{(t, r) \in I\!R^2 \mid r > r_0\}$ with the semi-Riemannian metric $ds^2 = -hdt^2 + h^{-1}dr^2$, where h denotes the function $h(r, t) := 1 - r_0/r$.

Show that the maps $(t, r) \mapsto (\pm t + b, r)$ are isometries. Moreover, calculate the Christoffel symbols and show that the r-lines are always geodesics. Show also that for the geodesics, written $\gamma(s) = (t(s), r(s))$, the quantity $h(\gamma(s))t'(s)$ is a constant. The constant r_0 corresponds to the Schwarzschild radius, which depends on the mass of a black hole, which one should imagine is situated at $r = 0$.

21. Suppose we are given coordinates in (M, g) such that in these coordinates the metric tensor has diagonal form, i.e., $g_{ij} = 0$ for $i \neq j$. Show that the system of equations for geodesics is as follows:
$$\frac{d}{ds}\left(g_{kk}\frac{dx^k}{ds}\right) = \frac{1}{2}\sum_{i=1}^{n}\frac{\partial g_{ii}}{\partial x^k}\left(\frac{dx^i}{ds}\right)^2 \qquad (k = 1, \ldots, n).$$

22. Let the *Schwarzschild metric* be given as follows:
$$ds^2 = -h \cdot dt^2 + h^{-1} \cdot dr^2 + r^2\left(\sin^2 \vartheta d\varphi^2 + d\vartheta^2\right),$$
where $h = h(r) = 1 - \frac{2M}{r}$. The Schwarzschild metric is a model for a universe in which there is precisely one rotationally symmetric star. Show that every geodesic c satisfies the following equations with constants E and L:
(a) $h \cdot \frac{dt}{ds} = E$,
(b) $r^2 \sin^2 \vartheta \cdot \frac{d\varphi}{ds} = L$,
(c) $\frac{d}{ds}\left(r^2 \cdot \frac{d\vartheta}{ds}\right) = r^2 \sin\vartheta \cos\vartheta \left(\frac{d\varphi}{ds}\right)^2$.
Now suppose that c is parametrized by arc length τ, which describes a freely falling particle (in particular, this implies it is not a light particle, for which $g(c', c') \neq 0$ holds), with the initial condition that it is falling equatorially, i.e., satisfies $\vartheta(0) = \frac{\pi}{2}$ and $\frac{d\vartheta}{ds}(0) = 0$. Then we have
(a′) $h \cdot \frac{dt}{d\tau} = E$,
(b′) $r^2 \frac{d\varphi}{d\tau} = L$,
(c′) $\vartheta = \frac{\pi}{2}$.
Hint: Exercise 21.

23. Calculate the exponential map of the flat torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2$ with the induced locally euclidean metric, cf. Example (v) after 5.10. Find out which geodesics are closed curves.

Chapter 6

The Curvature Tensor

In the Gauss equation 4.15 or 4.18, we have on the left-hand side an expression which we called the curvature tensor. Its connection to the curvature (and thus the nomenclature) is clarified by the *Theorema Egregium* 4.16 and 4.20. For this it is of great importance that the left-hand side of the equation only depends on the first fundamental form or the covariant derivative, which follows from the equation

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

in the Koszul-style calculus, or

$$R_{ikj}^s = \frac{\partial \Gamma_{ij}^s}{\partial u^k} - \frac{\partial \Gamma_{ik}^s}{\partial u^j} + \Gamma_{ij}^r \Gamma_{rk}^s - \Gamma_{ik}^r \Gamma_{rj}^s$$

in Ricci calculus. (The more precise notation here would be $R^{s}_{\cdot ikj}$ instead of $R^{s}_{ikj\cdot}$.) This expression is well-defined for an arbitrary Riemannian manifold and is the foundation for all further information on curvature of Riemannian manifolds. In fact, all scalar curvature quantities can be obtained from this curvature tensor. Before we go into this, we make a brief digression with some general remarks on tensors.

6A Tensors

Tensors are operators which are not determined by the process of taking derivatives of other quantities (a local process), but rather through evaluation of known quantities at single points. An example is the Weingarten map of a surface. For the calculation of a derivative it is never sufficient to know the given quantity at a point; rather, it is imperative to know that quantity (in typical cases) at least all along a curve, as was needed for the definition of the covariant derivative in Sections 4.2 and 4.3. In comparison, the metric tensor (also known as the measure tensor) g of a Riemannian manifold measures the scalar product of two vectors X, Y, using only their values at a given point, with no need for taking derivatives. A similar statement holds for the tension and inertia tensors which occur in mechanics. It is of great importance for differential geometry that the curvature tensor is a tensor in the above sense. We already met this in 4.19, but at that point it was a simple consequence of the Gauss equation, whose right-hand side contains only the Weingarten map. But even without using the Gauss equation this fact is easy to see, as we now describe.

Let X, Y, Z be three vector fields. Then the evaluation of the above expression for the curvature tensor $R(X,Y)Z$ at a point p requires various covariant derivatives at p such as $\nabla_Y Z$. However, if one modifies the arguments X, Y, Z by scalar functions α, β, γ (that is, if one replaces X, Y, Z by $\alpha X, \beta Y, \gamma Z$) and then evaluates the curvature tensor at p, then no derivatives of α, β, γ come in. Instead, we have the following equation:

$$\big(R(\alpha X, \beta Y)(\gamma Z)\big)\Big|_p = \alpha(p)\beta(p)\gamma(p)\big(R(X,Y)Z\big)\Big|_p.$$

More precisely, the derivatives of α, β and γ cancel out. Thus one may view the result $R(X,Y)Z$ at the point p as depending only on X_p, Y_p, Z_p. As a specific consequence we obtain the following expression in terms of local coordinates:

$$\begin{aligned} R(X,Y)Z &= R\Big(\sum_i \xi^i \frac{\partial}{\partial x^i}, \sum_j \eta^j \frac{\partial}{\partial x^j}\Big) \sum_k \zeta^k \frac{\partial}{\partial x^k} \\ &= \sum_{i,j,k} \xi^i \eta^j \zeta^k \underbrace{R\Big(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\Big)\frac{\partial}{\partial x^k}}_{\text{independent of } X,Y,Z}, \end{aligned}$$

So one can evaluate the left hand side at p just by evaluating the coefficients ξ^i, η^j, ζ^k at p. This is precisely the property which defines what a tensor is supposed to be.

6.1. Definition. (Tensors, tensor fields)
A *covariant tensor of degree s* (briefly, a $(0,s)$*-tensor*) at a point p on a differentiable manifold M is a multilinear mapping

$$A_p\colon \underbrace{(T_pM)\times\cdots\times(T_pM)}_{s} \longrightarrow \mathbb{R}.$$

Similarly, a $(1,s)$-tensor at a point p is a multilinear mapping

$$A_p\colon \underbrace{(T_pM)\times\cdots\times(T_pM)}_{s} \longrightarrow T_pM.$$

A basis of the space of all $(0,s)$-tensors is given by the set

$$\Big(dx^{j_1}|_p\otimes\cdots\otimes dx^{j_s}|_p\Big)_{j_1,\dots,j_s=1,\dots,n},$$

where

$$\Big(dx^{j_1}\otimes\cdots\otimes dx^{j_s}\Big)\Big(\frac{\partial}{\partial x^{l_1}},\dots,\frac{\partial}{\partial x^{l_s}}\Big) := \delta^{j_1}_{l_1}\cdot\dots\cdot\delta^{j_s}_{l_s}.$$

As in the remark at the beginning of 5.10, a comparison of coefficients yields for the coefficients of

$$A_p = \sum A_{j_1\dots j_s}\cdot dx^{j_1}\otimes\cdots\otimes dx^{j_s},$$

after insertion into the basis, the equation

$$A_{j_1,\dots,j_s} = A_p\Big(\frac{\partial}{\partial x^{j_1}},\dots,\frac{\partial}{\partial x^{j_s}}\Big).$$

This explains the abbreviated notation of Ricci calculus used for a $(0,s)$-tensor, namely $A_{j_1\dots j_s}$, and the similar notation $A^i_{j_1\dots j_s}$ for a $(1,s)$-tensor. In the latter case, we have

$$\sum_i A^i_{j_1,\dots,j_s}\frac{\partial}{\partial x^i} = A_p\Big(\frac{\partial}{\partial x^{j_1}},\dots,\frac{\partial}{\partial x^{j_s}}\Big).$$

A *differentiable* $(0,s)$*- or* $(1,s)$*-tensor field* A is an association $p\longmapsto A_p$ such that the coefficient functions $A_{j_1,\dots,j_s}$ and $A^i_{j_1,\dots,j_s}$, respectively, in the representation above are differentiable, just as in Definition 5.10.

Convention: In what follows, all tensor fields occurring will be assumed to be differentiable.

REMARK 1: In this book we will use basically only tensors of the type $(0, s)$ or $(1, s)$. More generally one considers also mixed tensors, containing covariant as well as contravariant degrees, as follows.

An *s-covariant and r-contravariant tensor* (briefly, an *(r, s)-tensor*) at a point p on a differentiable manifold M is a multilinear mapping

$$A_p : \underbrace{(T_pM)^* \times \cdots \times (T_pM)^*}_{r} \times \underbrace{(T_pM) \times \cdots \times (T_pM)}_{s} \longrightarrow \mathbb{R},$$

where $(T_pM)^* := \mathrm{Hom}(T_pM; \mathbb{R})$ denotes the dual space of the tangent space. A basis of the space of all (r, s)-tensors is given by

$$\Big(\frac{\partial}{\partial x^{i_1}}\Big|_p \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}}\Big|_p \otimes dx^{j_1}\big|_p \otimes \cdots \otimes dx^{j_s}|_p\Big)_{i_1,\ldots,i_r,j_1,\ldots,j_s=1,\ldots,n},$$

where

$$\Big(\frac{\partial}{\partial x^{i_1}}\Big|_p \otimes \cdots \otimes dx^{j_s}|_p\Big)\Big(dx^{k_1}, \ldots, dx^{k_r}, \frac{\partial}{\partial x^{l_1}}, \ldots, \frac{\partial}{\partial x^{l_s}}\Big)$$

$$:= \delta^{k_1}_{i_1} \cdot \cdots \cdot \delta^{k_r}_{i_r} \cdot \delta^{j_1}_{l_1} \cdot \cdots \cdot \delta^{j_s}_{l_s}.$$

Again, comparison of coefficients yields for the coefficients of the tensor

$$A_p = \sum A^{i_1 \ldots i_r}_{j_1 \ldots j_s} \cdot \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s},$$

after insertion into the basis, the following equation:

$$A^{i_1,\ldots,i_r}_{j_1,\ldots,j_s} = A_p\Big(dx^{i_1}, \ldots, dx^{i_r}, \frac{\partial}{\partial x^{j_1}}, \ldots, \frac{\partial}{\partial x^{j_s}}\Big).$$

Thus, in this case the abbreviated notation used in Ricci calculus is $A^{i_1 \cdots i_r}_{j_1 \cdots j_s}$ for an (r, s)-tensor. The covariant indices are written as subscripts, while the contravariant indices are written as superscripts.

REMARK 2: In multilinear algebra, one interprets the above definition of tensor as defining elements of a space called a *tensor product*, by means of the following canonical isomorphisms, where "Mult" is the set of multilinear mappings and "Hom" denotes the set of homomorphisms (i.e., linear maps):

$$\mathrm{Mult}\Big((T_pM^*)^r, (T_pM)^s; \mathbb{R}\Big) \cong \mathrm{Hom}\Big((\bigotimes_{i=1}^{r} T_pM^*) \otimes (\bigotimes_{j=1}^{s} T_pM); \mathbb{R}\Big)$$

$$\cong \Big((\bigotimes_{i=1}^{r} T_pM^*) \otimes (\bigotimes_{j=1}^{s} T_pM)\Big)^* \cong (\bigotimes_{i=1}^{r} T_pM) \otimes (\bigotimes_{j=1}^{s} T_pM)^*$$

For details on multilinear algebra, see for example [**31**] or [**33**].

For the set of tensors of type $(1, s)$, the above definition lets us interpret them as multilinear mappings

$$A : \underbrace{T_pM \times \cdots \times T_pM}_{s} \longrightarrow T_pM$$

where we have used also the canonical isomorphism between T_pM and T_pM^{**}:

$$A(\cdot, X_1, \ldots, X_s) \in (T_pM)^{**} \cong T_pM.$$

Examples of tensors:

1. A vector field X is at the same time a $(1, 0)$-tensor field, denoted by the symbol ξ^i in Ricci calculus, where $X = \sum_i \xi^i \frac{\partial}{\partial x^i}$ is the representation of X in local coordinates.

2. A one-form in the sense of Section 4F is a $(0, 1)$-tensor, which is also referred to as a *covector field.* This is true in particular for the differential of a function f, which is written $df := \sum_i \frac{\partial f}{\partial x^i} dx^i$, and in Ricci calculus $f_i = \frac{\partial f}{\partial x^i}$. One then has $df(X) = \nabla_X f = X(f)$.

3. A scalar function is a $(0, 0)$-tensor field, and thus has no index in the Ricci calculus.

4. A Riemannian metric g is a $(0, 2)$-tensor field, cf. 5.10. In Ricci calculus, g is described by the symbol g_{ij}, which was introduced in Chapter 3.

5. For a fixed vector field Y the covariant derivative ∇Y is a $(1, 1)$-tensor field, defined by $\nabla Y(X) := \nabla_X Y$ (but the association $X, Y \longmapsto \nabla_X Y$ is not a $(1, 2)$-tensor field because of the product rule). The difference of two connections ∇ and $\widetilde{\nabla}$ is always a $(1, 2)$-tensor field.

6. The Weingarten mapping $L \colon T_pM \longrightarrow T_pM$ of a hypersurface is a $(1, 1)$-tensor field which is written in local coordinates h_i^k. The second fundamental form of a hypersurface is the corresponding $(0, 2)$-tensor field $I\!I(X, Y) = I(LX, Y) = I(X, LY)$, which is written in Ricci calculus as $h_{ij} = h_i^k g_{kj}$. Similarly, we have $h_i^k = h_{ij} g^{jk}$, cf. 3.10. Informally this prceedure of raising or lowering indices is also called *index gymnastics.*

7. A Riemannian metric g yields an isomorphism of T_pM and T_pM^* by

$$T_pM \ni X \longmapsto g(\,\cdot\,,X) \in T_pM^*.$$

In local coordinates this is precisely the procedure of lowering indices; a vector ξ^j becomes a covector $\xi_i = \xi^j g_{ij}$ and conversely.

8. The curvature tensor (= left-hand side of the Gauss equation) is a $(1,3)$-tensor:

$$X, Y, Z \longmapsto R(X,Y)Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

In the Ricci calculus, the components of the curvature tensor are given by the left-hand side of the Gauss equation 4.15, where the position of the indices is for historical reasons as follows (deviating from the convention of Definition 6.1, in accordance with [**19**], Ch. III, Sect.7, and Ch. V):

$$R\Big(\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^j}\Big)\frac{\partial}{\partial x^i} = \sum_s R^s_{ikj} \frac{\partial}{\partial x^s},$$

$$R^s_{ikj} = \frac{\partial \Gamma^s_{ij}}{\partial x^k} - \frac{\partial \Gamma^s_{ik}}{\partial x^j} + \Gamma^r_{ij}\Gamma^s_{rk} - \Gamma^r_{ik}\Gamma^s_{rj}.$$

By lowering the remaining upper index, we get the corresponding $(0,4)$-tensor

$$X, Y, Z, V \longmapsto g(R(X,Y)V, Z).$$

The switch of the two arguments Z, V is also a historical convention. In Ricci calculus this amounts to putting the lowered index of $R_{mikj} = g_{ms} R^s_{ikj}$ in the first spot:

$$\Big\langle \frac{\partial}{\partial x^m}, R\Big(\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^j}\Big)\frac{\partial}{\partial x^i} \Big\rangle = \sum_s g_{ms} R^s_{ikj} = R_{mikj}.$$

WARNING: In the literature you will often find the curvature tensor to have the opposite sign, i.e., $R(X,Y)Z := \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z - \nabla_{[Y,X]} Z$; similarly for the components R^s_{ikj}.

For an appropriate notion of the derivative of a tensor, the action of a $(1,1)$-tensor A on a vector Y is just like the action of a matrix on a vector. Thus, if you want to take the derivative of the expression $A(Y)$, then the product rule $(A(Y))' = A'(Y)+A(Y')$ must be accommodated. In other words, for the covariant derivative with respect to X this means

$$\nabla_X(A(Y)) = (\nabla_X A)(Y) + A(\nabla_X Y).$$

In the case of tensors of higher degree, because of the multilinearity there is an iterated product rule which must be taken account of, with a corresponding number of summands. This motivates the following definition.

6.2. Definition. (Derivatives of tensor fields)
Let A be a $(0,s)$-tensor field (resp. a $(1,s)$-tensor field), and let X be a fixed vector field. Then we define the *covariant derivative* of A in the direction of X by the formula

$$\begin{aligned}(\nabla_X A)\big(Y_1,\ldots,Y_s\big) &:= \nabla_X\big(A(Y_1,\ldots,Y_s)\big)\\ &- \sum_{i=1}^{s} A\big(Y_1,\ldots,Y_{i-1},\nabla_X Y_i, Y_{i+1},\ldots,Y_s\big).\end{aligned}$$

$\nabla_X A$ is then also a $(0,s)$-tensor (resp. a $(1,s)$-tensor), and ∇A is a $(0,s+1)$-tensor (resp. a $(1,s+1)$-tensor) by means of the formula

$$(\nabla A)(X,Y_1,\ldots,Y_s) := (\nabla_X A)(Y_1,\ldots,Y_s).$$

In Ricci calculus the following notation is used for this:

$$\begin{aligned}\nabla_i A_{j_1\ldots j_s} = \frac{\partial}{\partial x^i} A_{j_1\ldots j_s} &- \Gamma_{ij_1}^k A_{kj_2\ldots j_s} - \Gamma_{ij_2}^k A_{j_1kj_3\ldots j_s}\\ &- \cdots - \Gamma_{ij_s}^k A_{j_1\ldots j_{s-1}k},\end{aligned}$$

(resp.

$$\begin{aligned}\nabla_i A^m_{j_1\ldots j_s} = \frac{\partial}{\partial x^i} A^m_{j_1\ldots j_s} + \Gamma^m_{ir} A^r_{j_1\ldots j_s} &- \Gamma_{ij_1}^k A^m_{kj_2\ldots j_s} - \Gamma_{ij_2}^k A^m_{j_1kj_3\ldots j_s}\\ &- \cdots - \Gamma_{ij_s}^k A^m_{j_1\ldots j_{s-1}k}).\end{aligned}$$

To legitimize this definition, we have to show that

$$(\nabla_X A)(Y_1, \ldots, f \cdot Y_j, \ldots, Y_s) = f \cdot (\nabla_X A)(Y_1, \ldots, Y_s),$$

which means precisely that $(\nabla_X A)_p(Y_1, \ldots, Y_s)$ depends only on the values $Y_1|_p, \ldots, Y_{s|p}$. This equation is easily verified using the rule 5.15 (iv): the first term on the right-hand side (in the above definition) contains the derivative of f in the form $X(f) \cdot A(Y_1, \ldots, Y_s)$ and the j-th summand of the second expression contains the expression

$$A\big(Y_1, \ldots, Y_{j-1}, X(f)Y_j, Y_{j+1}, \ldots, Y_s\big).$$

These two terms cancel because of signs.

Special cases:

1. For a scalar function f (that is, for a $(0,0)$-tensor), the covariant derivative is nothing but the differential $df = Df = \nabla f$ with $\nabla f(X) = \nabla_X f = X(f)$. The *gradient* of f with respect to a metric g, written $\text{grad} f$ or, in more detail, $\text{grad}_g f$, is then the vector determined by the relation $g(\text{grad} f, X) := \nabla f(X)$ for all X. Note that here ∇f does not denote the gradient, even though this is a usual practice in much of the literature. Instead, ∇f denotes as in Definition 6.2 the differential of f as a $(0,1)$-tensor, while the gradient is a $(1,0)$-tensor. In local coordinates, the components f^i of the gradient result from the components $f_j = \frac{\partial f}{\partial x^j}$ by raising indices, $f^i = f_j g^{ji}$. In the standard chart of Euclidean space there is no noticeable difference between f^i and f_i, but in polar coordinates with

$$(g_{ij}(r,\phi)) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \quad \text{and} \quad (g^{ij}(r,\phi)) = \begin{pmatrix} 1 & 0 \\ 0 & r^{-2} \end{pmatrix},$$

one has $f_r = \frac{\partial f}{\partial r}, f_\phi = \frac{\partial f}{\partial \phi}$ and similarly $f^r = f_r g^{rr} = f_r, f^\phi = f_\phi g^{\phi\phi} = f_\phi r^{-2}$.

2. The second covariant derivative of f is given by $\nabla^2 f = \nabla\nabla f$. More explicitly, here this is

$$\begin{aligned}(\nabla^2 f)(X,Y) := \big(\nabla_X \nabla f\big)(Y) &:= \nabla_X(\nabla f(Y)) - \nabla f(\nabla_X Y) \\ &= \nabla_X \nabla_Y f - (\nabla_X Y)(f).\end{aligned}$$

   $\nabla^2 f$ is also referred to as the *Hesse form* or the *Hessian* of f.

3. The derivative of the gradient is the associated *Hesse* $(1,1)$-*tensor* $\nabla \mathrm{grad} f$. To see this, we calculate

$$\underbrace{\nabla_X(g(\mathrm{grad} f, Y))}_{\nabla_X \nabla_Y f} = g(\nabla_X \mathrm{grad} f, Y) + \underbrace{g(\mathrm{grad} f, \nabla_X Y)}_{\nabla_X Y(f)}$$

and note the relation $g(\nabla_X \mathrm{grad} f, Y) = \nabla^2 f(X,Y)$. In Ricci calculus we have $\nabla_i \nabla_j f = \nabla_i f_j = \frac{\partial^2 f}{\partial x^i \partial x^j} - \Gamma_{ij}^k f_k$ and $\nabla_i f^j = \nabla_i f_k g^{kj}$ for the Hessian $(0,2)$-tensor and the Hesse $(1,1)$-tensor.

4. For a $(0,1)$-tensor ω (or a one-form), the covariant derivative is defined by

$$\nabla \omega(X,Y) = (\nabla_X \omega)(Y) = \nabla_X(\omega(Y)) - \omega(\nabla_X Y).$$

In local coordinates (i.e., in Ricci calculus) we have

$$\nabla_i \omega_j = \frac{\partial \omega_j}{\partial x^i} - \Gamma_{ij}^k \omega_k.$$

REMARK: The *exterior derivative* $d\omega$ is defined as an alternating two-form by the relation $d\omega(X,Y) = \nabla\omega(X,Y) - \nabla\omega(Y,X)$, cf. Section 4E.

5. Let A be an arbitrary $(0,2)$-tensor; then the relation

$$\nabla A(X,Y,Z) = (\nabla_X A)(Y,Z)$$

$$= \nabla_X(A(Y,Z)) - A(\nabla_X Y, Z) - A(Y, \nabla_X Z)$$

is written in Ricci calculus as

$$\nabla_k A_{ij} = \frac{\partial}{\partial x^k} A_{ij} - \Gamma_{ki}^s A_{sj} - \Gamma_{kj}^s A_{is}.$$

In particular, for the Riemannian metric $A = g$ we have the equality $\nabla g(X,Y,Z) = \nabla_X g(Y,Z) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) = 0$ for all X,Y,Z; hence

$$\nabla g \equiv 0.$$

This state of affairs is also expressed by saying that the metric tensor g is *parallel* with respect to the corresponding Riemannian connection ∇ (the so-called *Ricci Lemma*). In Ricci calculus the notation is $\nabla_i g_{jk} = \nabla_i g^{lm} = 0$.

6. For the Weingarten map L, we have the relation $\nabla L(X,Y) = \nabla_X(LY) - L(\nabla_X Y)$. The Codazzi-Mainardi equation is then nothing but the symmetry of the $(1,2)$-tensor ∇L, cf. 4.19, in Ricci calculus $\nabla_i h^j_k = \nabla_k h^j_i$, cf. Exercise 23 in Chapter 4.
7. For the curvature tensor $R(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}$, viewed here as a $(1,3)$-tensor, the covariant derivative $\nabla_X R$ is a $(1,3)$-tensor, written as follows: $(\nabla_X R)(Y,Z)V = \nabla_X\big(R(Y,Z)V\big) - R(\nabla_X Y, Z)V - R(Y, \nabla_X Z)V - R(Y,Z)\nabla_X V$.

6B The sectional curvature

We again return to the Theorema Egregium 4.20, $\big\langle R(X,Y)Y,X\big\rangle = K$ for orthonormal X,Y, and more generally

$$\big\langle R(X,Y)Y,X\big\rangle = K\big(\langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2\big).$$

This relation holds for two-dimensional surfaces in $\mathbb{R}^3$. For two-dimensional Riemannian manifolds we can use this equation as a definition of the curvature K, which we then again refer to as the *Gaussian curvature* or the *intrinsic curvature of the metric g*. If the dimension of the manifold is greater than two, we can make similar considerations for every choice of two-dimensional submanifolds in the tangent space. This leads to the notion of *sectional curvature*, which is so to speak the curvature of two-dimensional sections of the manifold. For our investigation of the sectional curvature, we need some symmetry properties of the curvature tensor, which are not so immediate.

6.3. Lemma. (Symmetries of the curvature tensor)
For arbitrary vector fields X,Y,Z,V the following relations hold.

1. $R(X,Y)Z = -R(Y,X)Z$;
2. $R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0$; (1st *Bianchi identity*)
3. $(\nabla_X R)(Y,Z)V + (\nabla_Y R)(Z,X)V + (\nabla_Z R)(X,Y)V = 0$; (2nd *Bianchi identity*)
4. $\big\langle R(X,Y)Z,V\big\rangle = -\big\langle R(X,Y)V,Z\big\rangle$;
5. $\big\langle R(X,Y)Z,V\big\rangle = \big\langle R(Z,V)X,Y\big\rangle$.

Written in Ricci notation, these five equations are as follows. (Recall that $R_{ijkl} = g_{si}R^s_{jkl}$.)

1. $R^m_{ijk} = -R^m_{ikj}$;
2. $R^m_{ijk} + R^m_{jki} + R^m_{kij} = 0$;
3. $\nabla_i R^m_{ljk} + \nabla_j R^m_{lki} + \nabla_k R^m_{lij} = 0$;
4. $R_{ijkl} = -R_{jikl}$;
5. $R_{ijkl} = R_{klij}$.

Using the algebraic symmetries 1, 4 and 5, one can write these equations also in the following manner:

1. $R_{ijkl} = -R_{jikl} = -R_{ijlk} = R_{klij}$;
2. $3R_{[ijk]l} = R_{ijkl} + R_{jkil} + R_{kijl} = 0$;
3. $3\nabla_{[i}R_{jk]lm} = \nabla_i R_{jklm} + \nabla_j R_{kilm} + \nabla_k R_{ijlm} = 0$.

Remark: The nomenclature "the first and second Bianchi identity" for the above relations is historically not quite correct but has been used traditionally for some time. The second of these is the classical Bianchi identity; the first is a kind of Jacobi identity, cf. also 5.14. For a historical account of the second Bianchi identity (which is also attributed to G. Ricci), see [**21**], VII.5 (p. 182).

Proof:

1. This is clear by definition. In what follows, without restriction of generality we let X, Y, Z, V be basis fields, so that the Lie brackets of two of these three vanish, and for example $\nabla_X Y = \nabla_Y X$.

2. This sum is calculated as follows:

$$\begin{array}{lcl} R(X,Y)Z & = & \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z, \\ R(Y,Z)X = \nabla_Y\nabla_Z X - \nabla_Z\nabla_Y X & = & \nabla_Y\nabla_X Z - \nabla_Z\nabla_Y X, \\ R(Z,X)Y = \nabla_Z\nabla_X Y - \nabla_X\nabla_Z Y & = & \nabla_Z\nabla_Y X - \nabla_X\nabla_Y Z. \end{array}$$

When we add them up, the sum of the right-hand sides vanishes.

3. We have
$$\begin{aligned}(\nabla_X R)(Y,Z)V &= \nabla_X(\nabla_Y\nabla_Z V - \nabla_Z\nabla_Y V) - R(\nabla_X Y, Z)V\\ &\quad - R(Y,\nabla_X Z)V - \nabla_Y\nabla_Z\nabla_X V + \nabla_Z\nabla_Y\nabla_X V,\\ (\nabla_Y R)(Z,X)V &= \nabla_Y(\nabla_Z\nabla_X V - \nabla_X\nabla_Z V) - R(\nabla_Y Z, X)V\\ &\quad - R(Z,\nabla_Y X)V - \nabla_Z\nabla_X\nabla_Y V + \nabla_X\nabla_Z\nabla_Y V,\\ (\nabla_Z R)(X,Y)V &= \nabla_Z(\nabla_X\nabla_Y V - \nabla_Y\nabla_X V) - R(\nabla_Z X, Y)V\\ &\quad - R(X,\nabla_Z Y)V - \nabla_X\nabla_Y\nabla_Z V + \nabla_Y\nabla_X\nabla_Z V.\end{aligned}$$
The sum of the three right-hand sides vanishes, as was to be shown.

4. The skew-symmetry of a bilinear form $\omega(X,Y) = -\omega(Y,X)$ is equivalent to $\omega(X,X) = 0$ for all X, since
$$\omega(X+Y,X+Y) = \omega(X,X) + \omega(Y,Y) + \underbrace{\omega(X,Y) + \omega(Y,X)}_{=0}.$$
Thus, we have to show that $\langle R(X,Y)Z,Z\rangle = 0$ for all X,Y,Z. For this, we consider the equation $Y\langle Z,Z\rangle = 2\langle\nabla_Y Z,Z\rangle$ and take one more derivative:
$$X(Y\langle Z,Z\rangle) = 2X\langle\nabla_Y Z,Z\rangle = 2\langle\nabla_X\nabla_Y Z,Z\rangle + 2\langle\nabla_Y Z,\nabla_X Z\rangle.$$
From this, for the curvature tensor we get
$$\begin{aligned}2\langle R(X,Y)Z,Z\rangle &= 2\langle\nabla_X\nabla_Y Z,Z\rangle - 2\langle\nabla_Y\nabla_X Z,Z\rangle\\ &= XY(\langle Z,Z\rangle) - 2\langle\nabla_Y Z,\nabla_X Z\rangle\\ &\qquad -YX(\langle Z,Z\rangle) + 2\langle\nabla_X Z,\nabla_Y Z\rangle\\ &= \underbrace{[X,Y]}_{=0}(\langle Z,Z\rangle) = 0.\end{aligned}$$

5. This follows purely algebraically from 1, 2 and 4:
$$\langle R(X,Y)Z,V\rangle \overset{(1)}{=} -\langle R(Y,X)Z,V\rangle$$
$$\overset{(2)}{=} \langle R(X,Z)Y,V\rangle + \langle R(Z,Y)X,V\rangle,$$
$$\langle R(X,Y)Z,V\rangle \overset{(4)}{=} -\langle R(X,Y)V,Z\rangle$$
$$\overset{(2)}{=} \langle R(Y,V)X,Z\rangle + \langle R(V,X)Y,Z\rangle.$$

By adding we get the following equation:

$$2\langle R(X,Y)Z,V\rangle = \langle R(X,Z)Y,V\rangle + \langle R(Z,Y)X,V\rangle$$

$$+\langle R(Y,V)X,Z\rangle + \langle R(V,X)Y,Z\rangle.$$

Now switching X and Z as well as Y and V, we get

$$2\langle R(Z,V)X,Y\rangle = \langle R(Z,X)V,Y\rangle + \langle R(X,V)Z,Y\rangle$$

$$+\langle R(V,Y)Z,X\rangle + \langle R(Y,Z)V,X\rangle,$$

which is the same sum as before (after a further application of 1 and 4). □

Preliminary remarks on sectional curvature: From the Gauss equation

$$R(X,Y)Z = \langle LY,Z\rangle LX - \langle LX,Z\rangle LY$$

(where L denotes the Weingarten map) it follows that the curvature tensor of the unit sphere (where L is the identity) is given by

$$R_1(X,Y)Z := \langle Y,Z\rangle X - \langle X,Z\rangle Y,$$

which in Ricci notation is written $(R_1)_{ijkl} = g_{ik}g_{jl} - g_{il}g_{jk}$. From this it can be seen that for given orthonormal X, Y, the endomorphism $R_1(X,Y)$ (also called the *curvature transformation*) is a rotation of $90°$, following an orthogonal projection onto the X,Y-plane. It is therefore only natural to compare an arbitrary curvature tensor R with this R_1.

Note that the curvature tensor R_1 of the unit sphere is parallel:

$$\begin{aligned}(\nabla_X R_1)(Y,Z)V &= \nabla_X(\langle Z,V\rangle Y - \langle Y,V\rangle Z) - \langle \nabla_X Z,V\rangle Y \\ &\quad +\langle \nabla_X Y,V\rangle Z - \langle Z,\nabla_X V\rangle Y + \langle Y,\nabla_X V\rangle Z \\ &\quad -\langle Z,V\rangle \nabla_X Y + \langle Y,V\rangle \nabla_X Z = 0,\end{aligned}$$

because $\nabla_X g = 0$.

6.4. Definition. With respect to a given Riemannian metric $\langle\ ,\ \rangle$, the *standard curvature tensor* R_1 is defined by the relation $R_1(X,Y)Z := \langle Y,Z\rangle X - \langle X,Z\rangle Y$. We then set

$$\kappa_1(X,Y) := \big\langle R_1(X,Y)Y, X\big\rangle = \langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2,$$

$$\kappa(X,Y) := \langle R(X,Y)Y, X\rangle.$$

Let $\sigma \subset T_pM$ be a two-dimensional subspace, spanned by X, Y. Then the quantity

$$K_\sigma := \frac{\kappa(X,Y)}{\kappa_1(X,Y)}$$

is called the *sectional curvature* of the Riemannian manifold with respect to the plane σ.

Remark: If X, Y are orthonormal, then one has simply

$$K_\sigma = \big\langle R(X,Y)Y, X\big\rangle.$$

For the case $n = 2$ we recognize the Theorema Egregium with $K_\sigma = K$ (Gaussian curvature).

To legitimatize this definition, we must verify that K_σ depends only on σ, not on X, Y. To see this, let $\widetilde{X} = \alpha X + \beta Y$, $\widetilde{Y} = \gamma X + \delta Y$ with $\alpha\delta - \beta\gamma \neq 0$. It follows from this that $\kappa(\widetilde{X}, \widetilde{Y}) = (\alpha\delta - \beta\gamma)^2 \kappa(X,Y)$ and $\kappa_1(\widetilde{X}, \widetilde{Y}) = (\alpha\delta - \beta\gamma)^2 \kappa_1(X,Y)$.

In the case of an indefinite metric g, the sectional curvature is not defined for all planes σ, but only for the *non-degenerate* planes, i.e., those for which $\big\langle R_1(X,Y)Y, X\big\rangle \neq 0$ holds for at least one basis X, Y.

$\kappa(.\,,.)$ may be viewed as a *biquadratic form*, which is associated with the $(0,4)$-curvature tensor. It is symmetric by 6.3.1 and 6.3.5: $\kappa(X,Y) = \kappa(Y,X)$.

We recall that a symmetric bilinear form ϕ can be reconstructed from the associated quadratic form $\psi(X) := \phi(X,X)$ by means of the elementary relation (the *polarization*)

$$2\phi(X,Y) = \psi(X+Y) - \psi(X) - \psi(Y),$$

cf. [**31**], V, §7. A similar fact can be used for the curvature tensor to reduce the number of arguments from four to two.

6.5. Theorem. The curvature tensor R can be reconstructed from the biquadratic form κ (and therefore from the knowledge of all sectional curvatures).

PROOF: This turns out to be a purely algebraic consequence of the symmetries 6.3.1, 6.3.2, 6.3.4 and 6.3.5.

1*st step:* We first show that $R(X,Y)Z$ can be expressed purely through the terms of the type $R(X,Y)Y$.

$$\begin{aligned} R(X,Y+Z)(Y+Z) &= R(X,Y)Y + R(X,Y)Z + R(X,Z)Y + R(X,Z)Z, \\ -R(Y,X+Z)(X+Z) &= -R(Y,X)X + R(X,Y)Z + R(Z,Y)X - R(Y,Z)Z, \\ 0 &= R(X,Y)Z + R(Y,X)Z. \end{aligned}$$

When we add these three equations, three rows, the next to last column vanishes because of the Bianchi identity in 6.3, yielding

$$\begin{aligned} 3R(X,Y)Z = {} & R(X,Y+Z)(Y+Z) - R(Y,X+Z)(X+Z) \\ & -R(X,Y)Y - R(X,Z)Z + R(Y,X)X + R(Y,Z)Z. \end{aligned}$$

2*nd step:* By 6.3,

$$\langle R(X,Y)Y, Z\rangle = \langle R(Y,Z)X, Y\rangle = \langle R(Z,Y)Y, X\rangle,$$

hence $\langle R(.,Y)Y, .\rangle$ is for fixed Y a symmetric bilinear form. Thus, for every fixed Y, we get the equation

$$2\langle R(X,Y)Y, Z\rangle = \kappa(X+Z,Y) - \kappa(X,Y) - \kappa(Z,Y).$$

If we now combine the first and second steps, we get the formula

$$\begin{aligned} 6\langle R(X,Y)Z, V\rangle = {} & \kappa(X+V, Y+Z) - \kappa(X, Y+Z) - \kappa(V, Y+Z) \\ & -\kappa(Y+V, X+Z) + \kappa(Y, X+Z) + \kappa(V, X+Z) \\ & -\kappa(X+V, Y) + \kappa(X,Y) + \kappa(V,Y) \\ & -\kappa(X+V, Z) + \kappa(X,Z) + \kappa(V,Z) \\ & +\kappa(Y+V, X) - \kappa(Y,X) - \kappa(V,X) \\ & +\kappa(Y+V, Z) - \kappa(Y,Z) - \kappa(V,Z), \end{aligned}$$

an explicit formula for constructing R in terms of κ. □

6.6. Corollary. Suppose that the sectional curvature K_σ does not depend on the choice of σ, but only on the choice of the point p, meaning that it is a scalar function $K\colon M \to \mathbb{R}$. Then one has $R = K \cdot R_1$.

The proof is obtained immediately upon an application of 6.5, since by assumption $\kappa(X,Y) = K \cdot \kappa_1(X,Y)$ for all X, Y and since the formula for R in dependence on κ only contains additive terms of κ. Thus the equation $\kappa = K\kappa_1$ carries over to $R = KR_1$.

The assumption of 6.6 is satisfied in particular for $n = 2$. Thus there is a single curvature tensor R_1 for a two-dimensional Riemannian manifold, up to multiplication by the Gaussian curvature K. The latter is of course not necessarily constant. In contrast, in dimensions $n \geq 3$, one has the following result.

6.7. Theorem. (F. Schur 1886[1])
When the sectional curvature K_σ of a connected manifold of dimension $n \geq 3$ does not depend on the plane σ, but only on the point p at which it is calculated, then it is constant, i.e., does not depend on the point.

PROOF: First of all we have by 6.6 the relation $R(Y,Z)V = K \cdot R_1(Y,Z)V$ with a differentiable function $K\colon M \to \mathbb{R}$. By taking derivatives we get

$$\begin{aligned}(\nabla_X R)(Y,Z)V &= K \cdot (\nabla_X R_1)(Y,Z)V + X(K) \cdot R_1(Y,Z)V \\ &= X(K) \cdot R_1(Y,Z)V\end{aligned}$$

because $\nabla_X R_1 = 0$, cf. the examples in 6.2. We now wish to show that $X(K) = 0$ for all X. By cyclically permuting the arguments, we get

$$\begin{aligned}(\nabla_X R)(Y,Z)V &= X(K)\big(\langle Z,V\rangle Y - \langle Y,V\rangle Z\big),\\ (\nabla_Y R)(Z,X)V &= Y(K)\big(\langle X,V\rangle Z - \langle Z,V\rangle X\big),\\ (\nabla_Z R)(X,Y)V &= Z(K)\big(\langle Y,V\rangle X - \langle X,V\rangle Y\big).\end{aligned}$$

Now when we take the sum of these equations, the left-hand side vanishes because of the third equation in 6.3, and hence we have

$$\begin{aligned}0 = \;& \big(Z(K)\langle Y,V\rangle - Y(K)\langle Z,V\rangle\big)X\\ &+\big(X(K)\langle Z,V\rangle - Z(K)\langle X,V\rangle\big)Y\\ &+\big(Y(K)\langle X,V\rangle - X(K)\langle Y,V\rangle\big)Z\end{aligned}$$

[1] *Über den Zusammenhang der Räume konstanten Krümmungsmaßes mit den projektiven Räumen*, Math. Annalen **27**, 537–567 (1886).

for all X, Y, Z, V. By our assumption on the dimension there are three orthogonal vectors X, Y, Z. We first set $V = X$, yielding

$$0 = -Z(K)Y + Y(K)Z,$$

and consequently $Y(K) = Z(K) = 0$. Now we choose similarly $V = Y$, yielding

$$0 = Z(K)X - X(K)Z,$$

and then also $X(K) = 0$. Since at least one of the three vectors may be chosen arbitrarily, it follows that $X(K) = 0$ for every X. Thus K is locally constant, and by the connectedness of M it is globally constant. □

6.8. Definition. (Spaces of constant curvature)
If on a Riemannian manifold K_σ is a constant or, equivalently, if $R = K \cdot R_1$ with $K \in \mathbb{R}$, the manifold is called a *space of constant curvature.*

REMARK: By *scaling* one means the process of replacing a metric g by $\widetilde{g} := \lambda^2 g$, where $\lambda \neq 0$ is a constant. In this case one has $\widetilde{R}_1(X,Y)Z = \lambda^2 R_1(X,Y)Z$. On the other hand we have $\widetilde{\Gamma}_{ij}^k = \Gamma_{ij}^k$, $\widetilde{\nabla}_X Y = \nabla_X Y$ and, consequently, $\widetilde{R}(X,Y)Z = R(X,Y)Z$ as well as $\widetilde{K} = K\lambda^{-2}$. Hence in 6.8 there are (up to scaling) only three possible curvature tensors with constant curvature:

$$\begin{array}{ll} R = R_1 & (\text{with } K = 1), \\ R = 0 & (\text{with } K = 0), \\ R = R_{-1} := -R_1 & (\text{with } K = -1). \end{array}$$

Model spaces for these are the sphere S^n, the Euclidean space $\mathbb{E}^n$ and the hyperbolic space H^n, cf. Chapter 7 or (for $n = 2$) Section 3E. The de-Sitter space-time is an example of a space-time with constant negative curvature.

At this point we mention that (as will be proved later in Section 7B) the constancy of the curvature holds not only for the curvature tensor, but also locally uniquely determines the metric tensor, a generalization of Theorem 4.30 (for two-dimensional surface elements) to higher dimensions.

Theorem: Any two Riemannian metrics with the same constant sectional curvature (and the same dimension) are locally isometric to one another.

6C The Ricci tensor and the Einstein tensor

Building traces of objects is an important process in all of mathematics, as well as forming determinants, not only for algebraic problems. The divergence, which occurs in classical integral theorems, is a trace quantity, as is the Laplace operator, which occurs in important differential equations. For an orthogonal (3×3)-matrix A one can determine the angle of rotation φ from the trace alone, using the formula $\mathrm{Tr}(A) = 1 + 2\cos\varphi$. Similarly, the mean curvature is a trace. We have already discussed its importance in Section 3D. In addition, all averaged quantities formed from the curvature occur are traces, in particular traces of the curvature tensor. Therefore, it is of utmost importance to discuss at this point a general notion of *trace*. First we recall that the trace of a linear mapping of a vector space is independent of the basis chosen, by defining it as a coefficient of the characteristic polynomial, although it is usually defined as the sum of the diagonal elements of the corresponding matrix, cf. [**31**], III, §3.

6.9. Definition. (Trace of a tensor, divergence)

(i) Let A be a $(1,1)$-tensor, $A_p : T_pM \longrightarrow T_pM$. We define the *contraction* or *trace* CA by

$$CA|_p = \mathrm{Tr}(A_p) = \sum_i \big\langle A_p E_i, E_i \big\rangle,$$

where $E_1, \ldots, E_n$ is an ON-basis of T_pM. In an arbitrary basis $b_1, \ldots, b_n$ with $Ab_j = \sum_i A^i_j b_i$, the trace can be expressed by the formula $\sum_i A^i_i$ as usual.

(ii) Let A be a $(1,s)$-tensor. Then for every $i \in \{1, \ldots, s\}$ and fixed vectors $X_j, j \neq i$, $A(X_1, \ldots, X_{i-1}, -, X_{i+1}, \ldots, X_s)$ is a $(1,1)$-tensor, whose contraction (or trace) is denoted by C_iA:

$$C_iA(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_s)$$

$$= \sum_{j=1}^{n} \langle A(X_1, \ldots, X_{i-1}, E_j, X_{i+1}, \ldots, X_s), E_j \rangle.$$

C_iA is then a $(0, s-1)$-tensor.

(iii) The *divergence* of a vector field Y is defined as the trace of ∇Y, i.e.,

$$\mathrm{div}Y = C\nabla Y = \sum_i \langle \nabla_{E_i} Y, E_i \rangle.$$

(iv) The *divergence* of a symmetric $(0,2)$-tensor A is similarly defined as

$$(\mathrm{div}A)(X) = \sum_i (\nabla_{E_i} A)(X, E_i).$$

REMARKS: The *naive* taking of the trace of a matrix makes no sense in the case of $(0,2)$-tensors. For example, for the second fundamental form h_{ij} of surface elements, the expression $\sum_i h_{ii}$ is not invariant under parameter transformations, since it always vanishes in asymptotic line parameters on hyperbolic surface elements, cf. 3.18. Instead, one must take the trace of the associated $(1,1)$-tensor: let A be a $(0,2)$-tensor and let $A^\#$ be the associated $(1,1)$-tensor determined by the relation $A(X,Y) = \langle A^\# X, Y \rangle = g(A^\# X, Y)$. We then set $\mathrm{Tr}_g A := \mathrm{Tr} A^\#$. For this reason, one also often writes $\mathrm{Tr}_g A$ instead of just $\mathrm{Tr}A$ in order to indicate that the forming of the trace is with respect to the metric g. In particular one has $\mathrm{Tr}_g(g) = n$.

In Ricci notation, $\mathrm{Tr}(A^i_j)$ is simply denoted by A^i_i, with, as always, a summation over i. Similarly, the ith contraction of $A^r_{j_1 \ldots j_s}$ is denoted by $A^m_{j_1 \ldots j_{i-1} m j_{i+1} \ldots j_s}$ (in this case with summation over m). Also, for the divergence one writes $\mathrm{div}(\eta^j) = \nabla_i \eta^i$.

In the case of an indefinite metric, one needs to take account of the fact that in an ON-basis $E_1, \ldots, E_n$ with $\langle E_i, E_j \rangle = \delta_{ij}\epsilon_i$, a vector X has the representation $X = \sum_i \epsilon_i \langle X, E_i \rangle E_i$. Consequently, one gets as a formula for the trace

$$\mathrm{Tr}(A) = \sum_i A^i_i = \sum_i \epsilon_i \langle A E_i, E_i \rangle.$$

EXAMPLES:

1. Suppose that a vector field Y (which is nothing but a (1,0)-tensor) is defined on an open subset of Euclidean space. Then the above definition of the divergence coincides with the expression $\sum_i \frac{\partial Y^i}{\partial x^i}$, where Y^i is the i-th component of Y. This is the classical divergence in $\mathbb{R}^n$ which occurs in vector calculus [**29**].

2. In particular, for $Y = \mathrm{grad} f$, one calls $\Delta f := C(\nabla Y) = \mathrm{div} Y = \mathrm{div}(\mathrm{grad} f) = \nabla_i f^i$ the *Laplace-Beltrami operator* of f. The classical version of this in $\mathbb{R}^n$ is the *Laplacian* $\Delta f = \sum_i \frac{\partial^2 f}{(\partial x^i)^2}$.

3. In particular, the trace of the Weingarten mapping (or the trace of the second fundamental form with respect to the first fundamental form) is just nH (i.e., up to a multiplicative factor n, the mean curvature, cf. 3.13).

6.10. Definition. (Ricci tensor, scalar curvature)
The first contraction of the curvature tensor $R(X,Y)Z$ is given by the expression

$$(C_1R)(Y,Z) = \mathrm{Tr}\ (X \longmapsto R(X,Y)Z) = \sum_i \langle R(E_i,Y)Z, E_i \rangle$$

and is called the *Ricci tensor* $\mathrm{Ric}(Y,Z)$, or briefly, $\mathrm{Ric} = C_1R$. In Ricci notation one gets from the special order of the indices the equation $R_{jk} = R^i_{jik}$, which is formally in a sense the second contraction instead of the first. The Ricci tensor is symmetric, hence $\mathrm{Ric}(Y,Z) = \mathrm{Ric}(Z,Y)$, because of the symmetries of R, cf. the second step in the proof of 6.5.

The trace of the Ricci tensor is called the *scalar curvature* S. One has

$$S = \sum_{i,j} \langle R(E_i,E_j)E_j, E_i \rangle$$

(one could also view this as the second iterated trace of the curvature tensor). In Ricci notation we have $S = R^j_j = R_{jk}g^{kj} = R^i_{jik}g^{kj}$, which is why one also writes R instead of S. At this point that could be misleading, as the symbol R is also used for the curvature tensor in Koszul calculus.

REMARKS:

1. By construction, the Ricci tensor is a mean value of other curvatures, however without any kind of normalization factor as in the arithmetic mean. More precisely, for every unit vector X, the value $\mathrm{Ric}(X,X)$ is the sum of all $n-1$ sectional curvatures in planes which contain X and are orthogonal to one another.

The scalar curvature $S = \sum_{i \neq j} K_{ij}$ is then the sum of *all* sectional curvatures in the (i, j)-planes of an ON-basis. It describes the volume distortion of the exponential map, cf. Theorem 7.16. Note, however, that for $n \geq 4$ the collection of all these $\binom{n}{2}$ sectional curvatures K_{ij} for a fixed ON basis of each tangent space is *not* sufficient to determine the curvature tensor uniquely, cf. Exercise 19 at the end of the chapter.

2. In the local theory of hypersurfaces, we have already met the scalar curvature as the second elementary symmetric function (cf. 4.22) of the principal curvatures κ_i. Hence $S = \sum_{i \neq j} \kappa_i \kappa_j$. Similarly, we get $\mathrm{Ric}(E_i, E_i) = \kappa_i \sum_{j \neq i} \kappa_j$, where the E_i are the principal curvatures. All of this is based on the Gauss equation

$$K_{ij} = \big\langle R(E_i, E_j)E_j, E_i \big\rangle = \big\langle LE_j, E_j \big\rangle \big\langle LE_i, E_i \big\rangle$$
$$- \big\langle LE_i, E_j \big\rangle \big\langle LE_j, E_i \big\rangle = \kappa_i \kappa_j.$$

3. The solutions of the evolution equation $\frac{dg}{dt} = -2\mathrm{Ric}$ for the so-called *Ricci flow* modifies the metric g as a function of the "time" t in such a way that the curvature can be controlled, even for the transition $t \to \infty$. Recently this method led to spectacular consequences in the theory of 3-manifolds, in particular to Perelman's proof of the famous and long-standing Poincaré conjecture from 1904.[2]

6.11. Lemma. One has the commutativity $C_i(\nabla_X A) = \nabla_X (C_i A)$ for every $(1, s)$-tensor A. In Ricci calculus this is expressed by the same symbol $\nabla_k A^i_{j_1 \ldots i \ldots j_s}$ for both sides.

Proof: First let A be a $(1,1)$-tensor (i.e., $s = 1$). We start from the definitions

$$CA = \sum_i \langle AE_i, E_i \rangle,$$
$$\nabla_X(CA) = \sum_i \Big[\langle \nabla_X(AE_i), E_i \rangle + \langle AE_i, \nabla_X E_i \rangle \Big],$$

[2]see J.W.Morgan, *Recent progress on the Poincaré conjecture and the classification of 3-manifolds*, Bulletin AMS **42**, 57–78 (2005) and J.Stillwell, *Poincaré and the early history of 3-manifolds*, Bulletin AMS **49** , 555–576 (2012).

$$C(\nabla_X A) = \sum_i \Big\langle (\nabla_X A)(E_i), E_i \Big\rangle$$
$$= \sum_i \Big[\Big\langle \nabla_X (AE_i), E_i \Big\rangle - \Big\langle A(\nabla_X E_i), E_i \Big\rangle \Big].$$

Applying the connection forms ω_j^i (cf. Section 4F), we have the equation $\nabla_X E_i = \sum_j \omega_i^j(X) E_j$ with $\omega_j^i + \omega_i^j = 0$. Inserting this, we get

$$\sum_i (\langle AE_i, \nabla_X E_i \rangle + \langle A(\nabla_X E_i), E_i \rangle)$$
$$= \sum_i \langle AE_i, \sum_j \omega_i^j(X) E_j \rangle + \sum_i \Big\langle A\Big(\sum_j \omega_i^j(X) E_j\Big), E_i \Big\rangle$$
$$= \sum_{i,j} \omega_i^j(X) \langle AE_i, E_j \rangle + \sum_{j,i} \underbrace{\omega_j^i(X)}_{-\omega_i^j(X)} \langle AE_i, E_j \rangle = 0.$$

For tensors of higher degree ($s > 1$), simply keep the arguments which are not involved in the contraction fixed. The same reasoning as above then applies to the remaining $(1,1)$-tensor. For the proof in Ricci calculus one would have to distinguish between $(\nabla_k A)_i^i$ on the one hand and $\nabla_k(A_i^i)$ on the other. Then one gets

$$(\nabla_k A)_i^i = \frac{\partial A_i^i}{\partial x^k} + \Gamma_{kl}^i A_i^l - \Gamma_{ki}^m A_m^i = \frac{\partial A_i^i}{\partial x^k} = \nabla_k(A_i^i)$$

for a (1,1)-tensor; similarly for the others. □

6.12. Definition. (Einstein space)
A Riemannian manifold (M, g) is called an *Einstein space* (in which case g is referred to as an *Einstein metric*), if the Ricci tensor is a multiple of the metric g:

$$\mathrm{Ric}(X, Y) = \lambda \cdot g(X, Y)$$

for all X, Y, with a function $\lambda : M \to \mathbb{R}$. Taking the traces, we see that $S = n\lambda$.

The relation $\mathrm{Ric}(X, X) = \lambda g(X, X)$ for all X is equivalent to this, with a function $\lambda : M \to \mathbb{R}$. The expression

$$\mathrm{ric}(X) := \frac{\mathrm{Ric}(X, X)}{g(X, X)}$$

is also called the *Ricci curvature* in the direction X. Einstein spaces are also those Riemannian manifolds for which the Ricci curvature only depends on the point, but not on the direction X. Formulated differently, Einstein spaces are characterized by the property that all eigenvalues of the Ricci tensor with respect to the metric g are equal to one another.

EXAMPLES:

(i) Formally, for $n = 2$ every metric is an Einstein metric, because $R = K \cdot R_1$, from which it follows that

$$\mathrm{Ric}(X, X) = K \cdot \sum_{i=1}^{2} g(R_1(E_i, X)X, E_i) = K \cdot g(X, X).$$

(ii) Spaces of constant curvature K are also Einstein spaces for the same reason. According to 6.7 and 6.8 we have $R = K \cdot R_1$ and $C_1 R = K \cdot (n-1)g$ (note that $S = n(n-1)K$).

(iii) The cartesian product of two two-dimensional manifolds with constant and equal Gaussian curvature is a four-dimensional Einstein space. In particular this holds for $M = S^2 \times S^2$. The reason for this is the block-matrix structure of the Ricci tensor. For more details, see Chapter 8.

(iv) Any Riemannian manifold M is an Einstein space if the following condition is satisfied: For any two points p, q there is an isometry of M into itself carrying p into q, and for any two unit tangent vectors X, Y at p there is an isometry which fixes p and which carries X into Y. Such spaces are also called *isotropy-irreducible homogeneous spaces*, cf. Section 8C.

6.13. Theorem. Let (M, g) be a connected Einstein space of dimension $n \geq 3$ with $\mathrm{Ric} = \lambda \cdot g$. Then λ is constant, and if $n = 3$, then (M, g) is even a space of constant curvature.

This theorem is actually an amazing result. There is a certain analogy with the results 3.14 and 3.47 for surfaces which consist solely of umbilics. If all eigenvalues at every point coincide, then they are actually constant on the entire manifold. However there are quite different laws for the second fundamental form than there are for the

Ricci tensor. Still, a result like this is in a sense typical for differential geometry. It is based on hidden dependencies between the quantities considered which occur upon taking higher derivatives. This holds also for the Schur theorem in 6.7. For the proof of 6.13 we require first a few properties of the tensor $\nabla_X R$, which we formulate as a lemma. In connection with this we will be led to the divergence-freeness of the Einstein tensor.

6.14. Lemma. The algebraic symmetries 6.3 (1), (4) and (5) are preserved upon taking the covariant derivatives of the curvature tensor, i.e., the following equations hold:

$$(\nabla_X R)(Y,Z)V = -(\nabla_X R)(Z,Y)V;$$
$$\langle(\nabla_X R)(Y,Z)V,U\rangle = -\langle(\nabla_X R)(Y,Z)U,V\rangle;$$
$$\langle(\nabla_X R)(Y,Z)V,U\rangle = \langle(\nabla_X R)(V,U)Y,Z\rangle.$$

Moreover, for every X we have

$$\operatorname{Tr}\big(\nabla_X \mathrm{Ric}\big) = 2\cdot \operatorname{div}(\mathrm{Ric})(X).$$

Proof: The algebraic symmetries follow immediately from the definition of the operator $\nabla_X R$. For the trace we calculate from the equation $C_1(\nabla_X R) = \nabla_X(C_1 R)$ in 6.11, using the symmetries of the operators,

$$\begin{aligned}
\operatorname{Tr}(\nabla_X \mathrm{Ric}) &= \operatorname{Tr}(C_1(\nabla_X R)) = \sum_{i,j} \big\langle(\nabla_X R)(E_i,E_j)E_j,E_i\big\rangle \\
&\overset{6.3.3}{=} -\sum_{i,j}\Big(\big\langle(\nabla_{E_i} R)(E_j,X)E_j,E_i\big\rangle + \big\langle(\nabla_{E_j} R)(X,E_i)E_j,E_i\big\rangle\Big) \\
&= \sum_{i,j}\Big(\big\langle(\nabla_{E_i} R)(E_j,X)E_i,E_j\big\rangle + \big\langle(\nabla_{E_j} R)(E_i,X)E_j,E_i\big\rangle\Big) \\
&= 2\sum_{i,j}\big\langle(\nabla_{E_i} R)(E_j,X)E_i,E_j\big\rangle = 2\sum_i \underbrace{C_1(\nabla_{E_i} R)}_{=\nabla_{E_i}(C_1 R)}(X,E_i) \\
&= 2\cdot\sum_i(\nabla_{E_i}\mathrm{Ric})(X,E_i) = 2\cdot\operatorname{div}(\mathrm{Ric})(X). \qquad \square
\end{aligned}$$

6.15. Definition and Theorem. (Einstein tensor)
The *Einstein tensor* G is defined as $G = \mathrm{Ric} - \frac{S}{2}g$. On an arbitrary Riemannian manifold, the divergence of the Einstein tensor vanishes, i.e.,

$$\mathrm{div}(\mathrm{Ric}) = \mathrm{div}\Big(\frac{S}{2}g\Big).$$

In Ricci notation this is indicated by the equation $\nabla^i G_{ji} = \nabla_i G_{jk} g^{ki} = 0$ with $G_{jk} = R_{jk} - \frac{S}{2}g_{jk}$.

For space-times, the tensor G is also referred to as the *Einstein gravitation tensor*, see for example [**22**], p. 336. The Einstein tensor should not be confused with the *traceless Ricci tensor* $\mathrm{Ric} - \frac{S}{n}g$. These two coincide only in the case $n = 2$, in which case they both vanish identically. The divergence-freeness of G is trivial in case S is constant and if the Ricci tensor is a constant multiple of the metric, since the metric is divergence-free because g is parallel, that is, $\nabla g = 0$. The Einstein tensor is important for the theory of gravitation, as it occurs in the Einstein field equations, cf. section 8B. It arises as the gradient of the Hilbert-Einstein functional, see 8.2 and 8.6. A space with vanishing Einstein tensor is called a *special Einstein space*. By taking traces one gets in this case $\mathrm{Ric} = 0$ (provided $n \geq 3$). This equation is satisfied for example by the Schwarzschild metric, cf. the exercises at the end of Chapter 5. For more details on this see Chapter 13 in [**22**].

PROOF OF 6.15: By Lemma 6.14 one has $\big(\mathrm{div}(\mathrm{Ric})\big)(X) = \frac{1}{2}\mathrm{Tr}\nabla_X \mathrm{Ric}$ and, furthermore,

$$\frac{1}{2}\mathrm{div}(Sg)(X) = \frac{1}{2}\sum_i \Big(\nabla_{E_i}(Sg)\Big)(X, E_i)$$

$$= \frac{1}{2}\sum_i (\nabla_{E_i} S) g(X, E_i) = \frac{1}{2}(\nabla_X S).$$

As in 6.11 we obtain $\mathrm{Tr}\nabla_X \mathrm{Ric} = \nabla_X(\mathrm{Tr}(\mathrm{Ric})) = \nabla_X S$ because the ω_i^j are skew-symmetric in i and j and $\mathrm{Ric}(E_i, E_j)$ symmetric. Therefore the difference term $\sum_i (\nabla_X \mathrm{Ric})(E_i, E_i) - \sum_i \nabla_X \big(\mathrm{Ric}(E_i, E_i)\big)$ $= -\sum_{i,j} \omega_i^j(X)\big(\mathrm{Ric}(E_j, E_i) + \mathrm{Ric}(E_i, E_j)\big)$ vanishes. $\square$

Proof of 6.13: By 6.15 we have $\operatorname{div}(\mathrm{Ric}) = \operatorname{div}\big(\frac{S}{2}g\big)$. Inserting for the Ricci tensor $\mathrm{Ric} = \lambda g$ and consequently $S = n\lambda$, we get

$$\operatorname{div}(\lambda g)(X) = \operatorname{div}\Big(\frac{n\lambda}{2}g\Big)(X) = \frac{n}{2}\operatorname{div}(\lambda g)(X).$$

Because $\nabla g = 0$, the left-hand side is equal to $\sum_i (E_i(\lambda))g(X, E_i) = X(\lambda)$, so we get

$$X(\lambda) = \frac{n}{2}X(\lambda)$$

for arbitrary X, which implies that either $n = 2$ or $X(\lambda) = 0$ for all X (provided $n \geq 3$). Thus, for $n \geq 3$ the function λ is locally constant and, since M is assumed to be connected, even globally constant. In case $n = 3$, the Ricci tensor alone determines the sectional curvature. This can be seen as follows: in an ON-basis E_1, E_2, E_3 we have

$$\begin{aligned}
\mathrm{Ric}(E_1, E_1) &= K_{12} + K_{13}, \\
\mathrm{Ric}(E_2, E_2) &= K_{12} + K_{23}, \\
\mathrm{Ric}(E_3, E_3) &= K_{13} + K_{23}.
\end{aligned}$$

If our Ricci tensor on the left-hand side is given, then these are three equations for three indeterminants, namely for the sectional curvatures $K_{ij}, i < j$. These equations have a unique solution since the rank of the corresponding matrix is maximal. If the three left-hand sides are each equal to λ, we get $K_{12} = K_{13} = K_{13} = \frac{\lambda}{2}$. Since this holds in an arbitrary ON-basis, the sectional curvature is constant. $\square$

6.16. Special case. (Einstein hypersurfaces)
Let $M \subset \mathbb{R}^{n+1}$ be a connected hypersurface such that the first fundamental form is an Einstein metric, and let $n \geq 3$. Then at any point there are at most two distinct principal curvatures, and at most one of them is not zero. In any case M is either a part of a hypersphere S^n or it is isometrically developable into Euclidean space $\mathbb{R}^n$. In particular M has constant positive or vanishing sectional curvature.

Sketch of proof: By Theorem 6.13 we have $\mathrm{Ric} = \lambda g$ with a constant λ. Now let $E_1, \ldots, E_n$ be the principal curvature directions with corresponding principal curvatures $\kappa_1, \ldots, \kappa_n$. By 4.21 we calculate the sectional curvature K_{ij} in the (E_i, E_j)-plane as $K_{ij} = \kappa_i\kappa_j$. From this we obtain the Ricci tensor

$$\mathrm{Ric}(E_i, E_i) = \kappa_i(\kappa_1 + \cdots + \kappa_{i-1} + \kappa_{i+1} + \cdots + \kappa_n) = \lambda,$$

$$\mathrm{Ric}(E_i, E_j) = 0 \text{ for } i \neq j.$$

For the mean curvature H, i.e. $nH = \sum_i \kappa_i$, we conclude that the equation

$$\kappa_i nH - \kappa_i^2 = \lambda$$

holds for any κ_i. Therefore any κ_i satisfies at any point of M the quadratic equation

$$x^2 - nHx + \lambda = 0,$$

which has at most two distinct solutions. If there is only one solution in an open set then M is a part of a hyperplane or a hypersphere by 3.47. If we have two distinct principal curvatures $\kappa, \bar{\kappa}$ at a certain point (and thus in a certain neighborhood), then we have constant multiplicities p and q of κ and $\bar{\kappa}$ throughout that neighborhood, where $p + q = n$. It follows that

$$\kappa + \bar{\kappa} = nH, \quad \kappa\bar{\kappa} = \lambda, \quad p\kappa + q\bar{\kappa} = nH$$

and furthermore

$$(p-1)\kappa = -(q-1)\bar{\kappa}.$$

In particular we have $\lambda \leq 0$ in this case. If one of the principal curvatures vanishes, say $\bar{\kappa} = 0$, then we have $p = 1$ and $q = n - 1$. consequently the principal curvatures are

$$\kappa, 0, \ldots, 0.$$

Extrinsically it is true that through each point of M there is an $(n-1)$-dimensional affine-linear subspace (ruling) in M. In the case $n = 2$ this corresponds to the ruling of a ruled surface. Along these subspaces the Gauss map is constant, compare Exercise 24. By the same argument as in 3.24 it follows that M is isometrically developable into Euclidean n-space. We omit the details here.

If κ and $\bar{\kappa}$ both are non-zero, then we get $\bar{\kappa}/\kappa = -(p-1)/(q-1)$. Thus their quotient is constant. On the other hand, the product $\kappa\bar{\kappa} = \lambda$ is also constant, hence κ and $\bar{\kappa}$ are both constant, they have different signs, and their multiplicities satisfy $p, q \geq 2$ because of $(p-1)\kappa = -(q-1)\bar{\kappa}$. As a matter of fact (not proved here) such a hypersurface in $\mathbb{R}^{n+1}$ cannot exist, not even locally.

Exercises.

1. Let $\mathcal{X}(M)$ denote the set of all differentiable vector fields on M. Show that an $\mathbb{R}$-multilinear map A of $\mathcal{X}(M) \times \cdots \times \mathcal{X}(M)$ to the set of all scalar functions on M is a $(0, s)$-tensor field if and only if for an arbitrary scalar function $f_1, \ldots, f_s$ on M at each point p the equation

$$A(f_1 \cdot X_1, \ldots, f_s \cdot X_s)\big|_p = f_1(p) \cdot \cdots \cdot f_s(p) \cdot A(X_1, \ldots, X_s)\big|_p$$

holds. Conversely, every $(0, s)$-tensor field can be obtained in this manner.

2. Verify with the result from Exercise 1 that the curvature tensor is actually a tensor field.

3. Derive the transformation rules for the Christoffel symbols under a change of coordinates, and conclude from this that the association $X, Y \mapsto \nabla_X Y$ is *not* a tensor.

4. Show that the difference of two connections ∇ and $\widetilde{\nabla}$ on a manifold is a $(1, 2)$-tensor field.

5. Verify the symmetries of the curvature tensor R_{ijkl} listed in Lemma 6.3 directly using Ricci notation.

6. Check the equation in 6.15 using Ricci notation, i.e., show that $\nabla^i R_{ji} = \frac{1}{2}\nabla^i(S g_{ij})$.

7. Derive an equation to obtain the Ricci tensor from the Einstein tensor.

8. Let $\mathrm{ric}(X)$ denote the Ricci curvature in the direction of a unit vector $X \in S^{n-1} \subset T_pM$ (cf. 6.12). We endow this unit sphere with the usual (induced from Euclidean space) volume element dV, so that $\int_{S^{n-1}} dV = \mathrm{Vol}(S^{n-1})$. Let S denote the scalar curvature. Show that the scalar curvature at a point p is the averaged integral of all Ricci curvatures, and hence

$$S(p) \cdot \mathrm{Vol}(S^{n-1}) = \int_{S^{n-1}} \mathrm{ric}(X) dV.$$

Hint: Use an eigenbasis of the Ricci tensor.

9. Let $f\colon (M^n, g) \to (\widetilde{M}^{n+1}, \widetilde{g})$ be an isometric immersion, i.e., $\widetilde{g}(DF(X), Df(Y)) = g(X, Y)$ for all tangent vectors X, Y to M. Define a Gaussian normal map and a second fundamental form II as a symmetric bilinear form on the tangent space of M, and derive from this as in 4.18 a Gauss equation in the form

$$\begin{aligned} g(R(X,Y)Z,V) = \;& \widetilde{g}(\widetilde{R}(X,Y)Z,V) \\ & + II(Y,Z)II(X,V) - II(X,Z)II(Y,V). \end{aligned}$$

Which version of the Theorema Egregium follows from this for surfaces M^2 in the standard three-sphere $\widetilde{M} = S^3$?

10. Let a Riemannian manifold $(M, {}^*g)$ be given, as well as a smooth function $f\colon \mathbb{R} \longrightarrow (0, \infty)$. We consider the warped product $\mathbb{R} \times_{f^2} M$ with the metric

$$g(t, x^1, \ldots, x^n) = dt^2 + f^2(t) \cdot {}^*g(x^1, \ldots, x^n),$$

cf. Exercise 12 at the end of Chapter 5. Show the following:
   a) $\nabla_{\frac{\partial}{\partial t}} \frac{\partial}{\partial t} = 0.$
   b) $\nabla_{\frac{\partial}{\partial t}} \frac{\partial}{\partial x^i} = \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial t} = \frac{f'}{f} \frac{\partial}{\partial x^i}.$
   c) $\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = -\frac{f'}{f} g_{ij} \frac{\partial}{\partial t} + {}^*\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}.$

   Here ${}^*\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}$ is an abbreviated notation and denotes the application of the Riemannian connection ${}^*\nabla$ composed with the natural projection of $\mathbb{R} \times M$ to M.

11. Let R_1 denote the standard curvature tensor from Definition 6.4. For the curvature tensor of a warped product (cf. Exercise 10), one has:
   a) $R(\frac{\partial}{\partial t}, \frac{\partial}{\partial x^i}) \frac{\partial}{\partial t} = \frac{f''}{f} \frac{\partial}{\partial x^i}.$
   b) $R(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) \frac{\partial}{\partial t} = 0.$
   c) $R(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial t}) \frac{\partial}{\partial x^j} = \frac{f''}{f} g_{ij} \frac{\partial}{\partial t}.$
   d) $R(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) \frac{\partial}{\partial x^k} = {}^*R(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) \frac{\partial}{\partial x^k} - \frac{f'^2}{f^2} R_1(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) \frac{\partial}{\partial x^k}.$

   Hint: Use the Gauss equation of Exercise 9.

12. For which functions f does the warped product of Exercises 10 and 11 lead to a space of constant curvature or an Einstein space, respectively?
Hint: For 2-dimensional manifolds this is not interesting, because the metric tensor of a warped product looks like the first fundamental form of a surface of rotation, compare 3.17. The Einstein condition is trivial in this case. For dimensions $n \geq 3$ one can derive from 11 d) that $(M, {}^*g)$ itself is a space of constant curvature or an Einstein space, respectively. This is a necessary condition. Furthermore prove by 11 a) and 11 c) that f''/f is constant. Therefore f satisfies the differential equation $f'' + cf = 0$ of the harmonic oscillator with a constant c that depends only of the scalar curvature. The solutions of this ODE can be found in 3.17.

13. Conclude from the solution to Exercise 12: A 4-dimensional warped product is Einstein if and only if the sectional curvature is constant. This is not true for higher dimensions.

14. Prove that every principal curvature direction of a hypersurface element in $\mathbb{R}^{n+1}$ is also an eigenvector of the Ricci tensor.

15. For which 3-dimensional hypersurfaces of rotation in $\mathbb{R}^4$ is the first fundamental form an Einstein metric? A hypersurface of rotation is defined by rotation of a regular curve in some $\mathbb{R}^2 \subset \mathbb{R}^4$ around the orthogonal 2-plane. Thus every point of the curve will be replaced by a 2-dimensional sphere of a certain radius.
Hint: Show that the principal curvatures coincide with those of a surface of revolution in 3-space, generated by the same curve. However, the multipicities are different.

16. Calculate the Ricci tensor of the 4-dimensional catenoid from Exercise 9 in Chapter 4, for example as a square matrix in a suitable basis.

17. Let (M_1, g_1) and (M_2, g_2) be two Einstein spaces of dimensions n_1 and n_2 with scalar curvatures S_1 and S_2 and the Einstein constants $\lambda_1 = S_1/n_1$ and $\lambda_2 = S_2/n_2$. Compare Exercise 6 in Chapter 5.

Prove the following:

(a) The cartesian product $M = M_1 \times M_2$ with the product metric $g = g_1 \times g_2$ is Einstein is and only if $\lambda_1 = \lambda_2$ holds.

(b) In particular the cartesian product of an Einstein space (M, g) with a euclidean space $I\!R^n$ is Einstein if and only if (M, g) is Ricci flat, i.e., if $\mathrm{Ric} = 0$ holds.

Hint: Use the structure of a block matrix for the Ricci tensor of the product metric.

18. Prove that the cartesian product of two Einstein spaces has always a parallel Ricci tensor, i.e., $\nabla \mathrm{Ric} = 0$.

19. Consider cartesian product $M = M_1 \times M_2$ of the 2-sphere $M_1 = S^2$ (with curvature $K = 1$) and the hyperbolic plane $M_2 = H^2$ (with curvature $K = -1$). Let X_1, X_2 be an orthonormal basis in the tangent space of a point $q \in S^2$ and, similarly, Y_1, Y_2 an orthonormal basis at a point $r \in H^2$. Show the following:

(a) The vectors $E_1 = X_1 + Y_1$, $E_2 = X_1 - Y_1$, $E_3 = X_2 + Y_2$, $E_4 = X_2 - Y_2$ form a basis of $T_pM \cong T_qM_1 \oplus T_rM_2$ with $p = (q, r) \in M_1 \times M_2$.

(b) The sectional curvature in each of the six (E_i, E_j)-planes with $1 \leq i < j \leq 4$ vanishes but the curvature tensor does not vanish identically.

This example shows that in dimensions 4 and higher the curvature tensor is *not* uniquely determined by the sectional curvatures in the coordinate planes of a coordinate system (or an ON basis). This is not quite consistent with a certain claim in Riemann's famous Habilitationsvortrag.[3]

20. Prove that a Riemannian manifold (M, g) is an Einstein space if and only if the following $(0, 4)$-tensor vanishes identically:

$$(X, Y, Z, V) \longmapsto \mathrm{Ric}(R_1(X, Y)Z, V) + \mathrm{Ric}(R_1(X, Y)V, Z)$$

Here R_1 denotes the standard curvature tensor from Def. 6.4.
Hint: One direction is trivial by the symmetries of the curvature tensor. For the other direction consider an eigenbasis of the Ricci tensor with respect to the metric tensor and show that all eigenvalues have to coincide.

[3]Compare Antonio J. Di Scala, *On an assertion in Riemann's Habilitationsvortrag*, L'Enseignement Mathématique (2) **47**, 57–63 (2001).

21. Show that a parallel 1-form ω can be locally written as a differential $\omega = df$ of a scalar function. By analogy with Definition 5.17 a 1-form ω is called *parallel* if $\nabla\omega = 0$ holds.
Hint: Integrability conditions $d\omega = 0$ as for 1-forms in Section 4F. Show first that the exterior derivative $d\omega$ (as well as df) is independent of the Riemannian metric.

22. Show that the Hessian of a scalar function f on a Riemannian manifold is symmetric, i.e., the equation
$$\nabla^2 f(X,Y) = \nabla^2 f(Y,X)$$
holds for all tangent vectors X, Y.

23. Show that the Hesse $(1,1)$-tensor on a Riemannian manifold is self-adjoint with respect to the Riemannian metric. Consequently all its eigenvalues are real. Conclude that at a local minimum (or maximum, respectively) of the function f of the Hessian has only non-negative (or non-positive) eigenvalues.

24. Let $c\colon I \to \mathbb{R}^{n+1}$ be a differentiable curve and $X_2, \ldots, X_n$ a differentiable orthonormal frame in $\mathbb{R}^{n+1}$ along c. Then the expression $f(u, v^2, \ldots, v^n) = c(u) + \sum_{i=2}^{n} v^i X_i(u)$ defines a hypersurface with an $(n-1)$-dimensional ruling. It can be regarded as the span of $n-1$ ruled surfaces $c(u) + v^i X_i(u)$ with the same directrix c. Show the following:
(a) The hypersurface is regular in a neighborhood of $v^2 = \cdots = v^n = 0$ if the tangent $\dot{c}$ is never contained in the span of $X_2, \ldots, X_n$ (compare the tangent surfaces in 3.24).
(b) The rank of the second fundamental form is at most 2.
(c) The rank is at most 1 if and only if the unit normal ν is constant along the $(n-1)$-dimensional rulings.
This case occurs in the proof of 6.16 as the case of an Einstein hypersurface with two distinct principal curvatures. The case of rank 2 cannot occur for Einstein hypersurfaces since no ruled surface can have constant Gaussian curvature K unless $K = 0$.
Hint: Proof of (2) $\Leftrightarrow$ (3) in 3.24: $\frac{\partial\nu}{\partial v^i} = 0 \Longleftrightarrow h_{1i} = 0$.

Chapter 7

Spaces of Constant Curvature

For any given quantity derived from curvature, it is natural to inquire as to the meaning of this quantity reducing to a constant on a manifold. Therefore, this chapter is concerned with Riemannian manifolds whose sectional curvature K is constant, or, what amounts to the same thing, for which the curvature tensor satisfies an equation $R = KR_1$, where R_1 denotes the curvature of the unit sphere (cf. 6.8) and K is a constant. One is also led to these spaces when one considers the problem of free motion of rigid bodies, cf. 7.6. Helmholtz had postulated such a motion in the nineteenth century from a physical argument. Of course Euclidean space and the sphere are both spaces of this kind. There are in fact other examples (other than open sets of the two mentioned spaces). Determining all of these spaces is what is known as the *space form problem.* The problem of finding a space with sectional curvature $K = -1$ (as the natural complement to the case of the sphere) was unsolved for a long time; its solution is given by hyperbolic space. We now investigate this case, and define hyperbolic space as a hypersurface in a pseudo-Euclidean space. In dimension two, we have already done this in Section 3E. So we are now in the pleasant position of just having to extend the methods developed there to the n-dimensional case, a process which is greatly simplified by the Gauss equation and the theorems in Section 6B on the curvature tensor. A fundamental result, presented in Section 7B, is the local isometry of two arbitrary metrics with the same constant sectional curvature. In Sections 7C and 7D we take up the space form problem, in particular in dimensions two and three.

7A Hyperbolic space

7.1. Pseudo-Euclidean space $\mathbb{R}^n_k$.

The underlying manifold $\mathbb{R}^n$ can be endowed with metrics completely different from the Euclidean one. For example, we consider here the so-called *pseudo-Euclidean* metric (or pseudo-Euclidean inner product)

$$g(X,X) = \langle X,X\rangle_k = -\sum_{i=1}^{k} x_i^2 + \sum_{i=k+1}^{n} x_i^2$$

for a vector X with components $x_1,\ldots,x_n$, where $0 \le k \le n$ is a fixed number which is called the *index* or the *signature* of the inner product. The pair $(\mathbb{R}^n, g)$ is then called a *pseudo-Euclidean space* and is denoted by $\mathbb{R}^n_k$ or E^n_k. In particular, $E^n = \mathbb{R}^n_0$ is the usual Euclidean space. The decisive difference from Euclidean space is the fact that there are three different types of vectors, other than the zero vector, depending on the value of the pseudo-Euclidean inner product:

$$\begin{array}{lll}
X \text{ is called } \textit{space-like}, & \text{if} & g(X,X) > 0, \\
X \text{ is called } \textit{time-like}, & \text{if} & g(X,X) < 0, \\
X \text{ is called } \textit{light-like} \text{ or } \textit{null}, & \text{if} & g(X,X) = 0, \text{ but } X \neq 0.
\end{array}$$

In Cartesian coordinates one then has $g_{ij} = \epsilon_i \delta_{ij}$, with $\epsilon_i = -1$ for $i \le k$ and $\epsilon_i = +1$ for $i > k$. Thus all Christoffel symbols vanish, and consequently the curvature tensor $R(X,Y)Z$ vanishes identically. Metrics for which this holds are called *flat*.

7.2. The sphere S^n.

The sphere with its spherical metric is most easily defined as a hypersurface in Euclidean space with the associated first fundamental form, that is,

$$S^n := \Big\{ X \in \mathbb{R}^{n+1} \mid \langle X,X\rangle = \sum_i x_i^2 = 1 \Big\}.$$

Then $S^n \subseteq I\!R^{n+1}$ is a hypersurface with Weingarten mapping $L = \pm\mathrm{Id}$ (cf. 3.10): for a local parametrization $f(u)$ one has for the unit normal ν the relation $\nu(u) = \pm f(u)$, hence $D\nu = \pm Df$. The Gauss equation 4.19 then implies the relation $R(X,Y)Z = R_1(X,Y)Z$, and the sectional curvature is consequently $K_\sigma = +1$ at every point p and every plane $\sigma \subseteq T_pS^n$. For a sphere of radius r,

$$S^n(r) = \Big\{X \in I\!R^{n+1} \,\big|\, \sum_i x_i^2 = r^2\Big\},$$

one similarly has $R = \frac{1}{r^2}R_1$ and $K_\sigma = \frac{1}{r^2}$ for every plane σ. This spherical metric can also be given in local coordinates without any use of an ambient space, cf. 7.7.

7.3. Hyperbolic space H^n.

There are several different possible ways of describing hyperbolic space. In dimension two, the Poincaré upper half-plane $\{(x,y) \mid y > 0\}$ with the metric $g_{ij} = y^{-2}\delta_{ij}$ is an often used model, as is the so-called disc model, cf. Figure 7.2 at the end of Section 7A. Yet it is more satisfying and simpler to define hyperbolic space as a hypersurface, as was already done in Section 3E; this cannot be done in Euclidean space (for $n \geq 3$ not even locally, cf. Exercise 1), but it can be done in a pseudo-Euclidean space. More precisely, consider a "sphere with an imaginary radius"

$$\left\{X \in I\!R_1^{n+1} \,\Big|\, -x_0^2 + \sum_{i=1}^{n} x_i^2 = -1\right\}.$$

Viewed with Euclidean eyes, this is nothing but a two-sheeted hyperboloid, cf. Figure 3.23. The analogy with the case of the sphere considered in Section 7.2 becomes clearer if the above equation is written

$$\{X \in I\!R_1^{n+1} \,|\, \langle X, X\rangle_1 = -1\},$$

which can formally be interpreted as a sphere with the imaginary radius i in $I\!R_1^{n+1}$.

This is a regular hypersurface in $I\!R_1^{n+1}$ by the implicit function theorem (which is independent of any metric involved). The position vector is always time-like and has a constant length. By taking derivatives it follows that the tangent plane to this hypersurface is on the one hand perpendicular to the position vector (just as in the case of the Euclidean sphere) and on the other hand consists only of space-like vectors. Consequently, the first fundamental form of this hypersurface is positive definite everywhere. In addition we see that the "unit normal" ν must coincide with the position vector, perhaps only up to a sign. From this we can conclude, just as in the case of the Euclidean sphere, that the Weingarten mapping L is the identity, up to a sign. Here, the Weingarten map is defined literally as in the Euclidean case: $L = -D\nu \cdot (Df)^{-1}$, if f denotes a parametrization of the position vector.

For a hypersurface in $I\!R^{n+1}_1$ with unit normal ν and for which $\langle \nu, \nu \rangle = \epsilon \in \{+1, -1\}$ there is a *covariant derivative* defined (which is at the same time the Riemannian connection of Section 5.16), just as in the Euclidean case (4.3), which one gets upon decomposing the directional derivatives into a tangent and a normal component

$$D_Y Z = \nabla_Y Z + \epsilon \cdot \langle LY, Z \rangle \nu$$

for two tangent vectors Y, Z. The normal component here is the *vector-valued second fundamental form*, cf. 3.41. Similarly, one may view the scalar ν-component

$$\langle D_Y Z, \nu \rangle = -\langle Z, D_Y \nu \rangle = \langle Z, LY \rangle = \langle LY, Z \rangle$$

as the *second fundamental form*, independent of the sign ϵ. On the other hand, the Gauss equation

$$R(Y, Z)W = \epsilon \Big(\langle LZ, W \rangle LY - \langle LY, W \rangle LZ \Big)$$

again contains this sign ϵ. In this equation, Y, Z, W denote arbitrary tangent vectors. This can be seen by making the same computation as we carried out in Section 4.18, but including the sign ϵ. For our particular hypersurface,

$$\{X \in I\!R^{n+1}_1 \big| \langle X, X \rangle_1 = -1\},$$

we have $L = \pm \mathrm{Id}$, $\epsilon = -1$, and consequently

$$R(X, Y)Z = -R_1(X, Y)Z = -\Big(\langle Y, Z \rangle X - \langle X, Z \rangle Y \Big),$$

from which it in particular follows that the sectional curvature is constant, in fact $K_\sigma = -1$.

We define the n-dimensional *hyperbolic space* H^n as the component of $\{X \in I\!R_1^{n+1} | \langle X, X\rangle_1 = -1\}$ which contains the point $(+1, 0, \ldots, 0)$, that is, the upper component of the two-sheeted hyperboloid. The sectional curvature of hyperbolic space as defined in this manner is constant: $K = -1$.

Similarly, one has, for every r, the set of those X for which $\langle X, X\rangle_1 = -r^2$. In this case the sectional curvature satisfies $K = -\frac{1}{r^2}$, just as in the case of the Euclidean sphere of radius r.

7.4. Remark. (Pseudo-sphere, pseudo-hyperbolic space)

We can make the same considerations for other "spheres" $\{\langle x, x\rangle_k = \pm 1\}$ in an arbitrary pseudo-Euclidean space $I\!R_k^{n+1}$. This also leads to examples of hypersurfaces with constant sectional curvature. However, the sectional curvature is only defined for so-called *non-degenerate planes*, i.e., for those planes which satisfy $\langle X, X\rangle\langle Y, Y\rangle - \langle X, Y\rangle^2 \neq 0$ for at least one basis X, Y of σ, cf. 6.4. We get the following hypersurfaces:

The *pseudo-sphere* $S_k^n = \{x \in I\!R_k^{n+1} | \langle x, x\rangle_k = 1\}$ with sectional curvature $K = 1$ (is should not be confused with the pseudo-sphere with $K = -1$ in 3.17),

The *pseudo-hyperbolic space* $H_k^n = \{x \in I\!R_{k+1}^{n+1} | \langle x, x\rangle_{k+1} = -1\}$ with sectional curvature $K = -1$.

In particular we have the statement that S_k^n is anti-isometric to H_{n-k}^n and $I\!R_k^n$ is anti-isometric to $I\!R_{n-k}^n$, which simply means that the metric of the one space is the same as that of the other, multiplied by -1.

For details, see [**22**], page 111.

7.5. Symmetries of the spaces E^n, S^n, H^n.

From the very construction of the three standard spaces E^n, S^n, H^n, it is quite obvious that the corresponding symmetry groups (that is, the group of diffeomorphisms which preserve the metric, cf. 5.11) are the following:

(1) The group $\mathbf{E}(n)$ of Euclidean motions (the so-called *Euclidean group*) acts on E^n. This group contains in particular all translations (these form a subgroup which is isomorphic to $I\!R^n$, in fact

a normal subgroup of $\mathbf{E}(n)$) as well as the rotation group $\mathbf{O}(n)$, consisting of symmetries which leave a point invariant. In fact, $\mathbf{E}(n)$ is a semi-direct product of these two subgroups.

(2) The *orthogonal group*

$$\mathbf{O}(n+1) = \left\{ \begin{matrix} A\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1} \mid \\ A \text{ preserves the Euclidean inner product} \end{matrix} \right\}$$

acts on the sphere S^n. Here, A denotes a linear map. As is well-known, $A \in \mathbf{O}(n+1)$ holds if and only if $A^T = A^{-1}$. As a matter of fact, the orthogonal group acts on the entire space $\mathbb{R}^{n+1}$, but we can consider its action when restricted to the sphere and denote the group in the same way.

(3) The *Lorentz group*

$$\mathbf{O}(n,1) = \left\{ \begin{matrix} A\colon \mathbb{R}^{n+1}_1 \to \mathbb{R}^{n+1}_1 \mid \\ A \text{ preserves the pseudo-Euclidean inner product} \end{matrix} \right\}$$

acts on *Lorentz space* or on *Minkowski space* $\mathbb{R}^{n+1}_1$ and preserves the set $\widetilde{H} = \{X \mid \langle X, X \rangle_1 = -1\}$. The "positive" part of this set,

$$\mathbf{O}_+(n,1) := \big\{A \in \mathbf{O}(n,1) \mid A \text{ preserves } \widetilde{H} \cap \{x_0 > 0\}\big\}$$

then acts on hyperbolic space H^n and preserves its metric.

7.6. Theorem. (Free motions in E^n, S^n, H^n)
The three groups $\mathbf{E}(n), \mathbf{O}(n+1)$ and $\mathbf{O}_+(n,1)$ act on the corresponding space E^n, S^n and H^n, respectively, transitively on the points of these sets and in addition on orthonormal n-frames of directional vectors. This means that one can map an arbitrary point to any other point by means of an element of the group, and upon fixing some point one can map an arbitrary n-frame to any other. The geometric or intuitive meaning of this is that any object in one of these three geometries can be freely moved to any other position in the space, preserving the geometry.
Conversely, every Riemannian manifold for which this freedom of motion exists and which admits the corresponding local isometries is necessarily a space with constant sectional curvature.

Proof: The necessity of this condition is clear, since one can map an arbitrary point into any other and every plane to any other by means of a local isometry. It then follows that the sectional curvature is constant, as it is preserved by isometries.

We now show the free motion of the three standard spaces. This is obvious for Euclidean space: here any point p can be mapped to any other q simply by taking the translation by the vector $q-p$. Moreover, fixing a point, every unit tangent vector can be mapped to any other by means of a rotation. Fixing this vector, the second vector in the frame can be arbitrarily mapped, and so forth.

Similarly, for the sphere we can first use a rotation to map an arbitrary point p to any other point q, and fixing a point, the argument with the frame is identical to the Euclidean case.

In the case of hyperbolic space we first note that by a rotation in space (in the Euclidean sense) around the x_0-axis every orthonormal n-frame at the point $p_0 = (1, 0, \ldots, 0)$ can be mapped to any other. In the same way, every point $p \neq p_0$ can be rotated in such a way that it takes on the coordinates $(x_0, x_1, 0, \ldots, 0)$ with $x_0 > 0, x_0^2 = 1 + x_1^2$. By utilizing a Lorentz transformation of the form

$$\begin{pmatrix} \cosh\varphi & \sinh\varphi & 0 & \ldots & 0 \\ \sinh\varphi & \cosh\varphi & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix},$$

we can then easily map the point p to the point p_0 and conversely. For this we can choose $\cosh\varphi = x_0$ and $\sinh\varphi = -x_1$. Just as we can relate two arbitrary vectors in the Euclidean plane of the same length by a matrix of the type

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix},$$

we can do the same in the pseudo-Euclidean plane by means of a matrix of the type

$$\begin{pmatrix} \cosh\varphi & \sinh\varphi \\ \sinh\varphi & \cosh\varphi \end{pmatrix},$$

which maps two arbitrary vectors of the same non-vanishing length into each other. □

The transformation in $\mathbf{O}(1,1)$ determined by the last matrix is called a *boost* with angle φ (cf. [**22**], p. 236).

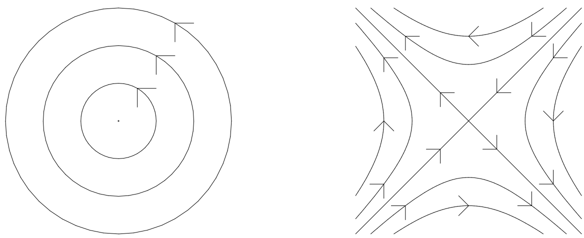


Figure 7.1. Euclidean rotation and boost

7.7. Other models for E^n, S^n, H^n in coordinates.

We begin by using the usual polar coordinates. In these coordinates, the variable r denotes the distance of a point from the origin, and there are $n-1$ further coordinates which are orthogonal to this. This leads to a metric tensor which can be schematically written as

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 g^*_{ij} \end{pmatrix}.$$

This can be done for an arbitrary Riemannian metric by using so-called *geodesic polar coordinates*, cf. 7.14 for this. In general the metric g^* will depend on the radius r and on the $n-1$ coordinates which are orthogonal to the curves of constant r-value. In the case of metrics of constant curvature, this is different, due to the increased symmetry of the situation. Indeed, in this case we can interpret the metric g^* as the metric g_1 of the unit sphere S^{n-1}. We introduce the more precise notation $g_1^{(n-1)}$ for the metric of the $(n-1)$-dimensional unit sphere. This will be further discussed in Section 7B for arbitrary spaces of constant curvature. By way of motivation we will do this here for the three standard spaces, and refer to Section 7B for the general construction. Here the calculations are easily carried out using Exercises 10-12 of Chapter 6. The individual cases now follow.

1. *The Euclidean metric g_0 in polar coordinates:*

$$g_0^{(n)} = dr^2 + r^2 \cdot g_1^{(n-1)}.$$

Here the parameter r runs through the interval $(0, \infty)$, and the r-curves are geodesics, parametrized by arc length. These coordinates have an apparent singularity at $r = 0$, just as we are familiar with from the polar coordinates in the plane.

2. *The spherical metric g_1 in polar coordinates:*

$$g_1^{(n)} = dr^2 + \sin^2 r \cdot g_1^{(n-1)}.$$

This means that the distance spheres (spheres at distance r from the north pole) have a radius of $\sin r$. The parameter r runs here through the interval $(0, \pi)$, and again the r-curves are geodesics which are parametrized by arc length. This is easiest to see in the standard model of the sphere, in which the r-curves are great circles through some fixed point (the north pole, say). Again these coordinates have an apparent singularity at $r = 0$ (at the north pole). In addition, we have the phenomenon that at the south pole ($r = \pi$) the coordinates again have an apparent singularity.

3. *The hyperbolic metric g_{-1} in polar coordinates:*

$$g_{-1}^{(n)} = dr^2 + \sinh^2 r \cdot g_1^{(n-1)}.$$

This time the variable r runs through the interval $(0, \infty)$, and again the r-curves are geodesics parametrized by arc length. The relation with the model described above in 7.3 is easiest to see if we measure the geodesic distance r from the point $(1, 0, \ldots, 0) \in \mathbb{R}^{n+1}_1$. The r-curve in the direction $(0, 1, 0, \ldots, 0)$ is then the geodesic

$$c(r) = (\cosh r, \sinh r, 0, \ldots, 0),$$

where we use the fact that

$$\langle c(r), c(r) \rangle = -1 \text{ and } \dot{c}(r) = (\sinh r, \cosh r, 0, \ldots 0)$$

with $\langle \dot{c}, \dot{c} \rangle = +1$. Because of the rotational symmetry with respect to the point $(1, 0, \ldots, 0)$ (cf. 7.6), the same holds also for every other direction. From 7.6 it follows in particular that every timelike or spacelike geodesic through every point has an infinite length, i.e., the arc length parameter is defined for arbitrarily large arguments. Null geodesics are defined for arbitrary values of the affine parameter, cf. 2.19.

Independently of these models, one can look for coordinates in which the metric becomes a scalar multiple of the Euclidean one and hence the measurement of angles coincides with that of Euclidean space (a so-called *conformal model* of the metric; compare this with the isothermal parameters of Section 3D). In the cases at hand we have the possibility of using the entire Euclidean space with its coordinates and look for an appropriately defined conformal factor. In the individual cases, we have the following metrics:

1. *The Euclidean metric:*
$$g_{ij} = \delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$
2. *The spherical metric:*
$$g_{ij} = \frac{4\delta_{ij}}{(1 + \|x\|^2)^2}$$
for all $x \in \mathbb{R}^n$.
3. *The hyperbolic metric:*
$$g_{ij} = \frac{4\delta_{ij}}{(1 - \|x\|^2)^2},$$
which is defined for all $x \in \mathbb{R}^n$ with $\|x\|^2 < 1$. This is the so-called *conformal disc model* of hyperbolic space, cf. Figure 7.2, left.

We omit the actual calculations showing that in the last two cases one does have sectional curvature equal to $+1$ and -1, respectively. For this one can use Cartesian coordinates in $\mathbb{R}^n$ and the equations from 8.27, which assume no further results from Chapter 8.

In the case of the sphere, the space $\mathbb{R}^n$ with the above metric only corresponds to a part of the sphere, since the former space is not compact. On the other hand, the open unit ball $D^n := \{x \in \mathbb{R}^n \mid \|x\| < 1\}$ together with the given hyperbolic metric (this is the conformal disc

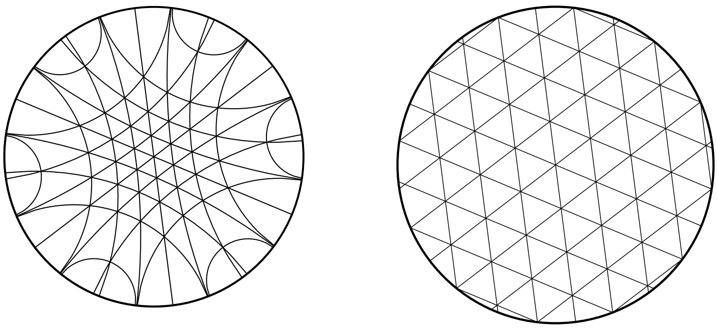


Figure 7.2. Conformal and projective disc model of the hyperbolic plane H^2 with geodesics

model)

$$g(x) = \frac{4g_0}{(1 - \|x\|^2)^2}$$

is *globally isometric* to hyperbolic space H^n, as defined in 7.3. This can be seen as follows: define a differentiable map $\Phi : D^n \to H^n$ by

$$\Phi(x) := (\lambda - 1, \lambda \cdot x) \in I\!R^{n+1}_1,$$

where $\lambda = \frac{2}{1-\|x\|^2}$. An easy calculation (exercise) then shows that:

1. $\langle \Phi(x), \Phi(x) \rangle = -1$, hence $\Phi(x) \in H^n$ because $\lambda - 1 \geq 1$;
2. Φ is bijective and isometric.

For the spherical metric one gets a similar picture through stereographic projection, cf. Exercise 2.

The conformal disc model should not be confused with the *projective disc model*, see Figure 7.2, right. For obtaining the projective disc model of hyperbolic space H^n, one regards each point on the hyperboloid in Minkowski $(n+1)$-space (see Section 7.3) as a point in projective n-space (ignoring the inner product here). Then the hyperbolic plane appears as an open disc, and all the geodesics appear as straight line segments, see [**51**, p.162]. However, the angles do not coincide with the Euclidean angles, so this model is not conformal.

7B Geodesics and Jacobi fields

In this section we come back to the geodesics, now considered on arbitrary Riemannian manifolds, and to start with we will make no assumption on the curvature. One of the goals here is to establish 7.21, the *local isometry* of two arbitrary spaces of the same constant sectional curvature. The tool used for this, the *Jacobi fields*, are quite interesting in themselves. They describe in a sense how the distance of neighboring geodesics changes along the path of a curve.

We let (M, g) be a Riemannian manifold, and $c\colon [a, b] \to M$ will denote a differentiable curve from $p = c(a)$ to $q = c(b)$. The *length* of this curve is given by

$$L(c) = \int_a^b \sqrt{g(\dot{c}, \dot{c})}dt.$$

Thus one always has $L(c) > 0$ for $p \neq q$. If g is indefinite, then we have to define the length instead as $L(c) = \int_a^b \sqrt{|g(\dot{c}, \dot{c})|}dt$, and then one has $L(c) > 0$ for $p \neq q$, provided that c is space-like or time-like.

Problem: For which curves c is the length $L(c)$ minimal? Which curves between two given points p and q realize a minimal possible arc length?

We already know from Chapter 4 that for surfaces in space, such minimal curves, if they exist at all, are necessarily geodesics. This actually holds for an arbitrary Riemannian manifold. In fact, the proof we have given in Chapter 4 can be literally adapted to the case at hand.

For this, consider again a one-parameter family of curves as a differentiable map

$$C\colon [a, b] \times (-\varepsilon, \varepsilon) \to M,$$

and view $c_s(t) := C(t, s)$ as a curve, depending on the additional parameter s.

7.8. Theorem. (First and second variation of arc length)
Let $C : [a,b] \times (-\varepsilon, \varepsilon) \longrightarrow M$ be a one-parameter family of curves. We may assume that $c_0(t) = C(t,0)$ is parametrized by arc length. Let $T(t,s)$ and $X(t,s)$ be the vector fields $T = \frac{\partial C}{\partial t}$, $X = \frac{\partial C}{\partial s}$. Then one has

$$\frac{dL}{ds}\bigg|_{s=0} = \langle X, T\rangle\bigg|_{(a,0)}^{(b,0)} - \int_a^b \langle X, \nabla_T T\rangle dt.$$

If, in addition, $\nabla_T T|_{(t,0)} = 0$, hence if c_0 is a geodesic, then one also has the relation

$$\frac{d^2L}{ds^2}\bigg|_{s=0} = \langle \nabla_X X, T\rangle\bigg|_{(a,0)}^{(b,0)}$$
$$+ \int_a^b \left(\langle \nabla_T \widetilde{X}, \nabla_T \widetilde{X}\rangle - \langle R(T,X)X, T\rangle\right)\bigg|_{(t,0)} dt,$$

where $\widetilde{X} = X - \langle X, T\rangle T$ is the component of X which is perpendicular to c_0.

Remark: For every geodesic c with tangent T one can consider the following quadratic form:

$$\operatorname{Ind}(X,X) := \int_a^b \Big(\langle \nabla_T X, \nabla_T X\rangle - \langle R(X,T)T, X\rangle\Big) dt,$$

where X denotes a vector field which is perpendicular to the geodesic. Because of the symmetries of the curvature tensor we get from this the following symmetric bilinear form:

$$\operatorname{Ind}(X,Y) := \int_a^b \Big(\langle \nabla_T X, \nabla_T Y\rangle - \langle R(X,T)T, Y\rangle\Big) dt,$$

which is called the *index form* of c. This form indicates how the lengths of neighboring geodesics behave. This can sensibly be viewed as a kind of analog of the Hessian of a function, which also indicates in which directions the function decreases and in which ones it increases, cf. 3.13.

Proof: The calculation of $\frac{d}{ds}\int_a^b \langle T, T\rangle^{\frac{1}{2}} dt$ has already been done in Section 4.13. There we saw that the following equivalence holds:

$$\frac{dL}{ds} = 0 \Longleftrightarrow \nabla_T T\Big|_{(t,0)} = 0.$$

This means that the first variation of the arc length vanishes precisely for geodesics, which proves the first statement.

As to the second statement, we make a computation, using the relation $\langle T,T\rangle\big|_{(t,0)} = 1$ as well as the symmetry $\nabla_X T = \nabla_T X$:

$$\begin{aligned}
\frac{d^2}{ds^2}\Big|_{s=0}\int_a^b \langle T,T\rangle^{\frac{1}{2}}\,dt &= \frac{d}{ds}\Big|_{s=0}\int_a^b X\Big|_{(t,0)}\Big(\langle T,T\rangle^{\frac{1}{2}}\Big)dt \\
= \frac{d}{ds}\Big|_{s=0}\int_a^b \frac{\langle\nabla_X T,T\rangle}{\langle T,T\rangle^{1/2}}\Big|_{(t,s)}dt &= \int_a^b X\Big|_{(t,0)}\Big(\frac{\langle\nabla_X T,T\rangle}{\langle T,T\rangle^{1/2}}\Big)dt \\
= \int_a^b \frac{1}{\langle T,T\rangle}\Big[\langle T,T\rangle^{1/2}\Big(&\langle\nabla_X\nabla_T X,T\rangle + \langle\nabla_T X,\nabla_X T\rangle\Big) \\
&-\langle\nabla_T X,T\rangle\cdot\frac{\langle\nabla_X T,T\rangle}{\langle T,T\rangle^{1/2}}\Big]\Big|_{(t,0)}dt \\
= \int_a^b \Big[-\langle R(T,X)X,T\rangle &+ \langle\nabla_T\nabla_X X,T\rangle \\
&+\langle\nabla_T X,\nabla_T X\rangle - \langle\nabla_T X,T\rangle^2\Big]\Big|_{(t,0)}dt \\
= \langle\nabla_X X,T\rangle\Big|_{(a,0)}^{(b,0)} + \int_a^b \Big[-\langle R(T,X)X,T\rangle &+ \langle\nabla_T\widetilde{X},\nabla_T\widetilde{X}\rangle\Big]dt.
\end{aligned}$$

The last equation in this chain of equalities follows, using the fact that $\widetilde{X} = X - \langle X,T\rangle T$, from the following auxiliary calculation: $\nabla_T T = 0$ for $s = 0$ implies $\nabla_T\langle X,T\rangle = \langle\nabla_T X,T\rangle$, hence $\nabla_T\widetilde{X} = \nabla_T X - \langle\nabla_T X,T\rangle T$ and

$$\begin{aligned}
\langle\nabla_T\widetilde{X},\nabla_T\widetilde{X}\rangle &= \langle\nabla_T X,\nabla_T X\rangle - 2\big\langle\nabla_T X,\langle\nabla_T X,T\rangle T\big\rangle \\
&\quad+\langle\nabla_T X,T\rangle^2\langle T,T\rangle \\
&= \langle\nabla_T X,\nabla_T X\rangle - \langle\nabla_T X,T\rangle^2.
\end{aligned}$$

Note that in the integrand the sectional curvature of M in the X,T-plane occurs, up to a normalization of the vectors X and T. This is the basis for many considerations on the influence of the curvature on the behavior of geodesics. □

7.9. Corollary.

(i) Every differentiable shortest curve joining p and q is a geodesic.

(ii) Let c be a geodesic from p to q, and assume that the sectional curvature of M is strictly negative in all planes which contain $\dot{c}$. Then c is strictly shorter than any other sufficiently close curve joining p and q.

PROOF: We consider a (fixed but arbitrary) one-parameter family of curves as above, but with fixed endpoints p, q, so that $X|(b,0) = X|(a,0) = 0$. Moreover we may assume that $\langle X, T\rangle = 0$.

For part (i) we have the simple fact that $\frac{dL}{ds} = 0$ holds for all such one-parameter families if and only if $\nabla_T T = 0$, that is, if the curve c_0 is a geodesic. This is the same conclusion as in the proof of 4.13.

For part (ii) we first consider the equation

$$\nabla_X X|_{(b,0)} = \nabla_X X|_{(a,0)} = 0,$$

which holds since the endpoints are taken to be fixed. Then we have

$$\frac{d^2L}{ds^2}\Big|_{s=0} = \int_a^b \Big(\underbrace{\langle \nabla_T X, \nabla_T X\rangle}_{\geq 0} - \underbrace{\langle R(T,X)X, T\rangle}_{<0}\Big)dt.$$

The integrand is thus strictly positive, and it follows that the integral is as well. Note that

$$K_{(X,T)} = \frac{\langle R(T,X)X, T\rangle}{\langle X, X\rangle}$$

is exactly the sectional curvature in the X,T-plane, which is by assumption strictly negative. Therefore we have $\frac{d^2L}{ds^2}\big|_{s=0} > 0$ and $\frac{dL}{ds}\big|_{s=0} = 0$ for all such X, hence for every one-parameter family in any direction. Consequently, the function has a strict local minimum at c. This implies that the neighboring curves (in this sense) are strictly longer than c. In the special case of hyperbolic space H^n, in fact *every* geodesic from p to q is strictly shorter than any other curve joining the two points. □

Remark: For a curve $c\colon [a,b] \to M$ the quantity

$$E(c) := \int_a^b \langle T, T\rangle dt$$

is called the *energy functional* of c. Under the same assumptions as in 7.8, one has $\frac{dE}{ds}\big|_{s=0} = 2\frac{dL}{ds}\big|_{s=0}$, so that the critical curves with respect to L coincide with those with respect to E (up to the parametrization).

In local coordinates we have the equation for a geodesic

$$\nabla_{\dot c}\dot c = 0 \iff \ddot c^k + \sum_{i,j} \dot c^i \dot c^j \Gamma^k_{ij} = 0 \quad \text{for } k = 1, \ldots, n.$$

The local existence of geodesics follows from this (cf. 4.12 and 5.18):

Theorem. (Existence of geodesics)
At a given point $p \in M$ and for a given vector $V \in T_pM$, $\langle V, V\rangle = 1$, there exists locally a unique geodesic $c_V^{(p)}$ with $c_V^{(p)}(0) = p$ and $\dot c_V^{(p)}(0) = V$

If one considers the set of all geodesics in all directions passing through some fixed point, one is lead to the *exponential mapping*. Recall the definition of this from 5.19.

7.10. Definition. (Exponential mapping)
For a fixed point $p \in M$ let $c_V^{(p)}$ denote the uniquely determined geodesic parametrized by arc length through p in the direction of the unit vector V. In a certain neighborhood U of $0 \in T_pM$, the following map is well defined:

$$T_pM \supseteq U \ni (p, tV) \longmapsto c_V^{(p)}(t).$$

Here we have chosen the parameter in such a way that $(p, 0) \mapsto p$. This mapping is called the *exponential mapping* at the point p, and it is denoted by $\exp_p\colon U \longrightarrow M$. For variable points p one can similarly define $\exp\colon \widetilde U \to M$ by setting $\exp(p, tV) = \exp_p(tV) = c_V^{(p)}(t)$, where $\widetilde U$ is an open set in the tangent bundle TM, for example $\widetilde U = \{(p, X) \mid ||X|| < \varepsilon\}$ for an appropriately chosen $\varepsilon > 0$, if M is compact and g is positive definite.

REMARK: $\exp_p$ maps the lines through the origin of the tangent space to geodesics, and this is done in an isometric manner. In all directions perpendicular to the geodesics through p the map $\exp_p$ is in general not isometric. In what follows it will be important to precisely describe how far this is from being isometric, in particular in the case of constant sectional curvature. We now verify that the exponential mapping can be used as a local parametrization in the first place.

7.11. Lemma. The exponential mapping $\exp_p$, restricted to a certain neighborhood U of the origin in T_pM, is a diffeomorphism

$$\exp_p : U \to \exp_p(U).$$

The inverse mapping $\exp_p^{-1}$ thus defines a chart at p. The corresponding coordinates are called *normal coordinates* or *Riemannian normal coordinates.*

PROOF: First we note that $\exp_p$ is differentiable by well-known results on the differentiable dependence of solutions of ordinary differential equations on the initial conditions, see for example [**27**], VI, §4.

We now show that $D(\exp_p)\big|_0 : T_0(T_pM) \to T_pM$ is a linear isomorphism. The statement of the theorem then follows from the theorem on inverse functions (cf. 1.4), which by using some local charts holds on a differentiable manifold just as it does in $\mathbb{R}^n$. Because $\dim T_0(T_pM) = n = \dim T_pM$, it suffices to show that the linear map $D(\exp_p)\big|_0$ is surjective. Thus, let $Y \in T_pM$ be an arbitrary vector with $||Y|| = 1$. We consider the line

$$\psi(t) := t \cdot Y$$

in T_pM, where for sufficiently small t the expression $\exp_p(\psi(t))$ is defined. Set $c(t) := \exp_p(\psi(t)) = \exp_p(tY) = c_Y^{(p)}(t)$. Then we have $\dot{c}(0) = \dot{c}_Y^{(p)}(0) = Y$, while at the same time

$$Y = \dot{c}(0) = D(\exp_p)\big|_0 \Big(\frac{d\psi}{dt}\Big|_0 \Big).$$

Thus, $D(\exp_p)\big|_0$ is a linear isomorphism, and the statement follows. $\square$

7.12. Lemma. (Normal coordinates)
Let $X_1, \ldots, X_n$ be an ON-basis in T_pM, and let

$$\exp_p \colon U \to \exp_p(U)$$

be the diffeomorphism of 7.11, defined on an open neighborhood of the origin $U \subset T_pM$. The associated coordinates are the normal coordinates, and we denote by ∂_i the elements of a basis of these coordinates on M, so that in particular $\partial_i|_p = (D\exp_p)\big|_0(X_i)$. Then all Christoffel symbols vanish for these coordinates at the point p.

Proof: Since the lines spanned by the X_i map to geodesics under $\exp_p$, the vector field ∂_i is tangent to a geodesic which leaves p radially, hence $\nabla_{\partial i}\partial_i\big|_p = 0$. A similar statement holds for the line in the direction $X_i + X_j$. Moreover,

$$D\exp_p(X_i + X_j)\big|_0 = D\exp_p(X_i)\big|_0 + D\exp_p(X_j)\big|_0 = \partial_i\big|_p + \partial_j\big|_p,$$

hence

$$0 = \nabla_{\partial_i+\partial_j}(\partial_i + \partial_j)\big|_p = \nabla_{\partial_i}\partial_j\big|_p + \nabla_{\partial_j}\partial_i\big|_p$$

$$= 2\nabla_{\partial_i}\partial_j\big|_p = 2\sum_k \Gamma^k_{ij}(p)\partial_k\big|_p.$$

Note that $\nabla_{\partial_i}\partial_j = \nabla_{\partial_j}\partial_i$ because these are coordinates; hence $[\partial_i, \partial_j] = 0$. It follows that at the point p, all Γ^k_{ij} must vanish. $\square$

The normal coordinates are thus optimally fitted to the Euclidean structure of the tangent space T_pM. In particular, the covariant derivative at the point p has vanishing Christoffel symbols, just as in the Euclidean case. This aspect is even more emphasized by introducing polar coordinates centered at p. This is brought out by the following lemma. It states that Euclidean polar coordinates in the tangent space are partially preserved under the exponential mapping $\exp_p$ in the sense that geodesic rays from the point p are perpendicular to the images of the distance spheres. The distortion of the metric is thus restricted to the direction perpendicular to the radial geodesics.

7.13. Lemma. (Gauss lemma)
Let $\exp_p : U \to \exp_p(U)$ be a diffeomorphism. Let W be a vector which is perpendicular to the line $t \longmapsto t \cdot V$ in some fixed direction V, $||V|| = 1$. The base point of W is arbitrary here. Then $D\exp_p(W)$ is perpendicular to the geodesic $c_V^{(p)}$.

PROOF: Let $c(s)$ be a differentiable curve in T_pM, so that on the one hand $c(0)$ is the base point of W and $\dot{c}(0) = W$, and on the other every point $c(s)$ has the same distance from the origin in T_pM. We denote by $\rho_s(t)$ the line from 0 to $c(s)$, parametrized by arc length $t \in [0, t_0]$, and define $\widetilde{c}(t,s) := \exp_p(\rho_s(t))$. By construction of $\exp_p$ the lengths of the curves $t \longmapsto \widetilde{c}(t,s)$ for every s are equal to the lengths of ρ_s, hence equal to t_0. It now follows that

$$\frac{d}{ds} L(\exp_p(\rho_s)) = 0;$$

consequently, by 7.8 with $s = 0$ and $T(t_0, 0) = \dot{c}_V^{(p)}$, the equation

$$X\Big|_{(t_0,0)} = D\exp_p(W), \quad \text{initial value: } X\Big|_{(0,0)} = 0$$

holds. Hence

$$\begin{aligned} 0 = \frac{dL}{ds}\Big|_{s=0} &= \langle X, T \rangle \Big|_{(0,0)}^{(t_0,0)} - \int_0^{t_0} \langle X, \nabla_T T \rangle dt \\ &= \langle X, T \rangle \Big|_{(t_0,0)} = \Big\langle D\exp_p(W), \dot{c}_V^{(p)}(t_0) \Big\rangle. \end{aligned}$$

□

7.14. Corollary. If one introduces polar coordinates in T_pM, then, under the exponential mapping $\exp_p$, these yield coordinates on M around p (so-called *geodesic polar coordinates*), which we denote by $r, \varphi_1, \ldots, \varphi_{n-1}$. In these coordinates, we have $\langle \frac{\partial}{\partial_r}, \frac{\partial}{\partial_r} \rangle = 1$ and $\langle \frac{\partial}{\partial_r}, \frac{\partial}{\partial \varphi_i} \rangle = 0$, hence

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & \ldots & \ldots & 0 \\ 0 & * & \ldots & \ldots & * \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & * & \ldots & \ldots & * \end{pmatrix},$$

where the submatrix indicated by the $*$'s is of order r^2 for $r \to 0$.

For $n = 2$ one has in particular

$$(g_{ij}(r,\varphi)) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 G(r,\varphi) \end{pmatrix}$$

for a bounded function G, so that the geodesic polar coordinates are (at least for $r \neq 0$) a special case of the geodesic parallel coordinates which were introduced in 4.27.

7.15. Lemma and Definition. (Jacobi fields)
Let $V, W \in T_pM$ be fixed vectors with $||V|| = ||W|| = 1$, which are moreover perpendicular to each other. Then $t \longmapsto t \cdot V$ describes a line in T_pM, and W is perpendicular to this line. In addition, the map defined by $t \longmapsto X(t) = t \cdot W \in T_{tV}(T_pM)$ is a (linear) vector field along this line. Now set

$$Y(t) := (D \exp_p)\big|_{tV}(X(t)).$$

Then $Y(t)$ is a vector field along $c_V^{(p)}$, which is perpendicular to $c_V^{(p)}$ by 7.13, and Y satisfies the so-called *Jacobi equation*

$$\nabla_T \nabla_T Y + R(Y,T)T = 0,$$

where T denotes the tangent to $c_V^{(p)}$. A vector field Y which satisfies this equation is referred to as a *Jacobi field*.

One can write the Jacobi equation in the abbreviated form $Y'' = -R(Y,T)T$, just like an ordinary differential equation of second order. In a nutshell, the Jacobi fields are the images of linear vector fields under the exponential mapping. They describe the mapping of two radial lines through the origin in the tangent space T_pM to the corresponding geodesics in M through p.

PROOF: Working directly in the tangent space, we can identify the tangent space $T_{tV}(T_pM)$ with T_pM by a canonical isomorphism, just like $\mathbb{R}^n$. We now set

$$X_s(t) = t \cdot V + t \cdot s \cdot W \in T_pM \cong T_{tV}(T_pM)$$

and

$$c(t,s) := \exp_p(X_s(t)).$$

Since s and t can then be viewed as local coordinates, the vectors $Y := \frac{\partial c}{\partial s}$ and $T := \frac{\partial c}{\partial t}$ commute, so that

$$\nabla_T Y = \nabla_Y T.$$

In addition, $\nabla_T T = 0$ always holds, since for fixed s the t-curves are geodesics (the parameter here is not arc length but a parameter which still is proportional to arc length). Thus, a direct calculation yields

$$R(Y,T)T = \nabla_Y \underbrace{\nabla_T T}_{=0} - \nabla_T \underbrace{\nabla_Y T}_{=\nabla_T Y} - \nabla_{\underbrace{[Y,T]}_{=0}} T = -\nabla_T \nabla_T Y. \qquad \square$$

7.16. Theorem. (Length distortion of the exponential map)
Let X, Y be as chosen as in 7.15. Then clearly $||X(t)||^2 = t^2$, since X is a linear vector field in the tangent space. Moreover,

$$||Y(t)||^2 = t^2 - \frac{1}{3} K t^4 + o(t^4),$$

where K is the sectional curvature in the (T, W)-plane, that is, the plane which contains the tangent to the curve and the vector W (which in turn determines Y).

Proof: We calculate the Taylor expansion of the function $t \longmapsto \langle Y(t), Y(t)\rangle$ at the point $t = 0$: first we have $Y(0) = 0$. As notation in this proof, we use $Y' := \nabla_T Y$, in particular $Y'' = -R(Y,T)T$, so that $Y''(0) = 0$. One then has $Y'(0) = W$, since the covariant derivative coincides with the usual partial derivative as here $\Gamma_{ij}^k\big|_p = 0$ (cf. Lemma 7.12). We then calculate the following expressions at the point $t = 0$:

$$\langle Y, Y\rangle' = 2\langle Y, Y'\rangle = 0;$$

$$\langle Y, Y\rangle'' = 2\langle Y', Y\rangle' = 2\langle Y'', Y\rangle + 2\langle Y', Y'\rangle = 2\langle W, W\rangle = 2;$$

$$\langle Y, Y\rangle''' = 2\langle Y'', Y\rangle' + 2\langle Y', Y'\rangle' = 2\langle Y''', Y\rangle + 6\langle Y'', Y'\rangle = 0;$$

$$\langle Y, Y\rangle'''' = 2\langle Y''', Y\rangle' + 6\langle Y'', Y'\rangle' = 2\langle Y'''', Y\rangle + 8\langle Y''', Y'\rangle + 6\langle Y'', Y''\rangle$$

$$= -8\big\langle \big(R(Y,T)T\big)'\big|_{t=0}, W\big\rangle$$

$$= -8\Big\langle (\nabla_T R)(Y,T)T + R(\nabla_T Y, T)T + R(Y, \nabla_T T)T + R(Y,T)\nabla_T T, W\Big\rangle$$

$$= -8\big\langle R(Y',T)T, W\big\rangle = -8\big\langle R(W,T)T, W\big\rangle = -8K_{(W,T)}. \qquad \square$$

The geometric interpretation of the formula in 7.16 is that the sectional curvature is in a sense the first non-trivial derivative of the metric tensor itself, provided one is working in geodesic polar coordinates.[1] Even more, Theorem 7.16 leads to a geometric interpretation of the Ricci curvature, just by taking the average over all planes containing the tangent T of the geodesics. In this way it follows that the Ricci curvature in direction T at p is the first non-trivial derivative of the volume element in that direction. Consequently the scalar curvature at p can be interpreted as the first non-trivial derivative of the volume distortion of the exponential map: It compares the volume of geodesic balls of radius t and centre at p with the volume of euclidean balls of the same radius.

7.17. Lemma. Let J_c be the set of all Jacobi fields along a geodesic $c\colon [a,b] \to M$ with $\dot{c} = T$ and $\| T \| = 1$. Then the following assertions are true:

(i) J_c is a real vector space.

(ii) $T \in J_c, t \cdot T \in J_c$, and every $X \in J_c$ has a unique orthogonal decomposition $X = \widetilde{X} + \kappa \cdot T + \lambda \cdot t \cdot T$ with $\langle \widetilde{X}, T \rangle = 0$ and with constants κ, λ. In other words: *linear combinations of T and $t \cdot T$ are the only Jacobi fields along c which are tangent to c.* The vector field $t \cdot T$ is more precisely described as the map $t \mapsto t \cdot T(t)$.

(iii) For $X, Y \in J_c$, the expression $\langle \nabla_T X, Y \rangle - \langle \nabla_T Y, X \rangle$ is constant along c; in particular, this holds for the expression $\langle \nabla_T X, T \rangle$ along c.

(iv) If to a given $X \in J_c$ there are two different parameters $\alpha, \beta \in [a,b]$ such that either X_α and X_β are orthogonal to c or X_α and $(\nabla_T X)_\beta$ are orthogonal to c, then X is orthogonal to c everywhere.

---

[1] This fact was already used by B. Riemann to determine the so-called "curvature measure" (that is, the sectional curvature K); for more on the history of this see [**53**] and [**7**], Vol. II, Chapter 4.

Proof: $X'' = -R(X,T)T$ is a linear differential equation, hence the space of solutions is a vector space, verifying (i).

(ii) can be seen as follows:

$$\langle X, T\rangle'' = \langle X', T\rangle' = \langle X'', T\rangle = -\langle R(X,T)T, T\rangle = 0,$$

hence $t \mapsto \langle X(t), T(t)\rangle$ is a linear function. Thus the space of Jacobi fields which are tangent to c is at most two-dimensional. But we already know two linearly independent elements, namely T and $t \cdot T$, where t is the arc length on c. In more detail,

$$T'' = \nabla_T \nabla_T T = 0 = -R(T,T)T,$$
$$(t \cdot T)'' = (tT' + T)' = tT'' + 2T' = 0 = -R(T,T)(tT).$$

T and tT are linearly independent as elements of J_c, although pointwise these two vectors point in the same direction.

For (iii) we calculate the derivative

$$\begin{aligned}\big(\langle X', Y\rangle - \langle Y', X\rangle\big)' &= \langle X'', Y\rangle + \langle X', Y'\rangle - \langle Y', X'\rangle - \langle Y'', X\rangle \\ &= -\langle R(X,T)T, Y\rangle + \langle R(Y,T)T, X\rangle = 0,\end{aligned}$$

where the last equality holds by the symmetries of the tensor R.

(iv) In the first case we assume that $X_\alpha \perp c, X_\beta \perp c$ and decompose $X = \widetilde{X} + \kappa \cdot T + \lambda \cdot tT$, from which it immediately follows that $\kappa = \lambda = 0$.

In the second case, $X_\alpha \perp c$ and $X'_\beta \perp c$, we observe that $\langle X', T\rangle = \langle X, T\rangle'$ is constant according to (iii). This constant must vanish, however, since this holds for the parameter β. □

7.18. Lemma. Given a point p on a geodesic c and given two vectors $Y_p, Z_p \in T_pM$, there is a uniquely determined Jacobi field Y along c with

$$Y(p) = Y_p \quad \text{and} \quad Y'(p) = \nabla_T Y\big|_p = Z_p.$$

Proof: We view Y_p, Z_p as the initial conditions for the differential equation $Y'' = -R(Y,T)T$. Let $X_1, \ldots, X_n$ be an ON-basis of T_pM. This basis can be uniquely extended to orthonormal vector fields $X_1, \ldots, X_n$ along c which are parallel along c, i.e., $X_i' = \nabla_T X_i = 0, i = 1, \ldots, n$. For a vector field $Y(t) = \sum_i \eta^i(t) X_i(t)$, the Jacobi equation $Y'' = -R(Y,T)T$ then takes the form

$$-R(Y,T)T = Y'' = \nabla_T\nabla_T Y = \nabla_T\nabla_T\Big(\sum_i \eta^i X_i\Big)$$

$$= \nabla_T\Big(\sum_i \eta^i \underbrace{\nabla_T X_i}_{=0} + \sum_i \dot\eta^i X_i\Big) = \sum_i \frac{d}{dt}(\dot\eta^i)X_i = \sum_i \ddot\eta^i X_i,$$

which in coordinates is

$$\ddot\eta^i = -\langle R(Y,T)T, X_i\rangle = -\sum_j \eta^j \underbrace{\langle R(X_j,T)T, X_i\rangle}_{\text{indepndent of } \eta^i}, \quad i = 1,\dots,n.$$

This is a system of ordinary differential equations in an open set of $\mathbb{R}^n$ with initial conditions at p for η^i and $\dot\eta^i$, $i = 1,\dots,n$. According to well-known results on the solutions of such systems there is a unique solution $\eta^i(t)$ along the entire curve c, a fact which we have already used in Section 4.10, compare [**27**], Chapter XIX. □

In 7.20 below we will explicitly solve this system of differential equations in the case of constant sectional curvature.

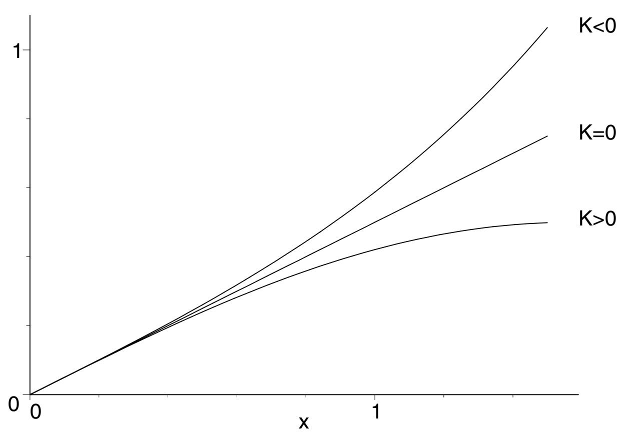


Figure 7.3. Length of Jacobi fields in spaces of constant sectional curvature K.

7.19. Corollary. The dimension of J_c on an n-dimensional manifold is $2n$. The map $Y \longmapsto (Y_p, \nabla_T Y|_p) \in (T_pM)^2$ is a linear isomorphism in this case. The dimension of $J_c^\perp = \{Y \in J_c \mid \langle Y, \dot{c}\rangle = 0\}$ is then $2(n-1)$ according to 7.17.

7.20. Corollary. (Jacobi fields on spaces of constant curvature) If M is a space of constant sectional curvature K, then

$$R(Y,T)T = K \cdot R_1(Y,T)T = K\big(\underbrace{\langle T,T\rangle}_{=1} Y - \langle Y,T\rangle T\big)$$

$$= \begin{cases} KY, & \text{if } \langle Y,T\rangle = 0, \\ 0 & \text{if } Y = \kappa \cdot T + \lambda \cdot (t \cdot T) \text{ with constants } \kappa, \lambda. \end{cases}$$

For parallel orthonormal vector fields $T, Y_1, \ldots, Y_{n-1}$, the Jacobi equation transforms into the system

$$\ddot{\eta}^i = -\sum_j \eta^j \underbrace{\langle R(Y_j,T)T, Y_i\rangle}_{=K\delta_{ij}}$$

$$\Longleftrightarrow \ \ddot{\eta}^i = -K \cdot \eta^i \text{ for } i = 1, \ldots, n-1.$$

Given initial conditions $\eta^i(0) = 0$, one gets the solution

$$\eta^i(t) = \begin{cases} \alpha \sin(\sqrt{K}t), & \text{if } K > 0, \\ \alpha t, & \text{if } K = 0, \\ \alpha \sinh(\sqrt{-K}t), & \text{if } K < 0, \end{cases}$$

in each case with arbitrary constants $\alpha \in \mathbb{R}$. Thus, one gets all Jacobi fields with $Y(0) = 0$ as a linear combination of

$$t \cdot T, \eta \cdot Y_1, \ldots, \eta \cdot Y_{n-1},$$

where η is a solution of $\ddot{\eta} + K\eta = 0$ with $\eta(0) = 0$.

Remark: If the sectional curvature is not constant, but is bounded by two bounds, then one can apply comparison theorems for the solutions of differential equations. For results of this type as well as other *comparison theorems*, see [**13**], Chapter 6 or [**18**].

7.21. Corollary. (Local isometry of spaces of constant curvature) In geodesic polar coordinates around an arbitrary (but fixed) point, the line element of the metric of a space of constant curvature K has the following form:

$$ds^2 = \begin{cases} dr^2 + \frac{1}{K}\sin^2(\sqrt{K}r)ds_1^2, & \text{if } K > 0, \\ dr^2 + r^2 ds_1^2, & \text{if } K = 0, \\ dr^2 + \frac{1}{-K}\sinh^2(\sqrt{-K}r)ds_1^2, & \text{if } K < 0. \end{cases}$$

Here, ds_1^2 denotes the line element on the standard sphere of radius 1 and r denotes the geodesic distance from a fixed point.

In particular, two Riemannian manifolds of the same constant curvature K (and the same dimension) are locally isometric to one another.

Proof: We first fix a point p on a Riemannian manifold of constant curvature K. In the direction which is tangential to the radial geodesic, the exponential mapping has no distortion of length, as we already noted in Definition 7.10. The Gauss Lemma 7.13 guarantees that the exponential mapping preserves the orthogonality of a vector to a radial geodesic. Hence we only need to calculate the length distortion of the exponential mapping $\exp_p$ in the orthogonal direction. According to 7.15, the orthogonal Jacobi fields occur via transport of linear fields in the tangent space T_pM by means of the exponential mapping $\exp_p$. If r denotes the arc length parameter on a radial geodesic (with $r > 0$), then the length of a linear vector field $r \mapsto rX$ in the tangent space is given by $r||X||$. By 7.20 the length of the corresponding Jacobi field $Y(r) = D\exp_p\big|_{rV}(rX)$ is equal to a constant α times

$$\begin{cases} \sin(\sqrt{K}r)||X||, & \text{if } K > 0, \\ r||X||, & \text{if } K = 0, \\ \sinh(\sqrt{-K}r)||X||, & \text{if } K < 0, \end{cases}$$

where α is independent of X. In fact, α is determined by comparison of the Taylor expansion of $||Y||$ in 7.16 with the Taylor expansion of the functions $\sin(\sqrt{K}r)$ and $\sinh(\sqrt{|K|}r)$. It follows that $\alpha = 1$ for $K = 0$ and $\alpha = 1/\sqrt{|K|}$ for $K \neq 0$.

If we do this for every unit vector $X \in T_pM$ which is perpendicular to the considered geodesic, then we get the length distortion of the exponential mapping in all directions perpendicular to the geodesics. This distortion is now clearly independent of the direction X and of the considered geodesic, since it only depends on the geodesic distance r and the curvature K. Therefore the Euclidean metric in polar coordinates in T_pM, which we denote by

$$dr^2 + r^2 ds_1^2,$$

is mapped under the exponential mapping to

$$dr^2 + \frac{1}{K}\sin^2(\sqrt{K}r)ds_1^2 \quad \text{or} \quad dr^2 + r^2 ds_1^2 \quad \text{or} \quad dr^2 + \frac{1}{-K}\sinh^2(\sqrt{-K}r)ds_1^2,$$

depending on the sign of K. □

7C The space form problem

Locally, there is by 7.21 only *one* metric of a given constant curvature, but this tells us nothing about the possibilities for manifolds of constant curvature in the large (global manifolds). The *Clifford-Klein space problem*[2] or *space form problem* is the question as to the global structure of Riemannian manifolds which have constant sectional curvature and are complete in an appropriate sense. Without this assumption on completeness, every open set of Euclidean space, for example, is a space of constant curvature, and the possible topological types of such spaces are just too vast for consideration.

7.22. Definition. (Completeness)
A Riemannian manifold (M, g) is called *(geodesically) complete*, if every geodesic which is parametrized by arc length is defined on all of $\mathbb{R}$ as a map $\gamma\colon \mathbb{R} \to M$. A *space form* is a complete Riemannian manifold with constant sectional curvature.

We remind the reader that the existence of a geodesic through a given point in a given direction is only guaranteed for a small, not prescribable, interval, cf. 4.12 and 5.18. On $M = \mathbb{R}^2 \setminus \{(0,0)\}$ with the Euclidean metric, for example, the geodesic through the point $p = (-1, 0)$ in the direction of the vector $V = (1, 0)$ ceases to exist, as

[2] H. Hopf, *Zum Clifford-Kleinschen Raumproblem*, Math. Annalen **95**, 313–339 (1926).

it would necessarily run through the origin. The maximal interval of definition for the arc length in this case would be the interval $[0, 1)$. To say a manifold is complete is thus roughly speaking to say that the manifold cannot be part of another, which would result through the addition of missing points. A compact Riemannian manifold is always complete, since every accumulation point on a geodesic again belongs to the manifold and one can therefore extend the geodesic around this point. Thus, a geodesic cannot cease to exist after the arc length runs through a finite interval.

7.23. Theorem.

(i) E^n, S^n, H^n are all geodesically complete.

(ii) Every n-dimensional Riemannian manifold of constant curvature $K = 0, +1, -1$ is locally isometric to an open set of one of E^n, S^n, H^n.

(iii) Every complete Riemannian manifold (M, g) with constant curvature $0, +1, -1$ is isometric to a quotient of E^n, S^n, H^n by a discrete and fixed-point free subgroup of one of $\mathbf{E}(n)$, $\mathbf{O}(n+1), \mathbf{O}_+(n, 1)$. This holds in particular for every compact Riemannian manifold of constant curvature.

For the proof of (i) it is sufficient by Theorem 7.6 to consider a geodesic through a fixed point, since all other points are equivalent under the respective symmetry group. It is clear that all lines in E^n are arbitrarily long in both directions and thus have arc length defined in all of $\mathbb{R}$.

It is just as clear that all great circles on the sphere S^n have the same property. For hyperbolic space we consider the geodesic γ through the point $p_0 = (1, 0, \ldots, 0)$ in the direction $(0, 1, 0, \ldots, 0)$. If this geodesic is parametrized by $\gamma(s) = (\cosh s, \sinh s, 0, \ldots, 0)$, then $\frac{d\gamma}{ds}$ is a unit vector and γ is defined in all of $\mathbb{R}$.

The proof of (ii) has already been carried out in 7.21 using geodesic polar coordinates and Jacobi fields. A different proof can be found in [**3**].

The proof of (iii) uses several notions of topology and the theory of group operations and coverings. We can therefore only sketch it here. A subgroup G of $\mathbf{E}(n)$ or $\mathbf{O}(n+1)$ or $\mathbf{O}_+(n,1)$ is called *discrete*, if for every point x the orbit $Gx = \{y \mid y = g(x) \text{ for a } g \in G\}$ of x is always a discrete set, that is, has no accumulation point. G is said to be *fixed-point-free*, or to *operate freely*, if there is no element of G other than the identity which fixes some point of the space.

The quotient space of E^n, S^n, H^n by such a discrete and fixed-point-free group can be formed as the space of cosets and endowed with a Riemannian metric in such a way that the projection onto the quotient is locally an isometry. In order to construct charts on this quotient, one only needs to verify that the projection is injective on the interior of these charts. This is done in principle just as in the case of the flat torus $\mathbb{R}^2/\mathbb{Z}^2$, cf. Example 3 in 5.1.

Conversely, one can construct for a given (M, g) the so-called *universal covering* and show that this space is again complete and, moreover, (globally) isometric to one of the spaces E^n, S^n, H^n. For this one applies 7.21 again. Then M is the quotient of E^n, S^n, or H^n by the group of *covering transformations*, and this group acts freely and fixed-point free. Details on coverings can be found in [**38**], §10.4.

Theorem 7.23 essentially reduces the classification of all space forms to the problem of determining all discrete and fixed-point-free subgroups of the symmetry groups $\mathbf{E}(n)$, $\mathbf{O}(n+1)$ and $\mathbf{O}_+(n,1)$. Instead of attempting to go into details on this algebraic problem (cf. [**20**]), we just present some examples here in dimensions two and three. The space forms are also given a quite readable presentation in [**49**], p. 254 ff.

7.24. Examples. (Two-dimensional space forms)

(i) The only complete two-manifolds with $K = 0$ are the following: the *plane* E^2, the *cylinder*, the *Möbius strip*, the *torus* and the *Klein bottle*. These occur by 7.23 as the quotients $\mathbb{R}^2/\Gamma$ of $\mathbb{R}^2$ by the following five subgroups $\Gamma_0, \Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4$ of the Euclidean symmetry group:

1. Γ_0 is the trivial group (consisting only of an identity element).

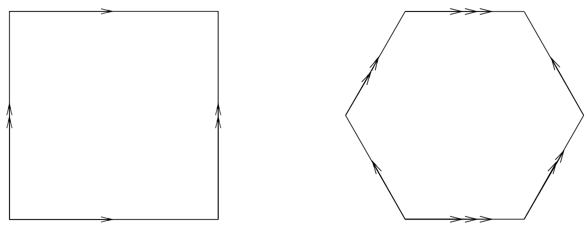


Figure 7.4. A flat torus, square and hexagonal

2. Γ_1 is generated by the pure translations t by a fixed vector X, for example $t(x, y) = (x + 1, y)$. The quotient E^2/Γ_1 is then an (abstract) *cylinder*.
3. Γ_2 is generated by Γ_1 and a translational reflection α, where $\alpha^2 = t$. In the special case above, we have $\alpha(x, y) = (x + \frac{1}{2}, -y)$. The quotient E^2/Γ_2 is an abstract *Möbius strip*, which can also be viewed as the quotient of the cylinder E^2/Γ_1 by α. The projection from the cylinder to the Möbius strip is then a double covering (i.e., a covering with two sheets).
4. Γ_3 is generated by two pure translations t_1, t_2 by two linearly independent vectors X_1, X_2, for example $t_1(x, y) = (x + 1, y)$, $t_2(x, y) = (x + a, y + b)$ with $b \neq 0$. The quotient is then a so-called *flat torus*. The standard situation is when $(a, b) = (0, 1)$ (the square torus); a different special case is when $(a, b) = (\frac{1}{2}, \frac{1}{2}\sqrt{3})$, the *hexagonal torus*, see Figure 7.4. One must identify pairs of points on opposite sides of the hexagon.

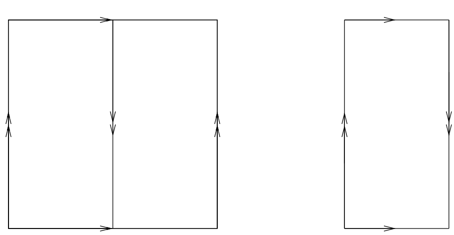


Figure 7.5. The flat Klein bottle

5. Γ_4 is generated by t_1, t_2, α with the same properties as above, i.e., $\alpha^2 = t_1$, for example

$$t_1(x,y) = (x+1,y),$$
$$t_2(x,y) = (x,y+1), \alpha(x,y) = (x+\frac{1}{2},-y).$$

The quotient E^2/Γ_4 is called a *flat Klein bottle*. Again, this may also be viewed as the quotient of a flat torus with $a=0, b=1$ by α. The projection is then a double cover. Alternatively, the Klein bottle can be viewed as a quotient of the Möbius strip. This becomes clearer when one considers the inclusions $\Gamma_1 \subset \Gamma_2 \subset \Gamma_4,\ \Gamma_1 \subset \Gamma_3 \subset \Gamma_4$.

(ii) The only complete two-manifolds with $K=1$ are the sphere S^2 itself and $I\!RP^2$, which is the quotient of S^2 by a group with two elements which is generated by the antipodal map $\sigma(x,y,z) = (-x,-y,-z)$. We get a model of $I\!RP^2$ if we take a copy of the closed upper hemisphere and imagine the identification of pairs of antipodal points along the equator.

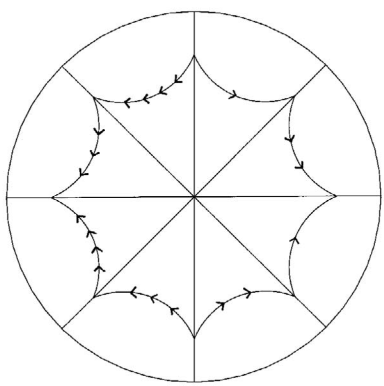


Figure 7.6. A geodesic octagon in the conformal disc model of H^2 with the necessary identifications

There are infinitely many compact surfaces with $K=-1$. Orientable examples are obtained from a regular $4g$-gon in H^2, which is chosen in such a way that the edges are all segments of geodesics of equal length and such that all inner angles are exactly $2\pi/4g$, see Figure

7.6 for the case $g = 2$. Looking at the identifications on the boundary in Figure 7.6, we see that the surface consists of two tori with a hole which are glued together.

This is made possible by choosing an appropriate size of the $4g$-gon, cf. the Gauss-Bonnet formulas in 4.39 and 4.40. For very small regular polygons the angles approximately coincide with the angles of Euclidean regular polygons, while as they grow in size, the angles become smaller and smaller. The genus g occuring here is an arbitrary number $g \geq 2$. By appropriate identifications on the sides one gets a closed surface of genus g (cf. [**50**, Ch.5] or [**47**]), which locally still has the hyperbolic metric, even along the loci of where identifications take place. These identifications have to be made in such a way that the result is a union of g tori, glued together after cutting. For more details on this "gluing recipe" see [**37**, Ch.12]. In Ch. 10 there one can also find more details on the geometry of the hyperbolic plane. To get non-oriented surfaces of genus g one can proceed similarly, starting with a $2g$-gon and identifying pairs of consecutive sides in the same direction. The result is a union of g copies of the projective plane, glued after cutting.

Together with the classification theorem for surfaces at the end of Section 4F, this yields a proof of the following theorem.

7.25. Theorem. There exists on an arbitrary two-dimensional compact manifold M a Riemannian metric with constant curvature K. The sign of K is by the Gauss-Bonnet theorem necessarily equal to the sign of the Euler characteristic of M, $\chi(M)$.

Compare Remark 2 after 4.47.

7D Three-dimensional Euclidean and spherical space forms

A three-dimensional analog of Theorem 7.25 is not possible, since there are three-dimensional manifolds which admit no metric with constant sectional curvature, for example $S^1 \times S^2$ (see Section 8.1 for this). In addition, in this dimension there are many more examples of

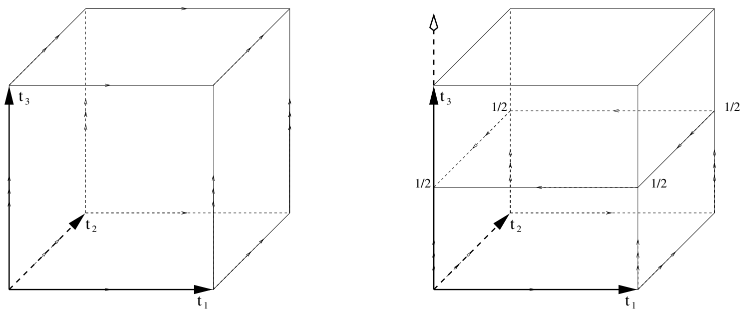


Figure 7.7. The groups Γ_1 and Γ_2

topologically different manifolds which have constant sectional curvature. This is already quite interesting in the cases $K = 0$ and $K = 1$. In what follows, we present a few examples.

7.26. Examples. (compact three-dimensional Euclidean space forms):

There are ten compact quotients E^3/Γ of E^3, of which six are orientable while four are non-orientable[3]. The orientable ones are the following, described by means of the respective groups $\Gamma_1, \ldots, \Gamma_6$ as in Figure 7.7.

1. Γ_1 is generated by three translations t_1, t_2, t_3 in the directions of three linearly independent vectors X_1, X_2, X_3. The quotient E^3/Γ_1 is called a *three-dimensional torus*. The most important special case is the situation where the X_i are the elements of the standard basis of $\mathbb{R}^3$. In this case, Γ_1 is just the translation group of the integer lattice consisting of all points in $\mathbb{Z}^3$:

$$\begin{aligned} t_1(x,y,z) &= (x+1,y,z), \\ t_2(x,y,z) &= (x,y+1,z), \\ t_3(x,y,z) &= (x,y,z+1). \end{aligned}$$

2. Γ_2 is generated by Γ_1 and a screw-motion α with $\alpha^2 = t_3$. Here we must make the assumption that the plane spanned by X_1

[3] W. Hantzsche & H. Wendt, *Dreidimensionale euklidische Raumformen*, Math. Annalen **110**, 593–611 (1935).

and X_2 is perpendicular to X_3. In the simplest case we thus have

$$\begin{aligned} t_1(x,y,z) &= (x+1,y,z),\\ t_2(x,y,z) &= (x,y+1,z),\\ t_3(x,y,z) &= (x,y,z+1),\\ \alpha(x,y,z) &= (-x,-y,z+\tfrac{1}{2}). \end{aligned}$$

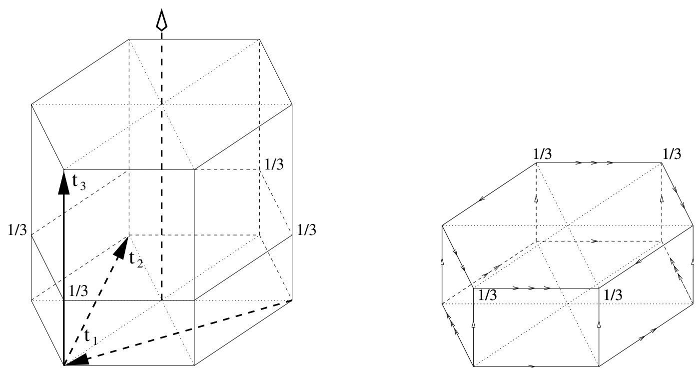


Figure 7.8. The group Γ_3

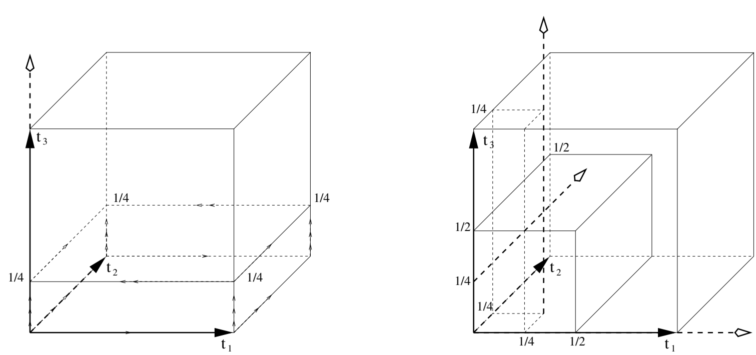


Figure 7.9. The groups Γ_4 and Γ_6

3. Γ_3 is generated by Γ_1 and a screw-motion α with $\alpha^3 = t_3$. Here we must assume that the (X_1, X_2)-plane is perpendicular to X_3 and, in addition, that t_1 and t_2 are compatible with a rotation of $2\pi/3$, for example

$$\begin{aligned} t_1(x,y,z) &= (x+1, y, z), \\ t_2(x,y,z) &= (x - \tfrac{1}{2}, y + \tfrac{1}{2}\sqrt{3}, z), \\ t_3(x,y,z) &= (x, y, z+1), \\ \alpha(x,y,z) &= (-\tfrac{1}{2}x - \tfrac{1}{2}\sqrt{3}y, \tfrac{1}{2}\sqrt{3}x - \tfrac{1}{2}y, z + \tfrac{1}{3}). \end{aligned}$$

4. Γ_4 is generated by $t_1(x,y,z) = (x+1, y, z)$, $t_2(x,y,z) = (x, y+1, z)$, $t_3(x,y,z) = (x, y, z+1)$ together with the screw-motion α with $\alpha^4 = t_3$, hence $\alpha(x,y,z) = (y, -x, z + \frac{1}{4})$.

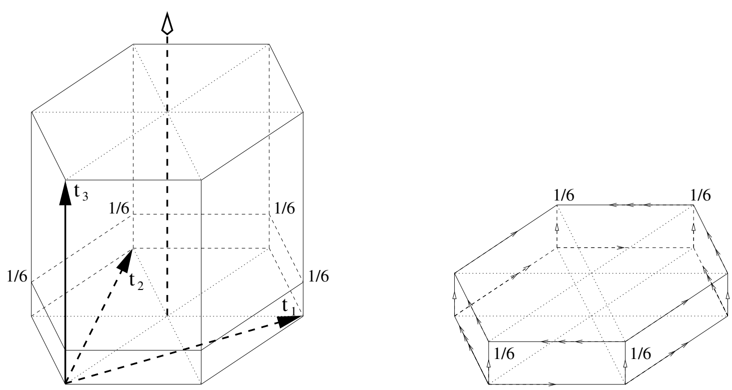


Figure 7.10. The group Γ_5

5. Γ_5 is defined like Γ_3, with the difference that here one has $\alpha^6 = t_3$, so that the screw-motion contains a rotation of $\pi/3$ instead of $2\pi/3$:

$$\alpha(x,y,z) = (\tfrac{1}{2}x - \tfrac{1}{2}\sqrt{3}y, \tfrac{1}{2}\sqrt{3}x + \tfrac{1}{2}y, z + \tfrac{1}{6}).$$

6. Γ_6 results from Γ_2 by adding two further screw-motions with an angle of π, so that here, there are screw-motions around the

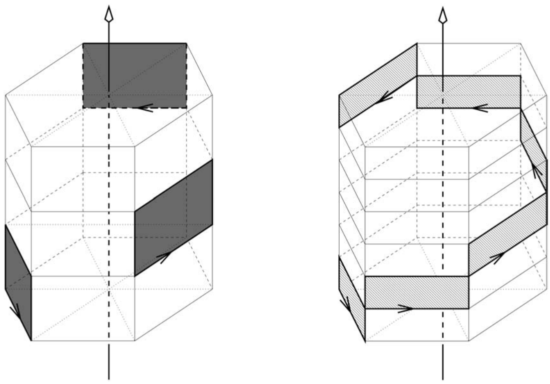


Figure 7.11. The screw-motion α in Γ_3 and Γ_5

$(x, 0, 0)$-axis, the $(0, y, \frac{1}{4})$-axis as well as the $(\frac{1}{4}, \frac{1}{4}, z)$-axis:

$$\begin{aligned}
t_1(x,y,z) &= (x+1, y, z),\\
t_2(x,y,z) &= (x, y+1, z),\\
t_3(x,y,z) &= (x, y, z+1),\\
\alpha(x,y,z) &= (x+\tfrac{1}{2}, -y, -z),\\
\beta(x,y,\tfrac{1}{4}+z) &= (-x, y+\tfrac{1}{2}, \tfrac{1}{4}-z),\\
\gamma(\tfrac{1}{4}+x, \tfrac{1}{4}+y, z) &= (\tfrac{1}{4}-x, \tfrac{1}{4}-y, z+\tfrac{1}{2}).
\end{aligned}$$

For the construction of *spherical space forms*, i.e., quotients of S^3 by finite groups, we can use the fact that S^3 is itself a group in a natural way. To explain this, recall that the *quaternion algebra* is defined on $\mathbb{R}^4$ as follows:

$$\mathbb{H} := \{a + bi + cj + dk \mid a, b, c, d \in \mathbb{R}\},$$

where the symbols i, j, k are so-called "imaginary units" for which the relations $i^2 = j^2 = k^2 = -1, ij = k = -ji, jk = i = -kj, ki = j = -ik$ hold. This defines a (non-commutative, but associative) multiplication on $\mathbb{H}$ which admits a well-defined division by every non-zero element. Such a structure is also called a *skew-field*. *Conjugation* is defined – similarly as with the complex numbers – by $\bar{z} = a - bi - cj - dk$ for $z = a + bi + cj + dk$. The quantity $z\bar{z}$ is then purely real, and in fact, calculates to $z\bar{z} = a^2 + b^2 + c^2 + d^2$.

To describe the three-dimensional spaces of constant positive curvature, it is convenient to view the unit sphere as a subset of the quaternions, since in this way the underlying space gets a multiplication, i.e., a group structure:

$$S^3 = \{z \in \mathbb{H} \mid z\bar{z} = 1\}.$$

The finite subgroups of this group are easily classified. There is also an interesting relation to the three-dimensional rotation group, which is described in the following lemma.

7.27. Lemma. There is a group homomorphism $S^3 \longrightarrow \mathbf{SO}(3)$ which identifies antipodal points on the sphere. In particular, the rotation group $\mathbf{SO}(3)$ is diffeomorphic to the projective space $\mathbb{R}P^3$, which implies in particular that as topological spaces they are identical.

PROOF: Let a unit quaternion $q \in S^3$ be given (i.e., a quaternion of length 1); then conjugation in the group by q defines a mapping $\mathbb{H} \longrightarrow \mathbb{H}$ by

$$x \longmapsto q \cdot x \cdot q^{-1}.$$

We will denote this mapping by $\widetilde{q}$; hence $\widetilde{q}(x) = qxq^{-1}$. This conjugation obviously fixes the real axis, so that $\widetilde{q}$ can be viewed as a linear map of the three-dimensional imaginary part E^3 (put $\mathbb{H} = \mathbb{R} \oplus E^3$). Since this map clearly preserves the norm, we have

$$\langle qxq^{-1}, qyq^{-1} \rangle = \langle x, y \rangle$$

for all x, y. It follows that $\widetilde{q} : E^3 \to E^3$ is an orthogonal map, i.e., the association $q \mapsto \widetilde{q}$ can be viewed as a map

$$\pi \colon S^3 \to \mathbf{SO}(3), \ \pi(q) = \widetilde{q}.$$

The equation

$$\widetilde{q_1 q_2} = \widetilde{q_1} \cdot \widetilde{q_2}$$

shows that π is a group homomorphism. It is clear that $\pi(q) = \pi(-q)$, and in fact

$$\pi(q_1) = \pi(q_2) \Longleftrightarrow q_1 = \pm q_2.$$

This can be seen as follows. Assume that q_1, q_2 are given and have the property that $q_1 x q_1^{-1} = q_2 x q_2^{-1}$ for all x. Then one also has $x q_1^{-1} q_2 = q_1^{-1} q_2 x$ for all x, so that $q_1^{-1} q_2$ commutes with arbitrary quaternions. Thus $q_1^{-1} q_2$ is necessarily real, and because of the absolute value, $q_1^{-1} q_2 = \pm 1$. From this it follows that the map π identifies exactly pairs of antipodal points. In topological language, $\pi : S^3 \to \mathbf{SO}(3)$ is a *double covering*, cf. [**38**]. In particular, the two manifolds $\mathbf{SO}(3) \cong \mathbb{R}P^3$ are diffeomorphic. $\square$

We can now determine the finite subgroups H of S^3 and the corresponding quotients S^3/H. In view of the double covering

$$\pi \colon S^3 \to \mathbf{SO}(3)$$

above, we set $\widetilde{G} := \pi^{-1}(G)$ for a finite subgroup $G \subset \mathbf{SO}(3)$. The latter groups are classified by the following theorem. For a proof, see [**52**], §62.

7.28. Theorem. The finite subgroups of $\mathbf{SO}(3)$ are the following

the *cyclic group* C_k or order k;

the *dihedral group* D_k of order $2k$;

the *tetrahedral group* T of order 12;

the *octahedral group* O of order 24;

the *icosahedral group* I of order 60.

These groups are defined as follows:

— C_k is the rotation group of a regular k-gon in $\mathbb{R}^2$;

— D_k is the rotation group of a regular k-gon in $\mathbb{R}^3$ (including a rotation in space);

— T is the rotation group of a regular tetrahedron;

— O is the rotation group of a regular octahedron (or cube);

— I is the rotation group of a regular icosahedron (or dodecahedron).

The *tetrahedron* is defined as the convex hull of the four points

$$(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)$$

in the three-dimensional hyperplane $x_1+x_2+x_3+x_4=1$ of $I\!R^4$, the *octahedron* is defined as the convex hull of the six points

$$(\pm 1,0,0),(0,\pm 1,0),(0,0,\pm 1)$$

in $I\!R^3$, and the *icosahedron* is defined as the convex hull of the 12 points

$$(0,\pm\tau,\pm 1),(\pm 1,0,\pm\tau),(\pm\tau,\pm 1,0),$$

where the number $\tau=\frac{1}{2}(1+\sqrt{5})=2\cos\frac{\pi}{5}\approx 1.618$ is also known as the *golden ratio* and satisfies the equation $\tau^2-\tau-1=0$. The number τ plays an important role in the esthetics of art and architecture regarding the optimal ratio between height and width.

The notation used to describe the last three groups mentioned is according to Coxeter the following:

$$\begin{array}{ccccc} T & = & [2,3,3]_+ & = & \langle A,B \mid A^3=B^3=(AB)^2=1\rangle; \\ O & = & [2,3,4]_+ & = & \langle A,B \mid A^3=B^4=(AB)^2=1\rangle; \\ I & = & [2,3,5]_+ & = & \langle A,B \mid A^3=B^5=(AB)^2=1\rangle. \end{array}$$

The right-hand sides of the equations contain two generating elements A,B and certain relations between them which determine the corresponding group completely, something called a *presentation* of the group by generators and relations, see [**34**].

7.29. Theorem. The finite subgroups of S^3 are the following:

1. The *cyclic group* C_k of odd order k;
2. The *cyclic group* $C_{2k}=\widetilde{C}_k=\pi^{-1}(C_k)$;
3. The *dicyclic group* (*binary dihedral group*) $\widetilde{D}_k=\pi^{-1}(D_k)$ of order $4k$;
4. The *binary tetrahedral group* $\widetilde{T}=\pi^{-1}(T)$ of order 24;
5. The *binary octahedral group* $\widetilde{O}=\pi^{-1}(O)$ of order 48;
6. The *binary icosahedral group* $\widetilde{I}=\pi^{-1}(I)$ of order 120.

This follows from 7.28 and 7.27, see [**50**], 3.8. The name *binary group* comes of course from the double cover π. Through π, every group element is doubled so to speak in a (+)-version and a (−)-version. In fact for all the groups in question we have $G=\widetilde{G}/\{\pm 1\}$. Again we have a presentation with two generators and relations, as follows:

$$\begin{array}{rcccl} \widetilde{T} & = & \langle 2,3,3\rangle & = & \langle A,B \mid A^3 = B^3 = (AB)^2\rangle; \\ \widetilde{O} & = & \langle 2,3,4\rangle & = & \langle A,B \mid A^3 = B^4 = (AB)^2\rangle; \\ \widetilde{I} & = & \langle 2,3,5\rangle & = & \langle A,B \mid A^3 = B^5 = (AB)^2\rangle. \end{array}$$

The difference in the presentations above is that one no longer requires the equality with unity. In fact, these elements are then equal to -1 (which is however not a part of the relation).

REMARK: The subgroup $\widetilde{I}$ may also be viewed as a set of points in $S^3 \subset \mathbb{R}^4$. The convex hull of these 120 points is the so-called 600-cell $\{3,3,5\}$, see [**48**, p.247]. This is a regular solid in 4-space with 120 vertices whose boundary consists of 600 regular tetrahedra.

Similarly, $\widetilde{T}$ is the set of vertices of the 24-cell $\{3,4,3\}$ and $\widetilde{O}$ is the set of vertices of the 24-cell plus the set of dual points (where duality here means vertices viewed as faces and faces viewed as vertices). The 24-cell is a regular solid in 4-space whose boundary consists of 24 ordinary octahedra, see [**48**, p.245]. Its vertices can be identified with the *unit Hurwitz quaternions* $\pm 1, \pm i, \pm j, \pm k, \frac{1}{2}(\pm 1 \pm i \pm j \pm k)$.

7.30. Theorem. (Three-dimensional spherical space forms)
The quotient $S^3/\widetilde{G}$ of the 3-sphere by any of the subgroups $\widetilde{G}$ in 7.29 is a three-dimensional spherical space form since in any case $\widetilde{G}$ operated on S^3 without fixed points. These spaces $S^3/\widetilde{G}$ are traditionally referred to in the following way[4], compare [**50**, 3.8]:

S^3/C_k	*lens space* (in the particular case when $k = 2$: *projective space*);
$S^3/\widetilde{D}_k$	*prism space* (for $k = 2$ also: *quaternion space*);
$S^3/\widetilde{T}$	*octahedral space*;
$S^3/\widetilde{O}$	*truncated cube space*;
$S^3/\widetilde{I}$	*(spherical) dodecahedral space.*

To get the spherical dodecahedral space one divides the sphere S^3 into 120 pieces (given by the 120-cell) and identifies sides in an appropriate manner. Every piece is a three-dimensional (solid) dodecahedron.

[4]We follow W. Threlfall & H. Seifert, *Topologische Untersuchung der Diskontinuitätsbereiche endlicher Bewegungsgruppen des dreidimensionalen sphärischen Raumes*, Math. Annalen **104**, 1–70 (1931). The complete classification is contained in part II of the paper *ibid.* **107**, 543–586 (1932). Here certain extensions of these groups as subgroups of $\mathbf{SO}(4)$ come in.

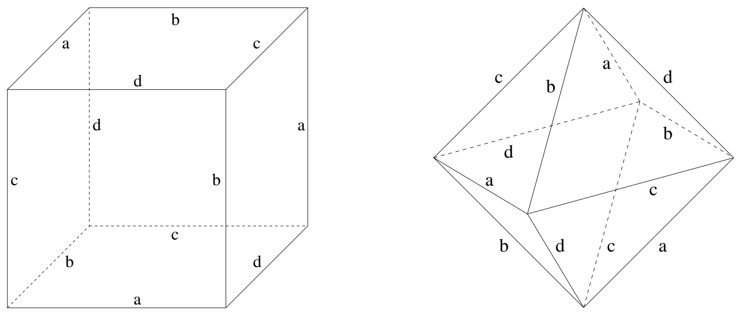


Figure 7.12. Quaternionic space and octahedral space

This can be visualized by starting with one (solid) dodecahedron and gluing the appropriate faces of the boundary, see Figure 7.13 (see also [**41**], p. 216). The spherical dodecaedral space is often called the *Poincaré dodecahedral space* or the *Poincaré homology sphere*. The pictures of the spherical space forms should actually be viewed as being contained in the three-sphere, that is, as spherical polyhedra with identifications along the boundaries. The angles between the edges are then much larger than they seem, and it is only because of insufficient technology that the pictures let them appear to look like Euclidean polyhedra. The difference is like that between a spherical triangle and a Euclidean triangle, cf. Figure 4.4. A special case is given by the *quaternion space*, which is the prism space in the case of $k = 2$, which in turn is also referred to as the *cube space*. The corresponding group $\widetilde{D}_2$ is best described as the subset of the quaternions $\widetilde{D}_2 = \{\pm 1, \pm i, \pm j, \pm k\} \subset I\!H$, the so-called *quaternion group* of order 8. The cube space results from a spherical three-dimensional cube by identifications along its boundary, and similarly for the octahedral space, see Figure 7.12. Surprisingly enough, the five spaces that occur in the statement of Theorem 7.30 are important not only in geometry and topology, but also in cosmology as models of a so-called multiconnected spherical universe.[5]

[5]See J. Weeks, *The Poincaré dodecahedral space and the mystery of the missing fluctuations*, Notices of the American Math. Society **51**, 610–619 (2004) and E. Gausmann, R. Lehoucq, J.-P. Luminet, J.-Ph. Uzan & J. Weeks, *Topological lensing in spherical spaces*, Classical and Quantum Gravity **18**, 5155–5168 (2001).

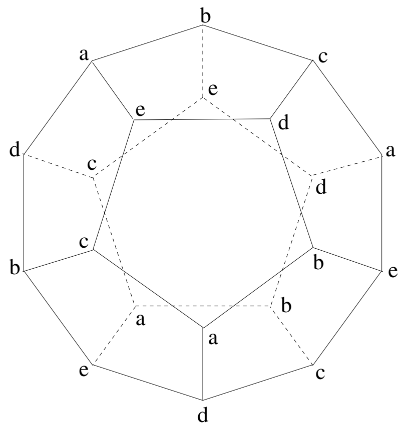


Figure 7.13. Spherical dodecahedral space

Exercises

1. Show that for $n \geq 3$, a metric with constant negative curvature on an n-manifold cannot be realized even locally as a hypersurface in Euclidean $\mathbb{R}^{n+1}$. Hint: Use the Gauss equation 4.21. There is consequently no higher-dimensional analog of the two-dimensional pseudo-sphere of 3.17.
2. Calculate the metric on $\mathbb{R}^n$ which arises from the standard metric on the sphere via the stereographic projection. In other words, calculate the metric on $\mathbb{R}^n$ with respect to which the stereographic projection becomes an isometry, see 7.7.
3. Show that the map Φ defined at the end of 7.7 between the two models of hyperbolic space is actually a globally defined isometry.
4. Show that the complex mapping
$$f(z) = i \cdot \frac{1-z}{1+z}$$
is a bijection between the open unit disc and the upper half-plane, and that moreover it is an isometry between the disc model of hyperbolic space (see 7.7) and the Poincaré upper half-plane (cf. the exercises at the end of Chapter 4).

5. Let ds_k^2 denote an $(n-1)$-dimensional metric of constant curvature k. Decide which sectional curvatures the following n-dimensional metrics have:

$$dt^2 + \cos^2(t)ds_1^2, \quad dt^2 + e^{2t}ds_0^2, \quad dt^2 + \cosh^2(t)ds_{-1}^2.$$

Hint: Exercises 10-12 at the end of Chapter 6.

6. Define geodesic polar coordinates for the standard metric of real projective space, which arises as the (locally isometric) quotient of the standard metric on the sphere.

7. Two points p, q on a geodesic c are said to be *conjugate* along c, if there is a Jacobi field along c which vanishes at p and q, but does not vanish identically. The dimension of the space of all such Jacobi fields is called the *multiplicity* of q with respect to p. Show that the multiplicity can have at most the value of $n-1$ if n is the dimension of the manifold.

8. Show that p and q are conjugate along some geodesic c, if there is a $V \in T_pM$ such that $D\exp_p|_V : T_V(T_pM) \to T_qM$ does not have maximal rank. The multiplicity is just the defect of $D\exp_p$. Hint: $D\exp_p$ transforms linear fields into Jacobi fields and conversely, according to 7.15.

9. Let $c : [a, b] \to M$ be a geodesic, and let $p = c(a)$ and $q = c(b)$ be *non*-conjugate along c. Show that a Jacobi field Y along c is uniquely determined by $Y(a)$ and $Y(b)$. Hint: Consider the difference of two Jacobi fields with the same "initial values".

10. Show that each of the compact Euclidean space forms in 7.26 is a quotient of a three-dimensional torus in such a way, that the fundamental group of the three-torus is a normal subgroup of the fundamental group of the space form. For the notion of the fundamental group compare [**38**, Ch.5].

11. Show that the holonomy group of a two- or three-dimensional Euclidean space form E^2/Γ or E^3/Γ is isomorphic to the quotient by Γ of the largest purely translation-subgroup of Γ (a normal subgroup). The order of the holonomy group is $1, 1, 2, 1, 2$ for the five examples in 7.24 and $1, 2, 3, 4, 6, 4$ for the six examples in 7.26.

12. Show that the tangent bundle of the three-sphere is globally diffeomorphic to the product manifold $S^3 \times I\!R^3$. Hint: Use the group structure on S^3 of 7.23 to obtain everywhere linearly independent vector fields. Is the same true of the rotation group $\mathbf{SO}(3)$?

13. Let M be a differentiable manifold and $\sigma\colon M \to M$ a differentiable involution without fixed points, i.e., assume $\sigma(\sigma(x)) = x$ and $\sigma(x) \neq x$ for all $x \in M$. Prove that we get a new differentiable manifold M_σ by identification of all pairs $\{x, \sigma(x)\}$. If g is a Riemannian metric on M ist and if σ is an isometry then M_σ carries an induced Riemannian metric in such a way, that the quotient map from M onto M_σ (which is a 2-fold covering map) becomes a local isometry. In particular, if M is a space form then so is M_σ.

14. Let M be a spherical space form, i.e., a compact manifold of constant positive curvature. Show that the compact manifold $S^1 \times M$ carries no metric of constant curvature. Hint: 7.23 (iii).

15. For this exercise, view the three-sphere S^3 as the subset of $\mathbb{C}^2$ given by
$$S^3 = \{(z, w) \in \mathbb{C}^2 \mid |z| = |w| = 1\}.$$
Define a group operation for relatively prime natural numbers p and q by the formula
$$(k, (z, w)) \mapsto (e^{2\pi ik/p}z, e^{2\pi iqk/p}w)$$
with $k = 0, 1, \ldots, p-1$. Determine the group, and show that this group operation is discrete and fixed-point-free and contained in the orthogonal group $\mathbf{SO}(4)$. What is the orbit of a point in S^3? The quotient by this operation is called the *lens space* $L(p, q)$. This space is important in topology, see [**38**], Chapter 4 and [**41**], §60.

Chapter 8

Einstein Spaces

The following question arises naturally for a given differentiable manifold M (initially considered without a Riemannian metric):

Is there a "best" metric whose curvature has the property of being most evenly distributed about the manifold?

For comparison, look at the surfaces of constant Gaussian curvature in 7.25 as well as at the minimal surfaces in Section 3D, for which the curvature is distributed in such a way that the mean curvature is everywhere vanishing. The mean curvature there was given as the gradient of a surface integral, see 3.28. This "variational principle" is a quite natural one and is often applied in the natural sciences. Similarly, there are physical reasons for considering a four-dimensional space-time with special curvature properties, in particular looking for a metric with optimal distribution of the curvature, where the distribution of mass and the resulting gravitational force are the motivating factors. In this way one is led to the Einstein field equations, in which the Einstein tensor occurs as the gradient of some functional. At any rate one is led to consider the so-called *Einstein metrics*, which represent an important and interesting chapter of Riemannian geometry. According to 6.13, Einstein metrics are those for which the Ricci curvature is constant. As a continuation of the end of Chapter 6, this Chapter 8 will give a brief introduction to some phenomena in the context of Einstein metrics. An in-depth source on the topic of Einstein manifolds is the book [**24**].

8.1. Remark. (Special metrics in dimensions $2, 3, 4$)

The following result, Theorem 7.25, summarizes the situation in dimension two:

On an arbitrary two-dimensional compact manifold there exists a Riemannian metric with constant curvature K.

The construction of this metric follows the line of thought of 7.24. One looks for

$$\text{quotients of } \begin{cases} E^2, & \text{if } K = 0, \\ S^2, & \text{if } K = 1, \\ H^2, & \text{if } K = -1, \end{cases}$$

and attempts to represent every possible topological type of a compact two-dimensional manifold as one of these quotients. The answer to the question initially posed is thus a clear "yes" in dimension $n = 2$.

Already in the case of $n = 3$, and all the more so in higher dimensions, the situation is fundamentally different. For the particular case of dimension $n = 3$ one has the following situation, which we quote without proof and which will not be used in the remainder of the chapter. It is just a good illustration of the phenomena which occur.

(1) Not every compact manifold of dimension three admits a Riemannian metric of constant curvature For example, there is no such metric on the product manifold $S^1 \times S^2$. $S^1 \times S^2$ does not admit such a metric since no covering of the space is E^3, S^3, or H^3 (there is no quotient map from one of these to $S^1 \times S^2$). The universal covering of $S^1 \times S^2$ is $\mathbb{R} \times S^2$ with the covering projection $(t, x) \longmapsto (e^{it}, x)$.

(2) According to W. Thurston[1], every three-manifold has a canonical decomposition into pieces, where each of the pieces carries one of eight standard metrics. Among these are $E^3, S^3, H^3, \mathbb{R} \times S^2, \mathbb{R} \times H^2$ and the so-called *Heisenberg group.*

(3) The more special class of all three-manifolds with a metric of constant negative curvature is already extremely rich.

[1] *Three-dimensional manifolds, Kleinian groups and hyperbolic geometry*, Bulletin of the American Math. Society **6**, 357–381 (1982).

In dimension $n = 4$ there is a similar state of affairs, insofar as $S^1 \times S^3$ and $S^2 \times S^2$ admit no metric of constant curvature, for the same reason as for $S^1 \times S^2$ above: the universal covering of $S^1 \times S^3$ is $\mathbb{R} \times S^3 \to S^1 \times S^3$, while the product $S^2 \times S^2$ is already simply connected, i.e., is its own universal cover.

If S^2 denotes the unit sphere, then the product metric on $S^2 \times S^2$ does not have constant curvature, but it is at least an Einstein metric. This can be seen as follows. Let E_1, E_2, E_3, E_4 denote an ON-basis, where E_1, E_2 are tangent to the first factor and E_3, E_4 are tangent to the second factor. Now calculate the corresponding sectional curvatures as $K_{12} = K_{34} = 1$ (curvature of S^2) and $K_{13} = K_{14} = K_{23} = K_{24} = 0$ (curvature of $\mathbb{R}^2$). It follows from this that

$$\begin{array}{rcccl} \mathrm{Ric}(E_1, E_1) & = & K_{12} + K_{13} + K_{14} & = & 1, \\ \mathrm{Ric}(E_2, E_2) & = & K_{21} + K_{23} + K_{24} & = & 1, \\ \mathrm{Ric}(E_3, E_3) & = & K_{31} + K_{32} + K_{34} & = & 1, \\ \mathrm{Ric}(E_4, E_4) & = & K_{41} + K_{42} + K_{43} & = & 1, \end{array}$$

and $\mathrm{Ric}(E_i, E_j) = 0$ for $i \neq j$. One gets $\mathrm{Ric} = g$, from which it follows that g is an Einstein metric.

Thus, by 6.13, dimension four is the smallest dimension for which non-trivial Einstein metrics can occur, i.e., Einstein metrics which are not already metrics of constant curvature. There is no local classification of Einstein metrics, but there is a classification in dimension four of those Einstein spaces which are homogenous.[2] At the same time, precisely this dimension is interesting, as it is on the one hand the dimension corresponding to two complex dimensions, while on the other hand it admits a duality (cf. Section 8E), and finally, it is the dimension in which the classical space-time of the general theory of relativity occurs. We will spend some time discussing this last aspect by considering the so-called Einstein field equations, which are motivated mathematically as well as physically.

---

[2] G. R. Jensen, *Homogeneous Einstein spaces of dimension four*, Journal of Differential Geometry **3**, 309–349 (1969).

8A The variation of the Hilbert-Einstein functional

From a mathematical point of view one can motivate the importance of Einstein metrics as those for which the distribution of the scalar curvature is optimal, meaning it is minimized in a certain sense. What we precisely mean by this is described in what follows by means of a variation of certain curvature functionals in the space of all Riemannian metrics. For this we will keep the underlying manifold M fixed for the following considerations.

8.2. Definition. Let (M, g) be a compact Riemannian (or pseudo-Riemannian) manifold which we assume is oriented. Let dV_g be the volume element (in coordinates $dV_g = \sqrt{\mathrm{Det} g_{ij}} dx_1 \wedge \cdots \wedge dx_n$, or $dV_g = \sqrt{|\mathrm{Det} g_{ij}|} dx_1 \wedge \cdots \wedge dx_n$ in the pseudo-Riemannian case). The following functionals are defined for fixed M and varying metric g:

$$\begin{array}{rcll} \mathbf{Vol}\,(g) & = & \int_M dV_g & \text{(volume of } g\text{);} \\ \mathbf{S}(g) & = & \int_M S_g dV_g & \text{(total scalar curvature of } g\text{).} \end{array}$$

The functional $\mathbf{S}$ is also referred to as the *Hilbert-Einstein functional* after A. Einstein and D. Hilbert, cf. 8.6.

In the particular case of dimension $n = 2$ one has according to the Gauss-Bonnet theorem 4.43

$$\mathbf{S}(g) = 2 \int_M K dV = 4\pi\chi(M),$$

which means the functional $\mathbf{S}(g)$ is constant if the manifold M is fixed.

Our goal in what follows is to calculate the "derivative" of $\mathbf{S}$ through consideration of the variational problem $\delta\mathbf{S} = 0$. Those metrics g for which $\delta\mathbf{S} = 0$ then naturally have a privileged geometric property. The method used here is quite similar to the variation of the length of curves used in 4.13 as well as the variation used in the study of minimal surfaces in Section 3D.

8.3. Definition. (Variation of the metric)

Recall the procedure from Section 3D used in the study of surfaces for which the surface area is minimal. There we started with a given surface element $f(u^1, u^2)$ as well as an arbitrary but fixed function $\varphi(u^1, u^2)$. The variation of the surface area was described by

$$f_\epsilon(u^1, u^2) := f(u^1, u^2) + \epsilon \cdot \varphi(u^1, u^2) \cdot \nu(u^1, u^2).$$

There we found, for the first fundamental form $I_\epsilon = g_\epsilon$,

$$g_\epsilon = g - 2\epsilon \cdot \varphi \cdot II + \epsilon^2(\dots),$$

where II is the second fundamental form of f. The *area functional* $\mathbf{A}$ was calculated to be

$$\mathbf{A}(g_\epsilon) = \int \sqrt{\mathrm{Det} g_\epsilon} du^1 \wedge du^2 = \int dV_{g_\epsilon}.$$

Its "derivative in the direction φ" was found to be

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} (\mathbf{A}(g_\epsilon)) = -\int 2H \cdot \varphi \cdot dV =: \big\langle -2H, \varphi \big\rangle_g,$$

where we view $\big\langle\ ,\ \big\rangle_g$ as an inner product on the space of scalar functions φ. The quantity $-2H = -\mathrm{Tr}_g II$ may then be viewed as the "gradient" of the functional $\mathbf{A}$. f is a minimal surface if and only if $\mathbf{A}$ is stationary (that is, if $\delta \mathbf{A} = 0$), which in turn means that $H \equiv 0$.

We proceed similarly for the functional $\mathbf{S}(g)$:

Let a manifold M with metric g be given. The *variation of the metric in the direction* h with the real parameter t is defined by

$$g_t := g + t \cdot h,$$

where h is an arbitrary, but fixed, symmetric (0,2)-tensor field.

Since g is certainly non-degenerate, g_t is also non-degenerate for sufficiently small $t \in (-\varepsilon, \varepsilon)$. This follows for example from the continuity of the determinant of g_t in local coordinates, as long as either M is compact or $h \equiv 0$ outside of a compact set.

The derivative of the real-valued function $t \longmapsto \mathbf{S}(g_t)$ at the point $t = 0$ can be viewed as a directional derivative of $\mathbf{S}$ in the direction h at the point g, and we can study the *variational problem* $\delta\mathbf{S}(g) = 0$, i.e., the condition

$$\delta\mathbf{S}(g) = 0 \quad \Longleftrightarrow \quad \frac{d}{dt}\mathbf{S}(g_t)\Big|_{t=0} = 0 \ \text{ for all } h.$$

One can also say that $\mathbf{S}$ is *stationary* for such a metric g. Assuming this is the case, we can try to define the "gradient" of $\mathbf{S}$ with respect to an appropriately defined inner product on the space of symmetric (0,2)-tensors.

But in order to evaluate this derivative of $\mathbf{S}$ in the direction h, we first have to calculate the individual parts occuring in 8.4 and 8.5, which are

— the derivative of $S_g = \mathrm{Tr}(\mathrm{Ric}_g)$ with respect to t,

— the derivative of the curvature tensor with respect to t,

— the derivative of the volume element with respect to t.

8.4. Lemma. (Variation of the volume form)
Let dV_t be the volume element of $g_t = g + t \cdot h$ with $dV_0 = dV_g$. Then

$$\frac{d}{dt}\Big|_{t=0}(dV_t) = \frac{1}{2}\mathrm{Tr}_g h \cdot dV_g$$

Proof: In local coordinates we calculate:

$$dV_t = \sqrt{\mathrm{Det}(g_{ij}^{(t)})}\; dx^1 \wedge \cdots \wedge dx^n,$$

$$\lim_{t\to 0}\frac{1}{t}(dV_t - dV_g) = \lim_{t\to 0}\frac{1}{t}\left(\sqrt{\mathrm{Det}(g_{ij}^{(t)})} - \sqrt{\mathrm{Det}(g_{ij})}\right) dx^1 \wedge \cdots \wedge dx^n$$

$$= \lim_{t\to 0}\frac{1}{2t}\left(\mathrm{Det}(g_{ij}^{(t)}) - \mathrm{Det}(g_{ij})\right) \cdot \frac{1}{\sqrt{\mathrm{Det}(g_{ij})}}\; dx^1 \wedge \cdots \wedge dx^n$$

$$= \lim_{t\to 0}\frac{1}{2t}\left(\mathrm{Det}\underbrace{\left(\sum_j g_{ij}^{(t)} g^{jk}\right)}_{\delta_i^k + t\sum_j h_{ij}g^{jk}} - 1\right)\underbrace{\sqrt{\mathrm{Det}(g_{ij})}\; dx^1 \wedge \cdots \wedge dx^n}_{dV_g}$$

$$= \frac{1}{2}\lim_{t\to 0}\frac{1}{t}\Big(1 + t\ \mathrm{Tr}_g h + t^2(\cdots) + \cdots + t^n(\cdots) - 1\Big)dV_g = \frac{1}{2}\ \mathrm{Tr}_g h \cdot dV_g.$$

A similar fact holds in the case of an indefinite matrix, if one replaces $\sqrt{\mathrm{Det}(g_{ij})}$ by $\sqrt{|\mathrm{Det}(g_{ij})|}$ throughout. □

In particular, dV is stationary if and only if $\mathrm{Tr}_g h \equiv 0$, a relation familar to us already from our work with minimal surfaces. In fact, the proof of 3.28 is in a sense already contained in the proof of 8.4 as a special case. There we have $h = -2II$, and the surface area is to be viewed as a two-dimensional volume. Summarizing this, we could say: "the derivative of the volume is the trace."

8.5. Lemma. (Variation of the curvature tensor)
Let $g_t = g + t \cdot h$ be a variation of the metric g, and let ∇^t be the Riemannian connection of g_t, R^t the curvature tensor of g_t with the natural notations $\nabla^0 = \nabla$ and $R^0 = R$. Then we have:

(i) $g\Big(\frac{d}{dt}\Big|_{t=0} \nabla^t_X Y, Z\Big)$
$$= \tfrac{1}{2}\Big((\nabla_X h)(Y,Z) + (\nabla_Y h)(Z,X) - (\nabla_Z h)(X,Y)\Big)$$

(ii) The map
$$X, Y \mapsto \nabla'_h(X,Y) := \frac{d}{dt}\Big|_{t=0} \nabla^t_X Y$$
is a $(1,2)$-tensor field. It is symmetric in X and Y.

(iii) $\frac{d}{dt}\Big|_{t=0}(R^t(X,Y)Z) = (\nabla_X \nabla'_h)(Y,Z) - (\nabla_Y \nabla'_h)(X,Z)$.

Proof: We apply the formula from 4.15 for the Riemannian connection ∇^t:
$$g_t(\nabla^t_X Y, Z) = \tfrac{1}{2}\Big(X(g_t(Y,Z)) + Y(g_t(Z,X)) - Z(g_t(X,Y))$$
$$-g_t(X,[Y,Z]) - g_t(Y,[X,Z]) - g_t(Z,[Y,X])\Big).$$
With $g_t - g = th$ we get
$$g_t(\nabla^t_X Y, Z) - g(\nabla_X Y, Z) = \tfrac{1}{2}t\Big(X(h(Y,Z)) + Y(h(Z,X)) - Z(h(X,Y))$$
$$-h(X, \nabla_Y Z - \nabla_Z Y) - h(Y, \nabla_X Z - \nabla_Z X) - h(Z, \nabla_Y X - \nabla_X Y)\Big)$$
$$= th(\nabla_X Y, Z) + \tfrac{1}{2}t\Big(X(h(Y,Z)) - h(Y, \nabla_X Z) - h(Z, \nabla_X Y)$$
$$+Y(h(Z,X)) - h(Z, \nabla_Y X) - h(X, \nabla_Y Z)$$
$$-Z(h(X,Y)) + h(X, \nabla_Z Y) + h(Y, \nabla_Z X)\Big).$$

Using the definition of ∇h from 6.2, we get the simple expression

$$\begin{aligned} & g(\nabla^t_X Y - \nabla_X Y, Z) + th(\nabla^t_X Y, Z) \\ = {} & th(Z, \nabla_X Y) + \frac{1}{2} t\Big(\nabla_X h(Y,Z) + \nabla_Y h(X,Z) - \nabla_Z h(X,Y))\Big), \end{aligned}$$

and consequently in the limit as $t \to 0$

$$\begin{aligned} & \lim_{t\to 0} \frac{1}{t} g\Big(\nabla^t_X Y - \nabla_X Y, Z\Big) \\ = {} & \frac{1}{2}\Big[(\nabla_X h)(Y,Z) + (\nabla_Y h)(X,Z) - (\nabla_Z h)(X,Y)\Big], \end{aligned}$$

from which the statement in (i) follows.

(ii) follows simply from the fact that the right-hand side of (i) is clearly tensorial in X, Y and Z (in the sense of 6.1). Hence we can introduce ∇'_h as a tensor, which will simplify the notation in what follows. The symmetry of ∇'_h is obvious.

For the proof of (iii) we calculate:

$$\begin{aligned} & R^t(X,Y)Z - R(X,Y)Z \\ = {} & \nabla^t_X \nabla^t_Y Z - \nabla^t_Y \nabla^t_X Z - \nabla^t_{[X,Y]} Z - \nabla_X \nabla_Y Z + \nabla_Y \nabla_X Z + \nabla_{[X,Y]} Z \\ = {} & \nabla^t_X (\nabla^t_Y Z - \nabla_Y Z) - \nabla^t_Y (\nabla^t_X Z - \nabla_X Z) \\ & + (\nabla^t_X - \nabla_X)(\nabla_Y Z) - (\nabla^t_Y - \nabla_Y)(\nabla_X Z) \\ & - \nabla^t_{[X,Y]} Z + \nabla_{[X,Y]} Z. \end{aligned}$$

In the limit as $t \to 0$ we get from this

$$\begin{aligned} & \lim_{t\to 0} \frac{1}{t}\Big(R^t(X,Y)Z - R(X,Y)Z\Big) \\ = {} & \nabla_X(\nabla'_h(Y,Z)) - \nabla_Y(\nabla'_h(X,Z)) + \nabla'_h(X, \nabla_Y Z) \\ & - \nabla'_h(Y, \nabla_X Z) - \nabla'_h([X,Y], Z) \\ = {} & (\nabla_X \nabla'_h)(Y,Z) - (\nabla_Y \nabla'_h)(X,Z). \end{aligned}$$

□

8.6. Theorem. (Variation of the total scalar curvature[3])
Let M be a compact, orientable manifold, and let $g_t = g + t \cdot h$ be the variation of the metric g. Let S_t denote the scalar curvature of g_t. Then

$$\frac{d}{dt}\Big|_{t=0} \mathbf{S}(g_t) = \frac{d}{dt}\Big|_{t=0} \int_M S_t dV_t = \left\langle \frac{S}{2} g - \mathrm{Ric}, h \right\rangle_g .$$

Here we have set, for two symmetric $(0,2)$-tensors A, B,

$$\left\langle A, B \right\rangle_g := \int_M \sum_{i,j} A(E_i, E_j) B(E_j, E_i) dV_g$$

with an ON-basis $E_1, \ldots, E_n$. At every point this expression is just the trace of the matrix $A \cdot B$, expressed in this basis.

Proof: To express the scalar curvature S_t as a trace, we must choose an orthonormal basis $E_1^t, \ldots, E_n^t$ with respect to the metric g_t. Then

$$S_t = \sum_j \mathrm{Ric}^t(E_j^t, E_j^t) = \sum_{i,j} g_t(R^t(E_i^t, E_j^t)E_j^t, E_i^t).$$

We want to differentiate this with respect to t. To do this, we begin with two considerations.

1. $g_t(E_i^t, E_j^t) = \delta_{ij}$ implies

$$\underbrace{\frac{dg_t}{dt}}_{=h} \left(E_i^t, E_j^t\right) + g_t\left(\frac{dE_i^t}{dt}, E_j^t\right) + g_t\left(E_i^t, \frac{dE_j^t}{dt}\right) = 0;$$

thus, after summing over i and h at the point $t = 0$,

$$\sum_{i,j} h\left(E_i, E_j\right) = -2 \sum_{i,j} g\left(\frac{dE_i^t}{dt}\Big|_{t=0}, E_j\right).$$

[3] D. Hilbert, *Die Grundlagen der Physik*, Nachrichten der Gesellschaft der Wissenschaften Göttingen, Math.-Phys. Klasse, (1915) 395-407. This paper appeared almost simultaneously with the fundamental work of A. Einstein "Zur allgemeinen Relativitätstheorie" in the Proceedings of the Prussian Academy of Sciences.

2. With $\nabla_X E_j = \sum_i \omega^i_j(X) E_i$ (cf. the connection form in 4.33) it follows that for every symmetric tensor A
$$\sum_j A(\nabla_X E_j, E_j) = \sum_{i,j} \omega^i_j(X) A(E_i, E_j) = 0,$$
as $A(E_i, E_j)$ is symmetric in i and j and ω^i_j is skew-symmetric in i and j.

With this, we now calculate

$$\left.\frac{dS_t}{dt}\right|_{t=0} = \left.\frac{d}{dt}\right|_{t=0} \sum_j \mathrm{Ric}^t(E^t_j, E^t_j)$$
$$= \left.\frac{d}{dt}\right|_{t=0} \sum_{i,j} g_t(R^t(E^t_i, E^t_j)E^t_j, E^t_i)$$
$$= \sum_j \left[\sum_i g\left(\left.\frac{dR^t}{dt}\right|_{t=0} ((E_i, E_j)E_j), E_i \right) + 2 \cdot \mathrm{Ric}\left(\left.\frac{dE^t_j}{dt}\right|_{t=0}, E_j \right) \right]$$
$$\underbrace{+\sum_{i,j} \left[2 \cdot g\left(R(E_i, E_j)E_j, \left.\frac{dE^t_i}{dt}\right|_{t=0} \right) + h\big(R(E_i, E_j)E_j, E_i\big) \right]}_{=0 \text{ by consideration 1}}$$
$$\overset{8.4}{=} \sum_{i,j} \left[g\Big((\nabla_{E_i} \nabla'_h)(E_j, E_j), E_i \Big) - g\Big((\nabla_{E_j} \nabla'_h)(E_i, E_j), E_i \Big) \right]$$
$$+2 \sum_{j,k} \mathrm{Ric}(E_k, E_j) \cdot g\left(\left.\frac{dE^t_j}{dt}\right|_{t=0}, E_k \right)$$
$$\overset{\text{Consid. 1}}{=} \sum_{i,j} \left[g\Big(\nabla_{E_i}(\nabla'_h(E_j, E_j)), E_i \Big) - 2g\Big((\nabla'_h(\nabla_{E_i} E_j, E_j)), E_i \Big) \right.$$
$$\left. -g\Big((\nabla_{E_j} \nabla'_h)(E_i, E_j), E_i \Big) \right] - \sum_{j,k} \mathrm{Ric}(E_k, E_j) \cdot h(E_j, E_k)$$
$$\overset{\text{Consid. 2}}{=} \sum_j \mathrm{div}\big(\nabla'_h(E_j, E_j)\big) - \sum_{i,j} g\Big((\nabla_{E_j} \nabla'_h)(E_i, E_j), E_i \Big)$$
$$- \sum_{j,k} \mathrm{Ric}(E_k, E_j) \cdot h(E_j, E_k).$$

We easily recognize the first of the three terms as the divergence of the vector field $\nabla'_h(E_j, E_j)$, summed over j. The second term is also a divergence, in fact that of the vector field $(C\nabla'_h)^\#$ which is associated to the $(0,1)$-tensor $C\nabla'_h$ by means of the relation

$$g((C\nabla'_h)^\#, X) = (C\nabla'_h)(X)$$

for all tangent vectors X. Here $(C\nabla'_h)(X) = \sum_i g(\nabla'_h(E_i, X), E_i)$ is the contraction of ∇'_h (because of symmetry there is but a *single* contraction). This can be seen as follows:

$$\sum_{i,j} g\Big((\nabla_{E_j}\nabla'_h)(E_i, E_j), E_i\Big)$$

$$= \sum_j \Big[\nabla_{E_j} \underbrace{\sum_i \langle \nabla'_h(E_i, E_j), E_i\rangle}_{=C\nabla'_h(E_j)} - \underbrace{\sum_i \langle \nabla'_h(E_i, \nabla_{E_j}E_j), E_i\rangle}_{=C\nabla'_h(\nabla_{E_j}E_j)}\Big]$$

$$- \sum_{i,j}\langle \nabla'_h(E_i, E_j), \nabla_{E_j}E_i\rangle - \sum_{i,j}\langle \nabla'_h(\nabla_{E_j}E_i, E_j), E_i\rangle$$

$$= \operatorname{div}(C\nabla'_h)^\# - \sum_{i,j,k}\Big[\langle \nabla'_h(E_i, E_j), \omega^i_k(E_j)E_k\rangle + \langle \nabla'_h(\omega^i_k(E_j)E_k, E_j), E_i\rangle\Big]$$

$$= \operatorname{div}(C\nabla'_h)^\#$$

because $\omega^i_k + \omega^k_i = 0$ (just exchange i and k in the last summand).

By the Gauss-Stokes theorem (see 8.7), the integral over the first two summands thus vanishes, and we get

$$\int_M \frac{dS_t}{dt}\Big|_{t=0} dV = -\int_M \sum_{j,k} \operatorname{Ric}(E_j, E_k)h(E_j, E_k),$$

and, applying 8.4, this results in

$$\frac{d}{dt}\Big|_{t=0}\int_M S_t dV_t = \int_M \Big(S\frac{d}{dt}\Big|_{t=0}(dV_t) + \frac{dS_t}{dt}\Big|_{t=0}dV\Big)$$

$$= \int_M \Big(S\cdot\frac{1}{2}\operatorname{Tr}_g h - \sum_{j,k}\operatorname{Ric}(E_j, E_k)h(E_j, E_k)\Big)dV = \Big\langle \frac{S}{2}g - \operatorname{Ric}, h\Big\rangle_g .$$

The same statement is obtained if M is compact with boundary, provided that h vanishes in a neighborhood of the boundary. $\square$

For a proof of 8.6 using Ricci calculus, see either the original paper by Hilbert (loc. cit.) or [**25**], p. 192. Yet another proof is contained in [**24**], Proposition 4.7 in connection with Theorem 1.174.

The Gauss-Stokes theorem is essential for the proof of 8.6, so we present a version of this theorem here. This version is more or less the classical one, holding for the integral of a divergence of a vector field.

8.7. Gauss–Stokes Theorem. Let M be a compact oriented manifold with or without boundary, and let ν denote the unit normal vector field along the boundary ∂M (if this is not empty), endowed with the orientation induced by that of the boundary ∂M. This means that ν is perpendicular to ∂M, while being tangent to M. Let X be an arbitrary vector field on M with divergence $\mathrm{div}(X)$. Then

$$\int_M \mathrm{div}(X) dV_M = \int_{\partial M} \langle X, \nu \rangle dV_{\partial M}.$$

In particular, the left-hand side vanishes if either $\partial M = \emptyset$ or if X vanishes on ∂M.

For the classical situation of compact sets in $\mathbb{R}^n$, one can find this theorem in [**28**]; for a formulation for compact manifolds see [**29**]. In the notation of differential forms, this theorem is just the famous formula $\int_M d\omega = \int_{\partial M} \omega$, cf. 4.36 or [**27**], Ch. XXI.

8B The Einstein field equations

One of the consequences of Theorem 8.6 is a global version of the Gauss-Bonnet theorem, with a proof which is independent of that given in Section 4F (except for the use of Stokes' theorem).

8.8. Corollary. (Theorem of Gauss–Bonnet)
For $n = 2$ one always has the relation $\frac{S}{2}g - \mathrm{Ric} \equiv 0$, so that the functional $\mathbf{S}(g)$ is locally constant. Fixing the manifold M and varying the metric g, requiring only that it be positive definite, $\mathbf{S}(g)$ is even globally constant, as any two Riemannian metrics can be smoothly perturbed into one another: $\lambda g_1 + (1-\lambda)g_2$ is positive definite for $0 \leq \lambda \leq 1$, provided both g_1 and g_2 are.

The actual value of the constant $\mathbf{S}(g)$ is obtained most easily in examples, for example using convex surfaces with glued-on handles of strictly non-positive curvature, cf. the tight surfaces in Section 4G. For an orientable surface of genus g_0 one gets the value $\mathbf{S}(g) = 2(2-2g_0)\cdot 2\pi = 4\pi\chi(M)$ (Euler-Poincaré characteristic of M), which means, taking $S = 2K$ into account, twice the total curvature occurring in the Gauss-Bonnet formula 4.43.

Remark: The Gauss-Bonnet formula also holds in the case of indefinite metrics on two-manifolds[4]. In fact, the argument of 8.8 shows that the functional $\mathbf{S}(g)$ is locally constant. But this does not yet allow a conclusion on the global value, as $\lambda g_1 + (1-\lambda)g_2$ can degenerate when g_1 and g_2 are indefinite.

The situation on n-dimensional manifolds with $n \geq 3$ is quite different, as the functional $\mathbf{S}$ in this case is almost never constant. Instead, if it is stationary one gets non-trivial Euler-Lagrange equations.

8.9. Corollary. (Euler–Lagrange equations for the functional $\mathbf{S}$)
Let M be a compact manifold of dimension $n \geq 3$, and let g be a fixed metric on M with a variation $g_t = g + th$, in which the symmetric

[4] G.S.Birman & K.Nomizu, *The Gauss-Bonnet theorem for two-dimensional spacetimes.*, Michigan Math. J. **31**, 77–81 (1984)

$(0,2)$-tensor h is arbitrary. Then we have:

(i) $\left.\dfrac{d}{dt}\right|_{t=0}\mathbf{S}(g_t) = 0$ for all h if and only if $\mathrm{Ric}_g \equiv 0$.

(ii) $\left.\dfrac{d}{dt}\right|_{t=0}\mathbf{S}(g_t)=0$ for all h with the constraint $\mathbf{Vol}(M)=\int_M dV_t=$ constant holds if and only if (M,g) is an Einstein space, that is, when $\mathrm{Ric}_g - \frac{S_g}{n} g \equiv 0$.

(iii) $\left.\dfrac{d}{dt}\right|_{t=0}\dfrac{\mathbf{S}(g_t)}{\big(\mathbf{Vol}(g_t)\big)^{(n-2)/n}} = 0$ for all h if and only if (M,g) is an Einstein space.

Proof: Ad (i): Here, according to 8.6, we must investigate the condition

$$\left\langle \frac{S}{2}g - \mathrm{Ric}, h\right\rangle_g = 0 \text{ for all } h.$$

Because of the non-degeneracy of the inner product $\langle\ ,\ \rangle_g$ this is equivalent to the vanishing of the Einstein tensor (cf. 6.15)

$$G = \mathrm{Ric} - \frac{S}{2}g.$$

From the vanishing of this quantity, one gets, upon taking traces,

$$S - \frac{n}{2}S = 0,$$

and from this $S = 0$ because $n \geq 3$. As a consequence,

$$0 = G = \mathrm{Ric}.$$

Conversely, the relation $\mathrm{Ric} = 0$ naturally implies $S = 0$ and $G = 0$.

Ad (ii): Using the rule of Lagrange multipliers, we have to investigate the linear dependency of the two gradients $\mathbf{S}$ and $\mathbf{Vol}$. Clearly

$$\left.\frac{d}{dt}\right|_{t=0}\mathbf{Vol}(g_t) \overset{8.4}{=} \frac{1}{2}\langle g, h\rangle_g,$$

so we get $\left\langle \mathrm{Ric} - \frac{S}{2}g, h\right\rangle_g = 0$ for all h with $\langle g, h\rangle_g = 0$ holds if and only if $\mathrm{Ric} - \frac{S}{2}g$ and g are linearly dependent as tensor fields. This is the case precisely when $\mathrm{Ric} = \lambda \cdot g$ for some function λ, hence if and only if g is an Einstein metric.

Ad (iii): This follows from the quotient rule for the variation

$$\delta\Big(\frac{\mathbf{S}}{\mathbf{Vol}^{(n-2)/n}}\Big) = \frac{\mathbf{Vol}^{(n-2)/n}\delta(\mathbf{S}) - \frac{n-2}{n}\mathbf{S}\delta(\mathbf{Vol})\mathbf{Vol}^{-2/n}}{\mathbf{Vol}^{2(n-2)/n}}.$$

If the numerator vanishes, then, because $\delta\mathbf{S} = \big\langle \frac{S}{2}g - \mathrm{Ric}, h\big\rangle_g$ and $\delta\mathbf{Vol} = \frac{1}{2}\big\langle g, h\big\rangle_g$, the Ricci tensor is a scalar multiple of g. Conversely, if g is an Einstein metric with $n \geq 3$, we can multiply g by a scalar α such that the volume of αg equals unity. Since $S_{\alpha g} = \alpha^{-2}S_g$ and $dV_{\alpha g} = \alpha^n dV_g$, the functional $\mathbf{S}/\mathbf{Vol}^{(n-2)/n}$ is invariant under scalings of this type. Hence, by (ii) every Einstein metric leads to a vanishing variation. □

8.10. Remark. (The case of an indefinite metric g)

All considerations of Section 8A as well as 8.9 remain valid also in the more general case of an indefinite metric g. Using Ricci calculus, nothing at all is changed, since one has $\mathrm{Tr}_g A = A^i_i = A_{ij}g^{ji}$, cf. 6.9. Using an ON-basis $E_1, \ldots, E_n$, one only has to take into account that the components of a vector X are given by the equations

$$X = \sum_i \epsilon_i \langle X, E_i\rangle E_i,$$

where $g(E_i, E_j) = \delta_{ij}\cdot\epsilon_i$ with a sign $\epsilon_i \in \{+1, -1\}$. This implies that the trace of a $(0,2)$-tensor has to be replaced by the expression

$$\mathrm{Tr}_g A := \sum_i \epsilon_i g(AE_i, E_i).$$

In particular one then has $\mathrm{Tr}_g\, g \;= \sum_i \epsilon_i \underbrace{g(E_i, E_i)}_{=\varepsilon_i} = n.$

One also must take appropriate changes in the trace of the matrix $A \cdot B$ introduced in 8.6 into account. Here we must introduce a sign $\epsilon_{ij} = \epsilon_i\epsilon_j$:

$$\sum_{i,j} \epsilon_{ij} A(E_i, E_j) B(E_j, E_i).$$

Moreover, one must note that the volume element is given in local coordinates by

$$dV = \sqrt{\big|\mathrm{Det}(g_{ij})\big|}dx_1 \wedge \cdots \wedge dx_n,$$

cf. [**22**], p. 195.

The functional $\mathbf{S}$ can also be considered for non-compact manifolds M, which is of importance in the general theory of relativity. In this case one must make the assumption that $\mathbf{S}(g)$ exists as an improper integral. For the variational problem in 8.6 it is convenient to also assume that $h \equiv 0$ and hence $g_t \equiv g$ outside some compact set. Then the boundary terms vanish in the application of Stokes theorem just as in the compact case.

8.11. Einstein field equations. From the considerations in Sections 8.6 and 8.9 we have a profound mathematical reason for considering the tensor

$$\mathrm{Ric} - \frac{S}{2} g$$

as the "gradient" of the functional $\mathbf{S}$. Beyond that, this tensor plays an important role in the general theory of relativity. There, this tensor is referred to as the *Einstein gravitational tensor*. We already showed in 6.15 that this tensor is divergence-free. Its physical importance comes from the Einstein field equations for four-dimensional spacetimes, i.e., four-dimensional Lorentz manifolds with a metric of type $(- + + +)$. How this comes about is best left to the originator of the idea, A. Einstein, as he described it in a lecture at Princeton in 1921:

If there is an analog of the Poisson equation in general relativity, then this must be a tensor equation for the gravitational potential tensor $g_{\mu\nu}$, on whose right-hand side we have the energy tensor of matter. On the left-hand side of the equation we need a differential tensor derived from $g_{\mu\nu}$. The goal is to determine this tensor precisely. It is completely determined by the following three conditions:

1. *The tensor in question should contain no higher than second derivatives of $g_{\mu\nu}$.*
2. *The tensor should depend linearly on the second derivatives.*
3. *The divergence of the tensor should vanish identically.*

(Translated from: A. Einstein, *Grundzüge der Relativitätstheorie*, Vieweg, 6. Aufl. 1990, p. 83)

The Einstein tensor satisfies all three conditions and is in a sense uniquely determined by this property. The Einstein field equations are then the following:

$$\mathrm{Ric} - \frac{S}{2} g = T,$$

or, written using Ricci calculus, $G_{ij} := R_{ij} - \frac{S}{2}g_{ij} = T_{ij}$, where the right-hand side is the *stress-energy tensor*, which must have vanishing divergence for physical reasons. Often the equation is written with an additional physical constant in front of T, whose size is not of interest to us here. The tensor g_{ij} is the *gravitational potential* of matter. In particular, one way of reading these equations is to view g_{ij} as a variable (not given), but with T_{ij} as given. In a vacuum there is no matter, meaning $T = 0$ and hence

$$\mathrm{Ric} - \frac{S}{2}g = 0.$$

Spaces of this kind are also referred to as *special Einstein spaces.* According to 8.9 they are necessarily Ricci flat: $\mathrm{Ric} = 0$.

But Einstein himself[5] as well as other also considered a variant of this field equation, by introducing a so-called *cosmological term* Λg_{ij} with a so-called *cosmological constant* Λ:

$$R_{ij} - \frac{S}{2}g_{ij} + \Lambda g_{ij} = T_{ij}.$$

A consequence of this is that the equation for the vacuum is satisfied if the metric g is an Einstein metric. This is again seen by taking the trace of of the left-hand side. This equation clearly implies the relation $R_{ij} = (\frac{S}{2} - \Lambda)g_{ij}$. On the other hand, the trace of the left-hand side is $S - 2S + 4\Lambda$. If we have $R_{ij} = \lambda g_{ij}$ with some function λ and $T = 0$, then it follows that $\lambda = \frac{S}{4} = \Lambda$. The cosmological constant is therefore coupled to the value of S.

8C Homogenous Einstein spaces

Besides the spaces of constant curvature, homogenous spaces are a very important class of spaces. These spaces are characterized by the property that a neighborhood of every point looks the same, i.e., is isometric to any other. The term homogenous is thus keyed to the fact that with respect to intrinsically defined geometric quantities there is but a single type of point. For this reason, it is sufficient to consider a single point, which leads to exceptionally clear statements and results.

[5] A. Einstein, *Über die formale Beziehung des Riemannschen Krümmungstensors zu den Feldgleichungen der Gravitation*, Math. Annalen **97**, 99–103 (1927).

In particular, one can relatively easily give a sufficient criterion for such a space to be Einstein. This criterion will be introduced in this section. It expresses, in addition to the homogeneity on points, a homogeneity of unit tangent vectors (cf. also 7.6). This leads to many interesting examples in a relatively clear and simple manner, see 8.16. The most elegant formulation uses isometry groups and subgroups of these. One requires only the additional fact that the group of all isometries on a Riemannian manifold is also a manifold.

8.12. Theorem. (Isometry group)

For an arbitrary Riemannian manifold (M, g), the set of all isometries $f\colon M \to M$ is again a differentiable manifold (a Lie group), whose dimension is at most $\binom{n+1}{2}$, where n denotes the dimension of M. If M is compact, then so too is the isometry group.

For a proof see [**46**], Chapter II, Theorem 1.2. For basic results on Lie groups in general see [**7**], Vol. I, Chapter 10, or [**42**].

8.13. Definition and Lemma. (Homogenous manifolds)
(M, g) is called *homogenous*, if for any two points $x, y \in M$ there is an isometry $f\colon M \to M$ with $f(x) = y$. (M, g) is called G-*homogenous* if, in addition, f can always be taken as $f \in G$, where G is a closed subgroup of the isometry group. If M is G-homogenous, then for $x \in M$ the subgroup $K_x := \{f \in G \mid f(x) = x\}$ is called the *isotropy group* of the point x. One has $K_x \simeq K_y$ for $x, y \in M$, and M is diffeomorphic to the space of cosets G/K_x. For any closed subgroup $K \subset G$ the quotient $M := G/K$ is often called a G-*homogeneous space* if the metric is invariant under the G-action. This avoids the use of Theorem 8.12.

The bijection between M and G/K_x is simply given as follows: every point $y \in M$ is identified with the K_x-coset of an isometry in G which maps x to y, compare [**42**, Ch.4]. Thanks to the homogeneity, the differentiability of this only has to be checked at a single point. But the differentiability at x is not difficult to check.

Standard examples are of course the spaces E^n, S^n, H^n of constant curvature with their isometry groups $\mathbf{E}(n), \mathbf{O}(n+1), \mathbf{O}_+(n, 1)$. These are also essentially all of those where the upper bound $\binom{n+1}{2}$ for the dimension is actually attained. The isometry group of the flat square torus (cf. 7.24) is only 2-dimensional: it is generated by all translations (modulo integers) as well as a rotation by $\pi/2$ and a reflection $(x, y) \mapsto (-x, y)$.

Both 8.12 and 8.13 are not really necessary for the proof of Theorem 8.15 below, at least as long as one views a homogenous space as a quotient G/K with a given group G which is not just a group, but a differentiable manifold (a Lie group). This is why we only sketch 8.12 and 8.13 above.

Example: The standard sphere S^n is G-homogenous, if we choose G as the special orthogonal group $\mathbf{SO}(n+1)$. In this case the isotropy group of a point is isomorphic to the standard subgroup $\mathbf{SO}(n)$, which consists of those rotations which fix a line (namely the line joining x with the origin). Then S^n is diffeomorphic to the quotient space $\mathbf{SO}(n+1)/\mathbf{SO}(n)$. However, nothing prevents our passing to a subgroup of $\mathbf{SO}(n+1)$, in which case of course also the isotropy group will be correspondingly smaller.

8.14. Definition. (Isotropy irreducible)
A *(faithful) representation* of a group G in a vector space V is an injective group homomorphism

$$G \to \mathrm{Aut}(V),$$

where $\mathrm{Aut}(V)$ denotes the group of linear automorphisms of V. A representation is said to be *irreducible*, if any *invariant subspace* $U \subseteq V$ (this means that for all $f \in G$, the set is mapped under f into itself, i.e., $f \in G, u \in U \Rightarrow f(u) \in U$) is either the trivial subspace $U = \{0\}$ or the entire space $U = V$. A homogenous manifold $M = G/K$ is called *isotropy irreducible*, if the corresponding isotropy representation $\chi\colon K \longrightarrow GL(n, \mathbb{R})$ of K, given by

$$K = K_x \ni f \stackrel{\chi}{\longmapsto} Df|_x \colon T_xM \to T_xM,$$

is irreducible for one (and hence for every) $x \in M$. In this case χ is automatically injective since an isometry f with a fixed point x is uniquely determined by $Df|_x$.

In the case of the square flat torus (cf. 7.24), the isotropy group is a finite group of order 8, which is nothing but the dihedral group D_4

which maps the square into itself. The flat standard torus $S^1 \times S^1 = I\!R^2/\mathbb{Z}^2$ is of course also G-homogenous if one takes G to be the group of pure translations modulo $\mathbb{Z}^2$, but it is not irreducible. Clearly the translations are isometric mappings of the torus to itself. Then the isotropy group is trivial, and consequently the x-axis as well as the y-axis are invariant subspaces of the tangent plane.

8.15. Theorem. (J. Wolf 1968[6])
Let $M = G/K$ be a G-homogenous Riemannian manifold, and let us suppose that it is isotropy irreducible. Then M is an Einstein space.

PROOF: We use the isotropy representation described above, $\chi\colon K \to GL(n, I\!R)$, which is irreducible by assumption. Moreover, the metric g is invariant under G, and hence we have, at every point $x \in M$,

$$g_{f(x)}(Df(X), Df(Y)) = g_x(X, Y)$$

for every $f \in G$ and every $X, Y \in T_xM$. If the metric is preserved, then so is the Ricci tensor; hence

$$\mathrm{Ric}_{f(x)}(Df(X), Df(Y)) = \mathrm{Ric}_x(X, Y)$$

for every f and every X, Y. Thus Ric_x is a K_x-invariant symmetric bilinear form on T_xM, so the eigenspaces of Ric_x with respect to g_x are invariant subspaces. Because of the assumption of irreducibility, this can only hold if each eigenspace coincides with the entire space, which means that at every point x the equation $\mathrm{Ric}_x = \lambda(x) \cdot g_x$ holds for some number $\lambda(x)$. But this is just the condition on the metric g for it to be Einstein. Of course $\lambda(x)$ must be constant in x because of the homogeneity. Note that for $n \geq 3$ this follows independently from 6.13. $\square$

8.16. Example. (Projective spaces)

For many standard spaces it is in fact easy to see that the assumptions of 8.14 are satisfied. For example, this is the case for real, complex, and quaternionic projective spaces. To see this, it is sufficient to check that the isometry group maps any point to any other, and in addition,

[6] *The geometry and structure of isotropy irreducible homogeneous spaces*, Acta Math. **120**, 59–148 (1968).

any direction in the tangent space to any other, i.e., an arbitrary unit tangent vector can be mapped to any other such vector, cf. also 7.6. This is for example an obvious state of affairs for the orthogonal group $\mathbf{O}(n)$, provided it acts in the standard fashion on $\mathbb{R}^n$. Besides this, one needs to note that on any Lie group G there exists a G-invariant metric, which can be constructed by starting with a fixed but arbitrary inner product on the tangent space at one point and transporting this inner product to any other tangent space by means of the left translation $x \mapsto g \cdot x$, where g runs through the entire group G. On a compact Lie group there is also a bi-invariant metric for which left translations and right translations are isometries. For the classical groups given in terms of matrices, one can alternatively take the first fundamental form given by the standard embedding in Euclidean space, identifying for example $\mathbf{SO}(3)$ as a subspace of $\mathbb{R}^9$.

1. The *sphere*
$$S^n = \mathbf{SO}(n+1)/\mathbf{SO}(n)$$
is such an example, as $\mathbf{SO}(n)$ acts irreducibly in all directions. Hyperbolic space, too, belongs to this series of spaces, viewing it as the quotient $H^n = \mathbf{O}_+(n,1)/\mathbf{O}(n)$, cf. 7.6.
2. The *real projective space*
$$\mathbb{R}P^n = \mathbf{O}(n+1)/\mathbf{O}(n) \times \mathbf{O}(1) = S^n/\pm,$$
which can also be viewed as the set of all lines in $\mathbb{R}^{n+1}$ passing through the origin, is another such example.
3. The *complex projective space*
$$\mathbb{C}P^n = \mathbf{U}(\mathbf{n}+\mathbf{1})/\mathbf{U}(n) \times \mathbf{U}(1),$$
which can again be viewed as the set of lines, this time complex lines in $\mathbb{C}^{n+1}$, through the origin, is again an example. Here, the *unitary group* is defined as
$$\mathbf{U}(n) := \{A\colon \mathbb{C}^n \to \mathbb{C}^n \mid A \cdot \overline{A}^T = E\}.$$
Again, the action of the isotropy group $\mathbf{U}(n)$ is transitive on the (real) directional vectors, since it is transitive on the complex unit vectors and since every real unit vector is contained in a complex line. From this it follows that the Ricci curva-

ture must be the same in every direction, cf. 8.14. Note that the manifold $\mathbb{C}P^n$ cannot possibly carry a metric of constant curvature, since it is compact and simply connected, see 7.23. In fact, the isotropy group is not transitive on the set of two-dimensional planes (on which the sectional curvature depends), since a two-dimensional real plane, which corresponds to a complex line (meaning it is invariant under multiplication by i), can never be rotated by a complex matrix to a two-dimensional real plane which doesn't correspond to a complex line. The sectional curvature of the standard metric $\mathbb{C}P^n$ varies from 1 to 4, assuming it is normed appropriately.

4. A further example is given by *quaternionic projective space*

$$\mathbb{H}P^n = \mathbf{Sp}(n+1)\big/\mathbf{Sp}(n) \times \mathbf{Sp}(1),$$

which can be viewed as the set of quaternionic lines in $\mathbb{H}^{n+1}$ through the origin. Here we use the notation

$$\mathbf{Sp}(n) := \left\{ \begin{pmatrix} A & -\overline{B} \\ B & \overline{A} \end{pmatrix} \in \mathbf{U}(2n) \right\}$$

for the group of quaternionic matrices. This uses the fact that a quaternion can also be viewed as a complex number over the complex numbers: $a + ib + jc + dk = (a + ib) + j(c - id)$.

5. Other examples are the higher *Grassmann manifolds* (or *Grassmannians*) of k–planes through the origin in $\mathbb{R}^n, \mathbb{C}^n$ and $\mathbb{H}^n$, respectively, as well as the *Cayley plane*, which is a homogenous space given as the quotient of two exceptional groups.

6. There is also a classification of all compact and simply connected homogenous Einstein spaces[7], but this is quite a bit more complicated. Here exceptional groups occur as well as exceptional cases, for example an Einstein metric on the 15-sphere which is not isometric to the standard metric.

---

[7] M. Wang & W. Ziller, *On normal homogeneous Einstein manifolds*, Annales Scientifiques de l'École Normale Supérieure **18**, 563–633 (1985).

8D The decomposition of the curvature tensor

In this section we discuss a different motivation for considering Einstein metrics, by considering the set of all possible curvature tensors. This latter space is a complicated vector space, and one can attempt to decompose it into simpler parts. One of these parts will be the traceless part of the Ricci tensor, given by $\mathrm{Ric} - \frac{S}{n}g$, which vanishes precisely for Einstein metrics. A part of a different component of the curvature is formed by the trace of the Ricci tensor. A decomposition of a vector space is given by a direct sum decomposition into subspaces, which is preferably orthogonal with respect to an appropriate inner product. Our first purpose will be to define this.

To give the reader a feeling for the principle to be used, we first consider the "trivial" case of symmetric (2×2)-matrices. These can be decomposed into their traceless parts and the part having a trace as follows:

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(a+c) & 0 \\ 0 & \frac{1}{2}(a+c) \end{pmatrix} + \begin{pmatrix} \frac{1}{2}(a-c) & b \\ b & \frac{1}{2}(c-a) \end{pmatrix}.$$

This decomposition is orthogonal with respect to the "inner product" $\langle A, B\rangle := \mathrm{Tr}(A \cdot B)$, which is defined on the space of all square matrices. If A is a scalar multiple of the unit matrix, then $\langle A, B\rangle = 0$ for every matrix B with $\mathrm{Tr}(B) = 0$. This can be done in the same way for symmetric $(0, 2)$-tensors, for example

$$\mathrm{Ric} = \frac{S}{n}g + \underbrace{\Big(\mathrm{Ric} - \frac{S}{n}g\Big)}_{\mathrm{Tr}=0}.$$

If one were to consider only $(0, 2)$-tensors, there would be no problem. But the curvature tensor is of type $(0, 4)$, which can also be described as a $(1, 3)$-tensor or possibly also as a $(2, 2)$-tensor, as is appropriate in the context at hand. In this case the linear algebra is more complicated, as we already saw in the investigation of the biquadratic form in Section 6.5. Thus, we require an algebraic tool known as *bivectors*. This is a special case of the more general concept of the exterior algebra over a vector space, see [**33**], Chapter 5. It is quite similar to the concept of differential forms which we used in Section 4F.

8.17. Definition. (Bivectors)

Suppose we are given a real vector space V with a basis $b_1, \ldots, b_n$. The *vectors* of this space can be expressed as linear combinations

$$X = \sum_i \alpha_i b_i, \quad \alpha_i \in \mathbb{R}.$$

Similarly, the *bivectors* can be expressed (formally) as linear combinations

$$\sum_{i<j} \alpha_{ij} b_i \wedge b_j, \quad \alpha_{ij} \in \mathbb{R},$$

in which we take the point of view that the elements

$$b_1 \wedge b_2,\ b_1 \wedge b_3,\ \ldots,\ b_1 \wedge b_n,\ b_2 \wedge b_3,\ b_2 \wedge b_4,\ \ldots,\ b_{n-1} \wedge b_n$$

form a basis. This is entirely similar to the tensor product $V \otimes V$, whose basis is

$$b_i \otimes b_j, \quad i, j = 1, \ldots, n.$$

For bivectors we agree that the relation $b_i \wedge b_j = -b_j \wedge b_i$ should hold identically, just as it does for alternating two-forms as in Section 4F. Think of a bivector $b_i \wedge b_j$ as a surface (area) element in the plane spanned by b_i, b_j. Formally, however, a bivector is dual to such an element. We then define $\bigwedge^2 V$ as the *space of all bivectors* over V, whose dimension is

$$\dim(\bigwedge^2 V) = \binom{n}{2} = \frac{n(n-1)}{2}.$$

For two vectors $X, Y \in V$ we can then define an exterior product $X \wedge Y \in \bigwedge^2 V$ by

$$X \wedge Y = \Big(\sum_i \alpha_i b_i\Big) \wedge \Big(\sum_j \beta_j b_j\Big) = \sum_{i,j} \alpha_i \beta_j (b_i \wedge b_j)$$

$$= \sum_{i<j} (\alpha_i \beta_j - \alpha_j \beta_i) b_i \wedge b_j.$$

This is formally quite similar to the vector product in $\mathbb{R}^3$. Two vectors X, Y are linearly independent if and only if $X \wedge Y = 0$. The algebraic properties of the space of bivectors are quite similar to those of the space of all alternating two-forms, cf. Section 4F. The following duality holds:

$$
\begin{aligned}
(\bigwedge^2(V))^* &= \{\omega : \bigwedge^2(V) \to I\!R \mid \omega \text{ is linear}\} \\
&= \{\omega : V \otimes V \to I\!R \mid \omega \text{ is linear and skew-symmetric}\} \\
&= \{\omega : V \times V \to I\!R \mid \omega \text{ is bilinear and skew-symmetric}\}.
\end{aligned}
$$

For more details, see [**33**], in particular Chapter 5. The reason we have made this diversion at this point in the presentation is that the curvature tensor $\big\langle R(X,Y)V, Z\big\rangle$ is skew-symmetric in both X, Y and Z, V. For this reason, one can view the curvature operator $R(X,Y)$ with fixed vectors X and Y as a linear mapping

$$R(X,Y)\colon \bigwedge^2(T_pM) \longrightarrow I\!R.$$

The other two arguments X, Y can also be viewed as a bivector. The goal of these considerations is to think of R as a symmetric (i.e., self-adjoint) endomorphism of $\bigwedge^2(T_pM) = \bigwedge^2_p$.

8.18. Lemma. Let $\bigwedge^2_p$ denote the space of all bivectors over T_pM, and let $\langle\ ,\ \rangle$ denote a Riemannian metric (which can also be indefinite).

1. Then an inner product on $\bigwedge^2_p$ is defined by
$$
\begin{aligned}
\langle\!\langle X \wedge Y, Z \wedge V \rangle\!\rangle &:= \langle R_1(X,Y)V, Z\rangle \\
&= \langle X, Z\rangle\langle Y, V\rangle - \langle Y, Z\rangle\langle X, V\rangle.
\end{aligned}
$$
2. An ON-basis $E_1, \ldots, E_n$ in T_pM induces an ON-basis $E_i \wedge E_j, i < j$, in $\bigwedge^2_p$.
3. On the space of all symmetric (self-adjoint) endomorphisms with respect to $\bigwedge^2_p$ there is an inner product defined by
$$\langle\!\langle\!\langle A, B\rangle\!\rangle\!\rangle := \mathrm{Tr}(A \circ B).$$

Proof: 1. The bilinearity and the symmetry of $\langle\!\langle\ ,\ \rangle\!\rangle$ are trivially satisfied. The non-degeneracy follows from the non-degeneracy of the biquadratic form $k_1(X,Y) := \langle R_1(X,Y)Y, X\rangle = \langle X, X\rangle\langle Y, Y\rangle - \langle Y, X\rangle\langle X, Y\rangle$. For a given $X \neq 0$, the null space, consisting of all

Y such that $k_1(X,Y) = 0$, is trivial. This also holds if X is a null vector (isotropic vector), since in that case there is at least one Y with $\langle X, Y\rangle = 1$. In the case of a positive definite metric $\langle\ ,\ \rangle$, $\langle\!\langle\ ,\ \rangle\!\rangle$ is also positive definite because of the relation $k_1(X,Y) > 0$, which holds for any two linearly independent X, Y, and hence for every $X \wedge Y \neq 0$.

2. This follows directly from 1, because $\langle R_1(E_i, E_j)E_j, E_i\rangle = 1$ for $i < j$ and $\langle R_1(E_i, E_j)E_k, E_l\rangle = 0$ if there are 3 or 4 distinct indices among i, j, k, l.

3. The trace of an endomorphism A with respect to an ON-basis $E_i \wedge E_j, i < j$, is, according to 6.9,

$$\mathrm{Tr}A := \sum_{i<j} \langle\!\langle A(E_i \wedge E_j), E_i \wedge E_j\rangle\!\rangle.$$

It follows from this that $\langle\!\langle\!\langle\ ,\ \rangle\!\rangle\!\rangle$ is bilinear. The symmetry property $\mathrm{Tr}(A \circ B) = \mathrm{Tr}(B \circ A)$ holds quite generally for endomorphisms and matrices. The same is true of the non-degeneracy. The positive definiteness of the metric $\langle\ ,\ \rangle$ implies that of $\langle\!\langle\!\langle\ ,\ \rangle\!\rangle\!\rangle$ because of the relation

$$\begin{aligned}\langle\!\langle\!\langle A, A\rangle\!\rangle\!\rangle = \mathrm{Tr}(A^2) &= \sum_{i<j} \langle\!\langle A^2(E_i \wedge E_j), E_i \wedge E_j\rangle\!\rangle \\ &= \sum_{i<j} \langle\!\langle A(E_i \wedge E_j), A(E_i \wedge E_j)\rangle\!\rangle > 0.\end{aligned}$$

□

8.19. Definition. The Riemannian curvature tensor

$$R(X,Y,Z,V) := \big\langle R(X,Y)V, Z\big\rangle$$

can be interpreted at every point p as a symmetric endomorphism

$$\widehat{R} : \textstyle\bigwedge_p^2 \longrightarrow \bigwedge_p^2$$

by virtue of the equation

$$\langle\!\langle \widehat{R}(X \wedge Y), Z \wedge V\rangle\!\rangle := R(X,Y,Z,V).$$

Note that the skew-symmetries of the curvature tensor in 6.3 are already part of the definition of $\widehat{R}$ in $\bigwedge_p^2$. The symmetry in 6.3.5 is nothing but the self-adjoint property of $\widehat{R}$. For a variable point p we omit the subscript and write simply $\bigwedge^2$.

The (purely algebraic) first Bianchi identity 6.3.2 must be required in addition to the above, if one wants to determine the space of all possible candidates for curvature tensors. This motivates the following definition.

Let $\mathcal{R}$ (resp. $\widehat{\mathcal{R}}$) be the set of all $(0,4)$-tensors (resp. endomorphisms of $\bigwedge^2$) which satisfy all algebraic symmetries of the curvature tensor, including the first Bianchi identity. Then we have the following correspondences:

$$\begin{array}{rcl} \mathcal{R} & \longleftrightarrow & \widehat{\mathcal{R}}, \\ R & \longleftrightarrow & \widehat{R}, \\ R_1 & \longleftrightarrow & \widehat{R_1} = \mathrm{Id}. \end{array}$$

On the left-hand side of these relations, we have the *Riemannian curvature tensor* in the sense of Definition 8.19, with the ordering of the arguments as described there. This explains the equation $\widehat{R_1} = \mathrm{Id}$.

8.20. Definition and Lemma. (Products of $(0,2)$ tensors)
Let A, B be symmetric (0,2) tensors. We define a product $A \bullet B$ by

$$\begin{aligned} (A \bullet B)(X,Y,Z,T) \quad := \quad & A(X,Z)B(Y,T) + A(Y,T)B(X,Z) \\ & -A(X,T)B(Y,Z) - A(Y,Z)B(X,T). \end{aligned}$$

For this product we have the following properties: $A \bullet B \in \mathcal{R}$, the symmetry $A \bullet B = B \bullet A$ holds, and we have the product rule $\nabla_X(A \bullet B) = (\nabla_X A) \bullet B + A \bullet (\nabla_X B)$.

PROOF: The symmetry of $A \bullet B$ is clear by definition. The product rule is easy to verify, simply by writing down the derivatives of all terms. The first Bianchi identity can be directly verified by the following calculation:

$$\begin{aligned}
A \bullet B(X,Y,Z,T) &+ A \bullet B(Y,Z,X,T) + A \bullet B(Z,X,Y,T) \\
&= A(X,Z)B(Y,T) + A(Y,T)B(X,Z) - A(X,T)B(Y,Z) \\
&\quad - A(Y,Z)B(X,T) + A(Y,X)B(Z,T) + A(Z,T)B(Y,X) \\
&\quad - A(Y,T)B(Z,X) - A(Z,X)B(Y,T) + A(Z,Y)B(X,T) \\
&\quad + A(X,T)B(Z,Y) - A(Z,T)B(X,Y) - A(X,Y)B(Z,T) \\
&= 0.
\end{aligned}$$

In particular we have

$$\begin{aligned}
g \bullet g(X,Y,Z,T) &= 2\langle X,Z\rangle\langle Y,T\rangle - 2\langle X,T\rangle\langle Y,Z\rangle \\
&= 2\big\langle R_1(X,Y)T, Z\big\rangle = 2R_1(X,Y,Z,T),
\end{aligned}$$

from which it follows that $\widehat{g \bullet g} = 2\widehat{R_1} = 2 \cdot \mathrm{Id}$. The equation $\nabla_X R_1 = 0$, which we met in Section 6B, follows here from the product rule, if we take the relation $\nabla_X g = 0$ into account. $\square$

In more recent literature, this product is sometimes referred to as the *Kulkarni-Nomizu product*, for example in [**24**], Definition 1.110. In Ricci calculus this product was traditionally referred to as the "double transvection", and written as follows:

$$(A \bullet B)_{ikjl} = 4A_{[i[j}B_{k]l]} := A_{ij}B_{kl} + A_{kl}B_{ij} - A_{il}B_{jk} - A_{jk}B_{il},$$

cf. [**15**], Kapitel I, §8 (unfortunately this is not contained in the English version [**16**]).

8.21. Theorem. In $\widehat{\mathcal{R}}$ there is a decomposition into three subspaces $\widehat{\mathcal{R}} = \widehat{\mathcal{U}} \oplus \widehat{\mathcal{Z}} \oplus \widehat{\mathcal{W}}$, which is orthogonal with respect to $\langle\!\langle\!\langle\ ,\ \rangle\!\rangle\!\rangle$, in which $\widehat{\mathcal{U}}$ is generated by the identity and $\widehat{\mathcal{Z}}$ is generated by all $\widehat{A \bullet g}$ with symmetric A, $\mathrm{Tr}_g A = 0$.

Alternatively, there is a decomposition $\mathcal{R} = \mathcal{U} \oplus \mathcal{Z} \oplus \mathcal{W}$, where $\mathcal{U}$ is generated by R_1 (or $g \bullet g$) and $\mathcal{Z}$ is generated by all $A \bullet g$, where A is symmetric and $\mathrm{Tr}_g A = 0$. In particular, (M,g) has constant curvature if and only if the $\mathcal{Z}$-part and the $\mathcal{W}$-part both vanish.

Proof: It is only necessary to show that $\widehat{\mathcal{U}}$ is orthogonal to $\widehat{\mathcal{Z}}$, since we can then define $\widehat{\mathcal{W}}$ as the orthogonal complement. Let $E_1, \ldots, E_n$

be an ON-basis in T_pM; then $E_i \wedge E_j, i < j$, is an ON-basis in $\bigwedge^2$. Then we have

$$\langle\!\langle\!\langle \mathrm{Id}, \widehat{A \bullet g} \rangle\!\rangle\!\rangle = \mathrm{Tr}(\widehat{A \bullet g})$$

$$= \sum_{i<j} \langle\!\langle \widehat{A \bullet g}(E_i \wedge E_j), E_i \wedge E_j \rangle\!\rangle = \sum_{i<j} A \bullet g(E_i, E_j, E_i, E_j)$$

$$= \sum_{i<j} \Big[A(E_i, E_i) \cdot 1 + A(E_j, E_j) \cdot 1 - A(E_i, E_j)\delta_{ij} - A(E_j, E_i)\delta_{ij} \Big]$$

$$= (n-1)\mathrm{Tr}A = 0.$$

Hence $\widehat{\mathcal{U}}$ and $\widehat{\mathcal{Z}}$ are orthogonal to one another. □

Our next goal is to calculate the parts of $R = U + Z + W$ or $\widehat{R} = \widehat{U} + \widehat{Z} + \widehat{W}$ in the orthogonal decomposition $\mathcal{U} \oplus \mathcal{Z} \oplus \mathcal{W}$. This is essentially now just a problem of choosing a correct normalization, as the identity Id is not a unit vector in $\widehat{\mathcal{R}}$. Since $\binom{n}{2}$ is the dimension of $\bigwedge^2$, we have

$$\langle\!\langle\!\langle \mathrm{Id}, \mathrm{Id} \rangle\!\rangle\!\rangle = \mathrm{Tr}(\mathrm{Id}) = \binom{n}{2}.$$

Moreover,

$$\langle\!\langle\!\langle \widehat{R}, \mathrm{Id} \rangle\!\rangle\!\rangle = \mathrm{Tr}\widehat{R} = \sum_{i<j} \langle\!\langle \widehat{R}(E_i \wedge E_j), E_i \wedge E_j \rangle\!\rangle$$

$$= \sum_{i<j} \big\langle R(E_i, E_j)E_j, E_i \big\rangle = \frac{1}{2}S.$$

Hence

$$\widehat{U} = \frac{\langle\!\langle\!\langle \widehat{R}, \mathrm{Id} \rangle\!\rangle\!\rangle}{\sqrt{\binom{n}{2}}} \cdot \frac{\mathrm{Id}}{\sqrt{\binom{n}{2}}} = \frac{S}{n(n-1)} \cdot \mathrm{Id}.$$

8.22. Lemma. The maps

$$A \overset{\Psi}{\longmapsto} A \bullet g \ (\in \mathcal{U} \oplus \mathcal{Z}) \quad \text{and} \quad R \overset{\Psi^*}{\longmapsto} C_{Ric}R \ (= \ \mathrm{Ric}\ )$$

are (formally) adjoint to one another, i.e., we have

$$\langle\!\langle\!\langle \widehat{\Psi A}, \widehat{R} \rangle\!\rangle\!\rangle = \langle\!\langle A, \Psi^* R \rangle\!\rangle.$$

Here, C_{Ric} denotes the *Ricci contraction* which forms the Ricci tensor from the curvature tensor, hence $C_{Ric}R(X,Y) := \sum_i R(E_i, X, E_i, Y)$, and the inner product between two symmetric $(0,2)$ tensors A, B is defined by (cf. 8.6)

$$\langle\!\langle A, B\rangle\!\rangle := \sum_{i,j} A(E_i, E_j)B(E_j, E_i).$$

PROOF: If $E_1, \ldots, E_n$ is an ON-basis of A with eigenvalues $\lambda_1, \ldots, \lambda_n$, then, as is easily checked, $E_i \wedge E_j, i < j$, is an ON-basis of $\widehat{A \bullet g}$ with eigenvalues $\lambda_i + \lambda_j$. It follows that

$$\langle\!\langle\!\langle \widehat{A \bullet g}, \widehat{R} \rangle\!\rangle\!\rangle = \operatorname{Tr}\big(\widehat{R} \circ \widehat{A \bullet g}\big) = \sum_{i<j} \big\langle\!\big\langle \widehat{R}\big(\widehat{A \bullet g}(E_i \wedge E_j)\big), E_i \wedge E_j \big\rangle\!\big\rangle$$

$$= \sum_{i<j} (\lambda_i + \lambda_j)\langle R(E_i, E_j)E_j, E_i\rangle = \sum_i \lambda_i \operatorname{Ric}(E_i, E_i)$$

$$= \sum_{i,j} \underbrace{A(E_i, E_j)}_{\lambda_i \delta_{ij}} \operatorname{Ric}(E_i, E_j) = \langle\!\langle A, \operatorname{Ric}\rangle\!\rangle.$$

□

8.23. Corollary.

1. $\mathcal{W}$ is the kernel of the mapping Ψ^*. Hence the $\mathcal{W}$-part of R satisfies the equation $C_{Ric}W = 0$.
2. The $\mathcal{U} \oplus \mathcal{Z}$-part of R is equal to $C \bullet g$ with
$$C = \frac{1}{n-2}\Big(\operatorname{Ric} - \frac{S}{2(n-1)} g\Big).$$
This tensor is referred to as the *Schouten tensor*.

PROOF: The first part follows directly from the adjointness relation in 8.22, since the image of Ψ always lies in $\mathcal{U} \oplus \mathcal{Z}$. For the second part, note that every element of $\mathcal{U} \oplus \mathcal{Z}$ can be written in the form $A \bullet g$ with some symmetric tensor A. Thus we are led to the Ansatz

$$R = \underbrace{A \bullet g}_{\in \mathcal{U} \oplus \mathcal{Z}} + \underbrace{W}_{\in \mathcal{W}},$$

with A still to be determined. By forming traces, we get

$$\mathrm{Ric} = C_{Ric} R = C_{Ric}(A \bullet g) + \underbrace{C_{Ric} W}_{=0} = C_{Ric}(A \bullet g),$$

and consequently

$$\begin{aligned}\mathrm{Ric}(X,Y) &= \sum_i A \bullet g(E_i, X, E_i, Y) \\ &= \sum_i \Big[A(E_i,E_i)\langle X,Y\rangle + A(X,Y)\cdot 1 \\ &\qquad -A(E_i,Y)\langle X,E_i\rangle - A(X,E_i)\langle E_i,Y\rangle\Big] \\ &= (\mathrm{Tr}\,A)\cdot\langle X,Y\rangle + A(X,Y)\cdot n - 2A(X,Y).\end{aligned}$$

From this it follows that

$$\mathrm{Ric} = (\mathrm{Tr}\,A)\cdot g + (n-2)\cdot A,$$

or, equivalently,

$$A = \frac{1}{n-2}(\mathrm{Ric} - \mathrm{Tr}\,A \cdot g).$$

In order to determine A completely, we must calculate the trace of A:

$$\mathrm{Tr}\,A = \frac{1}{n-2}(S - \mathrm{Tr}\,A\cdot n), \quad \text{hence} \quad \mathrm{Tr}\,A = \frac{S}{2(n-1)}.$$

This verifies the equation $A = C$. □

8.24. Theorem. The components of R in the decomposition $R = U + Z + W$ are given as follows:

$$U = \frac{S}{n(n-1)} R_1;$$

$$Z = \frac{1}{n-2}\Big(\mathrm{Ric} - \frac{S}{n} g\Big) \bullet g;$$

$$W = R - U - Z = R - C \bullet g = R - \frac{1}{n-2}\Big(\mathrm{Ric} - \frac{S}{2(n-1)} g\Big) \bullet g.$$

In Ricci calculus, the decomposition $R_{abcd} = U_{abcd} + Z_{abcd} + W_{abcd}$ corresponds to the following components:

$$\begin{aligned}
U_{abcd} &= \frac{S}{n(n-1)}\Big(g_{ac}g_{bd} - g_{ad}g_{bc}\Big);\\
Z_{abcd} &= \frac{1}{n-2}\Big(R_{ac}g_{bd} + R_{bd}g_{ac} - R_{ad}g_{bc} - R_{bc}g_{ad}\Big)\\
&\quad -\frac{2S}{n(n-2)}\Big(g_{ac}g_{bd} - g_{ad}g_{bc}\Big);\\
W_{abcd} &= R_{abcd} - \frac{1}{n-2}\Big(R_{ac}g_{bd} + R_{bd}g_{ac} - R_{ad}g_{bc} - R_{bc}g_{ad}\Big)\\
&\quad +\frac{S}{(n-1)(n-2)}\Big(g_{ac}g_{bd} - g_{ad}g_{bc}\Big).
\end{aligned}$$

Note that the coefficient $\frac{S}{n(n-1)}$ in the first term is nothing but the *normalized scalar curvature*, which is unity for the unit sphere. The term $\mathrm{Ric} - \frac{S}{n}g$ which occurs in Z is the traceless part of the Ricci tensor. Thus we are witness to the occurrence of a double trace (the scalar curvature) in the component U and a single trace (the Ricci contraction) in the component Z. The remaining component W is traceless. The decomposition of $\mathcal{R}$ in the three subspaces $\mathcal{U}, Z, W$ is in addition *irreducible* with respect to the (simultaneous) action of the orthogonal group $\mathbf{O}(n)$ on the four arguments of the tensor.

PROOF: Above we have already seen that

$$U = \frac{S}{n(n-1)}R_1 \quad \text{and} \quad \widehat{U} = \frac{S}{n(n-1)}\mathrm{Id}.$$

According to 8.23 we have

$$Z = C \bullet g - U = \frac{1}{n-2}\Big(\mathrm{Ric} - \frac{S}{2(n-1)}g\Big) \bullet g - \frac{S}{n(n-1)} \cdot \frac{1}{2} g \bullet g$$

$$= \frac{1}{n-2}\Big(\mathrm{Ric} - \Big[\frac{Sn}{2n(n-1)} + \frac{S(n-2)}{2n(n-1)}\Big]g\Big) \bullet g = \frac{1}{n-2}\Big(\mathrm{Ric} - \frac{S}{n}g\Big) \bullet g$$

as well as

$$W = R - C \bullet g. \qquad \square$$

8E The Weyl tensor

The component W of the curvature tensor R is according to 8.24 just the difference of R and the components U and Z. Looked at this way, it does not appear to have any particular geometric significance. However, of all components of the curvature, W, which is also known as the *Weyl tensor*, is the most important. It is what is referred to as the *conformal curvature*, that is, it is the component of the curvature which depends only on the conformal structure defined by g. First we deduce the following simple consequence of 8.24:

8.25. Corollary. For every n-dimensional Riemannian manifold (M, g) with $n \geq 3$ we have:
(i) g has constant curvature $\iff Z = W = 0$;
(ii) g is an Einstein metric $\iff Z = 0$;
(iii) g has vanishing scalar curvature $\iff U = 0$;
(iv) $\mathrm{Ric}_g = 0 \iff U = Z = 0$;
(v) $n = 3 \implies W = 0$.

Furthermore, as we shall see below, one has

(vi) g is locally conformally flat $\implies W = 0$;

for $n \geq 4$ also the converse of this statement is true (Theorem of Schouten, 8.31). In particular, in this case, if g is locally a conformally flat Einstein metric then g has constant curvature.[8]

PROOF: Parts (i) to (iv) follow immediately from 8.24. The case of dimension two was already shown in 6.6: the curvature tensor is always a multiple of the standard curvature tensor R_1, so $R = U$. For the proof of (v), recall that in dimension three the curvature tensor is already determined by the Ricci tensor. In an ON-basis E_1, E_2, E_3

[8] This goes back to J. A. Schouten & D. Struik, *On some properties of general manifolds relating to Einstein's theory of gravitation*, American Journal of Math. **43**, 213–216 (1921).

we have

$$\begin{aligned}\mathrm{Ric}(E_1,E_1) &= K_{12}+K_{13},\\ \mathrm{Ric}(E_2,E_2) &= K_{21}+K_{23},\\ \mathrm{Ric}(E_3,E_3) &= K_{31}+K_{32}.\end{aligned}$$

If the Ricci tensor is given, the left-hand side yields three equations for three indeterminants, namely for K_{12}, K_{13}, K_{23}. This system of equations can be uniquely solved, as the rank of the relevant matrix is maximal. Thus the Ricci tensor uniquely determines the sectional curvatures, which according to 6.5 uniquely determine the curvature tensor. Thus, the Ricci tensor uniquely determines the curvature tensor. In 6.13 we saw similarly that a three-dimensional Einstein space must have constant curvature. On the other hand, the Ricci tensor is completely determined by U and Z. It follows that in this case W must vanish, as otherwise there would be some tensor which is not algebraically determined by U and Z. □

Note that in dimension four the same argument leads to four equations in six indeterminants $K_{ij}, i<j$, so that in this dimension there are two degrees of freedom, and so there are non-trivial solutions. If we consider the decomposition of the curvature tensor $R=U+Z+W$ and the corresponding spaces

$$\mathcal{R}=\mathcal{U}\oplus\mathcal{Z}\oplus\mathcal{W},$$

the dimensions of the subspaces $\mathcal{U},\mathcal{Z},\mathcal{W}$ are of course of interest. They are as follows:

dim	$\mathcal{R}$	$\mathcal{U}$	$\mathcal{Z}$	$\mathcal{W}$
2	1	1	0	0
3	6	1	5	0
4	20	1	9	10
n	$\frac{1}{12}n^2(n^2-1)$	1	$\frac{1}{2}n(n+1)-1$	(difference)

8.26. Definition. The $\mathcal{W}$–component W of the curvature tensor is called the *Weyl tensor*, or sometimes also the *conformal curvature tensor*. The latter terminology comes from the fact that W is conformally invariant, see Lemma 8.30 below.

Two metrics $g, \widetilde{g}$ on one and the same manifold (cf. 3.29 and 5.11) are said to be *conformally equivalent*, if the measurement of angles is the same in both metrics, i.e., if $\widetilde{g} = e^{-2\varphi} g$ holds for some scalar function φ. For these two metrics, let

$$\nabla, \widetilde{\nabla}, R, \widetilde{R}, S, \widetilde{S}, \mathrm{Ric}, \widetilde{\mathrm{Ric}}, U, \widetilde{U}, W, \widetilde{W}, Z, \widetilde{Z}, \text{etc.}$$

denote the corresponding curvature quantities.

8.27. Lemma. For $\widetilde{g} = e^{-2\varphi} g = \psi^{-2} g$ one has the following equations between the corresponding quantities for the two metrics:

(i) $\widetilde{\nabla}_X Y = \nabla_X Y - (X\varphi)Y - (Y\varphi)X + \langle X, Y\rangle \mathrm{grad}\varphi$;

(ii) $$\widetilde{R}(X,Y)Z = R(X,Y)Z - \langle \nabla_X \mathrm{grad}\varphi, Z\rangle Y + \langle \nabla_Y \mathrm{grad}\varphi, Z\rangle X$$
$$-\langle X, Z\rangle \nabla_Y \mathrm{grad}\varphi + \langle Y, Z\rangle \nabla_X \mathrm{grad}\varphi + (Y\varphi)(Z\varphi)X$$
$$-(X\varphi)(Z\varphi)Y - \langle \mathrm{grad}\varphi, \mathrm{grad}\varphi\rangle \cdot R_1(X,Y)Z$$
$$+\Big((X\varphi)\langle Y, Z\rangle - (Y\varphi)\langle X, Z\rangle\Big) \cdot \mathrm{grad}\varphi;$$

(iii) $\widetilde{\mathrm{Ric}} = \mathrm{Ric} \ + (\Delta\varphi - (n-2)||\mathrm{grad}\varphi||^2)g + (n-2)e^{-\varphi}\nabla^2(e^{\varphi})$;

(iv) $\widetilde{S} = \psi^2 S + 2(n-1)\psi\Delta\psi - n(n-1)||\mathrm{grad}\psi||^2$.

PROOF: Part (i) follows simply by applying the Koszul formula of 5.16 to both metrics. For g we have

$$2\langle \nabla_X Y, Z\rangle = X\langle Y, Z\rangle + Y\langle X, Z\rangle - Z\langle X, Y\rangle$$
$$-\langle Y, [X,Z]\rangle - \langle X, [Y,Z]\rangle - \langle Z, [Y,X]\rangle$$

and for $\widetilde{g}$

$$2e^{-2\varphi}\langle \widetilde{\nabla}_X Y, Z\rangle = X\big(e^{-2\varphi}\langle Y, Z\rangle\big) + Y\big(e^{-2\varphi}\langle X, Z\rangle\big) - Z\big(e^{-2\varphi}\langle X, Y\rangle\big)$$
$$-e^{-2\varphi}\langle Y, [X,Z]\rangle - e^{-2\varphi}\langle X, [Y,Z]\rangle - e^{-2\varphi}\langle Z, [Y,X]\rangle.$$

The difference of the two left-hand sides is $2e^{-2\varphi}\langle \widetilde{\nabla}_X Y - \nabla_X Y, Z\rangle$

$$= X\big(e^{-2\varphi}\big)\langle Y, Z\rangle + Y\big(e^{-2\varphi}\big)\langle X, Z\rangle - Z\big(e^{-2\varphi}\big)\langle X, Y\rangle$$
$$= 2e^{-2\varphi}\Big(- X\varphi\langle Y, Z\rangle - Y\varphi\langle X, Z\rangle + \langle Z, \mathrm{grad}\varphi\rangle\langle X, Y\rangle\Big).$$

This implies (i).

Part (ii) follows from (i) by applying these formulas twice to the terms of the form $\nabla_X \nabla_Y Z$ and $\widetilde{\nabla}_X \widetilde{\nabla}_Y Z$. In addition to the first derivatives of φ, second derivatives of the form of the Hesse tensor $\nabla_X \mathrm{grad}\varphi$ appear, cf. 6.2 and Exercise 8 with its solution on p. 379.

The relation between φ and ψ is given by the chain rule $X(\psi) = X(e^{\varphi}) = (X\varphi) \cdot e^{\varphi}$:

$$\mathrm{grad}\psi = \psi \mathrm{grad}\varphi, \quad \Delta\psi = \psi\Delta\varphi + \psi||\mathrm{grad}\varphi||^2,$$

$$\nabla^2\psi(Y,Z) = \psi(\nabla^2\varphi(Y,Z) + (Y\varphi)(Z\varphi)), \quad \nabla^2\psi = \psi(\nabla^2\varphi + \nabla\varphi \cdot \nabla\varphi).$$

Using these equations we can prove part (iii) from (ii) by taking the trace with respect to an ON basis E_i for g and $\widetilde{E}_i = \psi E_i$ for $\widetilde{g}$:

$$\widetilde{\mathrm{Ric}}(Y,Z) = \mathrm{Ric}(Y,Z) - \sum_i \langle \nabla_{E_i} \mathrm{grad}\varphi, Z\rangle \langle Y, E_i\rangle$$

$$+ \sum_i \langle \nabla_Y \mathrm{grad}\varphi, Z\rangle\langle E_i, E_i\rangle - \sum_i \langle E_i, Z\rangle\langle \nabla_Y \mathrm{grad}\varphi, E_i\rangle$$

$$+ \sum_i \langle Y, Z\rangle \langle \nabla_{E_i}\mathrm{grad}\varphi, E_i\rangle + (Y\varphi)(Z\varphi)n - (Y\varphi)(Z\varphi)$$

$$-||\mathrm{grad}\varphi||^2 \cdot (n-1)\langle Y,Z\rangle + \sum_i (E_i\varphi)^2\langle Y,Z\rangle - \sum_i (Y\varphi)\langle E_i, Z\rangle E_i\varphi$$

$$= \mathrm{Ric}(Y,Z) + (n-2)\nabla^2\varphi(Y,Z) + (n-2)(Y\varphi)(Z\varphi)$$

$$+\Delta\varphi\langle Y,Z\rangle - (n-2)||\mathrm{grad}\varphi||^2\langle Y,Z\rangle$$

$$= \mathrm{Ric}(Y,Z) + (n-2)\psi^{-1}\nabla^2\psi(Y,Z) + \Big(\Delta\varphi - (n-2)||\mathrm{grad}\varphi||^2\Big) g(Y,Z).$$

Part (iv) follows from (iii) by taking again the trace:

$$\widetilde{S} = \mathrm{Tr}_{\widetilde{g}}\widetilde{\mathrm{Ric}} = \sum_i \widetilde{\mathrm{Ric}}(\widetilde{E}_i, \widetilde{E}_i)$$

$$= \psi^2 S + \psi^2\Big(n\big(\psi^{-1}\Delta\psi - (n-1)\psi^{-2}||\mathrm{grad}\psi||^2\big) + (n-2)\psi^{-1}\Delta\psi\Big)$$

$$= \psi^2 S + 2(n-1)\psi\Delta\psi - n(n-1)||\mathrm{grad}\psi||^2. \qquad \square$$

The somewhat complicated formula in part (ii) can be written more succinctly for the corresponding (0,4)-tensors as

$$\langle \widetilde{R}(X,Y)Z,T\rangle = \langle R(X,Y)Z,T\rangle - \frac{1}{2}\langle \mathrm{grad}\varphi, \mathrm{grad}\varphi\rangle (g\bullet g)(X,Y,T,Z)$$

$$+(\nabla^2\varphi \bullet g)(X,Y,T,Z) + (\nabla\varphi\cdot\nabla\varphi)\bullet g(X,Y,T,Z),$$

and consequently

$$\boxed{e^{2\varphi}\widetilde{R} = R - \frac{1}{2}\langle \mathrm{grad}\varphi, \mathrm{grad}\varphi\rangle g\bullet g + (\nabla^2\varphi)\bullet g + (\nabla\varphi)^2\bullet g.}$$

Here $\nabla^2\varphi = \nabla(\nabla\varphi)$ denotes the Hessian of φ as a (0,2)-tensor, while R denotes the Riemannian curvature tensor as a (0,4)-tensor (Definition 8.19).

8.28. Corollary. In every dimension $n \geq 3$ one has:

1. A metric g is conformally equivalent to an Einstein metric if and only if
$$e^{\varphi}\mathrm{Ric} + (n-2)\nabla^2(e^{\varphi})$$
is a scalar multiple of g for an appropriately chosen function φ.
2. If g is an Einstein metric, then $\widetilde{g} = e^{-2\varphi}g$ is an Einstein metric if and only if $\nabla^2(e^{\varphi}) = \lambda g$ for some scalar function λ.

The proof is easy, using equation (iii) in 8.27. Note the factor $(n-2)$ in front of $\nabla^2(e^{\varphi})$, which of course implies that part (1) of 8.28 is trivial and part (2) is no longer true in dimension $n = 2$. The differential equation in part (2) of 8.28 can be explicitly solved by reducing it to an ordinary differential equation $y'' + cy = 0$, where c is a constant which only depends on the scalar curvature and the dimension. This same is true for conformal transformations between metrics of constant sectional curvature.

8.29. Corollary. For two-dimensional Riemannian metrics $g, \widetilde{g} = e^{-2\varphi}g$ one has:

(i) $\quad R = K\cdot R_1 = \frac{1}{2}Kg\bullet g, \quad \widetilde{R} = \frac{1}{2}\widetilde{K}\widetilde{g}\bullet\widetilde{g} = \frac{1}{2}e^{-4\varphi}\widetilde{K}g\bullet g;$

(ii) g is conformally equivalent to a (flat) Euclidean metric if and only if there is a function φ with $\Delta_g \varphi = -K$. Here Δ_g denotes the Laplace-Beltrami operator with respect to g, see the examples in 6.9.

Proof: (i) follows immediately from the fact we already know that R and $\widetilde{R}$ are both scalar multiples of the standard curvature tensor. For (ii) we use 8.27:

$$e^{-2\varphi}\widetilde{K} g \bullet g = K g \bullet g + \big(2\nabla^2\varphi + 2(\nabla\varphi)^2 - \langle \mathrm{grad}\varphi, \mathrm{grad}\varphi\rangle g\big) \bullet g,$$

so that the Gaussian curvatures $K, \widetilde{K}$ are calculated in an ON-basis as $K = \big\langle R(X,Y)Y, X\big\rangle$, $\qquad \widetilde{K} = e^{2\varphi}\big\langle \widetilde{R}(X,Y)Y, X\big\rangle$,

$$e^{-2\varphi}\widetilde{K} = K + \mathrm{Tr}\nabla^2\varphi + (X\varphi)^2 + (Y\varphi)^2 - (X\varphi)^2 - (Y\varphi)^2 = K + \Delta_g\varphi.$$

The equation $\widetilde{K} = 0$ is thus equivalent to $K + \Delta_g \varphi = 0$. $\square$

Consequence. Every two-dimensional Riemannian metric is locally conformally Euclidean (or locally conformally flat). Thus, isothermal parameters always exist, cf. 3.29.

This follows when we solve (locally) the partial differential equation (known as the *potential equation*)

$$\Delta_g \varphi = -K$$

for a given function K as the Gaussian curvature of g. In the case of the Euclidean metric, this is also known as the *Poisson equation.* In the more general situation, the equation is given in local coordinates in the form $\Delta_g\varphi = \nabla_i \varphi^i = -K$ or $\frac{\partial}{\partial u^i}\big(\varphi_j g^{ji}\big) + \Gamma^i_{il}\varphi_j g^{jl} = -K$. Locally a solution always exists by general results on elliptic differential operators, see [**30**]. Then $\widetilde{K} = 0$ and hence also $\widetilde{R} = 0$, so that $e^{-2\varphi} \cdot g$ is flat (Euclidean). The expression "conformally flat" is used in the literature mostly to mean "locally conformally flat". This raises the interesting question

Problem: Which Riemannian metrics are locally conformally flat for $n \geq 3$?

8.30. Lemma. (H. Weyl[9])
The $\mathcal{W}$-component of the curvature tensor is conformally invariant, i.e., for a conformally equivalent metric $\widetilde{g} = e^{-2\varphi} g$ one has $\widetilde{W} = W$ for the corresponding $(1,3)$-tensors and $\widetilde{W} = e^{-2\varphi} W$ for the corresponding $(0,4)$-tensors. In particular, $W = 0$ whenever g is locally conformally flat.

PROOF: For this one just has to insert the equation of 8.27 into the expression for W following 8.24 and evaluate the terms (exercise). □

Question: Is the necessary condition $W = 0$ also sufficient for the conformal flatness of a metric? The **answer** is: "no" for $n = 3$ and "yes" for $n \geq 4$.

8.31. Theorem. (J. A. Schouten[10])
For $n \geq 4$ the metric g is conformally flat if and only if $W = 0$.
For $n = 3$ the metric g is conformally flat if and only if the relation

$$(\nabla_X C)(Y, Z) = (\nabla_Y C)(X, Z)$$

holds for all X, Y, Z. Here C and W denote the Schouten tensor and the Weyl tensor with $R = C \bullet g + W$.

PROOF: All calculations which follow are local. First of all, g is conformally flat if and only if $\widetilde{R} = 0$ holds for an appropriately chosen function φ, i.e., if and only if

$$R + \Big(-\frac{1}{2}\langle \mathrm{grad}\varphi, \mathrm{grad}\varphi \rangle g + \nabla^2\varphi + (\nabla\varphi)^2 \Big) \bullet g = 0$$

for an appropriately chosen function φ. Because of the orthogonal decomposition $R = C \bullet g + W$, this is equivalent to the relation

$$C - \frac{1}{2}\langle \mathrm{grad}\varphi, \mathrm{grad}\varphi \rangle g + \nabla^2\varphi + (\nabla\varphi)^2 = 0$$

for some φ, and in addition $W = 0$. This in turn holds if and only if $C - \frac{1}{2} \parallel \alpha \parallel^2 \cdot g + \nabla\alpha + \alpha \cdot \alpha = 0$ for a one-form $\alpha = d\varphi$ and $W = 0$. The integrability condition for the last equation $\alpha = d\varphi$ is

$$d\alpha = 0 \Longleftrightarrow \nabla\alpha \text{ is symmetric} \Longleftrightarrow C \text{ is symmetric.}$$

[9] *Reine Infinitesimalgeometrie*, Math. Zeitschrift **2**, 384–411 (1918).
[10] *Über die konforme Abbildung n-dimensionaler Mannigfaltigkeiten mit quadratischer Maßbestimmung auf eine Mannigfaltigkeit mit euklidischer Maßbestimmung*, Math. Zeitschrift **11**, 58–88 (1921), cf. also [**16**], Chapter VI, §5.

This symmetry is by definition of C always satisfied, see 8.23. Thus, we only have to show the integrability condition for the equation

$$C - \tfrac{1}{2} \parallel \alpha \parallel^2 \cdot g + \nabla\alpha + \alpha \cdot \alpha = 0 \text{ plus the condition } W = 0.$$

Note that the integrability conditions amount to symmetries of the next-higher derivatives. This can be expressed with the help of the "exterior derivative"

$$d^\nabla A(X, Y, Z) := (\nabla_X A)(Y, Z) - (\nabla_Y A)(X, Z)$$

for an arbitrary symmetric $(0, 2)$-tensor. We have

$$\begin{aligned} d^\nabla \nabla^2\varphi(X, Y, Z) &= \langle R(X, Y)\mathrm{grad}\varphi, Z\rangle \\ d^\nabla(\nabla\varphi \cdot \nabla\varphi)(X, Y, Z) &= (Y\varphi)\nabla^2\varphi(X, Z) - (X\varphi)\nabla^2\varphi(Y, Z) \\ d^\nabla(\tfrac{1}{2}||\mathrm{grad}\varphi||^2 \cdot g)(X, Y, Z) &= \nabla^2\varphi(X, \mathrm{grad}\varphi)\langle Y, Z\rangle \\ &\quad -\nabla^2\varphi(Y, \mathrm{grad}\varphi)\langle X, Z\rangle \end{aligned}$$

and, similarly,

$$\begin{aligned} d^\nabla \nabla\alpha(X, Y, Z) &= -\alpha\big(R(X, Y)Z\big) \\ d^\nabla(\alpha \cdot \alpha)(X, Y, Z) &= d\alpha(X, Y)\alpha(Z) \\ &\quad +\alpha(Y)\nabla\alpha(X, Z) - \alpha(X)\nabla\alpha(Y, Z) \\ d^\nabla(\tfrac{1}{2}||\alpha||^2 \cdot g)(X, Y, Z) &= \langle\nabla_X\alpha, \alpha\rangle\langle Y, Z\rangle - \langle\nabla_Y\alpha, \alpha\rangle\langle X, Z\rangle. \end{aligned}$$

We now evaluate d^∇ for the left hand side of the equation above as follows:

$$\begin{aligned} & d^\nabla\big(C + \nabla^2\varphi + (\nabla\varphi)^2 - \tfrac{1}{2}||\mathrm{grad}\varphi||^2 g\big)(X, Y, Z) \\ &= d^\nabla C(X, Y, Z) + R(X, Y, Z, \mathrm{grad}\varphi) + (Y\varphi)\nabla^2\varphi(X, Z) \\ &\quad -(X\varphi)\nabla^2\varphi(Y, Z) - \nabla^2\varphi(X, \mathrm{grad}\varphi)\langle Y, Z\rangle + \nabla^2\varphi(Y, \mathrm{grad}\varphi)\langle X, Z\rangle \\ &= d^\nabla C(X, Y, Z) + C \bullet g(X, Y, Z, \mathrm{grad}\varphi) \\ &\quad +Y\varphi\big[\tfrac{1}{2}||\mathrm{grad}\varphi||^2\langle X, Z\rangle - (\nabla\varphi)^2(X, Z) - C(X, Z)\big] \\ &\quad -X\varphi\big[\tfrac{1}{2}||\mathrm{grad}\varphi||^2\langle Y, Z\rangle - (\nabla\varphi)^2(Y, Z) - C(Y, Z)\big] \\ &\quad -\langle Y, Z\rangle\big[\tfrac{1}{2}||\mathrm{grad}\varphi||^2\langle X, \mathrm{grad}\varphi\rangle - (\nabla\varphi)^2(X, \mathrm{grad}\varphi) - C(X, \mathrm{grad}\varphi)\big] \\ &\quad +\langle X, Z\rangle\big[\tfrac{1}{2}||\mathrm{grad}\varphi||^2\langle Y, \mathrm{grad}\varphi\rangle - (\nabla\varphi)^2(Y, \mathrm{grad}\varphi) - C(Y, \mathrm{grad}\varphi)\big] \\ &= d^\nabla C(X, Y, Z) = (\nabla_X C)(Y, Z) - (\nabla_Y C)(X, Z), \end{aligned}$$

similarly for α instead of $\nabla\varphi$. Therefore $d^\nabla C = 0$ is the integrability condition for the equation above. Thus, $\widetilde{R} = 0$ if and only if

$$d^\nabla C = 0 \quad \text{and} \quad W = 0.$$

For $n = 3$ this verifies the statement of the theorem, as in this case $W = 0$ always holds according to 8.25. For $n \geq 4$ it can be shown that the equation $W = 0$ implies the remaining equation $d^\nabla C = 0$, as follows. From $W = 0$ we get the equality $R = C \bullet g$, which means, taking the product rule 8.20 into account, that $\nabla_X R = \nabla_X(C \bullet g) = (\nabla_X C) \bullet g$. We now insert this relation into the second Bianchi identity

$$\langle \nabla_X R(Y,Z)T, V\rangle + \langle \nabla_Y R(Z,X)T, V\rangle + \langle \nabla_Z R(X,Y)T, V\rangle = 0$$

and form the trace over Y and V, i.e., we set $Y = V = E_i$ for an ON-basis E_i and form the sum:

$$\begin{aligned}
0 \;&=\; \textstyle\sum_i \nabla_X C(E_i,E_i)g(Z,T) + \sum_i \nabla_X C(Z,T)g(E_i,E_i) \\
&\quad - \textstyle\sum_i \nabla_X C(E_i,T)g(Z,E_i) - \sum_i \nabla_X C(Z,E_i)g(E_i,T) \\
&\quad + \textstyle\sum_i \nabla_{E_i} C(Z,E_i)g(X,T) + \sum_i \nabla_{E_i} C(X,T)g(Z,E_i) \\
&\quad - \textstyle\sum_i \nabla_{E_i} C(Z,T)g(X,E_i) - \sum_i \nabla_{E_i} C(X,E_i)g(Z,T) \\
&\quad + \textstyle\sum_i \nabla_Z C(X,E_i)g(E_i,T) + \sum_i \nabla_Z C(E_i,T)g(X,E_i) \\
&\quad - \textstyle\sum_i \nabla_Z C(X,T)g(E_i,E_i) - \sum_i \nabla_Z C(E_i,E_i)g(X,T) \\
&=\; (n-3)\big(\nabla_X C(Z,T) - \nabla_Z C(X,T)\big) \\
&+\; \big(\mathrm{div}C(Z) - \mathrm{Tr}\nabla_Z C\big)g(X,T) - \big(\mathrm{div}C(X) - \mathrm{Tr}\nabla_X C\big)g(Z,T) \\
&=\; (n-3)d^\nabla C(X,Z,T).
\end{aligned}$$

Thus for $n \geq 4$ it follows that $d^\nabla C = 0$. The expressions

$$\mathrm{div}C(Z) - \mathrm{Tr}\nabla_Z C \quad \text{and} \quad \mathrm{div}C(X) - \mathrm{Tr}\nabla_X C$$

vanish because $\mathrm{div}(\mathrm{Ric})(X) = \frac{X(S)}{2}$, and consequently

$$\mathrm{div}C(X) - \mathrm{Tr}\nabla_X C$$

$$= \frac{1}{n-2}\Big(\frac{X(S)}{2} - \frac{X(S)}{2(n-1)} - \mathrm{Tr}\nabla_X \mathrm{Ric} + \frac{1}{2(n-1)}\mathrm{Tr}\nabla_X(Sg)\Big) = 0.$$

$\square$

For further aspects of conformal geometry in connection with Riemannian geometry, have a look at the volume "Conformal Geometry", editors R. S. Kulkarni and U. Pinkall, Vieweg, 1988.

8F Duality for four-manifolds and Petrov types

There are many reasons why dimension four is very special. One is that it is the smallest dimension in which non-trivial Einstein metrics can occur. Another is that it is the dimension of classical space-time (as the physicists say, it is $(3+1)$-dimensional). Finally, there is a duality in this dimension (and only in this dimension) between two-dimensional subspaces, which always occur in orthogonal pairs. Thus one has a duality operator for two-dimensional subspaces (or, fixing the orientation, for bivectors), which associates to each such subspace its orthogonal complement. This leads to the following additional structure, the so-called *Hodge duality.*

8.32. Definition. (Duality in dimension four)

Suppose we are given an oriented four-dimensional vector space V with inner product, and we denote by $\bigwedge^2 = \bigwedge^2(V)$ the space of bivectors over V. In a fixed ON-basis E_1, E_2, E_3, E_4, we define the *Hodge operator*

$$*\colon \bigwedge^2 \longrightarrow \bigwedge^2$$

by $*(E_i \wedge E_j) = E_k \wedge E_l$, where E_i, E_j, E_k, E_l is positively oriented in the sense that $E_i \wedge E_j \wedge E_k \wedge E_l = E_1 \wedge E_2 \wedge E_3 \wedge E_4$. Then the square $*^2 = * \circ *$ is the identity. More precisely, we have

$$\begin{aligned}
*(E_1 \wedge E_2) &= E_3 \wedge E_4, \\
*(E_1 \wedge E_3) &= E_4 \wedge E_2, \\
*(E_1 \wedge E_4) &= E_2 \wedge E_3, \\
*(E_2 \wedge E_3) &= E_1 \wedge E_4, \\
*(E_2 \wedge E_4) &= E_3 \wedge E_1, \\
*(E_3 \wedge E_4) &= E_1 \wedge E_2.
\end{aligned}$$

Because

$$\langle\!\langle *(E_i \wedge E_j), E_k \wedge E_l \rangle\!\rangle = \langle\!\langle E_i \wedge E_j, *(E_k \wedge E_l) \rangle\!\rangle,$$

the Hodge operator $*$ is self-adjoint, and thus the only eigenvalues of $*$, taking the relation $*^2 = \mathrm{Id}$ into consideration, are $+1$ and -1. We define the corresponding eigenspaces as follows:

$$\begin{array}{rcl} \bigwedge^2_+ & = & \{V \in \bigwedge^2 \,|\, *V = V\}, \\ \bigwedge^2_- & = & \{V \in \bigwedge^2 \,|\, *V = -V\}. \end{array}$$

These form an orthogonal decomposition

$$\bigwedge^2 = \bigwedge^2_+ \oplus \bigwedge^2_-$$

with $\dim \bigwedge^2_+ = \dim \bigwedge^2_- = 3$. In case we are on an oriented four-manifold, it now gets quite interesting to compare the self-adjoint endomorphism $*$ with the endomorphism

$$\widehat{R}\colon \bigwedge^2(T_pM) \to \bigwedge^2(T_pM),$$

which is itself self-adjoint.

8.33. Theorem. (A. Einstein, 1927[11], rediscovered by I. M. Singer and J. Thorpe in 1969[12])

For an oriented Riemannian four-manifold (M, g), the following conditions are equivalent:

1. (M, g) is an Einstein space.
2. $* \circ \widehat{R} = \widehat{R} \circ *$.
3. The sectional curvature in any two planes which are orthogonal to one another coincides, i.e., $K_\sigma = K_{\sigma^\perp}$.

Proof: First we show the equivalence of points 1 and 3. In an ON-basis $E_1, \ldots, E_n$ with associated sectional curvatures K_{ij} in the E_i, E_j-planes, we have

$$\begin{array}{rcl} \mathrm{Ric}(E_1, E_1) & = & K_{12} + K_{13} + K_{14}, \\ \mathrm{Ric}(E_2, E_2) & = & K_{21} + K_{23} + K_{24}, \\ \mathrm{Ric}(E_3, E_3) & = & K_{31} + K_{32} + K_{34}, \\ \mathrm{Ric}(E_4, E_4) & = & K_{41} + K_{42} + K_{43}. \end{array}$$

[11] *Über die formale Beziehung des Riemannschen Krümmungstensors zu den Feldgleichungen der Gravitation*, Math. Annalen **97**, 99–103 (1927).
[12] *The curvature of* 4*-dimensional Einstein spaces*, "Global Analysis", Papers in honor of K. Kodaira, 355–365, Princeton Univ. Press, 1969.

For an Einstein metric the left-hand sides all coincide, thus so must also the right-hand sides. Thus, from the first two equations, it follows that $K_{13}+K_{14} = K_{23}+K_{24}$, and from the last two it similarly follows that $K_{13} + K_{23} = K_{14} + K_{24}$. This implies

$$\begin{aligned} K_{14} - K_{23} &= K_{24} - K_{13}, \\ K_{14} - K_{23} &= K_{13} - K_{24}, \end{aligned}$$

thus both sides necessarily vanish. This holds in an arbitrary ON-basis, thus for every pair of planes which are orthogonal to one another.

Conversely, from $K_{12} = K_{34}, K_{13} = K_{24}, K_{14} = K_{23}$, the Einstein condition

$$\mathrm{Ric}(E_1, E_1) = \mathrm{Ric}(E_2, E_2) = \mathrm{Ric}(E_3, E_3) = \mathrm{Ric}(E_4, E_4)$$

follows, as well as the equation $\mathrm{Ric}(E_i, E_j) = 0$ for $i \neq j$, the latter using polarization and the equation

$$\mathrm{Ric}(E_i + E_j, E_i + E_j) = \mathrm{Ric}(E_i - E_j, E_i - E_j).$$

For the equivalence of these to condition 2, we represent the curvature endomorphism $\widehat{R}$ and $*$ in a convenient basis, using $E_1 \wedge E_2, E_3 \wedge E_4, E_1 \wedge E_3, E_4 \wedge E_2, E_1 \wedge E_4, E_2 \wedge E_3$, in that order. The corresponding matrices of $\widehat{R}$ will be momentarily denoted $A_{ab}, 1 \leq a, b \leq 6$. From the self-adjoint property we have $A_{ab} = A_{ba}$. The matrix of the duality operator $*$ is clearly

$$B = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

By calculating the products AB and BA we see that $AB = BA$ if and only if

$$A_{11} = A_{22},\ A_{33} = A_{44},\ A_{55} = A_{66},$$
$$A_{13}=A_{24},\ A_{23}=A_{14},\ A_{15}=A_{26},\ A_{25}=A_{16},\ A_{35}=A_{46},\ A_{45}=A_{36}.$$

Comparing this with the sectional curvatures K_{ij}, we have

$$A_{11} = K_{12},\ A_{22} = K_{34},\ A_{33} = K_{13},$$
$$A_{44} = K_{24},\ A_{55} = K_{14},\ A_{66} = K_{23}.$$

Moreover,

$$\mathrm{Ric}(E_1, E_4) = R_{2142} + R_{3143} = -A_{14} + A_{23},$$

and so forth. The equations above are thus equivalent to the Einstein condition. Thus 1 and 2 are equivalent. Note that we are using the first Bianchi identity, which appears here in the form $A_{12}+A_{34}+A_{56} = 0$. The scalar curvature S is of course just the trace $\sum_i A_{ii}$ of the matrix A. □

In considering the dimensions of the individual subspaces in the decomposition $\mathcal{R} = \mathcal{U} \oplus \mathcal{Z} \oplus \mathcal{W}$, note that the nine equations above define the space $\mathcal{U} \oplus \mathcal{W}$, where $\mathcal{U}$ corresponds to the unit matrix. Altogether there are 21 degrees of freedom for the matrix A, which are reduced to 20 by the first Bianchi identity. These 20 dimensions are split as $1 + 9 + 10$, cf. 8.25. The space $\mathcal{W} = \mathcal{W}_+ \oplus \mathcal{W}_-$ splits in this respect into two five-dimensional spaces.

In what follows we consider the modifications which are necessary upon passage from a four-dimensional Riemannian manifold to a *space-time*, i.e., a four-dimensional Lorentz manifold (M, g), where g is pseudo-Riemannian of signature $(- + ++)$.

8.34. Definition. (Duality on a four-dimensional space-time)

As usual we let E_1, E_2, E_3, E_4 denote an ON-basis and $\varepsilon_i := \langle E_i, E_i\rangle$ with $\varepsilon_1 = -1$ and $\varepsilon_2 = \varepsilon_3 = \varepsilon_4 = +1$. Correspondingly, we have a 6-dimensional subspace $\bigwedge^2$ with an inner product of signature $(- + - + -+)$. More precisely, for $i \neq j$ we have:

$$\langle\!\langle E_i \wedge E_j, E_i \wedge E_j \rangle\!\rangle = \varepsilon_i \cdot \varepsilon_j =: \varepsilon_{ij}.$$

The Hodge operator $*\colon \bigwedge^2 \longrightarrow \bigwedge^2$ should again be self-adjoint with respect to $\langle\!\langle\ ,\ \rangle\!\rangle$. Because of possible signs, we have to be careful about this. Let $*(E_i \wedge E_j) = \pm E_k \wedge E_l$, where $(ijkl)$ denotes an even permutation. Then the required self-adjointness implies

$$\varepsilon_{kl} = \langle\!\langle \underbrace{*(E_i \wedge E_j)}_{\pm E_k \wedge E_l}, \underbrace{*(E_i \wedge E_j)}_{\pm E_k \wedge E_l} \rangle\!\rangle = \langle\!\langle E_i \wedge E_j, \underbrace{*^2(E_i \wedge E_j)}_{\pm E_i \wedge E_j} \rangle\!\rangle = \pm\varepsilon_{ij}.$$

But taking the relation $\varepsilon_i \cdot \varepsilon_j \cdot \varepsilon_k \cdot \varepsilon_l = -1$ into account, this equation can only be satisfied if $*^2(E_i \wedge E_j) = -E_i \wedge E_j$. Thus in this case, the self-adjointness of $*$ implies $*^2 = -\mathrm{Id}$.

The *Hodge operator* $*\colon \bigwedge^2(T_pM) \longrightarrow \bigwedge^2(T_pM)$ *of a space-time* is therefore defined by

$$\begin{array}{rclrcl}
*(E_1 \wedge E_2) & = & E_3 \wedge E_4, & *(E_3 \wedge E_4) & = & -E_1 \wedge E_2, \\
*(E_1 \wedge E_3) & = & E_4 \wedge E_2, & *(E_4 \wedge E_2) & = & -E_1 \wedge E_3, \\
*(E_1 \wedge E_4) & = & E_2 \wedge E_3, & *(E_2 \wedge E_3) & = & -E_1 \wedge E_4.
\end{array}$$

In these relations, we view the curvature tensor R as an endomorphism $\widehat{R}$ of the space $\bigwedge^2(T_pM)$ of bivectors. In order to formulate Theorem 8.33 for space-times, we must take account of the fact that the sectional curvature

$$K_\sigma = \frac{\langle R(X,Y)Y,X\rangle}{\langle R_1(X,Y)Y,X\rangle}$$

is not well-defined for all types of planes, but rather only for *non-degenerate planes*, i.e., planes for which $\langle R_1(X,Y)Y,X\rangle \neq 0$ holds for at least one basis $X, Y \in \sigma$.

8.35. Theorem. (Variant of Theorem 8.33 for space-times)
For an oriented four-manifold (M,g) with the signature $(-+++)$, the following conditions are equivalent:

1. (M,g) is an Einstein space.
2. $* \circ \widehat{R} = \widehat{R} \circ *$.
3. The sectional curvatures in two non-degenerate planes which are orthogonal to one another are equal, i.e., $K_\sigma = K_{\sigma^\perp}$.
4. $\widehat{R}$ can be viewed as a $\mathbb{C}$-linear endomorphism of the complexification $\bigwedge^2_{\mathbb{C}}(T_pM)$ in which the representing matrix (induced by an ON-basis in $\bigwedge^2(T_pM)$) is symmetric.

Proof: To see the equivalence of 2 and 4, note that in the basis $E_1 \wedge E_2, E_3 \wedge E_4, E_1 \wedge E_3, E_4 \wedge E_2, E_1 \wedge E_4, E_2 \wedge E_3$ (in this order), the duality operator $*$ is represented by the matrix

$$B = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

We again represent the endomorphism $\widehat{R}$ by a matrix A_{ij}. By calculating the products AB and BA we see that $AB = BA$ if and only if

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ -A_{12} & A_{11} & -A_{14} & A_{13} & -A_{16} & A_{15} \\ A_{13} & A_{14} & A_{33} & A_{34} & A_{35} & A_{36} \\ -A_{14} & A_{13} & -A_{34} & A_{33} & -A_{36} & A_{35} \\ A_{15} & A_{16} & A_{35} & A_{36} & A_{55} & A_{56} \\ -A_{16} & A_{15} & -A_{36} & A_{35} & -A_{56} & A_{55} \end{pmatrix},$$

that is, if the matrix A can be written with symmetric 2×2 blocks of the form

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

each of which represents a complex number $C_{ij} = a - ib$. Thus $AB = BA$ is equivalent to A being given as a complex 3×3 matrix:

$$C = \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{12} & C_{22} & C_{23} \\ C_{13} & C_{23} & C_{33} \end{pmatrix},$$

where

$$\begin{aligned} C_{11} &= A_{11} - iA_{12}, & C_{12} &= A_{13} - iA_{14}, & C_{13} &= A_{15} - iA_{16}, \\ C_{22} &= A_{33} - iA_{34}, & C_{23} &= A_{35} - iA_{36}, & C_{33} &= A_{55} - iA_{56}, \end{aligned}$$

We now show the equivalence of 1 and 2. If (M, g) is an Einstein space, then from Ric$= \lambda g$ in our ON-basis we get $\mathrm{Ric}(E_i, E_j) = 0$ for $i \neq j$. In addition,

$$-\mathrm{Ric}(E_1, E_1) = \mathrm{Ric}(E_2, E_2) = \mathrm{Ric}(E_3, E_3) = \mathrm{Ric}(E_4, E_4).$$

Note here that, when doing calculations with the ON-basis, the raising and lowering of indices (i.e, passing from vectors to covectors) changes the sign if the index is the time-like index 1, while there is no such sign factor for the three space-like indices 2,3,4.

Calculating the diagonal elements of the Ricci tensor, we get, for example,

$$\mathrm{Ric}(E_1, E_1) = \sum_i \varepsilon_i \langle R(E_i, E_1)E_1, E_i\rangle$$

$$= R_{2121} + R_{3131} + R_{4141} = -A_{11} - A_{33} - A_{55},$$

and similarly for the remaining diagonal elements:

$$\begin{aligned}
\mathrm{Ric}(E_2, E_2) &= A_{11} + A_{44} + A_{66}, \\
\mathrm{Ric}(E_3, E_3) &= A_{22} + A_{33} + A_{66}, \\
\mathrm{Ric}(E_4, E_4) &= A_{22} + A_{44} + A_{66}.
\end{aligned}$$

Thus we have $A_{11} = A_{22}$, $A_{33} = A_{44}$ and $A_{55} = A_{66}$. Relations among the entries above and below the diagonal are calculated as in the following case:

$$0 = \mathrm{Ric}(E_3, E_4) = \sum_i \varepsilon_i \langle R(E_i, E_3)E_4, E_i\rangle$$

$$= -R_{1314} + R_{2324} = -R_{1314} - R_{2342} = A_{35} - A_{64},$$

resulting in the following conditions on the matrix A:

$$\begin{aligned}
A_{31} &= A_{24}, & A_{41} &= A_{23}, \\
A_{51} &= A_{62}, & A_{61} &= A_{25}, \\
A_{53} &= A_{46}, & A_{63} &= A_{45}.
\end{aligned}$$

Together with the self-adjointness (or the symmetry) of R, we find that A must have the form as given in the lemma above; thus $*\widehat{R} = \widehat{R}*$. The implication $2 \Rightarrow 1$ is obtained by following the above calculations backward.

The implication $1 \Rightarrow 3$ follows just as in 8.33: The Einstein condition yields $K_{12} = K_{34}, K_{13} = K_{24}, K_{14} = K_{23}$. For the converse implication $3 \Rightarrow 1$, we first calculate the diagonal elements of the Ricci tensor. Of course

$$\operatorname{Ric}(E_i, E_i) = \sum_j \varepsilon_j \langle R(E_j, E_i)E_i, E_j\rangle$$

$$= \varepsilon_i \sum_j \varepsilon_i \varepsilon_j \langle R(E_j, E_i)E_i, E_j\rangle = \varepsilon_i \sum_{j\neq i} K_{ij}.$$

Thus one gets

$$\begin{aligned}
-\operatorname{Ric}(E_1, E_1) &= K_{12} + K_{13} + K_{14},\\
\operatorname{Ric}(E_2, E_2) &= K_{12} + K_{23} + K_{24},\\
\operatorname{Ric}(E_3, E_3) &= K_{13} + K_{23} + K_{34},\\
\operatorname{Ric}(E_4, E_4) &= K_{14} + K_{24} + K_{34}.
\end{aligned}$$

The right-hand sides of the four equations are by assumption all equal to one another; thus $\operatorname{Ric}(E_i, E_i) = \lambda\varepsilon_i$. By taking traces we get $\lambda = \frac{S}{4}$. For the elements above and below the diagonal we use an argument with polarizations. For $i, j \neq 1$ we can polarize in the usual way, as in these cases $E_i + E_j$ is space-like. But for the quantities $E_1 + E_i$ we must proceed differently, as $E_1 + E_i$ is light-like, which means that it cannot be obtained as an element of an orthogonal basis. We consider $E_1 + tE_i$, which is space-like for every $t > 1$. Then, with $\operatorname{Ric}(E_i, E_i) = \varepsilon_i \frac{S}{4}$, we have

$$\tfrac{S}{4}(-1 + t^2) = \operatorname{Ric}(E_1 + tE_i, E_1 + tE_i) = \tfrac{S}{4}(-1 + t^2) + 2t\operatorname{Ric}(E_1, E_i).$$

But this equation can only be satisfied for $t > 1$ if $\operatorname{Ric}(E_1, E_i) = 0$ holds. □

If, instead of taking the complete curvature tensor R, we take only the Weyl component W, then we certainly have $\widehat{W}* = *\widehat{W}$, as W has no Z-component, which, according to Corollary 8.25 (ii), corresponds precisely to the Einstein condition. In this way we can associate to every curvature tensor a unique complex (3×3)-matrix coming from W. By the above argument the trace of this matrix vanishes, as in the decomposition of the curvature tensor $R = U + Z + W$ the component U carries the entire scalar curvature. This means that the sum of the eigenvalues of C must vanish.

8.36. Corollary. (Petrov types (following [**23**]))
For every space-time (M, g) we use Theorem 8.35 to associate to the Weyl tensor a complex matrix with $\mathrm{Tr}(C) = 0$. The Jordan normal form of this matrix is then one of the following six possibilities, in which $\lambda \neq 0$ and $\mu \neq \lambda$ denote complex numbers:

$$I : \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & -\lambda-\mu \end{pmatrix}, \quad D : \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & -2\lambda \end{pmatrix}, \quad O : \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$II : \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & -2\lambda \end{pmatrix}, \qquad N : \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$III : \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The type I, II, III, D, O or N which occurs is called the *Petrov type* of the metric g.

The notation is meant to emphasize that the main types which occur are I, II and III. For the type I, C is diagonalizable, i.e., the sum of the dimensions of the eigenspaces is three (which is precisely the sum of the geometric multiplicities). For type II the corresponding sum is two, while in the case of type III it is just one. The types D and O are "subtypes" of type I, for which not all eigenvalues are distinct. Similarly, type N is a "subtype" of type II for which the two eigenvalues λ and -2λ coincide. There can be no such "subtypes" of type III, as is easy to see. The Petrov types are important in the literature on the general theory of relativity, cf. [**25**].

Exercises

1. Show that the product metric of the two standard metrics of curvatures 1 and -1, respectively, on $S^2 \times H^2$, is a metric whose scalar curvature vanishes. Can an Einstein metric be obtained from this?

2. Calculate the Weyl tensor for the product metric on $S^2 \times S^2$, the product of two unit spheres.

3. Let (S^3, ds_1^2) denote the standard metric on the three-sphere of unit radius. Show that the product manifold $S^1 \times S^3$ with the Riemannian metric $ds^2 = dt^2 + (2 + \sin t)ds_1^2$ has constant scalar curvature and admits a one-parameter group of (globally defined) conformal diffeomorphisms onto itself. Hint: The conformal diffeomorphisms preserve the three-spheres $\{t\} \times S^3$ and "push" them in the t-direction, with varying radii of the individual spheres.

4. Prove that the Gauss equation for a hypersurface in 4.15 and 4.18 can be written as $R = \frac{1}{2}(II \bullet II)$ Compare the equation $R_1 = \frac{1}{2}(g \bullet g)$ in 8.20 and Exercise 23 at the end of Chapter 4.

5. Find a basis for the subspaces $\bigwedge^2_+$ and $\bigwedge^2_-$ in 8.32.

6. Verify the details of (ii) in Lemma 8.27.

7. Verify Lemma 8.30 by applying the formulas in 8.27 and 8.24.

8. Since the three-sphere admits three linearly independent vector fields at every point (see the exercises at the end of Chapter 7), it also admits a Lorentz metric g. Applying stereographic projection, one gets from this a Lorentz metric $\widetilde{g}$ on $\mathbb{R}^3$. Is $\widetilde{g}$ conformally equivalent to Minkowski space $\mathbb{R}^3_1$?

9. Show that in a space-time (a four-dimensional Lorentz manifold) a plane σ is non-degenerate if and only if there is a plane $\sigma^\perp$ which is perpendicular to it and also non-degenerate.

10. Determine the Petrov types for the standard curvature tensor R_1 as well as for the product metric of constant curvature, where one metric is Riemannian and the other is Lorentzian.

11. Show that the Schwarzschild metric from Chapter 5, Exercise 22 is Ricci-flat, i.e., that $\mathrm{Ric} \equiv 0$. What is the Petrov type?

12. A so-called *pp-wave* is defined as a four-dimensional manifold with a metric of the form $ds^2 = H(u, y, z)du^2 + 2dudv + dy^2 + dz^2$. Show that the Petrov type of a *pp*-wave is N or O.

13. Determine the Weyl tensor W for the following metrics:

$$\text{(a)}\quad ds^2 = -dudv + e^{uv}(dx^2 + dy^2)$$

$$\text{(b)}\quad ds^2 = -dudv + e^{xv}(dx^2 + dy^2)$$

Hint: Use the ON basis $E_1 = \frac{1}{2}\sqrt{2}(\partial_u + \partial_v)$, $E_2 = \frac{1}{2}\sqrt{2}(\partial_u - \partial_v)$, $E_3 = \frac{1}{||\partial_x||}\partial_x$, $E_4 = \frac{1}{||\partial_y||}\partial_y$.
Prove in addition that the Petrov type is D in (a) and III in (b).

14. Show that a *pp*-wave with metric $ds^2 = H(u,x,y)du^2 + 2dudv + dx^2 + dy^2$ is Ricci flat (i.e., the metric satisfies the Einstein field equations for the vacuum) if and only if the spatial Laplacian

$$\Delta_{xy} H = H_{xx} + H_{yy}$$

satisfies the equation $\Delta_{xy}H = 0$. Examples are the functions $H = h(u)(x^2 - y^2)$, $H = h(u)(x^4 + x^3y - 6x^2y^2 - xy^3 + y^4)$, and $H = h(u)\log(x^2 + y^2)$ where h denotes an arbitrary smooth function of the variable u.

15. The special case of a *pp*-wave with $H = 0$ is nothing but the flat Lorentz-Minkowski space $\mathbb{R}^4_1$ in spatial coordinates x, y and isotropic coordinates u, v.
    (a) Describe explicitly the isometry between the metrics $2dudv + dx^2 + dy^2$ and $-dx_0^2 + dx_1^2 + dx_2^2 + dx_3^2$.
    (b) Prove that the "boost" from 7.6 (see Figure 7.1) in $\mathbb{R}^4_1$ can also be written as the transformation $(u,v,x,y) \mapsto (e^{\varphi}u, e^{-\varphi}v, x, y)$.
    (c) Show that

$$\Phi_t(u,v,x,y) = \frac{1}{1-2tu}\big(u, v(1-2tu) + t(x^2+y^2), x, y\big)$$

    defines a 1-parameter group of conformal mappings Φ_t with the identity Φ_0 and $\Phi_{t+s} = \Phi_t \circ \Phi_s$.
    (d) Show that the associated conformal vector field is

$$\frac{\partial}{\partial t}\Big|_{t=0}\Phi_t = 2\big(u^2, \frac{1}{2}(x^2+y^2), ux, uy\big).$$

16. Prove that a conformal change of the metric $\widetilde{g} = \psi^{-2}g = e^{-2\varphi}g$ preserves the Ricci tensor in the sense that $\widetilde{\mathrm{Ric}}(X,Y) = \mathrm{Ric}(X,Y)$ for all X, Y if and only if ψ satisfies the differential equation $2\psi\nabla^2\psi = ||\mathrm{grad}\psi||^2 g$.
Hint: Formula in Lemma 8.27 (iii).

Solutions to selected exercises

Chapter 2.

1. We start with the equation $\dot{c} = ||\dot{c}||c'$ from 2.2 and obtain $\ddot{c} = (||\dot{c}||c')^{\cdot} = (||\dot{c}||)^{\cdot}c' + ||\dot{c}||(c')^{\cdot} = (||\dot{c}||)^{\cdot}c' + ||\dot{c}||^2c''$ and $\dddot{c} = (||\dot{c}||)^{\cdot\cdot}c' + 3(||\dot{c}||)^{\cdot}||\dot{c}||c'' + ||\dot{c}||^3c'''$. For a plane curve this implies $\mathrm{Det}(\dot{c},\ddot{c}) = ||\dot{c}||^3\mathrm{Det}(c',c'')$. On the other hand we have $\mathrm{Det}(c',c'') = \mathrm{Det}(e_1,\kappa e_2) = \kappa$ by the Frenet equations.
   For a space curve we have similarly $||\dot{c}\times\ddot{c}|| = ||(||\dot{c}||c')\times(||\dot{c}||^2c'')|| = ||\dot{c}||^3||c'\times c''|| = ||\dot{c}||^3||e_1\times\kappa e_2|| = ||\dot{c}||^3\kappa$.
   In order to calculate the torsion, we start with

$$\begin{aligned}\mathrm{Det}(\dot{c},\ddot{c},\dddot{c}) &= \mathrm{Det}(||\dot{c}||c', ||\dot{c}||^2c'', ||\dot{c}||^3c''')\\ &= ||\dot{c}||^6\mathrm{Det}(e_1,\kappa e_2,\kappa' e_2 - \kappa^2 e_1 + \kappa\tau e_3) = ||\dot{c}||^6\kappa^2\tau.\end{aligned}$$

   Here we used the equation for c''' from 2.9. From the result above we have $||\dot{c}||^3\kappa = ||\dot{c}\times\ddot{c}||$. Inserting this leads to $||\dot{c}\times\ddot{c}||^2\tau = ||\dot{c}||^6\kappa^2\tau = \mathrm{Det}(\dot{c},\ddot{c},\dddot{c})$, as claimed.

3. We can assume that c is parametrized by arc length because the vanishing of the derivative is independent of the choice of the parameter. Then the tangent vector of γ is $\gamma' = c' - \frac{\kappa'}{\kappa^2}e_2 + \frac{1}{\kappa}(-\kappa e_1) = -\frac{\kappa'}{\kappa^2}e_2$. This implies $\gamma' = 0 \Leftrightarrow \kappa' = 0$. The tangent

T to γ at the point $\gamma(t_0)$ is $T(u) = \gamma(t_0) + u\gamma'(t_0) = c(t_0) + \frac{1}{\kappa(t_0)} e_2(t_0) - u\frac{\kappa'(t_0)}{\kappa^2(t_0)} e_2(t_0)$. For the parameter $u = \kappa(t_0)/\kappa'(t_0)$ the tangent meets the curve c at $c(t_0)$, and there it points into the e_2-direction which is perpendicular to c or $c' = e_1$, respectively.

6. Any circle of radius r that is tangent to the x-axis at the origin can be described by the parametrization $\big(-r\sin\varphi, r(1-\cos\varphi)\big)$. Here $\varphi = 0$ corresponds to the origin $(0,0)$, and running through the circle clockwise corresponds to increasing φ. If we move the circle in the x-direction by x_0, then its center is moved from $(0,r)$ to (x_0,r). Now if the circle is rolled in the direction of the positive x-axis, then the x-coordinate of every point increases just by $r\varphi$, whereas the y-coordinate does not change. This leads to the parametrization $c(\varphi) = \big(r(\varphi - \sin\varphi), r(1-\cos\varphi)\big)$. From $c' = (r(1-\cos\varphi), r\sin\varphi)$ we see that the curve is not regular for all $\varphi = 2\pi k$ with $k \in \mathbb{Z}$. Indeed there is a cusp, as can be seen in Figure 2.11, which shows the case $r = 1$.

7. We use the formulas from the remarks in 2.7. The given $\kappa(s) = 1/\sqrt{s}$ leads to $\int_0^\sigma \kappa(t)dt = 2\sqrt{\sigma}$. This improper integral converges, even though κ is not finite for $s = 0$. One could also start with a positive s_0. It remains only to calculate the integrals $\int_0^s \cos(2\sqrt{\sigma})\ d\sigma$ and $\int_0^s \sin(2\sqrt{\sigma})\ d\sigma$. This is possible by substitution and integration by parts. One gets $\int \cos(2\sqrt{\sigma})d\sigma = \frac{1}{2}\int u\cos u\ du$ and $\int \sin(2\sqrt{\sigma})d\sigma = \frac{1}{2}\int u\sin u\ du$ with $u = 2\sqrt{\sigma}$. Therefore in the case of $\kappa(s) = 1/\sqrt{s}$, a parametrization of the curve can be expressed in terms of elementary functions.

8. We regard the Frenet matrix as a matrix-valued function $K(s)$ and observe the commutativity

$$K(s_1)K(s_2) = -\kappa(s_1)\kappa(s_2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = K(s_2)K(s_1).$$

The integral $\mathbf{K}(s) = \int_0^s K(t)dt$ satisfies the equations $\frac{d}{ds}\mathbf{K}(s) = K(s)$ and $\mathbf{K}(s_1)\mathbf{K}(s_2) = \mathbf{K}(s_2)\mathbf{K}(s_1)$. Just as with the well-known rule $(f^n(x))' = nf^{n-1}(x)f'(x)$, we obtain the similar rule

$$\frac{d}{ds}(\mathbf{K}(s))^j = j \cdot (\mathbf{K}(s))^{j-1} \cdot K(s)$$

from the telescoping sum

$$(a^j - b^j) = (a-b)(a^{j-1} + a^{j-2}b + \cdots + ab^{j-2} + b^{j-1})$$

as follows:

$$\frac{d}{ds}\Big(\mathbf{K}(s)\Big)^j = \lim_{h\to 0}\frac{1}{h}\Big(\big(\mathbf{K}(s+h)\big)^j - \big(\mathbf{K}(s)\big)^j\Big)$$

$$= \lim_{h\to 0}\frac{1}{h}\int_s^{s+h} K(t)dt \cdot \sum_{i=0}^{j-1}\Big(\mathbf{K}(s+h)\Big)^i\Big(\mathbf{K}(s)\Big)^{j-i-1}$$

$$= K(s)\cdot j\cdot\Big(\mathbf{K}(s)\Big)^{j-1}.$$

Therefore the exponential series $\mathbf{F}(s) := \sum_{j=0}^{\infty}\frac{1}{j!}(\mathbf{K}(s))^j$ produces the derivative $\mathbf{F}'(s) := K(s)\cdot\sum_{j=1}^{\infty}\frac{1}{(j-1)!}(\mathbf{K}(s))^{j-1} = K(s)\mathbf{F}(s)$ and, therefore, satisfies the same ODE $F' = KF$ as the Frenet frame. Both solutions coincide since they have the same initial condition. Just as $x^0 = 1$ holds, so we have here $(\mathbf{K}(s))^0 = \begin{pmatrix}1&0\\0&1\end{pmatrix}$ and consequently $\mathbf{F}(0) = \begin{pmatrix}1&0\\0&1\end{pmatrix}$. Hence the matrix-valued function $\mathbf{F}(s)$ is nothing but the Frenet frame of the corresponding curve with $e_1(0) = (1\ 0)$ and $e_2(0) = (0\ 1)$ and with the given curvature function $\kappa(s)$.

9. In cartesian coordinates we have $c(\varphi) = (r(\varphi)\cos\varphi, r(\varphi)\sin\varphi)$ with $c' = (r'\cos\varphi - r\sin\varphi, r'\sin\varphi + r\cos\varphi)$ and $c'' = (r''\cos\varphi - 2r'\sin\varphi - r\cos\varphi, r''\sin\varphi + 2r'\cos\varphi - r\sin\varphi)$. Here r' and c' denote the derivative with respect to φ, although the parameter is not the arc length. By the formula in Exercise 1 we obtain $\kappa = \mathrm{Det}(c', c'')/||c'||^3$. On the other hand we have $||c'||^2 = r'^2 + r^2$ and $\mathrm{Det}(c', c'') = 2r'^2 - rr'' + r^2$ which implies the assertion.

10. By inserting the function $r(\varphi) = a\varphi$ into the result of Exercise 9, one directly obtains $\kappa = (2a^2 + a^2\varphi^2)/(a^2 + a^2\varphi^2)^{3/2} = (2+\varphi^2)/|a|(1+\varphi^2)^{3/2}$.

11. In cartesian coordinates this curve is given by $c(t) = \big(e^t\cos(at), e^t\sin(at)\big)$. From the tangent vector $\dot{c}(t) = \big(e^t(\cos(at) - a\sin(at)), e^t(\sin(at) + a\cos(at))\big)$ we calculate the length L in the interval $(-\infty, t)$ as $L = \int_{-\infty}^{t}\sqrt{1+a^2}e^\tau d\tau = \sqrt{1+a^2}e^t = \sqrt{1+a^2}r(t)$. With $\langle c, \dot{c}\rangle = e^{2t}$ the angle ϑ between c and $\dot{c}$ becomes the constant $\cos\vartheta = e^{2t}/(||c(t)||\ ||\dot{c}(t)||) = 1/\sqrt{1+a^2}$.

14. Let c be parametrized by arc length s. We denote the osculating cubic parabola by $\gamma(s)$. However, s is not the arc length parameter for γ. Nevertheless, for $s = 0$ both curves have the same Frenet 3-frame $e_1(0), e_2(0), e_3(0)$. One calculates the curvature κ_γ and the torsion τ_γ of γ at the point $s = 0$ by the formulas in Exercise 1:
$$\kappa_\gamma(0) = \frac{||\dot\gamma \times \ddot\gamma||}{||\dot\gamma||^3} = \frac{||c' \times c''||}{||c'||^3} = \frac{||e_1 \times \kappa(0)e_2||}{||e_1||^3} = \kappa(0) \quad \text{and}$$
$$\tau_\gamma(0) = \frac{\mathrm{Det}(\dot\gamma, \ddot\gamma, \dddot\gamma)}{||\dot\gamma \times \ddot\gamma||^2} = \frac{\mathrm{Det}(e_1, \kappa(0)e_2, \kappa(0)\tau(0)e_3)}{||e_1 \times \kappa(0)e_2||^2} = \tau(0).$$
Furthermore one has $\frac{d\kappa_\gamma}{ds}\big|_{s=0} = \frac{d}{ds}\big|_{s=0} \frac{||\dot\gamma\times\ddot\gamma||}{||\dot\gamma||^3} = 0$. This implies that at $s = 0$ the derivative with respect to the arc length parameter (which we do not have to calculate) is $\gamma' = e_1(0) = c', \gamma'' = \kappa(0)e_1(0) = c''$ and, finally, $\gamma''' = \kappa(0)(-\kappa(0)e_1(0) + \tau(0)e_3(0)) = c'''$, compare 2.9. The last equation $\gamma''' = c'''$ holds only if $\kappa'(0) = 0$ holds for c. Then we have contact of third order at that point.

15. We consider the curve
$$c(s) = (\cos\varphi(s)\cos\vartheta(s), \sin\varphi(s)\cos\vartheta(s), \sin\vartheta(s))$$
with the arc length parameter s. We calculate the quantity $J = \mathrm{Det}(c, c', c'')$ from 2.10 (iii) at the point $s = 0$ with $\vartheta(0) = 0$, $\vartheta'(0) = 0$ and $\varphi'(0) = 1$ as follows:
$$J(0) = \mathrm{Det}\begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\varphi'(0)\sin\varphi & \varphi'(0)\cos\varphi & \vartheta'(0) \\ * & * & \vartheta''(0) \end{pmatrix}$$
$$= \mathrm{Det}\begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ * & * & \vartheta''(0) \end{pmatrix}.$$
The equation $J(0) = \vartheta''(0)$ follows since the quantities denoted by $*$ are not necessary for the calculation of the determinant. The equation $\kappa(0) = \sqrt{1 + (\vartheta''(0))^2}$ follows.

16. We start with a slope line satisfying $\tau = A\kappa$ with a constant A. Then we investigate solutions of the ODE in 2.10 (ii) which now appears as $(\kappa'/A\kappa^3))' = A$. Hence $\kappa'/A\kappa^3$ is a linear function $As + B$. On the other hand κ'/κ^3 is the derivative of $-\frac{1}{2}\kappa^{-2}$.

Therefore κ^{-2} is a quadratic function $-A^2s^2 - 2ABs + C$. By renaming B, one obtains the assertion.

18. One easily obtains $D \times e_1 = \kappa e_2, D \times e_2 = \tau e_3 - \kappa e_1, D \times e_3 = -\tau e_2$. This implies

$$\begin{aligned} e_1' = D \times e_1 &\iff e_1' = \kappa e_2, \\ e_2' = D \times e_2 &\iff e_2' = -\kappa e_1 + \tau e_3, \\ e_3' = D \times e_3 &\iff e_3' = -\tau e_2. \end{aligned}$$

19. The first assertion follows directly from the result of Exercise 18 by $\langle D, e_i' \rangle = \langle D, D \times e_i \rangle = \mathrm{Det}(D, D, e_i) = 0$. This can be interpreted by saying that D is in the kernel of the Frenet matrix. For the normal form one calculates the characteristic polynomial of the Frenet matrix

$$P(\lambda) = \mathrm{Det} \begin{pmatrix} -\lambda & \kappa & 0 \\ -\kappa & -\lambda & \tau \\ 0 & -\tau & -\lambda \end{pmatrix} = -\lambda\big(\lambda^2 + (\kappa^2 + \tau^2)\big)$$

with the complex zeros $\lambda = 0$ and $\lambda = \pm i\sqrt{\kappa^2 + \tau^2}$. One eigenvector for the eigenvalue $\lambda = 0$ is the Darboux vector D itself. The other eigenvectors are complex. The complex normal form is a diagonal matrix, the real normal is the following:

$$\begin{pmatrix} 0 & \sqrt{\kappa^2 + \tau^2} & 0 \\ -\sqrt{\kappa^2 + \tau^2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

20. D is constant if and only if $0 = D' = \tau' e_1 + \tau e_1' + \kappa' e_3 + \kappa e_3' = \tau' e_1 + \kappa' e_3$. This holds if and only if κ and τ are constant. This in turn characterizes the circular helices, cf. 2.12. $D/||D||$ is constant if and only if D and D' are linearly dependent. By the equation above for D' this means that $\tau'/\tau = \kappa'/\kappa$ holds or, equivalently, $(\log \tau)' = (\log \kappa)'$. This in turn is equivalent to the constancy of τ/κ, which characterizes the slope lines by 2.11.

23. In the proof of 2.15 we have seen

$$c^{(i)} = (\text{linear combination of } \ e_1, \ldots, e_{i-1}) + \kappa_1 \kappa_2 \cdots \kappa_{i-1} e_i.$$

This implies $\mathrm{Det}(c', c'', c''', \ldots, c^{(n)})$

$$= \mathrm{Det}(e_1, \kappa_1 e_2, \kappa_1\kappa_2 e_3, \ldots, \kappa_1\kappa_2\cdots\kappa_{n-1}e_n)$$

$$= \kappa_1^{n-1}\kappa_2^{n-2}\kappa_3^{n-3}\cdots\kappa_{n-1}^2\kappa_{n-1} = \Pi_{i=1}^{n-1}\kappa_i^{n-i}.$$

26. We can join the curves $c_1(t)$ and $c_2(t)$ by a family of closed curves $c_\alpha(t) = \alpha c_1(t) + (1-\alpha)c_2(t)$ with $0 \le \alpha \le 1$. By assumption $c_\alpha(t)$ never meets the origin, hence the winding number of every c_α is well defined. This winding number changes continuously with the parameter α and is, therefore, constant.

28. For a space curve the Frenet matrix $K(s)$ does in general *not* satisfy the commutativity $K(s_1)K(s_2) = K(s_2)K(s_1)$, but in the special case of slope lines with $\tau(s) = c\cdot\kappa(s)$ this holds because $\kappa(s_1)\tau(s_2) = c\kappa(s_1)\kappa(s_2) = \kappa(s_2)\tau(s_1)$. Consequently we can argue as in the solution to Exercise 8 with the analogous result. As in 2.16 we can express the Frenet matrix by the exponential series with the term $\mathbf{K}(s) = \int_0^s K(t)dt$ instead of sK. In the case of a constant matrix K we have $\mathbf{K}(s) = sK$.

Chapter 3.

2. The second fundamental form of f can be calculated from the quantities f_{uu}, f_{uv}, f_{vv}. From $f_{vv} = 0$ one obtains $\mathrm{Det}(II) = -\langle \nu_u, N\rangle^2$, where $N = f_u \times f_v/||f_u \times f_v||$ denotes the unit normal of the ruled surface f, whereas ν is the unit normal of the given surface. Therefore one has $K = 0 \Leftrightarrow \mathrm{Det}(II) = 0 \Leftrightarrow \langle \nu_u, N\rangle = 0 \Leftrightarrow \langle \nu_u, f_u \times f_v\rangle = 0 \Leftrightarrow \langle \nu_u, (c' + v\nu_u)\times\nu\rangle = 0 \Leftrightarrow \mathrm{Det}(\nu_u, c', \nu) = 0$. On the other hand ν is perpendicular to both ν_u and c'. This means that the last equality is satisfied if and only if ν_u and c' are linearly dependent. This in turn means that $c'(u)$ is an eigenvector of the Weingarten map L_u. Therefore $K = 0$ holds for all u if and only if c is a curvature line.

3. First of all, the matrix in this system of equations is necessarily skew-symmetric by $\langle E_i, E_j\rangle' = 0$ (as in 2.13). Therefore we have to determine only $\langle E_1', E_2\rangle$, $\langle E_1', E_3\rangle$ and $\langle E_2', E_3\rangle$. The first expression $\langle E_1', E_2\rangle$ is just the part of $E_1' = c''$ that is tangential to the given surface. This is nothing but the geodesic curvature κ_g of the curve c, cf. 3.11 or 4.37. Furthermore $\langle E_1', E_3\rangle$ is the

normal part of c'' which is nothing but the normal curvature κ_ν. The remaining coefficient $\langle E_2', E_3\rangle$ can be interpreted as a kind of torsion, by analogy with the Frenet equations in 2.8. This quantity indicates the change of the (E_1, E_2)-plane when passing through the curve.

5. We can assume a surface element of the form $f(r, \varphi) = (r\cos\varphi, r\sin\varphi, h(r))$, with the arc element $ds^2 = (1+\dot{h}^2)dr^2 + r^2 d\varphi^2$. By introducing a new variable $t = t(r)$ we obtain $dt^2 = \dot{t}^2 dr^2$ and $ds^2 = \frac{1+\dot{h}^2}{\dot{t}^2} dt^2 + r^2 d\varphi^2$. These are isothermal parameters if and only if $1 + \dot{h}^2 = \dot{t}^2 r^2$ or, equivalently, $t(r) = \int_{r_0}^{r} \frac{1}{\rho}\sqrt{1 + \dot{h}^2(\rho)}\, d\rho$. Then there exists an inverse function $r = r(t)$, and the arc element is $ds^2 = (r(t))^2(dt^2 + d\varphi^2)$.

7. The assumption $K < 0$ implies the following: Through every point there are two distinct asymptotic lines with two tangents (so-called *asymptotic directions*) X and Y such that $\langle X, X\rangle = \langle Y, Y\rangle = 1$ and $II(X, X) = II(Y, Y) = 0$. The relation $II(X, Y) \neq 0$ follows because otherwise we would have $II = 0$. We now calculate the trace of the second fundamental form with respect to the first fundamental form in the orthonormal basis X, $(Y - \langle X, Y\rangle X)/||Y - \langle X, Y\rangle X||$ (compare the Gram-Schmidt procedure in 2.4). This trace is given by the expression

$$II(X, X) + II(Y - \langle X, Y\rangle X, Y - \langle X, Y\rangle X)/||Y - \langle X, Y\rangle X||^2$$
$$= -2\langle X, Y\rangle II(Y, X)/||\cdots||^2$$

which vanishes if and only if $X \perp Y$.

9. First of all we have $||f(u, \varphi||^2 = \frac{1}{\cosh^2 u}(1 + \sinh^2 u) = 1$, hence this is really parametrization of a part of the sphere. By differentiation we obtain $f_u = -\frac{\sinh u}{\cosh^2 u}(\cos\varphi, \sin\varphi, \sinh u) + (0, 0, 1)$ and $f_\varphi = \frac{1}{\cosh u}(-\sin\varphi, \cos\varphi, 0)$. This implies $\langle f_u, f_\varphi\rangle = 0$ and $\langle f_u, f_u\rangle = \frac{\sinh^2 u}{\cosh^4 u}(1+\sinh^2 u) - 2\frac{\sinh^2 u}{\cosh^2 u} + 1 = (1-2)\frac{\sinh^2 u}{\cosh^2 u} + 1 = \frac{1}{\cosh^2 u} = \langle f_\varphi, f_\varphi\rangle$. Thus f is angle preserving, i.e., conformal.

11. For the one-sheeted hyperboloid we set $c(u) = (\cos u, \sin u, 0)$, $X_1(u) = (\sin u, -\cos u, 1)$ and $X_2(u) = (-\sin u, \cos u, 1)$. Obviously X_1, X_2 are always linearly independent. Then each point (x, y, z) on one of the ruled surfaces $f_1(u, v) = c(u) + vX_1(u)$

and $f_2(u,v) = c(u) + vX_2(u)$ satisfies $x^2 + y^2 - z^2 = (\cos u \pm v \sin u)^2 + (\sin u \mp v \cos u)^2 - v^2 = \cos^2 u + \sin^2 u = 1$. Hence the images of f_1 and f_2 are both contained in the one-sheeted hyperboloid of revolution which, therefore, is a ruled surface in two different ways, i.e., with two linearly independent rulings. One of them is depicted in Figure 3.14, the other one is a mirror image. By $||X_i'|| = 1, \langle X_i', c'\rangle = 0$ we have standard parameters except for the wrong normalization $||X_i|| = \sqrt{2}$. Instead we have here $||c'|| = 1$. By passing from u to $\sqrt{2}u$ and by replacing $X_i(u)$ by $\frac{1}{\sqrt{2}}X_i(\sqrt{2}u)$, $i = 1, 2$, we obtain standard parameters. This implies $|F| = 1$, $J = 1$ and $|\lambda| = 1$ with $\lambda = -F$. The Gaussian curvature is $K = -1/(1 + v^2)^2$. It tends to zero for $v \to \pm\infty$.
For the hyperbolic paraboloid with the equation $x^2 - y^2 - 4z = 0$ we set $c(u) = (u, u, 0)$ and $X(u) = \frac{1}{\sqrt{u^2+2}}(1, -1, u)$. Then for each point $(x, y, z) = c(u) + vX(u)$ the equation $x^2 - y^2 - 4z = (x + y)(x - y) - 4z = 2u \cdot 2\frac{v}{\sqrt{u^2+2}} - \frac{4vu}{\sqrt{u^2+2}} = 0$ is satisfied. Furthermore we have $||X|| = 1$ and $\langle \dot{c}, X\rangle = \langle \dot{c}, \dot{X}\rangle = \langle \dot{c}, \ddot{X}\rangle = 0$ (without calculation, just consider the first and second component). As in the case of the one-sheeted hyperboloid above we have standard parameters except for the normalization. Thus we can proceed as follows: $X' = \dot{X}/||\dot{X}||$ and $c' = \dot{c}/||\dot{X}||$ together imply $\langle c', X'\rangle = 0$. The equation $\langle c', X''\rangle = 0 = \langle X', X''\rangle$ follows. This implies $J = \mathrm{Det}(X, X', X'') = \mathrm{Det}(X, \dot{X}/||\dot{X}||, \ddot{X}/||\dot{X}||) = 0$ because obviously $X, \dot{X}, \ddot{X}$ are contained in the $(x, -x, z)$-plane and are, therefore, linearly dependent. We have $F = \langle c', X\rangle = 0$. In view of 3.23 λ is certainly not constant. Indeed one obtains $c' = \dot{c}/||\dot{X}||$ and $\dot{X} = (u^2 + 2)^{-3/2}(-u, u, 2)$, $||\dot{X}|| = \sqrt{2}/(u^2 + 2)$ and, finally, $|\lambda| = |\mathrm{Det}(c', X, X')| = ||\dot{c}||/||\dot{X}|| = u^2 + 2$. The Gaussian curvature is $K = -(u^2 + 2)^2/\big((u^2 + 2)^2 + v^2\big)^2$. It tends to zero for $u \to \pm\infty$ and $v \to \pm\infty$ independently. By the equation $J = 0$ the spherical curve X is a part of a great circle. Indeed $X(u)$ lies in the plane $x + y = 0$ and the sphere $x^2 + y^2 + z^2 = 1$. As the one-sheeted hyperboloid, the hyperbolic paraboloid carries two distinct rulings. However, here we used one of the lines as the directrix c.

13. Let u be the arc length parameter of the curve c. We calculate $f_{uv} = D'$ and $f_u \times f_v = (c' + vD') \times D$. This implies

$$\begin{aligned}\mathrm{Det}(f_{uv}, f_u, f_v) &= \mathrm{Det}(D', c', D)\\ &= \mathrm{Det}(\tau' e_1 + \tau e_1' + \kappa' e_3 + \kappa e_3', e_1, \tau e_1 + \kappa e_3)\\ &= \mathrm{Det}(\tau\kappa e_2 - \kappa\tau e_2, e_1, \kappa e_3) = 0\end{aligned}$$

and, therefore, $\mathrm{Det}II = 0$. Consequently, the ruled surface is developable. For $v = 0$ one has $f_u = c' = e_1$ and $f_v = D$. This implies $\langle f_u, e_2\rangle = 0$ and $\langle f_v, e_2\rangle = \langle \tau e_1 + \kappa e_3, e_2\rangle = 0$. Therefore for $v = 0$ the tangent plane is perpendicular to e_2 and coincides with the rectifying plane.

15. We describe the plane curve of this surface of rotation by $c(z) = (r(z), z)$. The axis of rotation is the (vertical) z-axis. In the formulas of 3.16 we have $r(z)$ and $h(z) = z$. This leads to the principal curvatures $\kappa_1 = -r''/(r'^2 + 1)^{3/2}$ and $\kappa_2 = \frac{1}{r}/(r'^2 + 1)^{1/2}$. Therefore $-\kappa_1 = \kappa_2$ holds if and only if $r'' = (r'^2 + 1)\frac{1}{r}$ or $r''r - r'^2 = 1$. For the initial condition $r(0) = a > 0$ the unique solution is $r(z) = a\cosh\left(\frac{z}{a}\right)$. This is the profile curve of the catenary (cf. 2.39). Thus the surface is the catenoid.

16. The first fundamental form is $g_{11} = \langle f_u, f_u\rangle = b^2, g_{12} = \langle f_u, f_v\rangle = 0, g_{22} = \langle f_v, f_v\rangle = (a + b\cos u)^2$ with the determinant $g = g_{11}g_{22} - g_{12}^2 = b^2(a + b\cos u)^2$. One of the unit normals is $\nu = (-\cos u\cos v, -\cos u\sin v, -\sin u)$ with the second fundamental form $h_{11} = -\langle \nu_u, f_u\rangle = b, h_{12} = 0, h_{22} = (a + b\cos u)\cos u$. Consequently the principal curvatures are $\kappa_1 = h_{11}/g_{11} = 1/b$ and $\kappa_2 = h_{22}/g_{22} = \cos u/(a + b\cos u)$, and the mean curvature is $H = \frac{1}{2}(\kappa_1 + \kappa_2) = (a + 2b\cos u)/2b(a + b\cos u)$. Therefore the integral $\int H^2 dA$ equals

$$\begin{aligned}\int \frac{(a + 2b\cos u)^2}{4b^2(a + b\cos u)^2}\sqrt{g}dudv &= \frac{1}{4b}\int_0^{2\pi}\int_0^{2\pi}\frac{(a + 2b\cos u)^2}{a + b\cos u}dudv\\ = \frac{\pi}{2b}\int_0^{2\pi}\Big(\frac{a^2}{a + b\cos u} + 4b\cos u\Big)du &= \frac{\pi a^2}{2b}\int_0^{2\pi}\frac{du}{a + b\cos u}\\ &= \frac{\pi^2 a^2}{b\sqrt{a^2 - b^2}} = \frac{\pi^2}{x\sqrt{1 - x^2}}\end{aligned}$$

with $x = b/a$ where $0 < x < 1$. The optimal ratio x of b and a is attained at the maximum of the function $\Phi(x) = x\sqrt{1-x^2}$, in particular at a point with $\Phi'(x) = 0$. From the equation $\Phi'(x) = (1-2x^2)/\sqrt{1-x^2}$ one sees that $\Phi' = 0$ holds if and only if $x^2 = 1/2$, hence $a = \sqrt{2}b$. In this case the maximum value of $x\sqrt{1-x^2}$ is $1/2$, hence $\int H^2 dA \geq 2\pi^2$.

17. If the given surface consists only of umbilics, then by 3.14 it consists of pieces of a sphere or a plane. Obviously the same holds for the parallel surface, and all curvatures are constant. The assertion is easily verified. Now we assume that in a neighborhood of a non-umbilic we have curvature line parameters u_1, u_2 with $-\frac{\partial \nu}{\partial u_i} = L(\frac{\partial f}{\partial u_i}) = \kappa_i \frac{\partial f}{\partial u_i}$. By $\frac{\partial f_\varepsilon}{\partial u_i} = \frac{\partial f}{\partial u_i} + \varepsilon \frac{\partial \nu}{\partial u_i} = (1-\varepsilon\kappa_i)\frac{\partial f}{\partial u_i}$ the unit normals of f and f_ε coincide up to translation. On the other hand $\frac{\partial f_\epsilon}{\partial u_i}, \frac{\partial f}{\partial u_i}$ are linearly dependent. Therefore one has

$$L^{(\varepsilon)}(\frac{\partial f}{\partial u_i}) = -\frac{\partial \nu}{\partial u_i} = \kappa_i \frac{\partial f}{\partial u_i} = \kappa_i^{(\varepsilon)} \frac{\partial f_\varepsilon}{\partial u_i} = \kappa_i^{(\varepsilon)}(1-\varepsilon\kappa_i)\frac{\partial f}{\partial u_i}$$

and consequently $\kappa_i = (1-\varepsilon\kappa_i)\kappa_i^{(\varepsilon)}$, as claimed in (a).
In order to prove (b) we calculate with a constant $\varepsilon = \frac{1}{2H} = 1/(\kappa_1+\kappa_2)$

$$K^{(\varepsilon)} = \kappa_1^{(\varepsilon)}\kappa_2^{(\varepsilon)} = \frac{\kappa_1\kappa_2}{(1-\varepsilon\kappa_1)(1-\varepsilon\kappa_2)} = \frac{\kappa_1\kappa_2}{(1-\frac{\kappa_1}{\kappa_1+\kappa_2})(1-\frac{\kappa_2}{\kappa_1+\kappa_2})}$$

$$= \frac{(\kappa_1+\kappa_2)\kappa_1\kappa_2}{\kappa_2\kappa_1} = 2H.$$

21. The curve $c(r) = (\cosh r, \sinh r, 0)$ lies entirely in H^2. The parameter is the arc length since $c' = (\sinh r, \cosh r, 0)$ and $\langle c', c'\rangle_1 = -\sinh^2 r + \cosh^2 r = 1$. By the rotational symmetry around the point $(1,0,0)$ the same holds for for any rotated curve in any direction. Therefore in polar coordinates one has $f(r,\varphi) = (\cosh r, \sinh r\cos\varphi, \sinh r \sin\varphi)$. For any fixed $r > 0$ one has the circle $k(\varphi) = (\cosh r, \sinh r\cos\varphi, \sinh r\sin\varphi)$, and all these circles are perpendicular to the radial r-curves which are hyperbolas in euclidean geometry.
From $k' = (0, -\sinh r\sin\varphi, \sinh r\cos\varphi)$ we calculate the length $L(r)$ of the circle as $L(r) = \int_0^{2\pi} \sinh r\; d\varphi = 2\pi\sinh r$. Thus the first fundamental form in polar coordinates is given by

$ds^2 = dr^2 + \sinh^2 r\ d\varphi^2$. Of course, this can be calculated also from the parametrization with

$$f_r = (\sinh r, \cosh r \cos\varphi, \cosh r \sin\varphi),\ \langle f_r, f_r\rangle_1 = 1,$$
$$f_\varphi = (0, -\sinh r \sin\varphi, \sinh r \cos\varphi),$$
$$\langle f_r, f_\varphi\rangle_1 = 0 \text{ and } \langle f_\varphi, f_\varphi\rangle_1 = \sinh^2 r.$$

Compare the analogous situation for the polar coordinates of the unit sphere in $\mathbb{R}^3$: $f(r,\varphi) = (\cos r, \sin r \cos\varphi, \sin r \sin\varphi)$ and $ds^2 = dr^2 + \sin^2 r\ d\varphi^2$.

22. The ruling is obvious since the parameter v occurs only in linear form. The equation $H = 0$ is not hard to see. The first fundamental form of each f_i is a diagonal matrix, i.e., $g_{12} = 0$; whereas the second fundamental form is of the opposite type, $h_{11} = h_{22} = 0$. By taking the trace, the equation $H = 0$ follows. In the exceptional case f_4 these quantities are the following:
$$(f_4)_u = \big(a(u^2+1)+v, a(u^2-1)+v, 2au\big),\ (f_4)_v = (u,u,1),$$
$$(f_4)_{uu} = (2au, 2au, 2a),\ (f_4)_{vv} = 0.$$
One (non-normalized) normal is $N = \big(a(u^2+1)-v, a(u^2-1)-v, 2au\big)$. This implies $\langle (f_4)_{uu}, N\rangle_1 = -2a^2u - 2a^2u + 4a^2u = 0$. Notice that the directrix is a *null cubic curve*: $\langle (f_4)_u, (f_4)_u\rangle_1 = 0$ for $v = 0$. Similarly N is isotropic along the directrix: $\langle N, N\rangle_1 = 0$ for $v = 0$. In fact we have $(f_4)_u = N$ for $v = 0$.

23. The following descriptions are obvious:
$$f_1(u,v) = v(0\ \ 1\ \ 0)\cdot\begin{pmatrix}1 & 0 & 0\\ 0 & \cos u & \sin u\\ 0 & -\sin u & \cos u\end{pmatrix} + (au\ \ 0\ \ 0),$$
$$f_2(u,v) = v(0\ \ 1\ \ 0)\cdot\begin{pmatrix}\cosh u & \sinh u & 0\\ \sinh u & \cosh u & 0\\ 0 & 0 & 1\end{pmatrix} + (0\ \ 0\ \ au),$$
$$f_3(u,v) = v(1\ \ 0\ \ 0)\cdot\begin{pmatrix}\cosh u & \sinh u & 0\\ \sinh u & \cosh u & 0\\ 0 & 0 & 1\end{pmatrix} + (0\ \ 0\ \ au),$$
$$f_4(u,v) = v(0\ \ 0\ \ 1)\cdot\begin{pmatrix}1+\frac{u^2}{2} & \frac{u^2}{2} & u\\ -\frac{u^2}{2} & 1-\frac{u^2}{2} & -u\\ u & u & 1\end{pmatrix} + a\Big(\tfrac{u^3}{3}+u\ \ \tfrac{u^3}{3}-u\ \ u^2\Big).$$

One recognizes the three types of rotation matrices from 3.42 in transposed form, because f_1, f_2, f_3, f_4 are written as row vectors. The items f_2 and f_3 differ by the type of the ruling (space-like or time-like). In the first three cases one clearly recognizes the corresponding 1-parameter groups of helicoidal motions (cf. 2.3), in the last case one speaks of *cubic screw-motions.*

24. For a ruled surface $f(u,v) = c(u) + vX(u)$ with $\langle X, X\rangle = 0$ one calculates the first and second fundamental form

$$I = \begin{pmatrix} \langle c' + vX', c' + vX'\rangle_1 & \langle c', X\rangle_1 \\ \langle c', X\rangle_1 & 0 \end{pmatrix}$$

$$II = \begin{pmatrix} \langle c'' + vX'', \nu\rangle_1 & \langle X', \nu\rangle_1 \\ \langle X', \nu\rangle_1 & 0 \end{pmatrix}$$

with $K = \mathrm{Det}II/\mathrm{Det}I = \langle X', \nu\rangle_1^2/\langle c', X\rangle_1^2$ and

$$2H = \frac{1}{-\langle c', X\rangle_1^2}\Big(0 - 2\langle c', X\rangle_1\langle X', \nu\rangle_1 + 0\Big) = 2\langle X', \nu\rangle_1/\langle c', X\rangle_1$$

by the same formula as in 3.13. The assertion $H^2 = K$ follows.

Chapter 4.

1. The geodesics of a developable surface are preserved by the development map into the plane because they depend only on the first fundamental form, which is preserved. Consequently every geodesic is mapped into a straight line in this case. To see the geodesics, one only has to consider the inverse map that "redevelops" the plane back into the surface.

2. If we use the result of Exercise 5, then it is sufficient to see that the principal normal of a great circle is the negative of the position vector and hence one of the two unit normals. Furthermore there is exactly one great circle through any given point in any given direction.
Without using Exercise 5 one can verify the ODE of a geodesic in 4.12 in spherical coordinates. Without loss of generality we can consider the equator. There we have the parametrization $f(u,v) = (\cos u \cos v, \sin u \cos v, \sin v)$ for the surface and $c(u) = (\cos u, \sin u, 0)$ for the equator, consequently the coordinates of the curve are $u^1 = u$ and $u^2 = 0$. Therefore all $\ddot{u}^i$

vanish identically, and in the sum there is only one possibly non-vanishing term $\dot{u}^1\dot{u}^1\Gamma_{11}^k$. Hence the equation of a geodesic is satisfied if $\Gamma_{11,1}g^{11} = \Gamma_{11}^1 = 0$ and $\Gamma_{11,2}g^{22} = \Gamma_{11}^2 = 0$. This in turn follows directly from $g_{11} = \cos^2 v$ and $g_{12} = 0$.

4. The remark after the definition of the geodesic curvature in 4.37 tells us that $\nabla_{e_1} e_1 = \kappa_g e_2, \nabla_{e_1} e_2 = -\kappa_g e_1$ holds by analogy with the Frenet equations for plane curves. For a given initial condition $c(0)$ and $c'(0) = e_1$ the 2-frame e_1, e_2 is uniquely determined as a solution of an ODE. In coordinates one has to apply the formula for ∇ in 4.6. Then the curve is determined by $c(s) = \int e_1(s)ds$.

5. If c denotes the curve, parametrized by arc length, then the geodesic curvature κ_g vanishes if and only if the normal part of the vector c'' tangential to the surface vanishes, cf. 4.37. This holds if and only if c'' and the unit normal ν are linearly dependent. On the other hand c'' always points into the direction of the principal normal of the curve.

11. In the coordinates x, y any ray in the y-direction with a constant x is a geodesic because its unit tangent vector $y\frac{\partial}{\partial y}$ is parallel:
$$\nabla_{\frac{\partial}{\partial y}}\left(y\frac{\partial}{\partial y}\right) = \frac{\partial}{\partial y} + y \cdot \nabla_{\frac{\partial}{\partial y}}\frac{\partial}{\partial y} = \frac{\partial}{\partial y} + y\Big(\Gamma_{22}^1\frac{\partial}{\partial x} + \Gamma_{22}^2\frac{\partial}{\partial y}\Big) = 0.$$
The last equality holds by $\Gamma_{22}^1 = 0$ and $\Gamma_{22}^2 = -1/y$.
In ordinary polar coordinates $x = r\cos\varphi, y = r\sin\varphi$ the arc length element reads are $ds^2 = \frac{1}{r^2\sin^2\varphi}(dr^2 + r^2 d\varphi^2)$. Any half circle with center at the origin and with a constant radius r is a geodesic because similarly its unit tangent vector $\sin\varphi\frac{\partial}{\partial\varphi}$ is parallel:
$$\begin{aligned}\nabla_{\frac{\partial}{\partial\varphi}}\left(\sin\varphi\frac{\partial}{\partial\varphi}\right) &= \cos\varphi\frac{\partial}{\partial\varphi} + \sin\varphi\nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi}\\ &= \cos\varphi\frac{\partial}{\partial\varphi} + \sin\varphi\Big(\Gamma_{22}^1\frac{\partial}{\partial r} + \Gamma_{22}^2\frac{\partial}{\partial\varphi}\Big) = 0.\end{aligned}$$
The last equality holds by $\Gamma_{22}^1 = 0$ and $\Gamma_{22}^2 = -\cos\varphi/\sin\varphi$. Finally we observe that the horizontal translations $(x, y) \mapsto (x+x_0, y)$ are isometries since (g_{ij}) does not depend on x. Therefore all other half circles with a center on the x-axis are geodesics as well.

12. We evaluate the formula in 4.26 (ii) for the function $\lambda = 1/y^2$:

$$\begin{aligned} K &= -\tfrac{y^2}{2}\Delta\big(\log\tfrac{1}{y^2}\big) = -\tfrac{y^2}{2}\tfrac{\partial^2}{\partial y^2}(-2\log y) \\ &= y^2\tfrac{\partial}{\partial y}\tfrac{1}{y} = y^2\big(-\tfrac{1}{y^2}\big) = -1. \end{aligned}$$

13. Because these transformations are given as complex analytic functions, it is convenient to write the arc length element ds^2 in complex notation. We use the notations $z\bar{z} = (x+iy)(x-iy) = x^2+y^2$ and $dz = dx + idy$, $d\bar{z} = dx - idy$. This implies $dx^2 + dy^2 = dzd\bar{z}$ and consequently $ds^2 = \frac{1}{y^2}dzd\bar{z}$. Now let $w = (az+b)/(cz+d)$ with $w = \xi + i\eta$ and consider w (or ξ and η, respectively) as a new parametrization of the Poincaré upper half-plane. One has $w = (az+b)(c\bar{z}+d)/(cz+d)(c\bar{z}+d)$ $= (acz\bar{z} + bd + adz + bc\bar{z})/(cz+d)(c\bar{z}+d)$, which implies $\eta = \frac{ad-bc}{(cz+d)(c\bar{z}+d)}y$ by $\mathrm{Im}(adz + bc\bar{z}) = (ad-bc)y$. If $y > 0$ and $ad - bc > 0$ hold, then also $\eta > 0$. Notice that in the case $ad - bc < 0$ the upper and lower half-plane would be interchanged. Therefore the transformation $z \mapsto w$ is really a transformation of the Poincaré upper half-plane in itself. It is bijective because there is an inverse map: For given w the equation $w(cz+d) = az+b$ can be resolved by $z = (b-dw)/(cw-a)$. We now calculate the arc length element by the complex derivative

$$\frac{dw}{dz} = \frac{d}{dz}\frac{az+b}{cz+d} = \frac{(cz+d)a-(az+b)c}{(cz+d)^2} = \frac{ad-bc}{(cz+d)^2}.$$

By analogy one has

$$\frac{d\bar{w}}{d\bar{z}} = \frac{d}{d\bar{z}}\frac{a\bar{z}+b}{c\bar{z}+d} = \frac{(c\bar{z}+d)a-(a\bar{z}+b)c}{(c\bar{z}+d)^2} = \frac{ad-bc}{(c\bar{z}+d)^2}.$$

Finally one obtains

$$dwd\bar{w} = dzd\bar{z}\frac{(ad-bc)^2}{(cz+d)^2(c\bar{z}+d)^2} = dzd\bar{z}\frac{\eta^2}{y^2},$$

and for the arc length element one gets the same expression $\frac{1}{y^2}dzd\bar{z} = \frac{1}{\eta^2}dwd\bar{w}$ on either side. This means that the transformation $z \mapsto w$ is an isometry.

Additional remark: Notice that one can map any given point z by such a transformation into any other point w and that the matrices of the type

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

fix the point $z = w = i$. Therefore these matrices define (isometric) rotations in the Poincaré upper half-plane around this point. Multiplying a, b, c, d by the same constant does not change the transformation. Consequently the group of all (proper) hyperbolic motions is isomorphic to the 3-dimensional group $\mathbf{SL}(2, \mathbb{R})$.

16. Because all entries g_{ij} are constant, all Christoffel symbols vanish and, consequently, the left hand side of the Gauss equation vanishes. By $\mathrm{Det}(h_{ij}) = 0$ the Gauss equation 4.15 (i) is satisfied. In contrast, the Codazzi-Mainardi equation 4.15 (ii) is not satisfied for $i = j = 2$ and $k = 1$: One has $(h_{22})_u - (h_{21})_v = 1 \neq 0$. Therefore there is no such surface element.

17. Considering the Gaussian curvature K of such a (hypothetical) surface element, one has necessarily $K = \mathrm{Det}(h_{ij})/\mathrm{Det}(g_{ij}) = \tan^2 u$. On the other hand the first fundamental form is given in geodesic parallel coordinates, hence by 4.28 the equation $K = 1$ holds. Consequently there is no such surface element.

18. The torus of rotation is a compact submanifold M of $\mathbb{R}^3$, cf. the example after 3.1. With the formulas in Exercise 16 in Chapter 3 one obtains $K = \kappa_1\kappa_2 = \cos u/b(a + b\cos u)$ and $dA = b(a + b\cos u)dudv$. This leads to the integral

$$\int_M |K|dA = \int_0^{2\pi}\int_0^{2\pi} |\cos u|dudv = 2\pi \int_0^{2\pi} |\cos u|du$$
$$= 2\pi \cdot 4 \int_0^{\pi/2} \cos u du = 8\pi.$$

Therefore the torus of rotation is tight by 4.47 and $\chi(M) = 0$. One can also directly see that the positive part $\int_{M_+} KdA$ equals 4π, cf. 4.46 (ii).

20. For $\kappa_1 \neq 0, \kappa_2 \neq 0$ the equation $\alpha(\kappa_1 + \kappa_2) + \beta\kappa_1\kappa_2 = 2\alpha H + \beta K = 0$ is equivalent with $\alpha(\frac{1}{\kappa_1} + \frac{1}{\kappa_2}) = -\beta$. Assuming that there is a point with $\kappa_1 = 0, \kappa_2 \neq 0$ as an accumulation point of elliptic points with $\kappa_1, \kappa_2 > 0$, this last equation leads to a

contradiction because $\frac{1}{\kappa_1}$ tends to infinity. The same holds if there is such a point with $\kappa_1 = \kappa_2 = 0$ because $\frac{1}{\kappa_1}$ and $\frac{1}{\kappa_2}$ both are positive. This shows that a connected compact submanifold cannot have a point with $K = 0$. Therefore one has $K > 0$ everywhere. This enables us to apply Lemma 4.52 because in the equation above the sum $\frac{1}{\kappa_1} + \frac{1}{\kappa_2}$ is constant, and one principal curvature has a maximum at the same point where the other has a minimum.

22. The essential difference between the two formulas $K = -r''/r$ and $H = (r\sqrt{1-r'^2})'/(r^2)'$ is the following: In the first case this expression is invariant under isometries, in the second case it is not. Notice that the function r itself is not a quantity that is invariant under isometries. This can be seen from the surfaces of rotation with a constant Gaussian curvature in 3.17. Therefore an expression containing this function has *a priori* no invariant meaning. In other words: If we realize the same first fundamental form by distinct surfaces of rotation (i.e., with distinct functions $r(t)$), then we can obtain distinct mean curvatures. This is different for the formula for the Gaussian curvature because it can be calculated from the curvature tensor by the Theorema Egregium 4.16 and 4.20, and this is invariant under isometries. Then necessarily the same holds also for the expression $-r''/r$.

23. For part (a) the equivalence follows directly from the Gauss equation 4.15 (i) by $\sum_j g_{ij} g^{jk} = \delta_j^k$. For part (b) the equivalence follows from 4.18 (ii) in connection with the "product rule" $\nabla_X(LY) = (\nabla_X L)(Y) + L(\nabla_X Y)$.

Chapter 5.

2. For arbitrary charts $\varphi_1 \colon U_1 \to \mathbb{R}^k$ and $\varphi_2 \colon U_2 \to \mathbb{R}^l$ with $U_1 \subset M_1, U_2 \subset M_2$, we obtain a chart in the product $M_1 \times M_2$ by the cartesian product $\varphi_1 \times \varphi_2 \colon U_1 \times U_2 \to \mathbb{R}^k \times \mathbb{R}^l \cong \mathbb{R}^{k+l}$ of the two mappings. This notation means $(\varphi_1 \times \varphi_2)(p_1, p_2) = (\varphi_1(p_1), \varphi_2(p_2))$. The union of all such $U_1 \times U_2$ covers $M_1 \times M_2$ entirely, all $\varphi_1 \times \varphi_2$ are injective, and the coordinate transformations are calculated componentwise: $(\psi_1 \times \psi_2) \circ (\varphi_1 \times \varphi_2)^{-1} = (\psi_1 \circ \varphi_1^{-1}) \times (\psi_2 \circ \varphi_2^{-1})$.

3. As usual we denote the tangent bundle of a differentiable manifold M by TM. For an atlas consisting of charts $\varphi_i \colon M_i \to \mathbb{R}^n$ we do not only have $\bigcup_i M_i = M$ but also $\bigcup_i TM_i = TM$, because every tangent vector at a point p belongs to some TM_i whose associated M_i contains the point p. The associated bundle charts Φ_i are defined in the exercise. Obviously $\Phi_i(TM_i) = \varphi_i(M_i) \times \mathbb{R}^n$ is open in $\mathbb{R}^{2n}$, and all Φ_i are injective. For a coordinate transformation $\Phi_j \circ \Phi_i^{-1}$ we introduce the notation $\varphi_i(p) = (x^1(p), \ldots, x^n(p))$ and $\varphi_j(p) = (\tilde{x}^1(p), \ldots, \tilde{x}^n(p))$. This leads to

$$\begin{aligned}\Phi_j \circ \Phi_i^{-1}&\big(\varphi_i(p), \xi^1(p), \ldots, \xi^n(p)\big)\\ &= \Phi_j\Big(p, \sum_k \xi^k \frac{\partial}{\partial x^k}\big|_p\Big) = \big(\varphi_j(p), \eta^1(p), \ldots, \eta^n(p)\big)\end{aligned}$$

with $X = \sum_k \xi^k \frac{\partial}{\partial x^k}|_p) = \sum_l \eta^l \frac{\partial}{\partial \tilde{x}^l}|_p)$. From the basis transformation $\frac{\partial}{\partial x^k}|_p = \sum_l \frac{\partial \tilde{x}^l}{\partial x^k}\frac{\partial}{\partial \tilde{x}^l}|_p$ one obtains $\eta^l(p) = \sum_k \xi^k \frac{\partial \tilde{x}^l}{\partial x^k}|_p$. Consequently the quantities $\eta^1, \ldots, \eta^n$ depend differentiably on $\xi^1, \ldots, \xi^n$, and $\varphi_j \circ \varphi_i^{-1}$ is differentiable by assumption.

4. For the manifold $\mathbb{R}^n$ *one* chart is sufficient (the identical map), and every tangent vector X is identified with the n-tuple of its components which is a vector of $\mathbb{R}^n$. Hence one has $T\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$, in accordance with 1.6.

6. Obviously the tangent space of the product manifold $M_1 \times M_2$ at a point (p_1, p_2) is $T_{(p_1,p_2)}(M_1 \times M_2) = T_{p_1}M_1 \times T_{p_2}M_2 \cong T_{p_1}M_1 \oplus T_{p_2}M_2$. For two given Riemannian metrics g_1, g_2 on M_1, M_2, the product metric can be defined as follows: Every tangent vector X of $M_1 \times M_2$ admits a unique decomposition $X = X_1 + X_2$ with $X_i \in T_{p_i}M_i$. If we define $g(X,Y) = g_1(X_1, Y_1) + g_2(X_2, Y_2)$, then g has all properties that are required for a Riemannian metric. Assume that in local coordinates g_1 is given by $g_{ij}^{(1)}$ and g_2 by $g_{kl}^{(2)}$. Then the components g_{rs} of g have the form of the following block matrix:

$$(g_{rs}) = \begin{pmatrix} g_{ij}^{(1)} & 0 \\ 0 & g_{kl}^{(2)} \end{pmatrix}.$$

9. We consider a geodesic triangle Δ with corners A, B, C and associated exterior angles α, β, γ (counterclockwise). Then we observe the parallel transport of a tangent vector X of one of the three sides at a certain point (the choice of the point does not matter) along the three sides. At the beginning X has an angle $\varphi = 0$ against the first side. At each corner this angle increases by the corresponding exterior angle. After one complete run through the triangle (counterclockwise) the new vector Y has an angle $\varphi = \alpha + \beta + \gamma$ against the tangent vector X at the beginning. The sign is determined in such a way that one has to turn Y by φ (with sign) for getting the original position of X. Therefore the rotation from X to Y requires the complementary angle $2\pi - \alpha - \beta - \gamma$ (with sign). On the other hand by the Theorema Elegantissimum in 4.40 one has $2\pi - \alpha - \beta - \gamma = \int_\Delta K dA$. Therefore the rotation angle of the parallel transport is nothing but the total curvature of the triangle. In the euclidean case both quantities vanish, and in the case of positive curvature the angle is positive (counterclockwise rotation), in the case of negative curvature it is negative (clockwise rotation).

10. By the rotational symmetry of the sphere it is sufficient to determine the holonomy group of the north pole N. For a given angle φ let c_1 and c_2 be two great circles through N meeting at N under the angle φ. Furthermore let c_3 be the equatorial circle. Let P and Q be two intersection points of $c_1 \cap c_3$ and $c_2 \cap c_3$ such that φ equals the angle PNQ. Then the parallel transport along the geodesic triangle with corners P, N, Q realizes a rotation by φ: Let X be a vector tangent to c_1 at N, then the parallel transport is also tangent to c_1 at P and, consequently, perpendicular to c_3. The parallel transport along k_3 to Q is still perpendicular to c_3 and is therefore tangential to c_2. Finally, the parallel transport along c_2 back to N produces a vector Y with an angle φ against X.

    By using the result of Exercise 9, one can argue as follows: Because the surface area of the entire sphere is 4π it is clear that for any given number ψ between 0 and 4π there is a geodesic triangle whose surface area equals ψ. By Exercise 9 ψ equals the rotation angle of the parallel transport along the boundary of the triangle.

11. By using the result of Exercise 9, one can argue as in Exercise 10: For a small geodesic triangle Δ the parallel transport along its boundary is a rotation by the (negative) angle $\psi = \int_\Delta K dA = -\int_\Delta dA$ which is nothing but the negative surface area of Δ. It follows that the holonomy group contains all rotations by small angles and, consequently, all rotations that can be composed by them. (As a matter of fact, the surface area of a geodesic triangle can attain any value between 0 and π, compare 4.41.)

12. Let $t = x^0$, then for the components $x^i(t)$ of the t-lines one has $\dot{x}^0 = 1, \dot{x}^i = 0$ for $i \geq 1$. It follows $\ddot{x}^i = 0$ for all i and $\Gamma_{00}^k = \sum_j \Gamma_{00,j} g^{jk} = \Gamma_{00,0} g^{00} = 0$. Therefore $\ddot{x}^k + \sum_{i,j} \dot{x}^i \dot{x}^j \Gamma_{ij}^k = 0$ holds for all k, hence every t-line is a geodesic.
Now assume that a geodesic in M_* is given with components x^i (and constant t), then in particular $\dot{x}^0 = 0$. The equation of a geodesic in M_* is $\ddot{x}^k + \sum_{i,j\geq 1} \dot{x}^i \dot{x}^j \Gamma_{ij}^{*k} = 0$ for any $k \geq 1$. We have $\Gamma_{ij,k} = f^2 \Gamma_{ij,k}^*$ and $\Gamma_{ij}^k = \Gamma_{ij}^{*k}$ for all $i, j, k \geq 1$ since t is constant. Above we have seen that $\Gamma_{00}^k = 0$, similarly $\Gamma_{i0,0} = 0$ holds. Furthermore we have $\Gamma_{ij,0} = -\frac{1}{2}\frac{\partial}{\partial t} g_{ij} = -\frac{1}{2}\frac{\partial}{\partial t}(f^2 g_{ij}^*) = -f\dot{f} g_{ij}^*$ and $\Gamma_{ij}^0 = -f\dot{f} g_{ij}^*$ by $g^{00} = 1$ and $g^{k0} = 0$. The equation $\ddot{x}^k + \sum_{i,j\geq 0} \dot{x}^i \dot{x}^j \Gamma_{ij}^k = 0$ for a geodesic in M is satisfied if and only if in addition $\cot f = \frac{df}{dt} = 0$ for the corresponding value of t, i.e., if f is stationary there.

Compare this with the analogous result for surfaces of rotation (the examples after 4.12), where only those φ-lines (i.e., the circles arising from the rotation) are geodesics that have a stationary radius r.

13. and 14. and 15. These exercises are solved in B.O'NEILL, *Semi-Riemannian Geometry* at the end of Chapter 1 (pp. 30 ff.).

18. Assuming the convergence of all series under consideration, one has $(A^k)^T = (A^T)^k = (-A)^k$ and, consequently,

$$(\exp A)^T = \Big(\sum_k A^k/k!\Big)^T = \sum_k (-A)^k/k! = \exp(-A).$$

For the product the equation

$$(\exp A)(\exp A)^T = (\exp A)(\exp(-A)) = \exp(A - A) = \exp 0 = E$$

follows (cf. Example 3 after 5.19). This exponential rule itself can be verified just as the formula $1 = e^{x-x} = e^x e^{-x}$ in calculus by multiplication of two power series.

19. The power series of the ordinary logarithm function $\log(1+x) = \sum_{n\geq 1}(-1)^{n+1}x^n/n$ leads to the analogous approach

$$\log(E + B) = \sum_{n\geq 1}(-1)^{n+1}B^n/n$$

for a matrix B and the unit matrix E in such a way that $E + B$ is orthogonal, i.e., $(E + B)(E + B^T) = E$. The real power series converges for all $|x| < 1$, hence the series for matrices converges at least for all matrices B with sufficiently small entries. It follows $\mathrm{Det}(E + B) = 1$. Just as the equation $\log(xy) = \log x + \log y$ in calculus, one can verify the equation

$$\begin{aligned} 0 &= \log((E + B)(E + B^T)) \\ &= \log(E + B) + \log(E + B^T) = \log(E + B) + (\log(E + B))^T. \end{aligned}$$

This just means that $\log(E+B)$ is always skew-symmetric. The fact that both mapping are inverse to one another can be seen from the power series since $\exp(\log y) = y$ and $\log(\exp x) = x$ in calculus. Similarly we have $\exp(\log B) = B$ and $\log(\exp A) = A$, because inserting a power series into another is a purely formal procedure and, therefore, leads to the same result whenever the same rules are applied.

21. By 4.12 and 5.18 the equation of a geodesic is

$$\ddot{x}^k + \sum_{i,j=1}^{n} \dot{x}^i \dot{x}^j \Gamma_{ij}^k = 0 \quad \text{for} \quad k = 1, \ldots, n.$$

By the diagonal matrix (g_{ij}) of the metric the only non-vanishing Christoffel symbols are the following:

$$\Gamma_{ii}^i = \frac{1}{2}(g_{ii})_i/g_{ii}, \quad \Gamma_{ii}^k = -\frac{1}{2}(g_{ii})_k/g_{kk} \quad \text{for } i \neq k,$$

$$\Gamma_{kj}^k = \frac{1}{2}(g_{kk})_j/g_{kk} \quad \text{for } j \neq k, \quad \Gamma_{ik}^k = \frac{1}{2}(g_{kk})_i/g_{kk} \quad \text{for } i \neq k.$$

Therefore in the sum $\sum_{i,j=1}^{n} \dot{x}^i \dot{x}^j \Gamma_{ij}^k$ there remain only the terms with $i = j = k, i = j \neq k, i = k \neq j, j = k \neq i$. The left hand side of this equation is

$$\ddot{x}^k + \frac{1}{2}(\dot{x}^i)^2 \frac{(g_{ii})_i}{g_{ii}} - \frac{1}{2}\sum_{i\neq k}(\dot{x}^i)^2 \frac{(g_{ii})_k}{g_{kk}} + 2 \cdot \frac{1}{2}\sum_{i\neq k} \dot{x}^k \dot{x}^i \frac{(g_{kk})_i}{g_{kk}}.$$

Hence we can write the left hand side as

$$\ddot{x}^k + (\dot{x}^i)^2 \frac{(g_{ii})_i}{g_{ii}} + \sum_{i\neq k} \dot{x}^k \dot{x}^i \frac{(g_{kk})_i}{g_{kk}} - \frac{1}{2}\sum_{i=1}^{n}(\dot{x}^i)^2 \frac{(g_{ii})_k}{g_{kk}}.$$

Consequently the equation of a geodesic is equivalent to

$$\ddot{x}^k + \sum_{i=1}^{n} \dot{x}^k \dot{x}^i \frac{(g_{kk})_i}{g_{kk}} = \frac{1}{2}\sum_{i=1}^{n}(\dot{x}^i)^2 \frac{(g_{ii})_k}{g_{kk}}$$

for $k = 1, \dots, n$, and that in turn is equivalent to the assertion by the chain rule $\dot{g}_{kk} = \sum_i (g_{kk})_i \dot{x}^i$.

Chapter 6.

4. For two given connections ∇ and $\widetilde{\nabla}$ the equation

$$\begin{aligned}\nabla_{fX} hY - \widetilde{\nabla}_{fX} hY &= f\big(h\nabla_X Y + X(h)Y\big) - f\big(h\widetilde{\nabla}_X Y + X(h)Y\big)\\ &= fh\big(\nabla_X Y - \widetilde{\nabla}_X Y\big)\end{aligned}$$

holds. From the result of Exercise 1 it follows that $A(X, Y) = \nabla_X Y - \widetilde{\nabla}_X Y$ is a $(1, 2)$-tensor field.

7. We start with the equation $G = \mathrm{Ric} - \frac{S}{2}g$ for the Einstein tensor and assume $n \geq 3$ because otherwise $G \equiv 0$. For a given G one has $\mathrm{Tr}_g G = S\frac{2-n}{2}$ and, consequently, $S = \frac{2}{2-n}\mathrm{Tr}_g G$. The equation $\mathrm{Ric} = G + \frac{1}{2-n}\mathrm{Tr}_g G \cdot g$ follows.

9. By assumption $Df(T_pM)$ is a hyperplane in $T_{f(p)}\widetilde{M}$. This has a unique unit normal ν in each point (up to sign). Locally we can always introduce a Gauss map $\nu \colon M \to T\widetilde{M}$ as a normal vector field along M, globally this is possible if both M and $\widetilde{M}$ are orientable. Let $\nabla, \widetilde{\nabla}$ denote the Riemannian connections of $M, \widetilde{M}$. As in 3.9 for any tangential $Df(X)$ the derivative $\widetilde{\nabla}_{Df(X)}\nu$ is again tangential, and we can introduce a Weingarten map by $L(Df(X)) = -\widetilde{\nabla}_{Df(X)}\nu$ and a second

fundamental form by $I\!I(X,Y) = \widetilde{g}(L(Df(X)), Df(Y))$. By assumption the first fundamental form coincides with the metric g, i.e., $I(X,Y) = \widetilde{g}(Df(X), Df(Y)) = g(X,Y)$. Then one has for the two Riemannian connections the following decomposition into the tangential part and the normal part, by analogy with 4.3 in the case $\widetilde{M} = R^{n+1}$: $\widetilde{\nabla}_{Df(X)} Df(Y) = Df(\nabla_X Y) + I\!I(X,Y) \cdot \nu$ or, briefly, $\widetilde{\nabla}_X Y = \nabla_X Y + I\!I(X,Y) \cdot \nu$.
As in the proof of 4.18 we have the decomposition in to tangential part and normal part
$\widetilde{\nabla}_X \widetilde{\nabla}_Y Z = \widetilde{\nabla}_X(\nabla_Y Z + I\!I(Y,Z)\nu)$
$\quad = \nabla_X \nabla_Y Z + I\!I(X, \nabla_Y Z)\nu + (\widetilde{\nabla}_X I\!I(Y,Z))\nu + I\!I(Y,Z)\widetilde{\nabla}_X \nu$
and
$\widetilde{\nabla}_{[X,Y]} Z = \nabla_{[X,Y]} Z + I\!I([X,Y],Z)\nu$.
If V is another tangent vector, then $\langle V, \nu\rangle = 0$ implies
$\widetilde{g}(\widetilde{R}(X,Y)Z,V) = \widetilde{g}(\widetilde{\nabla}_X \widetilde{\nabla}_Y Z - \widetilde{\nabla}_Y \widetilde{\nabla}_X Z - \widetilde{\nabla}_{[X,Y]} Z, V)$
$= g(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z, V)$
$\quad + I\!I(Y,Z)\widetilde{g}(\widetilde{\nabla}_X \nu, V) - I\!I(X,Z)\widetilde{g}(\widetilde{\nabla}_Y \nu, V)$
$= g(R(X,Y)Z,V) - I\!I(Y,Z)I\!I(X,V) + I\!I(X,Z)I\!I(Y,V)$.
This is the assertion. For the sake of a short notation throughout the considerations above $\widetilde{g}(\widetilde{R}(X,Y)Z,V)$ were used instead of the more correct $\widetilde{g}(\widetilde{R}(Df(X), Df(Y))Df(Z), Df(V))$, similarly $\widetilde{\nabla}_X \nu$ instead of $\widetilde{\nabla}_{f(X)} \nu$). For an embedded hypersurface $M \subset \widetilde{M}$ the notation f is superfluous, compare B.O'Neill, *Semi-Riemannian Geometry*, Ch. 4 (pp. 97–102).
If $\widetilde{M}$ is the unit sphere $S^3(1)$ with the standard metric $\langle .,. \rangle$, then we obtain

$$\widetilde{R}(X,Y)Z = R_1(X,Y)Z = \langle Y,Z\rangle X - \langle X,Z\rangle Y \text{ and}$$
$$\langle R(X,Y)Y,X\rangle = \langle R_1(X,Y)Y,X\rangle + I\!I(Y,Y)I\!I(X,X) - I\!I(X,Y)^2.$$

Similarly we have the equation $K = 1 + \kappa_1\kappa_2 = 1 + \mathrm{Det}(L)$. where K denotes the inner Gaussian curvature of the surface and κ_1, κ_2 the two principal curvatures (eigenvalues of L).
Example: For the equator $S^2(1) \subset S^3(1)$, one has $K = 1$ and $\kappa_1 = \kappa_2 = 0$; for the Clifford torus $S^1(\frac{1}{\sqrt{2}}) \times S^1(\frac{1}{\sqrt{2}}) \subset S^3(1)$, one has $K = 0$ and $\kappa_1 = 1, \kappa_2 = -1$. Therefore this is an intrinsically flat minimal surface in the 3-sphere: $\kappa_1 + \kappa_2 = 0$.

10. We regard t as the coordinate x^0 and calculate the Christoffel symbols for $i, j \geq 1$:

$$\Gamma_{00,0} = \frac{1}{2}(g_{00})_t = 0, \Gamma_{0i,0} = \frac{1}{2}(g_{00})_i = 0,$$

$$\Gamma_{0i,j} = -\Gamma_{ij,0} = \frac{1}{2}(g_{ij})_t = ff'{}^*g_{ij}.$$

This implies $\Gamma^0_{00} = \Gamma^i_{00} = \Gamma^0_{0i} = 0, \Gamma^j_{0i} = \frac{f'}{f}\delta^j_i$ which in turn implies (a) and (b).
Furthermore one has

$$\begin{aligned}\Gamma_{ij,k} &= \tfrac{1}{2}(-(g_{ij})_k + (g_{ik})_j + (g_{kj})_i) \\ &= \tfrac{1}{2}f^2(-({}^*g_{ij})_k + ({}^*g_{ik})_j + ({}^*g_{kj})_i) = f^2\,{}^*\Gamma_{ij,k}.\end{aligned}$$

This implies $\Gamma^0_{ij} = -\frac{f'}{f}g_{ij}$ and $\Gamma^k_{ij} = {}^*\Gamma_{ij,k}$ which is essentially part (c) of the assertion. If X, Y as perpendicular to the t-lines then these equations can be written without coordinates as $\nabla_X \frac{\partial}{\partial t} = \frac{f'}{f}X$ and $\nabla_X Y = {}^*\nabla_X Y - \frac{f'}{f}g(X,Y)\frac{\partial}{\partial t}$.

11. We use the formulas from Exercise 10. Part (b) there tells us that $\nabla_X \frac{\partial}{\partial t} = \frac{f'}{f}X$ for all X that are perpendicular on the t-lines. This implies $R(X,Y)\frac{\partial}{\partial t} = \nabla_X\nabla_Y\frac{\partial}{\partial t} - \nabla_Y\nabla_X\frac{\partial}{\partial t} - \nabla_{[X,Y]}\frac{\partial}{\partial t} = \frac{f'}{f}(\nabla_X Y - \nabla_Y X - [X,Y]) = 0$, which is part (b). Part (a) is obtained with $R(\frac{\partial}{\partial t}, \frac{\partial}{\partial x^i})\frac{\partial}{\partial t} = \nabla_{\frac{\partial}{\partial t}}\nabla_{\frac{\partial}{\partial x^i}}\frac{\partial}{\partial t} - 0 = \nabla_{\frac{\partial}{\partial t}}(\frac{f'}{f}\frac{\partial}{\partial x^i}) = \left((\frac{f'}{f})' + \frac{f'^2}{f^2}\right)\frac{\partial}{\partial x^i} = \frac{f''}{f}\frac{\partial}{\partial x^i}$. For part (d) let X, Y, Z a basis of coordinate vector fields orthogonal to the t-lines (with mutually vanishing Lie brackets):

$$\begin{aligned}\nabla_X\nabla_Y Z &= \nabla_X({}^*\nabla_Y Z - \frac{f'}{f}g(Y,Z)\frac{\partial}{\partial t}) \\ &= {}^*\nabla_X{}^*\nabla_Y Z - \frac{f'}{f}X(g(Y,Z))\frac{\partial}{\partial t} \\ &\quad - \frac{f'}{f}g(Y,Z)\nabla_X\frac{\partial}{\partial t} - \frac{f'}{f}g(X,\nabla_Y Z)\frac{\partial}{\partial t} \\ &= {}^*\nabla_X{}^*\nabla_Y Z - \frac{f'^2}{f^2}g(Y,Z)X \\ &\quad - \frac{f'}{f}\big(g(X,\nabla_Y Z) + g(\nabla_X Y, Z) + g(Y,\nabla_X Z)\big)\frac{\partial}{\partial t}.\end{aligned}$$

From this equation one obtains

$$\begin{aligned} R(X,Y)Z &= {}^*R(X,Y)Z - \frac{f'^2}{f^2}\big(g(Y,Z)X - g(X,Z)Y\big) \\ &= {}^*R(X,Y)Z - \frac{f'^2}{f^2}R_1(X,Y)Z. \end{aligned}$$

Part (c) follows from part (d) and the symmetries of the curvature tensor:

$$\begin{aligned} \Big\langle R\Big(\frac{\partial}{\partial x_i},\frac{\partial}{\partial t}\Big)\frac{\partial}{\partial x_j},\frac{\partial}{\partial t}\Big\rangle &= \Big\langle R\Big(\frac{\partial}{\partial t},\frac{\partial}{\partial x_i}\Big)\frac{\partial}{\partial t},\frac{\partial}{\partial x_j}\Big\rangle \\ &= \frac{f''}{f}\Big\langle \frac{\partial}{\partial x_i},\frac{\partial}{\partial x_j}\Big\rangle = \frac{f''}{f}g_{ij}\Big\langle\frac{\partial}{\partial t},\frac{\partial}{\partial t}\Big\rangle \end{aligned}$$

as well as

$$\Big\langle R\Big(\frac{\partial}{\partial x_i},\frac{\partial}{\partial t}\Big)\frac{\partial}{\partial x_j},\frac{\partial}{\partial x_k}\Big\rangle = \Big\langle R\Big(\frac{\partial}{\partial x_j},\frac{\partial}{\partial x_k}\Big)\frac{\partial}{\partial x_i},\frac{\partial}{\partial t}\Big\rangle = 0.$$

This determines $R(\frac{\partial}{\partial x_i},\frac{\partial}{\partial t})\frac{\partial}{\partial x_j}$ uniquely as $R(\frac{\partial}{\partial x_i},\frac{\partial}{\partial t})\frac{\partial}{\partial x_j} = \frac{f''}{f}g_{ij}\frac{\partial}{\partial t}$.

14. With the principal curvatures $\kappa_1,\dots,\kappa_n$ and the corresponding principal curvature directions $X_1,\dots,X_n$, the Ricci tensor is given by $\mathrm{Ric}(X_i,X_i) = \sum_j\langle R(X_j,X_i)X_i,X_j\rangle = \sum_{j\neq i}K_{ij} = \sum_{j\neq i}\kappa_i\kappa_j$ and $\mathrm{Ric}(X_i,X_j)=0$ for $i\neq j$, cf. 6.16. For the associated $(1,1)$-tensor r with $\langle r(X),Y\rangle = \mathrm{Ric}(X,Y)$, this implies $r(X_i) = \sum_{j\neq i}\kappa_i\kappa_j X_i$ for any fixed i. Consequently X_i is an eigenvector of the Ricci tensor to the eigenvalue $\kappa_i\sum_{j\neq i}\kappa_j$.

16. By the rotational symmetry the principal curvatures of this hypersurface are κ_1 and $\kappa_2=\kappa_3=\kappa_4=-\kappa_1$. In a fixed point we can choose an orthonormal eigenbasis E_1,E_2,E_3,E_4 of the Weingarten map: $LE_1=\kappa_1E_1, LE_i=-\kappa_1E_i$ for $i=2,3,4$. Therefore the sectional curvature in the (i,j)-plane is $-\kappa_1^2$ for $i=1$ and κ_1^2 for $i,j\geq 2$. Then the result of Exercise 14 implies $\mathrm{Ric}(E_1,E_1)=-3\kappa_1^2, \mathrm{Ric}(E_i,E_i)=\kappa_1^2$ for $i=2,3,4$ and $\mathrm{Ric}(E_i,E_j)=0$ for $i\neq j$. In this basis the Ricci tensor appears as a diagonal matrix with the entries $-3\kappa_1^2,\kappa_1^2,\kappa_1^2,\kappa_1^2$ along the diagonal. One can easily see that the trace (which is the scalar curvature) vanishes.

21. The exterior derivative df of a scalar function does not depend on the Riemannian metric since $df(X) = \nabla_X f = X(f)$ for any Riemannian connection. The exterior derivative $d\omega$ is defined by $d\omega(X,Y) = \nabla\omega(X,Y) - \nabla\omega(Y,X)$, cf. the special cases after 6.2. For a parallel ω one has $d\omega = 0$. On the other hand $d\omega(X,Y) = \nabla_X(\omega(Y)) - \omega(\nabla_X Y) - \nabla_Y(\omega(X)) + \omega(\nabla_Y X) = X(\omega(Y)) - Y(\omega(X)) - \omega([X,Y])$, and the last expression does not depend on the Riemannian metric. Therefore we can consider the equations $d\omega = 0$ and $\omega = df$ in a chart (that is, in $I\!R^n$), and there we have the ordinary integrability condition $d\omega = 0$ (symmetry of the derivatives) of the equation $\omega = df$ for a locally defined function f, cf. the examples after 4.33.

22. The Hessian $\nabla^2 f$ is defined by $\nabla^2 f(X,Y) = \nabla_X\nabla_Y f - (\nabla_X Y)(f)$. This implies $\nabla^2 f(X,Y) - \nabla^2 f(Y,X) = \nabla_X\nabla_Y f - (\nabla_X Y)(f) - \nabla_Y\nabla_X f + (\nabla_Y X)(f) = X(Yf) - Y(Xf) - [X,Y](f) = 0$ by the definition of the Lie bracket.

23. From $g(\nabla_X \mathrm{grad} f, Y) = \nabla^2 f(X,Y)$ we see that the self-adjointness of the Hesse $(1,1)$-tensor is equivalent to the symmetry of the Hessian, and this holds by Exercise 22. For an eigenvalue λ with eigenvector X one has $\nabla_X \mathrm{grad} f = \lambda X$. Let $c(t)$ be a geodesic with $c'(0) = X$. Then at a maximum $c(0)$ of f one has $(f \circ c)'(0) = 0$ and $\lambda = g(\nabla_{c'(0)} \mathrm{grad} f, c'(0)) = \nabla^2 f(c'(0), c'(0)) = (f \circ c)''(0) - \nabla_{c'} c' = (f \circ c)''(0) \leq 0$, similarly $\lambda = (f \circ c)''(0) \geq 0$ at a minimum.

Chapter 7.

1. Let $f \colon U \to I\!R^{n+1}$ be a hypersurface element with $n \geq 3$. There are n principal curvatures $\kappa_1, \ldots, \kappa_n$ with associated principal curvature directions $X_1, \ldots, X_n$. The Gauss equation 4.18 (i) implies that the sectional curvature in the (X_i, X_j)-plane equals $K_{ij} = \langle R(X_i, X_j)X_j, X_i\rangle = \kappa_i\kappa_j$, cf. 4.21 and the remarks after 6.10. The assumption of strictly negative sectional curvatures implies $\kappa_i\kappa_j < 0$ for any $i \neq j$. Therefore all κ_i are non-zero and have pairwise distinct signs. However, there are only two signs $(+)$ and $(-)$. This leads to a contradiction if $n \geq 3$. In particular this holds for the case of constant negative sectional curvature.

3. First of all one has $\langle \Phi(x), \Phi(x) \rangle_1 = -(\lambda - 1)^2 + \lambda^2 ||x||^2 = -\lambda^2(1 - ||x||^2) + 2\lambda - 1 = -\frac{4}{1-||x||^2} + \frac{4-1+||x||^2}{1-||x||^2} = -1$, hence Φ maps D^n really into H^n, i.e., $\Phi(x) = (\xi_0, \xi)$ with $\xi_0 = \lambda - 1 \geq 1, \xi = \lambda x \in I\!R^n, -\xi_0^2 + ||\xi||^2 = -1$. Furthermore Φ is bijective since there is an inverse map: $\Phi^{-1}(\xi_0, \xi) = \xi/(\xi_0 + 1)$, where $\lambda = \xi_0 + 1$ and $x = \xi/\lambda$. For any $\xi_0 \geq 1$ we obtain $\lambda \geq 2$ and $||\frac{\xi}{\xi_0+1}||^2 = \frac{\xi_0-1}{\xi_0+1} < 1$. A calculation of the partial derivatives

$$\frac{\partial \Phi}{\partial x_i} = \frac{1}{(1 - ||x||^2)^2}\Big(4x_i, 4x_1x_i, \ldots, 4x_{i-1}x_i, 2(1 - ||x||^2) + 4x_i^2, 4x_{i+1}x_i, \ldots, 4x_nx_i\Big)$$

leads to a comparison of $\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_i} \rangle = \frac{4}{(1-||x||^2)^2}$ and $\langle \frac{\partial \Phi}{\partial x_i}, \frac{\partial \Phi}{\partial x_i} \rangle_1$. The result is

$$\begin{aligned}\Big\langle \frac{\partial \Phi}{\partial x_i}, \frac{\partial \Phi}{\partial x_i} \Big\rangle_1 &= \frac{1}{(1 - ||x||^2)^4}(-16x_i^2 + 16x_1^2x_i^2 + \cdots + 16x_i^4 \\ &\quad + \cdots + 16x_n^2x_i^2 + 4(1 - ||x||^2)^2 + 16x_i^2(1 - ||x||^2)) \\ &= \frac{4}{(1 - ||x||^2)^2}.\end{aligned}$$

Similarly one has $\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \rangle = 0 = \langle \frac{\partial \Phi}{\partial x_i}, \frac{\partial \Phi}{\partial x_j} \rangle_1$ for $i \neq j$. This shows that Φ is an isometric mapping.

4. From the equation $i\frac{1-z}{1+z} = i\frac{(1-z)(1+\bar{z})}{(1+z)(1+\bar{z})} = i\frac{1-z\bar{z}-z+\bar{z}}{(1+z)(1+\bar{z})}$, we conclude that the imaginary part is $\frac{1-z\bar{z}}{(1+z)(1+\bar{z})}$, which is a positive real number for any z with $z\bar{z} < 1$. Therefore the transformation

$$z \mapsto w = i\frac{1 - z}{1 + z}$$

maps the unit disc into the Poincaré upper half-plane. This mapping is invertible, because for any given w the equation $w = i\frac{1-z}{1+z}$ admits the unique solution $z = \frac{i-w}{i+w}$. Moreover it follows that $z\bar{z} = \frac{i-w}{i+w} \cdot \frac{-i-\bar{w}}{-i+\bar{w}} = \frac{1+w\bar{w}+i(w-\bar{w})}{1+w\bar{w}-i(w-\bar{w})} < 1$ since $i(w - \bar{w})$ is a negative real number if $\mathrm{Im}(w) > 0$. By the considerations above for $w = i\frac{1-z}{1+z}$, one has $\mathrm{Im}(w) = \frac{1-z\bar{z}}{(1+z)(1+\bar{z})}$. Now we transform the hyperbolic arc length element $ds^2 = \frac{4}{(1-z\bar{z})^2}dzd\bar{z}$ of the conformal disc model into the new variable w of the Poincaré upper half-plane, by analogy with the solution to Exercise 13 in Chapter 4 and by using the equations $\frac{dw}{dz} = \frac{-2i}{(1+z)^2}$ and $\frac{d\bar{w}}{d\bar{z}} = \frac{2i}{(1+\bar{z})^2}$. This leads to $dwd\bar{w} = \frac{4}{(1+z)^2(1+\bar{z})^2}dzd\bar{z} = \frac{4}{(1-z\bar{z})^2}(\mathrm{Im}(w))^2 dzd\bar{z}$. Therefore both arc length elements coincide:

$$ds^2 = \frac{4}{(1-z\bar{z})^2}dzd\bar{z} = \frac{1}{(\mathrm{Im}(w))^2}dwd\bar{w}.$$

This just means that the transformation $z \mapsto w$ is an isometry.

7. By 7.19 the space of *all* Jacobi fields along a geodesic c starting at p is $2n$-dimensional. The Jacobi fields that vanish at p form a subspace of dimension n. One of them is tangential and does certainly not vanish at any other point q, namely, the field $t \cdot T$ (cf. 7.17.(ii)). Thus there are $n-1$ linearly independent Jacobi fields that are orthogonal to c and vanish at p. Therefore the multiplicity cannot exceed $n-1$.

8. We use the notations as in 7.15: $Y(0) = 0 = X(0)$, $X(t) = t \cdot W$ and $p = c(0)$. Then 7.15 implies $Y(t) = D\exp_p|_{tV}(X(t))$. Now if a point $q = c(t_0)$ is conjugate to p, then there is such a Jacobi field Y with $Y(t_0) = 0$. Consequently $X(t_0)$ lies in the kernel of $D\exp_p|_{t_0V}$. The converse holds as well.

9. Let Y_1, Y_2 be Jacobi fields along c with $Y_1(a) = Y_2(a)$ and $Y_1(b) = Y_2(b)$. Then $Y_1 - Y_2$ vanishes at p and at q. If p and q are not conjugate then this is impossible unless $Y_1 - Y_2 \equiv 0$.

12. If one regards the 3-sphere S^3 as the set of unit quaternions $\{a + bi + cj + dk \mid a^2 + b^2 + c^2 + d^2 = 1\} \subset \mathbb{H}$, then the tangent space at any point $p \in S^3$ can be identified with the hyperplane perpendicular to the position vector of p. In the particular case $p = 1$ this is the hyperplane spanned by i, j, k. The quaternionic multiplication with a fixed $p \in S^3$ maps i, j, k into three linearly independent and orthonormal tangent vectors pi, pj, pk at p because the multiplication by p is an isometry of the sphere. Consequently any tangent vector $(p, X) \in T_pS^3$ can be written as $(p, X) = (p, p(xi + yj + zk))$ in a unique way that does not depend on the choice of a basis. Hence the mapping $\Phi\colon TS^3 \to S^3 \times \mathbb{R}^3$ defined by $\Phi(p, X) = (p, x, y, z)$ is globally defined, bijective and differentiable. Its inverse map is also differentiable by $\Phi^{-1}(p, x, y, z) = (p, p(xi + yj + zk))$.

Therefore TS^3 and $S^3 \times I\!R^3$ are globally diffeomorphic to one another. This is also expressed by saying: TS^3 *is parallelizable.* This principle holds more generally not only for the rotation group $\mathbf{SO}(3)$ but also for any other Lie group: Choose a basis in the tangent space at the unit element and transfer that by the differential of the multiplication from left (the so-called *left translation*) into all the other tangent spaces. The 3-sphere can be interpreted as the Lie group Spin(3) $\cong$ Sp(1) of unit quaternions.

Chapter 8.

4. By Definition 8.20 the following equation holds:

$$(II \bullet II)(X,Y,Z,T) = 2II(X,Z)II(Y,T) - 2II(X,T)II(Y,Z).$$

The Gauss equation 4.18 (i) is

$$\langle R(X,Y)T,Z\rangle = II(Y,T)II(X,Z) - II(X,T)II(Y,Z).$$

This implies $R = \frac{1}{2}II \bullet II$ (note the convention $R(X,Y,Z,T) = \langle R(X,Y)T,Z\rangle$). In Ricci calculus we can see that from Exercise 23 in Ch. 4: $R_{ijkl} = h_{ik}h_{jl} - h_{il}h_{jk} = 2h_{[i[j}h_{k]l]}$.

5. Obviously the three bivectors

$$B_1 = E_1 \wedge E_2 + E_3 \wedge E_4,$$
$$B_2 = E_1 \wedge E_3 + E_4 \wedge E_2,$$
$$B_3 = E_1 \wedge E_4 + E_2 \wedge E_3$$

are contained in $\bigwedge^2_+$. They are linearly independent because by 8.18 all $E_1 \wedge E_j$ with $i < j$ form a basis of $\bigwedge^2$. Similarly the three bivectors

$$B_4 = E_1 \wedge E_2 - E_3 \wedge E_4,$$
$$B_5 = E_1 \wedge E_3 - E_4 \wedge E_2,$$
$$B_6 = E_1 \wedge E_4 - E_2 \wedge E_3$$

are contained in Λ^2_- and are linearly independent. By the known dimension of the space, B_1, B_2, B_3 are a basis of $\bigwedge^2_+$ and B_4, B_5, B_6 are a basis of $\bigwedge^2_-$.

6. By applying the equation in 8.27 (i) twice one obtains

$$\begin{aligned}
\widetilde{\nabla}_X\widetilde{\nabla}_Y Z &= \widetilde{\nabla}_X\Big(\nabla_Y Z - (Y\varphi)Z - (Z\varphi)Y + \langle Y,Z\rangle \mathrm{grad}\varphi\Big)\\
&= \nabla_X\nabla_Y Z - (X\varphi)\nabla_Y Z - ((\nabla_Y Z)\varphi)X + \langle X,\nabla_Y Z\rangle \mathrm{grad}\varphi\\
&\quad - X(Y\varphi)Z - (Y\varphi)\Big(\nabla_X Z - (X\varphi)Z - (Z\varphi)X + \langle X,Z\rangle \mathrm{grad}\varphi\Big)\\
&\quad - X(Z\varphi)Y - (Z\varphi)\Big(\nabla_X Y - (X\varphi)Y - (Y\varphi)X + \langle X,Y\rangle \mathrm{grad}\varphi\Big)\\
&\quad + X\langle Y,Z\rangle \mathrm{grad}\varphi\\
&\quad + \langle Y,Z\rangle\Big(\nabla_X \mathrm{grad}\varphi - (X\varphi)\mathrm{grad}\varphi\\
&\qquad\qquad - ||\mathrm{grad}\varphi||^2 X + \langle X,\mathrm{grad}\varphi\rangle\mathrm{grad}\varphi\Big)
\end{aligned}$$

and

$$\widetilde{\nabla}_{[X,Y]}Z = \nabla_{[X,Y]}Z - ([X,Y]\varphi)Z - (Z\varphi)[X,Y] + \langle [X,Y],Z\rangle \mathrm{grad}\varphi.$$

After skew-symmetrizing with respect to X and Y, one gets

$$\begin{aligned}
\widetilde{R}(X,Y)Z &= R(X,Y)Z + \langle Y,Z\rangle \nabla_X \mathrm{grad}\varphi - \langle X,Z\rangle\nabla_Y\mathrm{grad}\varphi\\
&\quad + ||\mathrm{grad}\varphi||^2\big(\langle X,Z\rangle Y - \langle Y,Z\rangle X\big) + (Y\varphi)(Z\varphi)X - (X\varphi)(Z\varphi)Y\\
&\quad + Y(Z\varphi)X - \big((\nabla_Y Z)\varphi\big)X - X(Z\varphi)Y + \big((\nabla_X Z)\varphi\big)Y\\
&\quad + \big(X\varphi\langle Y,Z\rangle - Y\varphi\langle X,Z\rangle\big)\mathrm{grad}\varphi.
\end{aligned}$$

This coincides with the claim in 8.27 (ii) if one inserts the equations

$$\begin{aligned}
Y(Z\varphi) - (\nabla_Y Z)\varphi &= \nabla_Y\langle Z,\mathrm{grad}\varphi\rangle - \langle\nabla_Y Z,\mathrm{grad}\varphi\rangle\\
&= \langle Z,\nabla_Y\mathrm{grad}\varphi\rangle \text{ and}\\
X(Z\varphi) - (\nabla_X Z)\varphi &= \nabla_X\langle Z,\mathrm{grad}\varphi\rangle - \langle\nabla_X Z,\mathrm{grad}\varphi\rangle\\
&= \langle Z,\nabla_X\mathrm{grad}\varphi\rangle.
\end{aligned}$$

7. We verify the assertion for the $(0,4)$-tensors R and $\widetilde{R}$ of g and $\widetilde{g} = e^{-2\varphi}g$. The corresponding Weyl tensors W and $\widetilde{W}$ are given by Theorem 8.24 as

$$W = R - \frac{1}{n-2}\big(\mathrm{Ric} - \frac{S}{2(n-1)}g\big) \bullet g$$

and

$$\widetilde{W} = \widetilde{R} - \frac{1}{n-2}\big(\widetilde{\mathrm{Ric}} - \frac{\widetilde{S}}{2(n-1)}\widetilde{g}\big) \bullet \widetilde{g}.$$

For R and $\widetilde{R}$ we have the equation after 8.27 (framed formula on page 345)

$$e^{2\varphi}\widetilde{R} = R - \frac{1}{2}||\mathrm{grad}\varphi||^2 g \bullet g + (\nabla^2\varphi) \bullet g + (\nabla\varphi)^2 \bullet g,$$

and for the associated Ricci tensors one has by 8.27

$$\widetilde{\mathrm{Ric}} = \mathrm{Ric} + \big(\Delta\varphi - (n-2)||\mathrm{grad}\varphi||^2\big)g + (n-2)e^{-\varphi}\nabla^2(e^{\varphi}),$$

where $e^{-\varphi}\nabla^2(e^{\varphi}) = \nabla^2\varphi + (\nabla\varphi)^2$. For the scalar curvatures $S, \widetilde{S}$ part (iv) of 8.27 states the equation

$$\widetilde{S} = e^{2\varphi}S + ne^{2\varphi}\big(\Delta\varphi - (n-2)||\mathrm{grad}\varphi||^2\big) + (n-2)e^{\varphi}\Delta(e^{\varphi}),$$

where $e^{\varphi}\Delta(e^{\varphi}) = e^{2\varphi}\Delta\varphi + e^{2\varphi}||\mathrm{grad}\varphi||^2$.
We insert these equations into the equation for $\widetilde{W}$ and obtain

$$\begin{aligned}
e^{2\varphi}\widetilde{W} &= e^{2\varphi}\widetilde{R} - e^{2\varphi}\frac{1}{n-2}\big(\widetilde{\mathrm{Ric}} - \frac{\widetilde{S}}{2(n-1)}\widetilde{g}\big) \bullet \widetilde{g} \\
&= R - \frac{1}{2}||\mathrm{grad}\varphi||^2 g \bullet g + (\nabla^2\varphi) \bullet g + (\nabla\varphi)^2 \bullet g \\
&\quad - \frac{1}{n-2}\Big(\mathrm{Ric} + \big(\Delta\varphi - (n-2)||\mathrm{grad}\varphi||^2\big)g \\
&\quad + (n-2)\nabla^2\varphi + (n-2)(\nabla\varphi)^2\Big) \bullet g \\
&\quad + \frac{1}{2(n-1)(n-2)}\Big(S + n(\Delta\varphi - (n-2)||\mathrm{grad}\varphi||^2\big) \\
&\quad + (n-2)\big(\Delta\varphi + ||\mathrm{grad}\varphi||^2\big)\Big) g \bullet g \\
&= R - \frac{1}{n-2}\mathrm{Ric} \bullet g + \frac{S}{2(n-1)(n-2)} g \bullet g \\
&\quad + \frac{1}{2}||\mathrm{grad}\varphi||^2\big(-1 + 2 - \frac{n}{n-1} + \frac{1}{n-1}\big) g \bullet g \\
&\quad + \frac{\Delta\varphi}{(n-1)(n-2)}\big(-(n-1) + \frac{n}{2} + \frac{n-2}{2}\big) g \bullet g \\
&= R - \frac{1}{n-2}\big(\mathrm{Ric} - \frac{S}{2(n-1)} g\big) \bullet g = W.
\end{aligned}$$

Bibliography

Textbooks on Differential Geometry

[1] M. do Carmo, "Differential Geometry of Curves and Surfaces", Prentice Hall, Englewood Cliffs, NJ, 1976.

[2] W. Klingenberg, "A Course in Differential Geometry", Springer, New York, 1978.

[3] D. Laugwitz, "Differential and Riemannian Geometry", Academic Press, New York, 1965.

[4] D. J. Struik, "Lectures on Classical Differential Geometry", Addison-Wesley 1950/1961, reprinted by Dover, New York, 1988.

[5] E. Kreyszig, "Differential Geometry", Dover, New York, 1991.

[6] J. J. Stoker, "Differential Geometry", Wiley-Interscience, New York, 1969, reprint 1989.

[7] M. Spivak, "A Comprehensive Introduction to Differential Geometry", Volumes I – V, Publish or Perish, Wilmington, Del., 1979.

[8] T.G.Feeman, *Portraits of the Earth: A Mathematician Looks at Maps* American Mathematical Society, 2002. Paperback, 123 pages Mathematical World, vol 18 MAWRLD/18

With special emphasis on computer graphics:

[9] A. Gray, "Modern Differential Geometry of Curves and Surfaces", 2nd edition, CRC Press, Boca Raton, 1998.

[10] J. Oprea, "The Mathematics of Soap Films: Explorations with Maple", Student Math. Library Vol. 10, AMS, Providence, 2000.

On minimal surfaces:

[11] U. Dierkes, S. Hildebrandt, A. Küster, O. Wohlrab, "Minimal Surfaces I", Springer, 1992.

[12] D. Hoffmann & H. Karcher, *Complete embedded minimal surfaces of finite total curvature*, Geometry V (R. Osserman, ed.), Encycl. Math. Sci. **90**, 5–93, Springer 1997.

Textbooks on Riemannian Geometry

[13] P. Petersen, "Riemannian Geometry", 2nd edition, Springer, New York, 2006.

[14] M. do Carmo, "Riemannian Geometry", Birkhäuser, Boston-Basel-Berlin, 1992.

[15] J. A. Schouten, "Der Ricci-Kalkül", Springer, Heidelberg, 1924, reprint Springer, Berlin-New York, 1978.

[16] J. A. Schouten, "Ricci-calculus", Springer, Heidelberg, 1954.

[17] T. J. Willmore, "Total Curvature in Riemannian Geometry", Ellis Horwood, 1982.

[18] J. Cheeger, D. G. Ebin, "Comparison Theorems in Riemannian Geometry", North Holland, 1975.

[19] S. Kobayashi, K. Nomizu, "Foundations of Differential Geometry" I, Wiley-Interscience, 1963.

[20] J. Wolf, "Spaces of Constant Curvature", Publish or Perish, Boston, 1974.

Including the general theory of relativity:

[21] T. Levi-Civita, "The Absolute Differential Calculus (Calculus of Tensors)", Dover, New York, 1977.

[22] B. O'Neill, "Semi-Riemannian Geometry", Academic Press, San Diego, 1983.

[23] A. Petrov, "Einstein Spaces", Pergamon Press, Oxford, 1969.

[24] A. Besse, "Einstein Manifolds", Springer, Heidelberg-New York, 1987.

[25] F. de Felice & C. J. S. Clarke, "Relativity on Curved Manifolds", Cambridge University Press, 1990.

[26] R. K. Sachs & H. Wu, "General Relativity for Mathematicians", Springer, New York, 1977.

Other textbooks

On analysis:

[27] S. Lang, "Undergraduate Analysis", 2nd ed., Springer, New York, 1997.

[28] S. Lang, "Calculus of Several Variables", Addison-Wesley, 1973.

[29] M. Spivak, "Calculus on Manifolds", Benjamin, New York, 1965.

[30] L. C. Evans, "Partial Differential Equations", Graduate Studies in Math. 19, AMS, Providence, 1998.

On algebra:

[31] S. Lang, "Linear Algebra", 3rd ed., Springer, New York, 1987.

[32] W. Greub, "Linear Algebra", 3rd ed., Springer, Heidelberg, 1967.

[33] W. Greub, "Multilinear Algebra", Springer, Heidelberg, 1967.

[34] H. S. M. Coxeter & W. O. J. Moser, "Generators and Relations for Discrete Groups", 4^{th} ed., Springer, Heidelberg, 1980.

On the theory of functions of a complex variable:

[35] J. Bak & D. J. Newman, "Complex Analysis", Springer, New York, 1982.

[36] R. V. Churchill, J. W. Brown, R. F. Verhey, "Complex Variables and Applications", McGraw-Hill, New York 1974.

On topology:

[37] R. E. Schwartz, "Mostly Surfaces", Student Math. Library Vol. 60, AMS, Providence, 2011.

[38] M. A. Armstrong, "Basic Topology", Springer, 1983.

[39] T. Bröcker & K. Jänich, "Introduction to Differential Topology", Cambridge University Press, 1982.

[40] H. Schubert, "Topology", MacDonald Technical and Scientific, 1968.

[41] H. Seifert & W. Threlfall, "Lehrbuch der Topologie", Teubner 1934 (reprint Chelsea 1980); Engl. transl.: "Seifert and Threlfall: A Textbook of Topology". Academic Press, New York, 1980.

On Lie groups:

[42] S. Arvanitoyeorgos, "An Introduction to Lie Groups and the Geometry of Homogeneous Spaces", Student Math. Library Vol. 22, AMS, Providence, 2003.

[43] H. Pollatsek, "Lie Groups: A problem-oriented introduction via matrix groups", Math. Association of America, 2009

[44] F. W. Warner, "Foundations of Differentiable Manifolds and Lie Groups", Springer, Heidelberg-New York, 1983.

[45] J. J. Duistermaat & J. A. C. Kolk, "Lie Groups", Springer, Heidelberg, 2000.

[46] S. Kobayashi, "Transformation Groups in Differential Geometry", Springer, Heidelberg, 1972.

On geometry:

[47] J. Stillwell, "Geometry of Surfaces", Springer, New York, 1992.

[48] H. S. M. Coxeter, "Regular Polytopes", Dover, New York, 1948.

[49] F. Klein, "Vorlesungen über Nicht-Euklidische Geometrie", Springer 1928 (reprint 1968).

[50] J. M. Montesinos, "Classical Tessellations and Three-manifolds", Springer, Heidelberg-New York, 1987.

[51] P. J. Ryan, "Euclidean and Non-Euclidean Geometry", Cambridge Univ. Press, 1986.

[52] M. A. Armstrong, "Groups and Symmetry", Springer, New York, 1988.

On the history of mathematics:

[53] F. Klein, "Development of Mathematics in the 19th Century", Lie groups: History, Frontiers and Appl. IX, Math. Sci. Press, Brookline MA, 1979.

[54] P. Dombrowski, "150 years after Gauss' *Disquisitiones generales circa superficies curvas*", With the original text of Gauss. Astérisque, 62, Société Mathématique de France, Paris, 1979.

List of notation

$\mathbb{Z}, \mathbb{R}$ integers, real numbers

$\mathbb{R}^n$ real vector space, also Euclidean space with fixed origin

E^n Euclidean space without fixed origin

S^n n-dimensional unit sphere in $\mathbb{R}^{n+1}$

$\mathbb{R}^n_1$ Minkowski space or Lorentzian space

H^n hyperbolic space

$\mathbb{C}, \mathbb{H}$ complex numbers, quaternions

$\langle\ ,\ \rangle$ Euclidean scalar product, in Chapters 5 to 8 also a Riemannian metric

$\langle\ ,\ \rangle_1$ Lorentzian metric in Minkowski space $\mathbb{R}^3_1$

I, II, III first, second and third fundamental forms

g_{ij}, h_{ij}, e_{ij} first, second and third fundamental forms in local coordinates

g^{ij} inverse matrix to g_{ij}

$h_i^k = \sum_j h_{ij} g^{jk}$ Weingarten mapping in local coordinates

E, F, G Gaussian symbols for the first fundamental form $E = g_{11}, F = g_{12}, G = g_{22}$

g Riemannian metric

κ curvature of a plane or space curve

τ torsion of a space curve

$e_1, \ldots, e_n$ Frenet n-frame of a Frenet curve

$\kappa_1, \ldots, \kappa_{n-1}$ Frenet curvatures of a Frenet curve in $\mathbb{R}^n$ (in Ch. 2)

$\dot{c} = \frac{dc}{dt}$ tangent vectors to a curve with parameter t

$c' = \frac{dc}{ds}$ tangent vectors to a curve with arc length parameter s

U_c index of a closed plane curve c

κ_N normal curvature of a curve on a surface

κ_g geodesic curvature of a curve on a surface

ν Gaussian normal mapping, Gauss map

L Weingarten mapping

κ_1, κ_2 principal curvatures of a surface element in $\mathbb{R}^3$

$\kappa_1, \ldots, \kappa_n$ principal curvatures of a hypersurface in $\mathbb{R}^{n+1}$ (in Ch. 3)

λ parameter of distribution of a ruled surface

dA area element of a two-dimensional surface element

dV volume element in higher dimensions

H mean curvature

K Gaussian curvature

K_i ith mean curvature (on hypersurface elements)

D directional derivative in $\mathbb{R}^n$

∇ covariant derivative or Riemannian connection

$[X, Y]$ Lie bracket of two vector fields X, Y

$\Gamma_{ij}^k, \Gamma_{ij,m}$ Christoffel symbols

$R(X, Y)Z$ curvature tensor

R_{ijk}^s, R_{ijkl} curvature tensor in local coordinates

$\mathrm{Ric}(X, Y)$ Ricci tensor

$\mathrm{ric}(X)$ Ricci curvature in the direction X

R_{ij} Ricci tensor in local coordinates

S scalar curvature

W, C Weyl and Schouten tensors

$\exp_p$ exponential mapping at a point p

Index

acceleration vector, 141
angle, 2
angle preserving, 99, 127
apex, 12
arc element, 60
arc length, 9
Archimedean spiral, 51
asymptotic curve, 83, 126
atlas, 200, 202

Banchoff, T., 185
Beltrami, E., 84, 93
Bertrand curve, 54
Bianchi identity, 243, 335
binary dihedral group, 303
binary icosahedral group, 303
binary octahedral group, 303
binary tetrahedral group, 303
binormal, 17
biquadratic form, 246
bivector, 332
Bonnet, O., 153
boost, 272

canal surface, 77
Cardan angles, 202
Cartan, É., 165
Catalan, E. C., 111, 128
catenary, 11, 109, 193
catenoid, 109, 128, 156, 193, 262
Cauchy-Riemann equations, 101
Cayley map, 201
Cayley plane, 330
Cayley ruled surface, 132
chain rule, 210
chart, 5, 198, 199, 202
Christoffel symbols, 139, 166, 223
circle, 9
Clifford torus, 32
Codazzi-Mainardi equation, 147, 154, 168, 196, 242
Cohn-Vossen, S., 188
complex manifold, 203
complex projective space, 329
complex structure, 204
cone, 88, 90, 92
conformal, 99, 103, 216, 274
conformal curvature, 341
conformally flat, 124, 346, 347
conic type, 81
conjugate point, 307
conjugate surface, 110
connection, 221
connection form, 166
constant curvature, 25, 36, 80, 161, 189, 249, 269, 289, 310
constant mean curvature, 191
contact of kth order, 12
contraction, 250
contravariant tensor, 236

convergence, 204
convex, 43, 182
convex hull, 45, 182, 304
coordinate transformation, 200
Cornu spiral, 16
cosmological constant, 325
countability axiom, 218
covariant derivative, 136, 138, 166, 221, 239, 268
covariant tensor, 236
covector field, 237
covering, 293, 302, 310
Coxeter, H.S.M., 303
CR equations, 101
critical point, 187
cubical parabola, 20
curvature, 14, 17, 20, 36, 70
curvature tensor, 150, 169, 243, 247, 315
curve, 7, 8
curve, closed, 37
curve, length of, 8
curve, simply closed, 37, 46
cyclic group, 302, 303
cycloid, 50
cylinder, 88, 90

Darboux equations, 26, 53
Darboux vector, 26, 53
derivative, 3, 6, 210
developable surface, 88, 119
dicyclic group, 303
diffeomorphic, 204
differentiable, 3
differentiable manifold, 199
differentiable structure, 200
differential, 6, 210, 240
differential form, 166
dihedral group, 302
Dini, U., 93
directional derivative, 98, 135, 206, 207, 221
directional vector, 62
directrix, 84
distance, 2
divergence, 251, 257, 319, 320
double point, 37, 49
double tangent, 49

dual basis, 212
duality, 350
Dupin indicatrix, 75

eigenvalue, 71
eigenvector, 71
Einstein field equations, 324
Einstein space, 254, 263, 328, 351
Einstein tensor, 257, 324
Einstein, A., 312, 317, 351
ellipse, 75
ellipsoid, 130
elliptic point, 72, 193
elongated sphere, 81
energy functional, 280
Enneper, A., 84, 111
equations of Gauss and Weingarten, 140, 146
Euclidean motion, 269
Euler angles, 202
Euler characteristic, 177, 179, 180, 188, 218
evolute, 15
exponential mapping, 226, 280
exterior derivative, 167

Fabricius-Bjerre, Fr., 49
Fenchel, W., 47
Fermi coordinates, 160
first fundamental form, 59
Flamm's paraboloid, 95
flow, 230
focal curve, 15
four vertex theorem, 46
free motion, 270
Frenet curvature, 27
Frenet curve, 13
Frenet equations, 14, 17, 27, 36
Frenet matrix, 25, 27, 37
Frenet n-frame, 13
Frobenius, G., 155

Gauss equation, 147, 150, 154, 168, 196, 261, 268, 359
Gauss formula, 140
Gauss lemma, 283
Gauss map, 63, 66, 100, 116, 122
Gauss, C. F., 148, 320
Gauss-Bonnet formula, 172, 321

Gauss-Kronecker curvature, 123
Gaussian curvature, 72, 117, 148, 193, 242
geodesic, 71, 121, 141, 216, 225, 280, 307
geodesic curvature, 23, 71, 125, 171
geodesic parallel coordinates, 160
geodesic polar coordinates, 272, 283, 290
geodesic torsion, 125
geodesic triangle, 176, 230
geometric linearization, 8, 55
golden ratio, 303
gradient, 4, 98, 240, 313
Gram determinant, 181
Gram-Schmidt orthogonalization, 13
graph, 58, 73

harmonic function, 100
Hausdorff separation axiom, 203
helicoid, 86, 109, 132, 156
helicoidal motion, 10, 131
helicoidal ruled surface, 86, 93
helix, 10, 20
Henneberg, L., 111
Hesse tensor, 240
Hessian, 73, 240, 264, 345
Hessian matrix, 73
hexagonal torus, 294
Hilbert, D., 162, 190, 312, 317
Hilbert-Einstein functional, 312
Hodge operator, 350, 353
holomorphic, 101, 105
holonomy group, 228
homogenous space, 326
Hopf, H., 41, 180
Hurwitz quaternions, 304
hyperbola, 35, 75
hyperbolic plane, 119, 130, 196
hyperbolic point, 72
hyperbolic space, 267, 269, 292
hyperboloid, 73, 83, 85, 113, 127, 267
hyperboloid type, 81
hyperplane, 124
hypersphere, 124
hypersurface element, 122

icosahedral group, 302
icosahedron, 303
immersion, 3, 5, 8, 55
implicit function, 3
index, 124
index form, 277
inflection point, 14, 49
inner product, 2, 98, 214
integrability conditions, 146, 147, 149, 155, 168, 264, 348
inverse mapping, 4
irreducible, 327
isometric, 59, 110, 161, 216
isometry, 216
isometry group, 326
isothermal, 99
isotropic, 34, 35, 114, 131
isotropy group, 326

Jacobi determinant, 64
Jacobi equation, 284
Jacobi field, 284, 286, 307
Jacobi identity, 220, 243
Jacobian, 3

Killing field, 219
Klein bottle, 201, 218
Koszul formula, 343
Kuiper, N. H., 185

Lagrange multiplier, 72
Laplace-Beltrami operator, 252
Laplacian, 252, 360
length preserving, 161
lens space, 304, 308
level point, 72, 104, 105
Levi-Civita connection, 221
Lie algebra, 227
Lie bracket, 137, 219, 231
Lie derivative, 219
Lie group, 227
Lie, S., 219
Liebmann, H., 47, 189, 191
light-cone, 34, 113
light-like, 34
light-like line, 35
line, 9, 84, 121
line of curvature, 76

lines of curvature parameters, 76, 103
locally compact, 204
locally isometric, 249
logarithmic spiral, 51
Lorentz group, 270
Lorentz rotation, 117
Lorentz space, 33, 270
Lorentz transformation, 271
Lorentzian metric, 214, 218

manifold, 199
Maurer-Cartan equations, 168
mean curvature, 72, 98, 123, 129
mean curvature vector, 100
measure tensor, 234
Mercator projection, 127
meridian curve, 77
meromorphic, 105, 204
metric tensor, 234
Meusnier, M., 71
minimal surface, 98, 131
Minkowski space, 33, 113, 214, 270
Möbius strip, 65
Möbius, A., 65
Monge coordinates, 73, 75, 124
Monge surface, 132
monkey saddle, 73
multilinear, 236
multiplicity, 307
mylar balloon, 95

Neil parabola, 19
non-Euclidean geometry, 120
norm, 2
normal coordinates, 281
normal curvature, 71, 171
normal plane, 19
normal section, 71
normal space, 6, 56
normal variation, 96
normal vector, 14, 57, 66, 115, 122
null cubic, 371
null vector, 34, 113, 214, 266
null-cone, 113

oblate sphere, 81
octahedral group, 302
octahedral space, 304
octahedron, 303
orientability, 63
orthogonal group, 201, 270
osculating plane, 19, 71
osculating sphere, 20
ovaloid, 182

parabola, 19
parabola of contact, 50
parabolic point, 72
paraboloid, 73, 127
parallel, 141, 225, 263
parallel displacement, 142, 225
parallel surface, 65, 129
parameter, 56
parameter of distribution, 86
parameter transformation, 64
parametrization, 5, 56, 198
parametrized curve, 8
partition of unity, 217
Petrov type, 358
Pfaffian form, 166
Poincaré upper half-plane, 194, 216, 267, 306
polar angle function, 39, 42
polar coordinates, 38, 272, 273
polarization, 246
position vector, 62
potential equation, 346
primitive, 103
principal curvature, 71, 123, 253, 258
principal normal, 17, 193
prism space, 304
product rule, 137, 206, 221, 239
profile curve, 77
projective plane, 201, 216
pseudo-Euclidean space, 37, 266
pseudo-hyperbolic space, 269
pseudo-Riemannian metric, 214
pseudo-sphere, 82, 269

quadratic integral, 60
quaternion algebra, 300
quaternion group, 305
quaternion space, 304, 305
quaternionic projective space, 330
quaternions, 300

rank, 3
rank theorem, 4
rectifying developable, 128
rectifying plane, 20
relativity theory, 214
relativity, special, 33
Ricci calculus, 207
Ricci curvature, 255, 260, 286
Ricci flow, 253
Ricci tensor, 252, 331
Ricci, G., 207, 241
Riemann sphere, 204
Riemann, B., 197, 263
Riemannian connection, 221
Riemannian manifold, 213
Riemannian metric, 213
Rodrigues, O., 72
rotation group, 301
rotation index, 39, 40
rotation matrix, 202
rotational torus, 58, 129
ruled surface, 77, 84, 92, 119
ruling, 84, 132, 259, 264, 368

saddle point, 75
scalar curvature, 152, 193, 252, 260, 317, 331
scaling, 249
Scherk, H. F., 111
Schmidt orthogonalization, 13
Schouten tensor, 347
Schouten, J. A., 347
Schur, F., 248
Schwarz, H. A., 112
Schwarzschild metric, 231, 359
screw-motion, 10, 87, 131, 372
scroll, 77
second fundamental form, 68, 116, 237, 261
sectional curvature, 151, 242, 245–247, 285, 351
self-adjoint, 67, 264, 333
semi-Riemannian metric, 214
shape operator, 67
shortest path, 144, 279
singularity, 56, 80, 82, 90, 95, 102
slope line, 23, 52
space curve, 17
space form, 291
space-like, 34, 114, 266
space-time, 309, 311, 324, 353
sphere, 57, 66, 73, 75, 80, 81, 95, 123, 127, 195, 266
spherical coordinates, 60, 123
spherical curve, 20
spherical dodecahedral space, 304
spiral, 51
square torus, 294
standard parameters, 84
Stiefel manifold, 14
Stokes, G., 170, 320
striction line, 85, 86
structural equations, 168
structure, 202
submanifold, 5, 57
submersion, 3, 5
surface, 55, 58, 179, 180
surface area, 61, 98
surface classification, 179
surface element, 56
surface integral, 61
surface of revolution, 77
surface of rotation, 77, 117
symmetries, 269

tangent, 8
tangent bundle, 5, 229, 308
tangent developable, 88
tangent hyperplane, 122
tangent plane, 56
tangent space, 5, 6, 57, 205
tangent surface, 90
tangent vector, 8, 17, 57, 206
Taylor expansion, 12, 18
Tchebychev grid, 126
tensor, 235, 236
tensor field, 235
tensor product, 236
tetrahedral group, 302
tetrahedron, 303
theorem on turning tangents, 41, 173
Theorema Egregium, 148, 151, 157, 233
Theorema Elegantissimum, 176, 378

theory of relativity, 317
third fundamental form, 68
tightness, 184, 186, 187
time-like, 34, 114, 266
topological manifold, 203
topology, 2, 203
torse, 88
torsion, 17, 20, 27, 36
torsion tensor, 221
torus, 58, 129, 201, 215, 218
torus knot, 32
torus of revolution, 58, 129
total absolute curvature, 43, 45, 47, 182
total curvature, 38, 40, 185, 321
total mean curvature, 129
totally umbilical, 75, 124
trace, 250
tractrix, 11, 82
transition function, 200
truncated cube space, 304
tubular surface, 132

umbilic, 72, 75, 124

variation, 98
variation of a metric, 313
variation of arc length, 277
vector field, 62, 212, 237
vector space, 2
vertex, 46

warped product, 195, 230, 261
wedge product, 167
Weierstrass representation, 106
Weingarten equation, 140
Weingarten map, 67, 116, 122, 237, 242, 268
Weingarten surface, 93, 130, 132
Wente-torus, 191
Weyl tensor, 341, 342, 347
Weyl, H., 347
Willmore conjecture, 129
winding number, 39
Wolf, J., 328

SELECTED PUBLISHED TITLES IN THIS SERIES

77 **Wolfgang Kühnel,** Differential Geometry: Curves — Surfaces — Manifolds, Third Edition, 2015

76 **John Roe,** Winding Around, 2015

75 **Ida Kantor, Jiří Matoušek, and Robert Šámal,** Mathematics++, 2015

74 **Mohamed Elhamdadi and Sam Nelson,** Quandles, 2015

73 **Bruce M. Landman and Aaron Robertson,** Ramsey Theory on the Integers, Second Edition, 2014

72 **Mark Kot,** A First Course in the Calculus of Variations, 2014

71 **Joel Spencer,** Asymptopia, 2014

70 **Lasse Rempe-Gillen and Rebecca Waldecker,** Primality Testing for Beginners, 2014

69 **Mark Levi,** Classical Mechanics with Calculus of Variations and Optimal Control, 2014

68 **Samuel S. Wagstaff, Jr.,** The Joy of Factoring, 2013

67 **Emily H. Moore and Harriet S. Pollatsek,** Difference Sets, 2013

66 **Thomas Garrity, Richard Belshoff, Lynette Boos, Ryan Brown, Carl Lienert, David Murphy, Junalyn Navarra-Madsen, Pedro Poitevin, Shawn Robinson, Brian Snyder, and Caryn Werner,** Algebraic Geometry, 2013

65 **Victor H. Moll,** Numbers and Functions, 2012

64 **A. B. Sossinsky,** Geometries, 2012

63 **María Cristina Pereyra and Lesley A. Ward,** Harmonic Analysis, 2012

62 **Rebecca Weber,** Computability Theory, 2012

61 **Anthony Bonato and Richard J. Nowakowski,** The Game of Cops and Robbers on Graphs, 2011

60 **Richard Evan Schwartz,** Mostly Surfaces, 2011

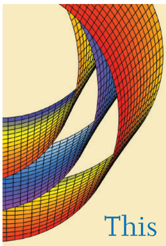
59 **Pavel Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina,** Introduction to Representation Theory, 2011

58 **Álvaro Lozano-Robledo,** Elliptic Curves, Modular Forms, and Their L-functions, 2011

57 **Charles M. Grinstead, William P. Peterson, and J. Laurie Snell,** Probability Tales, 2011

56 **Julia Garibaldi, Alex Iosevich, and Steven Senger,** The Erdős Distance Problem, 2011

For a complete list of titles in this series, visit the AMS Bookstore at **www.ams.org/bookstore/stmlseries/**.



This carefully written book is an introduction to the beautiful ideas and results of differential geometry. The first half covers the geometry of curves and surfaces, which provide much of the motivation and intuition for the general theory. The second part studies the geometry of general manifolds, with particular emphasis on connections and curvature. The text is illustrated with many figures and examples. The prerequisites are undergraduate analysis and linear algebra. This new edition provides many advancements, including more figures and exercises, and—as a new feature—a good number of solutions to selected exercises.

"This new edition is an improved version of what was already an excellent and carefully written introduction to both differential geometry and Riemannian geometry. In addition to a variety of improvements, the author has included solutions to many of the problems, making the book even more appropriate for use in the classroom."

– Colin Adams, Williams College

"This book on differential geometry by Kühnel is an excellent and useful introduction to the subject. . . . There are many points of view in differential geometry and many paths to its concepts. This book provides a good, often exciting and beautiful basis from which to make explorations into this deep and fundamental mathematical subject."

– Louis Kauffman, University of Illinois at Chicago

ISBN: 978-1-4704-2320-9

9 781470 423209

STML/77

For additional information and updates on this book, visit

www.ams.org/bookpages/stml-77

AMS on the Web
www.ams.org