DIFFERENTIAL GEOMETRY

Series of Lecture Notes and Workbooks for Teaching Undergraduate Mathematics

Algoritmuselmélet
Algoritmusok bonyolultsága
Analitikus módszerek a pénzügyekben
Bevezetés az analízisbe
Differential Geometry
Diszkrét optimalizálás
Diszkrét matematikai feladatok
Geometria
Igazságos elosztások
Interaktív analízis feladatgyűjtemény matematika BSc hallgatók számára
Introductory Course in Analysis
Matematikai pénzügy
Mathematical Analysis-Exercises 1-2
Mértékelmélet és dinamikus programozás
Numerikus funkcionálanalízis
Operációkutatás
Operációkutatási példatár
Optimális irányítások
Parciális differenciálegyenletek
Példatár az analízishez
Szimmetrikus kombinatorikai struktúrák
Többváltozós adatelemzés

Balázs Csikós

DIFFERENTIAL GEOMETRY

Eötvös Loránd University
Faculty of Science

Typotex

2014

Referee: Árpád Kurusa

ISBN 978 963 279 221 7
Prepared under the editorship of Typotex Publishing House
(http://www.typotex.hu)
Responsible manager: Zsuzsa Votisky
Technical editor: József Gerner

Made within the framework of the project Nr. TÁMOP-4.1.2-08/2/A/KMR-2009-0045, entitled "Jegyzetek és példatárak a matematika egyetemi oktatásához" (Lecture Notes and Workbooks for Teaching Undergraduate Mathematics).

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Regionális Fejlesztési Alap társfinanszírozásával valósul meg.

KEY WORDS: Curve, Frenet frame, curvature, torsion, hypersurface, fundamental forms, principal curvature, Gaussian curvature, Minkowski curvature, manifold, tensor field, connection, geodesic curve

SUMMARY: The aim of this textbook is to give an introduction to differential geometry. It is based on the lectures given by the author at Eötvös Loránd University and at Budapest Semesters in Mathematics. In the first chapter, some preliminary definitions and facts are collected, that will be used later. The classical roots of modern differential geometry are presented in the next two chapters. Chapter 2 is devoted to the theory of curves, while Chapter 3 deals with hypersurfaces in the Euclidean space. In the last chapter, differentiable manifolds are introduced and basic tools of analysis (differentiation and integration) on manifolds are presented. At the end of Chapter 4, these analytical techniques are applied to study the geometry of Riemannian manifolds.

Contents

Chapter 1

Preliminaries

In this chapter, we collect some definitions and facts that will be used later in the text.

1.1 Categories and Functors

We shall often use the term *natural map* or *natural isomorphism* between two sets carrying certain structures. The concept of naturality can be properly defined within the framework of category theory. Category theory yields a unified way to look at different areas of mathematics and their constructions.

Definition 1.1.1. A *category* $\mathcal{C}$ consists of

- a class $\mathrm{Ob}_{\mathcal{C}}$ of objects;
- an assignment of a set $\mathrm{Mor}_{\mathcal{C}}(X, Y)$ to any pair of objects $X, Y \in \mathrm{Ob}_{\mathcal{C}}$, the elements of which are called *morphisms, arrows* or *maps* from X to Y;
- a composition operation $\mathrm{Mor}_{\mathcal{C}}(X, Y) \times \mathrm{Mor}_{\mathcal{C}}(Y, Z) \to \mathrm{Mor}_{\mathcal{C}}(X, Z)$, $(f, g) \mapsto g \circ f$ for any three objects X, Y, Z.

These should satisfy the following axioms.

(i) For any four objects $X, Y, Z, W \in \mathrm{Ob}_{\mathcal{C}}$ and any morphisms $h \in \mathrm{Mor}_{\mathcal{C}}(X, Y)$, $g \in \mathrm{Mor}_{\mathcal{C}}(Y, Z)$ $f \in \mathrm{Mor}_{\mathcal{C}}(Z, W)$, we have

$$(f \circ g) \circ h = f \circ (g \circ h).$$

(ii) There is a (unique) morphism $1_X \in \mathrm{Mor}_{\mathcal{C}}(X, X)$ for any object X, called the identity morphism of X, such that for any morphisms $f \in$

$\mathrm{Mor}_{\mathcal{C}}(X,Y)$ and $g \in \mathrm{Mor}_{\mathcal{C}}(Y,X)$, the identities

$$f \circ 1_X = f \text{ and } 1_X \circ g = g$$

hold. ※

Definition 1.1.2. A morphism $f \in \mathrm{Mor}_{\mathcal{C}}(X,Y)$ is called an *isomorphism* if it has a two-sided inverse, that is a morphism $g \in \mathrm{Mor}_{\mathcal{C}}(Y,X)$ with $f \circ g = 1_Y$ and $g \circ f = 1_X$. Two objects are *isomorphic* if there is an isomorphism between them. The morphism f is an *endomorphism of* X if $Y = X$. The set $\mathrm{Mor}_{\mathcal{C}}(X,X)$ of endomorphisms of X is also denoted by $\mathrm{End}_{\mathcal{C}}(X)$. The morphism f is an *automorphism* if f is both an endomorphism and an isomorphism. Automorphisms of an object X form a group $\mathrm{Aut}_{\mathcal{C}}(X)$ with respect to the composition operation. ※

A basic example is the category of sets, in which the objects are the sets, $\mathrm{Mor}(X,Y)$ is the set of all maps from X to Y, $\circ$ is the ordinary composition of maps, 1_X is the identity map of X. Isomorphisms of this category are the bijective maps. Two sets are isomorphic in this category if and only if they have the same cardinality. $\mathrm{Aut}(X)$ is the group of *permutations* of the elements of X.

Definition 1.1.3. A *diagram* is a directed graph, the vertices of which are objects of a category and the edges are labeled by morphisms from the initial point of the edge to the endpoint. Any directed path in a diagram gives rise to morphism from the initial point of the path to its endpoint obtained as the composition of morphisms attached to the consecutive edges of the path. A diagram is said to be *commutative* when, for each pair of vertices X and Y and for any two directed paths from X to Y, the compositions of the edge labels of the paths are equal morphism from X to Y. ※

Definition 1.1.4. A *covariant functor* F from a category $\mathcal{C}$ to a category $\mathcal{D}$ associates to each object X of the category $\mathcal{C}$ an object $F(X) \in \mathrm{Ob}_{\mathcal{D}}$; and to each morphism $f \in \mathrm{Mor}_{\mathcal{C}}(X,Y)$ a morphism $F(f) \in \mathrm{Mor}_{\mathcal{D}}(F(X),F(Y))$ in such a way that

- $F(1_X) = 1_{F(X)}$ for any object $X \in \mathrm{Ob}_{\mathcal{C}}$;
- $F(g \circ f) = F(g) \circ F(f)$ for any $f \in \mathrm{Mor}_{\mathcal{C}}(X,Y)$ and $g \in \mathrm{Mor}_{\mathcal{C}}(Y,Z)$.※

Contravariant functors are defined similarly. The difference is that contravariant functors reverse arrows.

Definition 1.1.5. A *contravariant* functor F from a category $\mathcal{C}$ to a category $\mathcal{D}$ associates to each object X of the category $\mathcal{C}$ an object $F(X) \in \mathrm{Ob}_{\mathcal{D}}$; and to each morphism $f \in \mathrm{Mor}_{\mathcal{C}}(X,Y)$ a morphism $F(f) \in \mathrm{Mor}_{\mathcal{D}}(F(Y),F(X))$ in such a way that

- $F(1_X) = 1_{F(X)}$ for any object $X \in \mathrm{Ob}_{\mathcal{C}}$;
- $F(g \circ f) = F(f) \circ F(g)$ for any $f \in \mathrm{Mor}_{\mathcal{C}}(X, Y)$ and $g \in \mathrm{Mor}_{\mathcal{C}}(Y, Z)$.※

To each category $\mathcal{C}$, there is an *opposite category* $\mathcal{C}^{\mathrm{op}}$. This category has the same objects as $\mathcal{C}$ but for $A, B \in \mathrm{Ob}_{\mathcal{C}} = \mathrm{Ob}_{\mathcal{C}^{\mathrm{op}}}$, $\mathrm{Mor}_{\mathcal{C}^{\mathrm{op}}}(A, B) = \mathrm{Mor}_{\mathcal{C}}(B, A)$. The composition rule $\circ^{\mathrm{op}}$ in the opposite category is $f \circ^{\mathrm{op}} g = g \circ f$. With this construction, every contravariant functor from a category $\mathcal{C}$ to a category $\mathcal{D}$ can be thought of as a covariant functor form $\mathcal{C}^{\mathrm{op}}$ to $\mathcal{D}$ or from $\mathcal{C}$ to $\mathcal{D}^{\mathrm{op}}$.

Definition 1.1.6. Let F and G be two covariant functors from the category $\mathcal{C}$ to the category $\mathcal{D}$. Then a *natural transformation* Φ *from* F *to* G assigns to every $X \in \mathrm{Ob}_{\mathcal{C}}$ a morphism $\Phi_X \in \mathrm{Mor}_{\mathcal{D}}(F(X), G(X))$ so that for any morphism $f \in \mathrm{Mor}_{\mathcal{C}}(X, Y)$, the diagram

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ {\scriptstyle \Phi_X}\downarrow & & \downarrow{\scriptstyle \Phi_Y} \\ G(X) & \xrightarrow{G(f)} & G(Y) \end{array}$$

is commutative, i.e., $\Phi_Y \circ F(f) = G(f) \circ \Phi_X$. A *natural isomorphism between two functors* is a natural transformation Φ for which Φ_X is an isomorphism for all X objects of $\mathcal{C}$. ※

The definition of natural transformations can be extended also for the case, when one of the functors or both are contravariant. Then the contravariant functors should be substituted by their covariant counterparts going into the opposite categories. Natural transformations and natural isomorphisms can be defined by obvious modifications also for multivariable functors.
There is a principle in mathematics, that if there is an isomorphism between two objects of a category, then the two objects can be identified, as they behave in the same way and look like the same way from the view point of the category.
The importance of naturality is that if there is a natural isomorphism between two functors, then they can be identified with one another, or considered to be essentially the same.

1.2 Linear Algebra

1.2.1 Linear Spaces and Linear Maps

Definition 1.2.1. A set V is a *linear space*, or *vector space* over $\mathbb{R}$ if V is equipped with a binary operation $+$, and for each $\lambda \in \mathbb{R}$, the multiplication of elements of V with λ is defined in such a way that the following identities are satisfied:

(i) $(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z})$ (associativity);

(ii) $\exists\,\mathbf{0}\in V$ such that $\mathbf{x}+\mathbf{0}=\mathbf{x}$ for all $\mathbf{x}\in V$;

(iii) $\forall\,\mathbf{x}\in V\ \exists-\mathbf{x}\in V$ such that $\mathbf{x}+(-\mathbf{x})=\mathbf{0}$;

(iv) $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$ (commutativity);

(v) $\lambda(\mathbf{x}+\mathbf{y})=\lambda\mathbf{x}+\lambda\mathbf{y}$;

(vi) $(\lambda+\mu)\mathbf{x}=\lambda\mathbf{x}+\mu\mathbf{x}$;

(vii) $(\lambda\mu)\mathbf{x}=\lambda(\mu\mathbf{x})$;

(viii) $1\mathbf{x}=\mathbf{x}$. ※

Definition 1.2.2. A *linear combination* of some vectors $\mathbf{x}_1,\dots,\mathbf{x}_k\in V$ is a vector of the form

$$\lambda_1\mathbf{x}_1+\cdots+\lambda_k\mathbf{x}_k, \tag{1.1}$$

where the λ_i's are real numbers. The vectors $\mathbf{x}_1,\dots,\mathbf{x}_k\in V$ are *linearly independent* if the linear combination (1.1) can be $\mathbf{0}$ only if all of the coefficients λ_i vanish. A *basis* of V is a maximal set of linearly independent vectors. ※

It is known that any two bases of a linear space have the same cardinality.

Definition 1.2.3. The *dimension* $\dim V$ of the linear space V is the cardinality of a basis.

※

Definition 1.2.4. A map $L\colon V\to W$ between the linear spaces V and W is said to be *linear* if

$$L(\lambda\mathbf{x}+\mu\mathbf{y})=\lambda L(\mathbf{x})+\mu L(\mathbf{y})$$

for any $\mathbf{x},\mathbf{y}\in V$ and $\lambda,\mu\in\mathbb{R}$. A *linear isomorphism* is a bijective linear map. Two linear spaces are *isomorphic* if there is a linear isomorphism between them. ※

Linear spaces as objects and linear transformations as morphisms form a category. Two linear spaces are isomorphic if and only if they have the same dimension. The automorphism group of a linear space V is called the *general linear group of* V and it is denoted by $\mathrm{GL}(V)$.

If V is an n-dimensional linear space, and $(\mathbf{e}_1,\dots,\mathbf{e}_n)$ is a basis of V, then any vector $\mathbf{x}\in V$ can be written uniquely as a linear combination $\mathbf{x}=$

$x^1\mathbf{e}_1 + \cdots + x^n\mathbf{e}_n$ of the basis vectors. The numbers $(x^1, \dots, x^n)$ are called the *coordinates of* $\mathbf{x}$ *with respect to the basis* $(\mathbf{e}_1, \dots, \mathbf{e}_n)$.
The indices of the coordinates are not exponents, they are just upper indices. The reason why it is practical to use both upper and lower indices is the observation that if we position the indices properly in a linear algebraic formula, then usually summations go exactly over those indices that appear twice in a term, once as a lower index, once as an upper one. Therefore, if we take care of the right positioning of the indices, summation signs show redundant information and can be supressed. This leads to *Einstein's convention* which suggests us to position the indices properly and omit the summation signs. It is a rule for correct index positioning that if a single index appears on one side of an equation, then the same index must appear as a single index at the same (upper or lower) position on the other side as well. In this book, we shall pay attention to index positioning, but we shall not omit the summation signs.
Working with coordinates, linear transformations are represented by matrices. Let $L\colon V \to W$ be a linear map from the n-dimensional linear space V to the m-dimensional linear space W. Choose a basis $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ for V and a basis $(\mathbf{f}_1, \dots, \mathbf{f}_m)$ for W. Write $L(\mathbf{e}_i)$ as a linear combination $L(\mathbf{e}_i) = l_i^1\mathbf{f}_1 + \cdots + l_i^m\mathbf{f}_m$. Arranging the coefficients l_i^j into an $m \times n$ matrix

$$[L] = \begin{pmatrix} l_1^1 & \dots & l_n^1 \\ \vdots & \ddots & \vdots \\ l_1^m & \dots & l_n^m \end{pmatrix},$$

we obtain the *matrix of* L *with respect to the bases* $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ *and* $(\mathbf{f}_1, \dots, \mathbf{f}_m)$. If we arrange the coordinates of $\mathbf{x}$ with respect to the basis $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ into a column vector $[\mathbf{x}]$, then the column vector $[L(\mathbf{x})]$ of the coordinates of $L(\mathbf{x})$ with respect to the basis $(\mathbf{f}_1, \dots, \mathbf{f}_m)$ can be computed by the matrix multiplication $[L(\mathbf{x})] = [L][\mathbf{x}]$. For an endomorphism of V, we usually use the same basis for V and $W = V$.

Examples.

(1) Recall that $\mathbb{R}^n$ denotes the set of n-tuples of real numbers $\mathbb{R}^n = \{(x_1, \dots, x_n) | x_i \in \mathbb{R}\}$.

If $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ are two elements of $\mathbb{R}^n$ and $\lambda \in \mathbb{R}$ is a real number, then we define the *sum* and *difference* of $\mathbf{x}$ and $\mathbf{y}$ and the *scalar multiple* of $\mathbf{x}$ by

$$\begin{aligned} \mathbf{x} \pm \mathbf{y} &= (x_1 \pm y_1, \dots, x_n \pm y_n), \\ \lambda\mathbf{x} &= (\lambda x_1, \dots, \lambda x_n). \end{aligned}$$

It is clear that $\mathbb{R}^n$ is an n-dimensional linear space over the field of real numbers with respect to the operations defined above.

Let $\mathbf{e}_i$ denote the vector $(0,\dots,0,\overset{i}{1},0\dots0)$, the only non-zero coordinate of which is the ith one, being equal to 1. The n-tuple $(\mathbf{e}_1,\dots,\mathbf{e}_n)$ is a basis of $\mathbb{R}^n$ called the *standard basis of* $\mathbb{R}^n$.

(2) Let V and W be linear spaces, $\mathrm{Hom}(V,W)$ be the set of all linear maps from V to W. $\mathrm{Hom}(V,W)$ becomes a linear space if for $A,B\in\mathrm{Hom}(V,W)$ and $\lambda\in\mathbb{R}$, we define the maps $A+B$ and λA by

$$(A+B)(\mathbf{x})=A(\mathbf{x})+B(\mathbf{x}),\qquad (\lambda A)(\mathbf{x})=\lambda(A(\mathbf{x})).$$

(3) A *linear subspace* of a linear space V is a nonempty subspace $W\subseteq V$, which contains all linear combinations of its elements. Linear subspaces of a linear space are linear spaces themselves. The dimension of a linear subspace W of V is less than or equal to $\dim V$. If V is finite dimensional, then $\dim W=\dim V$ holds only if $V=W$.

The intersection of an arbitrary family of linear subspaces is a linear subspace, therefore, for any subset $S\subset V$ there is a unique smallest linear subspace among all linear subspaces containing S. We shall call this linear subspace *the linear subspace spanned* or *generated by* S, or simply the *linear hull of* S. We shall denote the linear hull of S by $\mathrm{lin}[S]$.

Definition 1.2.5. The *Grassmann manifold of k-dimensional linear subspaces of the linear space* V is the set $\mathrm{Gr}_k(V)$ of all k-dimensional subspaces of V. In the special case $k=1$, $\mathrm{P}(V)=\mathrm{Gr}_1(V)$ is also called the *projective space associated to* V. Later we shall introduce a topology and a manifold structure on $\mathrm{Gr}_k(V)$. Then the name Grassmann *manifold* will be justified. ※

(4) If $L\in\mathrm{Hom}(V,W)$ is a linear transformation, then the *image* $\mathrm{im}\,L=\{L(\mathbf{v})\mid \mathbf{v}\in V\}$ *of* L is a linear subspace in W, and the *kernel* $\ker L=\{\mathbf{v}\in V\mid L(\mathbf{v})=\mathbf{0}\}$ *of* L is a linear subspace in V. The *rank* $\mathrm{rk}\,L$ *of* L is the dimension of $\mathrm{im}\,L$.

(5) For an element $\mathbf{v}\in V$ of a linear space V, we define *translation* $T_{\mathbf{v}}$ *by* $\mathbf{v}$ as the map $T_{\mathbf{v}}\colon V\to V$, $T_{\mathbf{v}}(\mathbf{x})=\mathbf{x}+\mathbf{v}$. Translations by a non-zero vector are not linear transformations. If W is a linear subspace of V then two translates of W are either equal or disjoint. The set $V/W=\{T_{\mathbf{v}}(W)\mid \mathbf{v}\in V\}$ of translates of W carries a linear space structure defined by $\lambda T_{\mathbf{v}_1}(W)+\mu T_{\mathbf{v}_2}(W)=T_{\lambda\mathbf{v}_1+\mu\mathbf{v}_2}(W)$. (Check that the definition is correct.) The linear space V/W is called the *factor space of V with respect to the subspace* W. The surjective linear map $L\colon V\to$

V/W, $\mathbf{v} \mapsto T_{\mathbf{v}}(W)$ is the *factor map*. Every linear map $L\colon V \to W$ is a factor map $L\colon V \to V/(\ker L) \cong \operatorname{im} L$ onto its image. The dimension of the factor space or the image of a linear space can be computed by the formula

$$\begin{aligned} \dim(V/W) &= \dim(V) - \dim(W), \\ \operatorname{rk} L = \dim(\operatorname{im} L) &= \dim(V) - \dim(\ker L). \end{aligned} \tag{1.2}$$

(6) The linear space $V^* = \operatorname{Hom}(V, \mathbb{R})$ consisting of the linear functions on V is the *dual space of* V. Assigning the dual space V^* to a linear space V is a contravariant functor of the category of linear spaces into itself. This functor assigns to a linear map $L\colon V \to W$ the *adjoint map* $L^*\colon W^* \to V^*$, defined by $L^*(l) = l \circ L$, where $l \in V^*$.

If $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ is a basis of V, then the linear functions $\mathbf{e}^i \in V^*$, $(i = 1, \dots, n)$, defined by the equalities $\mathbf{e}^i(\mathbf{e}_j) = \delta^i_j$, where

$$\delta^i_j = \delta^j_i = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j \end{cases}$$

is the *Kronecker delta symbol*, form a basis of V^*. This basis is called the *dual basis of the basis* $(\mathbf{e}_1, \dots, \mathbf{e}_n)$.

We remark that the Kronecker delta symbol δ^j_i is also denoted by δ_{ij} and δ^{ij}. In formulae involving the Kronecker delta symbol, we always position the indices as dictated by the Einstein convention.

Though for a finite dimensional linear space V, the dual space V^* has the same dimension as V, there is no natural isomorphism between these two spaces. In other words, the identical functor on the category of finite dimensional linear spaces is not naturally isomorphic to the dual space functor. On the other hand, there is a natural transformation from the identical functor to the double dual space functor, given by the embeddings $\Phi_V\colon V \to V^{**}$, $(\Phi(\mathbf{v}))(l) = l(\mathbf{v})$, where $\mathbf{v} \in V$, $l \in V^*$. The restriction of this natural transformation onto the category of finite dimensional linear spaces is a natural isomorphism.
With the help of this natural isomorphism, elements of a finite dimensional linear space V can be identified with elements of V^{**}. With this identification, the dual basis of the dual basis of a basis of V will be equal to the original basis.

1.2.2 Determinant of Matrices and Linear Endomorphisms

Let us denote by $\mathfrak{S}_n$ the group of all permutations of the set $\{1, \dots, n\}$. For a permutation $\sigma \in \mathfrak{S}_n$, we denote by $\operatorname{sgn}\sigma$ the *sign of the permutation* σ, which can be defined by the formula

$$\operatorname{sgn}\sigma = \prod_{1\le i<j\le n} \frac{\sigma(j)-\sigma(i)}{j-i} \in \{-1,1\}.$$

Exercise 1.2.6. Show that the sign of a permutation is always equal to ± 1. Prove that $\operatorname{sgn}(\sigma_1 \circ \sigma_2) = \operatorname{sgn}(\sigma_1)\cdot\operatorname{sgn}(\sigma_2)$. ¤

Definition 1.2.7. The *determinant of an $n\times n$ matrix $A = (a_{ij})_{1\le i,j\le n}$* is the number

$$\det A = \det\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn}\end{pmatrix} = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}\sigma \cdot a_{1\sigma(1)}\cdots a_{n\sigma(n)}.$$ ※

Some important properties of the determinant are summarized in the following proposition.

Proposition 1.2.8. (1) *If all but one columns of a square matrix are fixed, then the determinant is a linear function of the varying column. This means that if we denote by $(\mathbf{a}_1,\dots,\mathbf{a}_n)$ the square matrix with column vectors $\mathbf{a}_1,\dots,\mathbf{a}_n\in\mathbb{R}^n$, then*

$$\det(\mathbf{a}_1,\dots,\lambda\mathbf{a}_j+\mu\bar{\mathbf{a}}_j,\dots,\mathbf{a}_n) = \\ \lambda\det(\mathbf{a}_1,\dots,\mathbf{a}_j,\dots,\mathbf{a}_n)+\mu\det(\mathbf{a}_1,\dots,\bar{\mathbf{a}}_j,\dots,\mathbf{a}_n).$$

(2) *When we permute the columns of a square matrix the determinant is multiplied by the sign of the permutation, i.e.,*

$$\det(\mathbf{a}_{\sigma(1)},\dots,\mathbf{a}_{\sigma(n)}) = \operatorname{sgn}\sigma\det(\mathbf{a}_1,\dots,\mathbf{a}_n) \text{ for all } \sigma\in\mathfrak{S}_n.$$

(3) *The value of the determinant does not change if an arbitrary multiple of a column is added to another column.*

(4) *The determinant of a matrix vanishes if and only if its columns are linearly dependent.*

(5) *A square matrix A and its transposition $A^\top$, that is the reflection of A in the main diagonal, have the same determinant*

$$\det A = \det A^\top.$$

As a consequence, properties (1)-(4) *hold also if the columns are replaced by rows.*

(6) *The determinant of an upper or lower triangular matrix is the product of the diagonal elements.*

Recall that the product AB of two $n \times n$ matrices A and B is also an $n \times n$ matrix. The determinants of A, B and AB are related to one another as follows.

Proposition 1.2.9.

$$\det(AB) = \det(A) \cdot \det(B).$$

Proposition 1.2.10. *Let $L\colon V \to V$ be a linear endomorphism of the linear space V. Choose two bases $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ and $(\mathbf{f}_1, \ldots, \mathbf{f}_n)$ of V and consider the matrices $[L]_{\mathbf{e}} = (l_i^j)_{1 \le i,j \le n}$ and $[L]_{\mathbf{f}} = (\tilde{l}_i^j)_{1 \le i,j \le n}$ with respect to these bases respectively. Then*

$$\det([L]_{\mathbf{e}}) = \det([L]_{\mathbf{f}}).$$

Proof. Let $S\colon V \to V$ be the invertible linear endomorphism which takes the basis vector $\mathbf{e}_i$ to the basis vector $\mathbf{f}_i$ for all i. Denote by $[S]_{\mathbf{e}} = (s_i^j)_{1 \le i,j \le n}$ the matrix of S with respect to the basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. Then

$$L(\mathbf{e}_i) = \sum_{j=1}^{n} l_i^j \mathbf{e}_j, \quad L(\mathbf{f}_i) = \sum_{j=1}^{n} \tilde{l}_i^j \mathbf{f}_j, \quad S(\mathbf{e}_i) = \mathbf{f}_i = \sum_{j=1}^{n} s_i^j \mathbf{e}_j,$$

and

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \tilde{l}_i^j s_j^k \mathbf{e}_k = \sum_{j=1}^{n} \tilde{l}_i^j S(\mathbf{e}_j) = L(\mathbf{f}_i) = L\left(\sum_{j=1}^{n} s_i^j \mathbf{e}_j\right) = \sum_{j=1}^{n} \sum_{k=1}^{n} s_i^j l_j^k \mathbf{e}_k.$$

Comparing the coefficients we obtain

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \tilde{l}_i^j s_j^k = \sum_{j=1}^{n} \sum_{k=1}^{n} s_i^j l_j^k \text{ for all } 1 \le i, k \le n, \tag{1.3}$$

which means that $[L]_{\mathbf{f}}[S]_{\mathbf{e}} = [S]_{\mathbf{e}}[L]_{\mathbf{e}}$. Taking the determinant of both sides we obtain

$$\det([L]_{\mathbf{f}}) \det([S]_{\mathbf{e}}) = \det([S]_{\mathbf{e}}) \det([L]_{\mathbf{e}}).$$

Since the columns of $[S]_{\mathbf{e}}$ are linearly independent, as they are the coordinate vectors of the basis vectors $\mathbf{f}_i$ with respect to the basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, the determinant of $[S]_{\mathbf{e}}$ is nonzero. Thus, equation (1.3) implies the proposition. □

Definition 1.2.11. The *determinant of a linear endomorphism* $L : V \to V$ is the determinant of the matrix of L with respect to an arbitrary basis of V.
※

The definition is correct according to the previous proposition.

Definition 1.2.12. The *complexification* $\mathbb{C} \otimes V$ *of a linear space* V *over* $\mathbb{R}$ is the set of formal linear combinations $\mathbf{v} + i\mathbf{w}$, where $\mathbf{v}, \mathbf{w} \in V$. ※

$\mathbb{C} \otimes V$ is a linear space of dimension $2 \dim V$, which contains V as a linear subspace. Elements of the complexification can also be multiplied by complex numbers as follows

$$(x + iy)(\mathbf{v} + i\mathbf{w}) = (x\mathbf{v} - y\mathbf{w}) + i(x\mathbf{w} + y\mathbf{v}).$$

A linear endomorphism of $L\colon V \to V$ can be extended to the complexification by the formula $L(\mathbf{v} + i\mathbf{w}) = L(\mathbf{v}) + iL(\mathbf{w})$.

Definition 1.2.13. A non-zero vector $\mathbf{z} = \mathbf{v}+i\mathbf{w} \neq \mathbf{0}$ of the complexification of a linear space V is an *eigenvector of the linear endomorphism* $L\colon V \to V$ if there is a complex number $\lambda \in \mathbb{C}$ such that $L(\mathbf{z}) = \lambda\mathbf{z}$. The number λ is called the *eigenvalue corresponding to the eigenvector* $\mathbf{z}$. A complex number is an eigenvalue of L if there is an eigenvector to which it corresponds. ※

Proposition 1.2.14. *A complex number* λ *is an eigenvalue of the linear transformation* $L\colon V \to V$ *if and only if* $\det(L - \lambda \operatorname{id}_V) = 0$*. An eigenvalue* λ *is real if and only if there is an eigenvector in* V *with eigenvalue* λ*.*

Definition 1.2.15. The *characteristic polynomial of a linear endomorphism* $L\colon V \to V$ is the polynomial $p_L(\lambda) = \det(L-\lambda \operatorname{id}_V)$. Similarly, the *characteristic polynomial of an* $n\times n$ *matrix* A is the polynomial $p_A(\lambda) = \det(A-\lambda I_n)$, where I_n is the $n \times n$ unit matrix. The *algebraic multiplicity of an eigenvalue* is its multiplicity as a root of the characteristic polynomial. ※

The coefficients of the characteristic polynomial of a matrix A can be expressed as polynomials of the matrix elements. They can also be expressed in terms of the eigenvalues $\lambda_1, \dots, \lambda_n$ using the factorization

$$p_A(\lambda) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda).$$

Comparing these expressions we can relate elementary symmetric polynomials of the eigenvalues to some matrix invariants. For example, the constant term of the characteristic polynomial of the matrix A is

$$\det A = p_A(0) = \lambda_1 \cdots \lambda_n.$$

The coefficient of $(-\lambda)^{n-1}$ in $p_A(\lambda)$ is equal to the sum of the diagonal elements of the matrix and also to the sum of the eigenvalues.

Definition 1.2.16. The *trace of a matrix* $A = (a_{ij})_{1\leq i,j\leq n}$ is the number

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii},$$

which is also equal to the sum of the eigenvalues of A. The *trace of a linear endomorphism* is the trace of its matrix with respect to an arbitrary basis. ※

1.2.3 Orientation of a Linear Space

Definition 1.2.17. Let $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ and $(\mathbf{w}_1, \ldots, \mathbf{w}_k)$ be two ordered bases of a linear space V. We say that they *have the same orientation* or they *define the same orientation* of V, if the $k \times k$ matrix (a_i^j) defined by the system of equalities

$$\mathbf{v}_i = \sum_{j=1}^{k} a_i^j \mathbf{w}_j \qquad \text{for } i = 1, 2, \ldots, k$$

has positive determinant. ※

"Having the same orientation" is an equivalence relation on ordered bases, and there are two equivalence classes. A choice of one of the equivalence classes, the elements of which will be called then *positively oriented bases*, is an *orientation of* V.

Definition 1.2.18. The *standard orientation of* $\mathbb{R}^n$ is the orientation defined by the ordered basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, where $\mathbf{e}_i = (0, \ldots, 0, \overset{i}{1}, 0 \ldots 0)$. ※

1.2.4 Tensor Product

Definition 1.2.19. Let V, W and Z be linear spaces. A map $B \colon V \times W \to Z$ is said to be a *bilinear map* if it is linear in both variables, i.e., if it satisfies the identities

$$\begin{aligned} B(\lambda\mathbf{v}_1 + \mu\mathbf{v}_2, \mathbf{w}) &= \lambda B(\mathbf{v}_1, \mathbf{w}) + \mu B(\mathbf{v}_2, \mathbf{w}) \text{ and} \\ B(\mathbf{v}, \lambda\mathbf{w}_1 + \mu\mathbf{w}_2) &= \lambda B(\mathbf{v}, \mathbf{w}_1) + \mu B(\mathbf{v}, \mathbf{w}_2). \end{aligned}$$

※

Definition 1.2.20. The *tensor product of the linear spaces V and W* is a linear space $V \otimes W$ together with a bilinear map $\otimes \colon V \times W \to V \otimes W$, $(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} \otimes \mathbf{w}$, such that for any bilinear map $B \colon V \times W \to Z$, there is a unique linear map $L \colon V \otimes W \to Z$ which makes the diagram

$$\begin{array}{ccc} V \times W & \xrightarrow{\otimes} & V \otimes W \\ \| & & \big\downarrow L \\ V \times W & \xrightarrow{B} & Z \end{array}$$

commutative. ※

One can consider the category of bilinear maps defined on $V \times W$, in which the objects are the bilinear maps, a morphism between the bilinear maps $B_1\colon V \times W \to Z_1$ and $B_2\colon V \times W \to Z_2$ is a linear map $L\colon Z_1 \to Z_2$, for which the diagram

$$\begin{array}{ccc} V \times W & \xrightarrow{B_1} & Z_1 \\ \| & & \downarrow L \\ V \times W & \xrightarrow{B_2} & Z_2 \end{array}$$

is commutative. In general, an object X of a category is called an *initial object*, if for any other object Y of the category, there is a unique morphism form X to Y. Using this terminology, the tensor product of the linear spaces V and W is the initial object of the category of bilinear maps on $V \times W$. It is a simple exercise playing with arrows, that up to isomorphism, a category can have at most one initial object. However, initial objects do not exist in all categories. Existence of initial objects are always shown by explicit constructions in the given category.

To construct the tensor product explicitly, one first considers the linear space $F_{V\times W}$ generated freely by the elements of $V \times W$. More explicitly, $F_{V\times W}$ is the linear space of all formal linear combinations $\lambda_1(\mathbf{v}_1, \mathbf{w}_1)+\cdots+\lambda_k(\mathbf{v}_k, \mathbf{w}_k)$ of some pairs $(\mathbf{v}_i, \mathbf{w}_i) \in V \times W$ with real coefficients $\lambda_i \in \mathbb{R}$. Then we take the smallest linear subspace Z of $F_{V\times W}$ that contains all elements of the form

$$(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) - (\mathbf{v}_1, \mathbf{w}) - (\mathbf{v}_2, \mathbf{w}), \quad (\mathbf{w}, \mathbf{v}_1 + \mathbf{v}_2) - (\mathbf{w}, \mathbf{v}_1) - (\mathbf{w}, \mathbf{v}_2),$$
$$\lambda(\mathbf{v}, \mathbf{w}) - (\lambda\mathbf{v}, \mathbf{w}), \quad \lambda(\mathbf{v}, \mathbf{w}) - (\mathbf{v}, \lambda\mathbf{w})$$

for all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in W$, $\lambda \in \mathbb{R}$. Set $V \otimes W = F_{V\times W}/Z$ and let $\otimes\colon V \times W \to V \otimes W$ be the composition of the embedding $V \times W \to F_{V\times W}$ and the factor map $F_{V\times W} \to F_{V\times W}/Z$. It is not difficult to check that the bilinear map $\otimes\colon V \times W \to V \otimes W$ is an initial object of the category of bilinear maps on $V \times W$.

It is known that if $\mathbf{e}_1, \dots, \mathbf{e}_n$ is a basis of V, $\mathbf{f}_1, \dots, \mathbf{f}_m$ is a basis of W, then the vectors $\{\mathbf{e}_i \otimes \mathbf{f}_j \mid 1 \le i \le n,\ 1 \le j \le m\}$ form a basis of $V \otimes W$. In particular, $\dim(V \otimes W) = \dim(V)\dim(W)$.

The tensor product construction can be thought of as a functor from the category of pairs of linear spaces to the category of linear spaces In the category of pairs of linear spaces a morphism from the pair (V_1, V_2) to the pair (W_1, W_2) is a pair (L_1, L_2) of linear maps, where $L_1\colon V_1 \to W_1$ and $L_2\colon V_2 \to W_2$. Tensor product as a functor assigns to a pair (V_1, V_2) the tensor product $V_1 \otimes V_2$ and to the pair of linear maps (L_1, L_2) the linear map $L_1 \otimes L_2\colon V_1 \otimes V_2 \to W_1 \otimes W_2$, where the *tensor product* $L_1 \otimes L_2$ *of the linear maps* L_1 *and* L_2 is defined as the unique linear map for which

$$L_1 \otimes L_2(\mathbf{v}_1 \otimes \mathbf{v}_2) = L_1(\mathbf{v}_1) \otimes L_2(\mathbf{v}_2).$$

Definition 1.2.21. Let V be an n dimensional linear space and V^* its dual space. The tensor product $T^{(k,l)}V = \underbrace{V^* \otimes \cdots \otimes V^*}_{k \text{ times}} \otimes \underbrace{V \otimes \cdots \otimes V}_{l \text{ times}}$ will be called the *linear space of tensors of type* (k,l). We agree that $T^{(0,0)}V$ is equal to the ground field $\mathbb{R}$. ※

If $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis of V, $\mathbf{e}^1, \ldots, \mathbf{e}^n$ is its dual basis, then the type (k,l) tensors $\mathbf{e}^{i_1 \ldots i_k}_{j_1 \ldots j_l} = \mathbf{e}^{i_1} \otimes \cdots \otimes \mathbf{e}^{i_k} \otimes \mathbf{e}_{j_1} \otimes \cdots \otimes \mathbf{e}_{j_l}$ form a basis of $T^{(k,l)}V$. In the special case $k = l = 0$ the basis vector $\mathbf{e}$ of $T^{(0,0)}$ is the unit element of $\mathbb{R}$. The direct sum $\bigoplus_{k,l=0}^{\infty} T^{(k,l)}V$ can be equipped with a bilinear associative tensor multiplication, which turns it into an associative algebra. Tensor product is defined on the basis vectors by the formula

$$\mathbf{e}^{i_1 \ldots i_k}_{j_1 \ldots j_l} \otimes \mathbf{e}^{p_1 \ldots p_r}_{q_1 \ldots q_s} = \mathbf{e}^{i_1 \ldots i_k p_1 \ldots p_r}_{j_1 \ldots j_l q_1 \ldots q_s}.$$

If $T = \sum_{i_1,\ldots,i_k,j_1,\ldots,j_l=1}^{n} T^{j_1 \ldots j_l}_{i_1 \ldots i_k} \mathbf{e}^{i_1 \ldots i_k}_{j_1 \ldots j_l}$ is a tensor of type (k,l), then the numbers $T^{j_1 \ldots j_l}_{i_1 \ldots i_k}$ are called the *coordinates* or *components of the tensor T with respect to the basis* $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Exercise 1.2.22. How are the coordinates of a tensor transformed when we change the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to another one $\mathbf{f}_1, \ldots, \mathbf{f}_n$, where $\mathbf{f}_i = \sum_{j=1}^{n} a_i^j \mathbf{e}_j$? Show that if (b_i^j) is the inverse matrix of the matrix (a_i^j), then

$$\mathbf{f}^i = \sum_{j=1}^{n} b_j^i \mathbf{e}^j, \quad \mathbf{f}^{i_1 \ldots i_k}_{j_1 \ldots j_l} = \sum_{p_1,\ldots,p_k,q_1,\ldots,q_l=1}^{n} a^{i_1}_{p_1} \ldots a^{i_k}_{p_k} \cdot b^{q_1}_{j_1} \ldots b^{q_k}_{j_k} \cdot \mathbf{e}^{p_1 \ldots p_k}_{q_1 \ldots q_l}.$$ ¤

Exercise 1.2.23. Find all type $(0,2)$ and type $(1,1)$ tensors, the coordinates of which do not depend on the choice of the basis. ¤

Exercise 1.2.24. The coordinates of a type $(0,2)$ or a type $(1,1)$ tensor can always be arranged into an $n \times n$ matrix. Do the trace and determinant of this matrix depend on the choice of the basis? ¤

Exercise 1.2.25. A type $(2,0)$ or a type $(0,2)$ tensor is said to be *non-degenerate* if the $n \times n$ matrix built from its coordinates with respect to a basis has non-zero determinant. Show that non-degeneracy does not depend on the choice of the basis. ¤

Exercise 1.2.26. Let ξ be a non-degenerate tensor of type $(0,2)$. Show that there is a unique type $(2,0)$ tensor η such that the $n \times n$ matrices built from the coordinates of ξ and η with respect to any basis are inverses of one another. ¤

Exercise 1.2.27. Construct natural isomorphisms between the following linear spaces:

(a) $(V \otimes W)^* \cong V^* \otimes W^*$;

(b) $\mathrm{Hom}(V, W) \cong V^* \otimes W$, in particular, $\mathrm{End}(V) \cong T^{(1,1)}V$ and $\mathrm{Hom}(V, V^*) \cong T^{(2,0)}V$;

(c) $(T^{(k,l)}V)^* \cong T^{(k,l)}V^* \cong T^{(l,k)}V$;

(d) $\{\underbrace{V \times \cdots \times V}_{k \text{ times}} \times \underbrace{V^* \times \cdots \times V^*}_{l \text{ times}} \to \mathbb{R}\ (k+l)\text{-linear functions}\} \cong T^{(k,l)}V$;

(e) $\{\underbrace{V \times \cdots \times V}_{k \text{ times}} \to W\ k\text{-linear maps into } W\} \cong T^{(k,0)}V \otimes W$;

(f) $\{\underbrace{V \times \cdots \times V}_{k \text{ times}} \to V\ k\text{-linear maps into } V\} \cong T^{(k,1)}V$;

(g) $\mathrm{Hom}(T^{(k,l)}V, T^{(p,q)}V) \cong T^{(l+p,k+q)}V$. ¤

Remark. We explain what naturality of an isomorphism means for case (b). For other cases a similar definition can be given. Consider the category of pairs of linear spaces, in which the objects are pairs (V, W) of linear spaces, the morphisms from (V_1, W_1) to (V_2, W_2) are pairs (Φ, Ψ) of *isomorphisms* $\Phi\colon V_1 \to V_2$ and $\Psi\colon W_1 \to W_2$. Both $F_1\colon (V, W) \mapsto \mathrm{Hom}(V, W)$ and $F_2\colon (V, W) \mapsto V^* \otimes W$ are functors from this category to the category of linear spaces. If (Φ, Ψ) is a morphism from (V_1, W_1) to (V_2, W_2), then the linear map $F_1(\Phi, \Psi)\colon \mathrm{Hom}(V_1, W_1) \to \mathrm{Hom}(V_2, W_2)$ assigns to the linear map $L\colon V_1 \to W_1$ the linear map $\Psi \circ L \circ \Phi^{-1} \in \mathrm{Hom}(V_2, W_2)$, while $F_2(\Phi, \Psi)\colon V_1^* \otimes W_1 \to V_2^* \otimes W_2$ is the linear map $(\Phi^{-1})^* \otimes \Psi$. The statement that there is a natural isomorphism between $\mathrm{Hom}(V, W)$ and $V^* \otimes W$ means that there is a natural isomorphism between the functors F_1 and F_2.
There is a more practical (but less formal) way to characterize natural isomorphisms. A natural isomorphism between two linear spaces is an isomorphism for which the image of an element can be described *uniquely* by a set of instructions or formulae. If the definition of an isomorphism involves the random choice of a basis, for example, then it may not be natural, since the image of an element may depend on the choice of the basis. On the other hand, the definition of a natural isomorphism is allowed to contain random choices, but to prove naturality, we have to check that the image of an element does not depend on the random variables.

1.2.5 Exterior Powers

Denote by $\mathfrak{S}_k$ the group of all permutations of the set $\{1, \dots, k\}$.

Definition 1.2.28. Let V and W be linear spaces, $k \in \mathbb{N}$. A k-linear map

$$K\colon V^k = \underbrace{V \times \cdots \times V}_{k \text{ times}} \to W$$

is said to be *alternating* if for any permutation $\sigma \in \mathfrak{S}_k$ and any k vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$, we have

$$K(\mathbf{v}_{\sigma(1)}, \dots, \mathbf{v}_{\sigma(k)}) = \operatorname{sgn}\sigma \cdot K(\mathbf{v}_1, \dots, \mathbf{v}_k).$$ ※

Alternating k-linear maps on a given linear space V form a category. A morphism from an alternating k-linear map $K_1\colon V^k \to W_1$ to another one $K_2\colon V^k \to W_2$ is a linear map $L\colon W_1 \to W_2$ such that $K_2 = L \circ K_1$.

Definition 1.2.29. The *kth exterior power of a linear space* V is an initial object of the category of alternating k-linear maps on V. In other words, it is a linear space $\Lambda^k V$ together with an alternating k-linear map $\wedge_k\colon V^k \to \Lambda^k V$, $(\mathbf{v}_1, \dots, \mathbf{v}_k) \mapsto \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$, such that for any alternating k-linear map $K\colon V^k \to W$, there is a unique linear map $L\colon \Lambda^k V \to W$ for which the diagram

$$\begin{array}{ccc} V^k & \xrightarrow{\wedge_k} & \Lambda^k V \\ \Big\| & & \Big\downarrow L \\ V^k & \xrightarrow{K} & W \end{array}$$

is commutative.

The element $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k \in \Lambda^k V$ is called the *exterior product* or *wedge product* of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. Elements of $\Lambda^k V$ are called k-vectors. The words *bivector* and *trivector* are also used for 2-vectors and 3-vectors respectively. ※

Uniqueness of the exterior power up to isomorphism follows from uniqueness of initial objects. Existence is proved by an explicit construction as follows. Consider the kth tensor power $T^{(0,k)}V$ and the k-linear map $\otimes_k\colon V^k \to T^{(0,k)}V$, $(\mathbf{v}_1, \dots, \mathbf{v}_k) \mapsto \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k$. By the universal property of the tensor product, this is an initial object of the category of k-linear maps on V. Denote by $W_k < T^{(0,k)}V$ the linear hull of the set of elements of the form

$$\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k - \operatorname{sgn}\sigma \cdot \mathbf{v}_{\sigma(1)} \otimes \cdots \otimes \mathbf{v}_{\sigma(k)}, \text{ where } \mathbf{v}_i \in V,\ \sigma \in \mathfrak{S}_k.$$

Let $\Lambda^k V$ be the factor space $T^{(0,k)}V/W_k$ and $\wedge_k\colon V^k \to \Lambda^k V$ be the composition of $\otimes_k$ with the factor map $\phi_k\colon T^{(0,k)}V \to T^{(0,k)}V/W_k$. It can be checked that $\wedge_k\colon V^k \to \Lambda^k V$ is indeed an initial object of the category of alternating k-linear maps on V.

Proposition 1.2.30. *Assume that* $\mathbf{e}_1, \dots, \mathbf{e}_n$ *is a basis of* V*. Then*

$$\{\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_k} \mid 1 \leq i_1 < \cdots < i_k \leq n\}$$

is a basis of $\Lambda^k V$*. In particular,* $\dim \Lambda^k V = \binom{n}{k}$.

The following formula has many applications.

Proposition 1.2.31. *Suppose that the vectors* $\mathbf{v}_1, \dots, \mathbf{v}_k$ *can be expressed as a linear combination of the vectors* $\mathbf{w}_1, \dots, \mathbf{w}_k$ *as follows*

$$\mathbf{v}_i = c_i^1 \mathbf{w}_1 + \cdots + c_i^k \mathbf{w}_k, \ (i = 1, \dots, k).$$

Then

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k = \det \begin{pmatrix} c_1^1 & \cdots & c_1^k \\ \vdots & \ddots & \vdots \\ c_k^1 & \cdots & c_k^k \end{pmatrix} \mathbf{w}_1 \wedge \cdots \wedge \mathbf{w}_k.$$

The next statement is a corollary of the previous two ones.

Proposition 1.2.32. *The vectors* $\mathbf{v}_1, \dots, \mathbf{v}_k$ *are linearly independent if and only if* $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k \neq \mathbf{0}$*. Two linearly independent* k*-tuples of vectors* $\mathbf{v}_1, \dots, \mathbf{v}_k$ *and* $\mathbf{w}_1, \dots, \mathbf{w}_k$ *span the same* k*-dimensional linear subspace if and only if* $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k = c \cdot \mathbf{w}_1 \wedge \cdots \wedge \mathbf{w}_k$ *for some* $c \in \mathbb{R} \setminus \{0\}$*. In addition, these two collections are bases of the same orientation in the linear space they span if and only if* $c > 0$*.*

Corollary 1.2.33. *The Grassmann manifold* $\mathrm{Gr}_k(V)$ *can be embedded into the projective space* $\mathrm{P}(\Lambda^k V)$ *by assigning to the* k*-dimensional subspace spanned by the linearly independent vectors* $\mathbf{v}_1, \dots, \mathbf{v}_k$ *the* 1*-dimensional linear space spanned by* $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$*. This embedding is called the* Plücker embedding.

Corollary 1.2.34. *There is a natural one-to-one correspondence between orientations of a* k*-dimensional linear subspace* W *of an* n*-dimensional linear space* V *and orientations of the* 1*-dimensional linear subspace of* $\Lambda^k(V)$ *assigned to* W *by the Plücker embedding.*

The direct sum $\Lambda^*(V) = \bigoplus_{k=0}^{\dim V} \Lambda^k(V)$ of all the exterior powers of V becomes an associative algebra with multiplication $\tilde{\wedge}$ defined uniquely by the rule

$$(\wedge_k(\mathbf{v}_1, \dots, \mathbf{v}_k)) \tilde{\wedge} (\wedge_l(\mathbf{v}_{k+1}, \dots, \mathbf{v}_{k+l})) = \wedge_{k+l}(\mathbf{v}_1, \dots, \mathbf{v}_{k+l})$$
$$\forall\, k, l \in \mathbb{N}; \mathbf{v}_1, \dots, \mathbf{v}_{k+l} \in V.$$

As we have

$$\wedge_k(\mathbf{v}_1, \dots, \mathbf{v}_k) = \mathbf{v}_1 \tilde{\wedge} \dots \tilde{\wedge} \mathbf{v}_k$$

for any k vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$, we may (and we shall) denote the multiplication $\tilde{\wedge}$ simply by $\wedge$ without causing confusion with the earlier notation $\wedge_k(\mathbf{v}_1, \dots, \mathbf{v}_k) = \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$.

Definition 1.2.35. The associative algebra $(\Lambda^*(V), \wedge)$ is the *exterior algebra* or *Grassmann algebra* of V. ※

Exterior Powers and Alternating Tensors

Let us analyze the factor map $\phi_k \colon T^{(0,k)}(V) \to \Lambda^k(V)$ to obtain another construction of the kth exterior power of V, which is, of course, naturally isomorphic to the first construction. The permutation group $\mathfrak{S}_k$ has a representation on the tensor space $T^{(0,k)}V$. The representation $\Phi \colon \mathfrak{S}_k \to \mathrm{GL}(T^{(0,k)}V)$, $\sigma \mapsto \Phi_\sigma$ is given on decomposable tensors as follows:

$$\Phi_\sigma(\mathbf{v}_{i_1} \otimes \cdots \otimes \mathbf{v}_{i_k}) = \mathbf{v}_{\sigma(i_1)} \otimes \cdots \otimes \mathbf{v}_{\sigma(i_k)}.$$

Definition 1.2.36. A tensor T of type $(0,k)$ is called *symmetric* if $\Phi_\sigma(T) = T$ for any permutation $\sigma \in \mathfrak{S}_k$. T is said to be *alternating* if $\Phi_\sigma(T) = \operatorname{sgn}(\sigma) \cdot T$ for all $\sigma \in \mathfrak{S}_k$. We shall use the notation $S_k(V) \subset T^{(0,k)}(V)$ for the linear space of symmetric tensors, and $A_k(V) \subset T^{(0,k)}(V)$ for the linear space of alternating tensors. ※

Exercise 1.2.37. Show that $T^{(0,2)}(V) = S_2(V) \oplus A_2(V)$. ¤

Exercise 1.2.38. Show that $\dim S_k(V) = \binom{n+k-1}{k}$, where $n = \dim V$. ¤

Hint: Find an isomorphism between $S_k(V)$ and homogeneous polynomials of degree k in n variables.

Exercise 1.2.39. Compute the dimension of $A_k(V)$. ¤

Define the linear map $\pi_k \colon T^{(0,k)}V \to T^{(0,k)}(V)$ by the formula

$$\pi_k(T) = \sum_{\sigma \in \mathfrak{S}_k} \operatorname{sgn}(\sigma) \Phi_\sigma(T).$$

Proposition 1.2.40. *The image of the map π_k is $A_k(V)$. The kernel of π_k is $W_k = \ker \phi_k$, where $\phi_k \colon T^{(0,k)}V \to T^{(0,k)}V/W_k = \Lambda^k V$ is the factor map we defined in the construction of $\Lambda^k V$. The map $\pi_k/k!$ is a projection onto $A_k(V)$.*

Proof. It is not difficult to check that $\operatorname{im} \pi_k$ contains only alternating tensors, and if T is alternating, then $\pi_k(T) = k!T$, so $\operatorname{im} \pi_k = A_k(V)$ and $(\pi_k)/k!$ is a projection of $T^{(0,k)}V$ onto $A_k(V)$.

If $\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k - \operatorname{sgn}\sigma \cdot \Phi_\sigma(\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k)$ is a generator of W_k, then denoting $\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k$ by T

$$\pi_k(T - \operatorname{sgn}\sigma \cdot \Phi_\sigma(T))) = \sum_{\sigma' \in \mathfrak{S}_k} \operatorname{sgn}\sigma' \cdot \Phi_{\sigma'}(T) - \sum_{\sigma'' = \sigma' \circ \sigma \in \mathfrak{S}_k} \operatorname{sgn}\sigma'' \cdot \Phi_{\sigma''}(T) = 0,$$

thus, $\ker \pi_k \supseteq W_k$. Conversely, if $T \in \ker \pi_k$, then

$$T = \frac{1}{k!} \sum_{\sigma \in \mathfrak{S}_k} (T - \operatorname{sgn}\sigma \cdot \Phi_\sigma(T)).$$

Since the image of the linear map $(I - \operatorname{sgn}\sigma \cdot \Phi_\sigma)$ is in W_k for any $\sigma \in \mathfrak{S}_k$ by the definition of W_k, the above expression for T shows that T is also in W_k. □

According to the proposition,

$$T^{(0,k)}(V) = A_k(V) \oplus W_k,$$

therefore, the factor space $\Lambda^k(V) = T^{(0,k)}(V)/W_k$ is naturally isomorphic to $A_k(V)$. There are two different ways to identify these two linear spaces. One is to identify them with the restriction $\beta_k = \phi_k|_{A_k(V)}$ of the factor map $\phi_k \colon T^{(0,k)(V) \to \Lambda^k(V)}$ onto $A_k(V)$.
We can also define another natural isomorphism using the universal property of the exterior power. Since the map

$$\pi_k \circ \otimes_k \colon V^k \to A_k(V), \quad (\mathbf{v}_1, \dots, \mathbf{v}_k) \mapsto \pi_k(\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k)$$

is an alternating k-linear map, by the universal property of the exterior power, there is an induced linear map $\alpha_k \colon \Lambda^k(V) \to A^k(V)$ such that

$$\alpha_k(\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_k) \mapsto \pi_k(\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k).$$

Since

$$\beta_k(\alpha_k(\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_k) = \sum_{\sigma \in \mathfrak{S}_k} \operatorname{sgn}\sigma \cdot \mathbf{v}_{\sigma(1)} \wedge \dots \wedge \mathbf{v}_{\sigma(k)} = k! \cdot \mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_k,$$

$\alpha_k^{-1} = \beta_k/k!$.
It is a matter of taste which isomorphism is used to identify wedge products of vectors with alternating tensors, but to avoid confusion, we should choose one of the identifications and then insist on that for the rest of the discussion. Making this decision, *from now on we use α_k to identify $\Lambda^k(V)$ and $A_k(V)$.* Nevertheless, one should be aware of the fact that some authors prefer the identification by β_k.

Taking the direct sum of the linear isomorphisms $\alpha_k \colon \Lambda^k(V) \to A_k(V)$, we obtain a linear isomorphism $\alpha_* \colon \Lambda^*(V) \to A_*(V)$ between the direct sum $A_*(V) = \bigoplus_{k=0}^{\dim V} A_k(V)$ and the Grassmann algebra of V. Using this isomorphism, we can define an associative multiplication $\wedge$ on $A_*(V)$ setting $T_1 \wedge T_2 = \alpha_*(\alpha_*^{-1}(T_1) \wedge \alpha_*^{-1}(T_2))$ for $T_1, T_2 \in A_*(V)$. More explicitly, if $T_1 \in A_k(V)$, $T_2 \in A_l(V)$, then $T_1 = \pi_k(T_1)/k!$, $T_2 = \pi_l(T_2)/l!$, and

$$\begin{aligned} T_1 \wedge T_2 &= \frac{1}{k! \cdot l!}\alpha_{k+l}(\beta_k(T_1) \wedge \beta_l(T_2)) = \frac{1}{k! \cdot l!}\alpha_{k+l}(\phi_k(T_1) \wedge \phi_l(T_2)) \\ &= \frac{1}{k! \cdot l!}\alpha_{k+l}(\phi_{k+l}(T_1 \otimes T_2)) \\ &= \frac{1}{k! \cdot l! \cdot (k+l)!}\alpha_{k+l}(\phi_{k+l}(\pi_{k+l}(T_1 \otimes T_2))) \\ &= \frac{1}{k! \cdot l! \cdot (k+l)!}\alpha_{k+l}(\beta_{k+l}(\pi_{k+l}(T_1 \otimes T_2))) = \frac{1}{k! \cdot l!}\pi_{k+l}(T_1 \otimes T_2). \end{aligned}$$

Exercise 1.2.41. How should we modify the formula for $T_1 \wedge T_2$ if we used the isomorphisms β_k to identify $A_*(V)$ with $\Lambda^*(V)$? ¤

Exercise 1.2.42. Find a natural isomorphism

$$A^k(V_1 \oplus V_2) \cong \bigoplus_{r+s=k} A^r(V_1) \otimes A^s(V_2).$$ ¤

Exterior Powers of the Dual Space and Alternating Forms

As we saw above, there is a natural identification $\alpha_* \colon A_*(V) \to \Lambda^*(V)$ for any finite dimensional linear space V. Let us apply this identification to the dual space V^* of V. To simplify notation, set $A^k(V) = A_k(V^*)$, $A^*(V) = A_*(V^*)$, and let α^k, α^*, β^k and β^* be the isomorphisms analogous to α_k, α_*, β_k and β_* respectively, obtained when V is replaced by the dual space V^*.

Proposition 1.2.43. *The linear space $T^{(k,0)}(V)$ is naturally isomorphic to the linear space K of k-linear functions from V^k to $\mathbb{R}$. Under this isomorphism $A^k(V) < T^{(k,0)}(V)$ corresponds to the linear space of alternating k-linear functions on V.*

Proof. The first part of the statement is a special case of Exercise 1.2.27. ((d)), and can be proved as follows. Assign to $(l_1, \dots, l_k) \in (V^*)^k$ the k-linear function $\psi^k(l_1, \dots, l_k) \in K$ given by the equality

$$\psi^k(l_1, \dots, l_k)(\mathbf{v}_1, \dots, \mathbf{v}_k) = l_1(\mathbf{v}_1) \cdots l_k(\mathbf{v}_k).$$

Since $\psi^k(l_1, \dots, l_k)$ depends on each $l_i \in V^*$ linearly, $\psi^k \colon (V^*)^k \to K$ is k-linear and induces a unique linear map $\Psi^k \colon T^{(k,0)}V \to K$ such that $\psi^k =$

$\Psi^k \circ \otimes_k$. We show that Ψ^k is an isomorphism. Choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of V. Then

$$\begin{aligned}\Psi^k(\mathbf{e}^{i_1\dots i_k})(\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_k}) &= \psi^k(\mathbf{e}^{i_1}, \dots, \mathbf{e}^{i_k})(\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_k}) \\ &= \mathbf{e}^{i_1}(\mathbf{e}_{j_1}) \cdots \mathbf{e}_{i_k}(\mathbf{e}_{j_k}) = \delta^{i_1}_{j_1} \cdots \delta^{i_k}_{j_k},\end{aligned}$$

where δ^i_j is the Kronecker delta symbol, thus,

$$\Psi^k\left(\sum_{i_1,\dots,i_k=1}^n T_{i_1\dots i_k}\mathbf{e}^{i_1\dots i_k}\right)(\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_k}) = \sum_{i_1,\dots,i_k=1}^n T_{i_1\dots i_k}\delta^{i_1}_{j_1} \cdots \delta^{i_k}_{j_k} = T_{j_1\dots j_k}.$$

Since a k linear function is uniquely determined by its values on k-tuples of basis vectors, this equality shows that the unique preimage of a k-linear function $\tau\colon V^k \to \mathbb{R}$ is the tensor

$$(\Psi^k)^{-1}(\tau) = \sum_{i_1,\dots,i_k=1}^n \tau(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k})\mathbf{e}^{i_1\dots i_k}.$$

If $\sigma \in \mathfrak{S}_k$ is a permutation, then

$$\begin{aligned}\Phi_\sigma((\Psi^k)^{-1}(\tau)) &= \sum_{i_1,\dots,i_k=1}^n \tau(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k})\mathbf{e}^{\sigma(i_1)\dots\sigma(i_k)} \\ &= \sum_{i_1,\dots,i_k=1}^n \tau(\mathbf{e}_{\sigma^{-1}(i_1)}, \dots, \mathbf{e}_{\sigma^{-1}(i_k)})\mathbf{e}^{i_1\dots i_k},\end{aligned}$$

which shows that $(\Psi^k)^{-1}(\tau)$ is an alternating tensor if and only if τ is an alternating k-linear function. □

Definition 1.2.44. Alternating k-linear functions on a linear space V are called *alternating k-forms* or shortly *k-forms* on V. ※

The composition of the isomorphisms Ψ^k and α^k yields a natural isomorphism between $\Lambda^k(V^*)$ and the linear space of alternating k-forms on V. Using this natural isomorphism we shall identify the elements of the two linear spaces. If $l_1, \dots, l_k \in V^*$ are linear functions on V, then $l_1 \wedge \cdots \wedge l_k$ as a k-form assigns to the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ the number

$$\begin{aligned}&l_1 \wedge \cdots \wedge l_k(\mathbf{v}_1, \dots, \mathbf{v}_k) = (\pi_k(l_1 \otimes \cdots \otimes l_k))(\mathbf{v}_1, \dots, \mathbf{v}_k) \\ &= \sum_{\sigma\in\mathfrak{S}_k} \operatorname{sgn}\sigma \cdot l_{\sigma(1)}(\mathbf{v}_1) \cdots l_{\sigma(k)}(\mathbf{v}_k) = \det\begin{pmatrix} l_1(\mathbf{v}_1) & \cdots & l_1(\mathbf{v}_k) \\ \vdots & \ddots & \vdots \\ l_k(\mathbf{v}_1) & \cdots & l_k(\mathbf{v}_k)\end{pmatrix}.\end{aligned}$$

1.2.6 Euclidean Linear Spaces

We know that k-linear functions on a linear space V are naturally identified with tensors of type $(k, 0)$ (cf. Exercise 1.2.27 ((d))). A k-linear function is symmetric if and only if the corresponding tensor is symmetric, or, equivalently, if the value of the function does not change when we permute its variables.

Definition 1.2.45. A symmetric bilinear function $\langle\,,\rangle\colon V \times V \to \mathbb{R}$ is called *positive definite* if $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$ for all $\mathbf{v} \in V$ and equality occurs only when $\mathbf{v} = \mathbf{0}$. ※

Definition 1.2.46. A *Euclidean linear space* is a finite dimensional linear space V equipped with a positive definite symmetric bilinear function $\langle\,,\rangle$, which is usually called the *inner product* or *dot product.* ※

For example, $\mathbb{R}^n$ with the standard dot product

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = x_1y_1 + \cdots + x_ny_n$$

on it is a Euclidean vector space.
The dot product enables us to define the length of a vector.

Definition 1.2.47. The *length* or *Euclidean norm of a vector* $\mathbf{v}$ in a Euclidean space V is

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}.$$

※

Proposition 1.2.48 (Cauchy–Schwarz Inequality)**.** *For any two vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *in a Euclidean space* V*, we have*

$$|\langle \mathbf{v}, \mathbf{w}\rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|.$$

Equality holds if and only if $\mathbf{v}$ *and* $\mathbf{w}$ *are linearly dependent.*

Proof. If $\mathbf{v} = \mathbf{0}$, then both sides are equal to 0 and the vectors are linearly dependent. If $\mathbf{v} \neq \mathbf{0}$, then consider the quadratic polynomial

$$P(t) = \|t\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2t^2 + 2\langle \mathbf{v}, \mathbf{w}\rangle t + \|\mathbf{w}\|^2.$$

Since $P(t) \geq 0$ for all t,

$$\begin{aligned} P\left(-\frac{\langle \mathbf{v}, \mathbf{w}\rangle}{\|\mathbf{v}\|^2}\right) &= \|\mathbf{v}\|^2 \cdot \frac{\langle \mathbf{v}, \mathbf{w}\rangle^2}{\|\mathbf{v}\|^4} - 2\langle \mathbf{v}, \mathbf{w}\rangle \cdot \frac{\langle \mathbf{v}, \mathbf{w}\rangle}{\|\mathbf{v}\|^2} + \|\mathbf{w}\|^2 \\ &= -\frac{\langle \mathbf{v}, \mathbf{w}\rangle^2}{\|\mathbf{v}\|^2} + \|\mathbf{w}\|^2 \geq 0, \end{aligned}$$

which gives the Cauchy–Schwarz inequality after rearrangement.

If the Cauchy–Schwarz inequality holds with equality, then

$$P\left(-\frac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{v}\|^2}\right) = \left\|-\frac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{v}\|^2}\mathbf{v}+\mathbf{w}\right\|^2 = 0,$$

which implies that $\mathbf{w} = \frac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{v}\|^2}\mathbf{v}$, so $\mathbf{v}$ and $\mathbf{w}$ are linearly dependent. It is also clear that if $\mathbf{w} = \lambda\mathbf{v}$, then both sides of the Cauchy–Schwarz inequality are equal to $\lambda\|\mathbf{v}\|^2$. □

Corollary 1.2.49. *For any two vectors* $\mathbf{v},\mathbf{w} \in V$*, we have*

$$\|\mathbf{v}+\mathbf{w}\| \le \|\mathbf{v}\| + \|\mathbf{w}\| \text{ and } \big|\|\mathbf{v}\| - \|\mathbf{w}\|\big| \le \|\mathbf{v}-\mathbf{w}\|.$$

Proof. The first inequality is equivalent to the inequality

$$\|\mathbf{v}+\mathbf{w}\| = \|\mathbf{v}\|^2 + 2\langle \mathbf{v},\mathbf{w}\rangle + \|\mathbf{w}\|^2 \le \|\mathbf{v}\| + \|\mathbf{w}\| = \|\mathbf{v}\|^2 + 2\|\mathbf{v}\|\cdot\|\mathbf{w}\| + \|\mathbf{w}\|^2,$$

which follows form the Cauchy–Schwarz inequality. The second inequality is a corollary of the first one applied to the pairs $(\mathbf{v},\mathbf{w}-\mathbf{v})$ and $(\mathbf{w},\mathbf{v}-\mathbf{w})$. □

Exercise 1.2.50. When do we have equality in the inequalities of Corollary 1.2.49? ¤

Corollary 1.2.51. *If* $\mathbf{v}$ *and* $\mathbf{w}$ *are nonzero vectors, then*

$$-1 \le \frac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{v}\|\cdot\|\mathbf{w}\|} \le 1.$$

Definition 1.2.52. If $\mathbf{v}$ and $\mathbf{w}$ are nonzero vectors, then there is a uniquely defined number $\alpha \in [0,\pi]$ for which

$$\cos\alpha = \frac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{v}\|\cdot\|\mathbf{w}\|}.$$

The number α is called the *angle enclosed by the vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *(measured in radians).* ※

Definition 1.2.53. Two non-zero vectors are *orthogonal* if the angle enclosed by them is $\pi/2$. Orthogonality of two non-zero vectors is equivalent to the condition $\langle \mathbf{v},\mathbf{w}\rangle = 0$. Since the latter equation is automatically fulfilled when $\mathbf{v} = \mathbf{0}$, we agree, that $\mathbf{0}$ *is said to be orthogonal to every vector.* ※

Definition 1.2.54. A collection of some vectors $\mathbf{e}_1,\dots,\mathbf{e}_k$ of a Euclidean linear space V is said to be an *orthonormal system* if $\langle \mathbf{e}_i,\mathbf{e}_j\rangle = \delta_{ij}$ for all $1 \le i,j \le k$, i.e., if the vectors have unit length and are mutually orthogonal to one another. ※

Theorem 1.2.55 (Gram–Schmidt Orthogonalization)**.** *Assume that the vectors* $\mathbf{f}_1, \dots, \mathbf{f}_k$ *of a Euclidean linear space* V *are linearly independent then there is a unique orthonormal system of vectors* $\mathbf{e}_1, \dots, \mathbf{e}_k$ *satisfying the following two properties:*

(i) $\mathrm{lin}[\{\mathbf{f}_1, \dots, \mathbf{f}_s\}] = \mathrm{lin}[\{\mathbf{e}_1, \dots, \mathbf{e}_s\}]$ *for* $s = 1, \dots, k$*;*

(ii) *For* $s = 1, \dots, k$*, the ordered* s*-tuples of vectors* $(\mathbf{f}_1, \dots, \mathbf{f}_s)$ *and* $(\mathbf{e}_1, \dots, \mathbf{e}_s)$ *are bases of the same orientation in the linear space they span.*

Proof. According to condition (i), each vector $\mathbf{e}_s$ must be a linear combination of $\mathbf{f}_1, \dots, \mathbf{f}_s$

$$\begin{aligned}
\mathbf{e}_1 &= \alpha_1^1 \mathbf{f}_1, \\
&\vdots \\
\mathbf{e}_s &= \alpha_s^1 \mathbf{f}_1 + \cdots + \alpha_s^s \mathbf{f}_s, \\
&\vdots \\
\mathbf{e}_k &= \alpha_k^1 \mathbf{f}_1 + \alpha_k^2 \mathbf{f}_2 + \cdots\cdots\cdots + \alpha_k^k \mathbf{f}_k.
\end{aligned} \tag{1.4}$$

Condition (ii) on the orientation is fulfilled if and only if

$$\det \begin{pmatrix} \alpha_1^1 & 0 & 0 & \dots & 0 \\ \alpha_2^1 & \alpha_2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \alpha_{s-1}^1 & \alpha_{s-1}^2 & \dots & \alpha_{s-1}^{s-1} & 0 \\ \alpha_s^1 & \alpha_s^2 & \dots & \dots & \alpha_s^s \end{pmatrix} = \alpha_1^1 \cdots \alpha_s^s > 0 \text{ for all } 1 \le s \le k, \tag{1.5}$$

which means that all the diagonal elements α_s^s must be positive.
We prove the theorem by induction on k and give an explicit recursive formula for the computation of the vector $\mathbf{e}_s$. For $k = 1$, since $\mathbf{e}_1$ must be a unit vector and α_1^1 must be positive, the only good choice for $\mathbf{e}_1$ is

$$\mathbf{e}_1 = \frac{\mathbf{f}_1}{\|\mathbf{f}_1\|}.$$

Suppose that the theorem is true for $k-1$. Then for $\mathbf{f}_1, \dots, \mathbf{f}_{k-1}$ we can find an orthonormal system $\mathbf{e}_1, \dots, \mathbf{e}_{k-1}$ satisfying (i) and (ii) for $1 \le s \le k-1$. Then $\mathbf{e}_k$ must be of the form

$$\mathbf{e}_k = (\beta_k^1 \mathbf{e}_1 + \cdots + \beta_k^{k-1} \mathbf{e}_{k-1}) + \alpha_k^k \mathbf{f}_k. \tag{1.6}$$

Taking the dot product of both sides of (1.6) with $\mathbf{e}_i$ $(1 \le i \le k-1)$, we obtain

$$0 = \beta_k^i + \alpha_k^k \langle \mathbf{f}_k, \mathbf{e}_i \rangle,$$

consequently,

$$\mathbf{e}_k = \alpha_k^k \big(\mathbf{f}_k - (\langle \mathbf{f}_k, \mathbf{e}_1\rangle \mathbf{e}_1 + \cdots + \langle \mathbf{f}_k, \mathbf{e}_{k-1}\rangle \mathbf{e}_{k-1})\big).$$

The parameter α_k^k must be used to normalize the vector which stands on the right of it. Thus,

$$\alpha_k^k = \pm \|\mathbf{f}_k - (\langle \mathbf{f}_k, \mathbf{e}_1\rangle \mathbf{e}_1 + \cdots + \langle \mathbf{f}_k, \mathbf{e}_{k-1}\rangle \mathbf{e}_{k-1})\|^{-1}.$$

Since $\alpha_k^k > 0$, the only possible choice of $\mathbf{e}_k$ is

$$\mathbf{e}_k = \frac{\mathbf{f}_k - (\langle \mathbf{f}_k, \mathbf{e}_1\rangle \mathbf{e}_1 + \cdots + \langle \mathbf{f}_k, \mathbf{e}_{k-1}\rangle \mathbf{e}_{k-1})}{\|\mathbf{f}_k - (\langle \mathbf{f}_k, \mathbf{e}_1\rangle \mathbf{e}_1 + \cdots + \langle \mathbf{f}_k, \mathbf{e}_{k-1}\rangle \mathbf{e}_{k-1})\|}.$$

It is not difficult to check that the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$ will satisfy the requirements. □

Applying the Gram-Schmidt orthogonalization to a basis we obtain the following corollary.

Corollary 1.2.56. *Every finite dimensional Euclidean linear space has an orthonormal basis.*

Definition 1.2.57. The *Gram matrix of a system of vectors* $\mathbf{f}_1, \ldots, \mathbf{f}_k$ of a Euclidean linear space V is the matrix

$$\mathcal{G}(\mathbf{f}_1, \ldots, \mathbf{f}_k) = \begin{pmatrix} \langle \mathbf{f}_1, \mathbf{f}_1\rangle & \ldots & \langle \mathbf{f}_1, \mathbf{f}_k\rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{f}_k, \mathbf{f}_1\rangle & \ldots & \langle \mathbf{f}_k, \mathbf{f}_k\rangle \end{pmatrix}.$$

※

Corollary 1.2.58. *Let $\mathcal{G}$ be the Gram matrix of the vectors $\mathbf{f}_1, \ldots, \mathbf{f}_k$ of a Euclidean linear space V. Then* $\det \mathcal{G} \geq 0$, and $\det \mathcal{G} = 0$ *if and only if the vectors $\mathbf{f}_1, \ldots, \mathbf{f}_k$ are linearly dependent.*

Proof. If there is a non-trivial linear relation $\sum_{i=1}^k \alpha^i \mathbf{f}_i = \mathbf{0}$, then the rows of $\mathcal{G}$ are linearly dependent with the same coefficients α^i therefore $\det \mathcal{G} = 0$. If $\mathbf{f}_1, \ldots, \mathbf{f}_k$ are linearly independent, then applying the Gram–Schmidt orthogonalization to them we obtain an orthonormal system $\mathbf{e}_1, \ldots, \mathbf{e}_k$ and we can express the vectors $\mathbf{f}_s$ as follows

$$\begin{aligned} \mathbf{f}_1 &= \tilde{\alpha}_1^1 \mathbf{e}_1 \\ &\vdots \\ \mathbf{f}_s &= \tilde{\alpha}_s^1 \mathbf{e}_1 + \cdots + \tilde{\alpha}_s^s \mathbf{e}_s \\ &\vdots \\ \mathbf{f}_k &= \tilde{\alpha}_k^1 \mathbf{e}_1 + \tilde{\alpha}_k^2 \mathbf{e}_2 + \cdots\cdots\cdots + \tilde{\alpha}_k^k \mathbf{e}_k. \end{aligned}$$

The lower triangular matrix $\tilde{T}$ put together from the coefficients $\tilde{\alpha}_i^j$ is the inverse of the matrix (1.5) coming from the decompositions (1.4), in particular

$$\tilde{\alpha}_s^s = \frac{1}{\alpha_s^s} = \left\| \mathbf{f}_s - \sum_{i=1}^{s-1} \langle \mathbf{f}_s, \mathbf{e}_i \rangle \right\| > 0 \text{ for } s = 1, \dots, k. \tag{1.7}$$

It is clear that $\mathcal{G} = \tilde{T} \cdot \tilde{T}^T$, so

$$\det \mathcal{G} = \det(\tilde{T} \cdot \tilde{T}^T) = \det(\tilde{T})^2 = (\tilde{\alpha}_1^1 \cdots \tilde{\alpha}_k^k)^2 > 0. \qquad \square$$

Corollary 1.2.59. *With the notation used in Corollary 1.2.58, the identity*

$$\det \mathcal{G}(\mathbf{f}_1, \dots, \mathbf{f}_s) = \left\| \mathbf{f}_s - \sum_{i=1}^{s-1} \langle \mathbf{f}_s, \mathbf{e}_i \rangle \mathbf{e}_i \right\|^2 \det \mathcal{G}(\mathbf{f}_1, \dots, \mathbf{f}_{s-1})$$

holds.

Corollary 1.2.60. *A symmetric bilinear function $\langle \, , \rangle$ on a linear space V is positive definite if and only if there is a basis $\mathbf{f}_1, \dots, \mathbf{f}_n$ such that the Gram matrices $\mathcal{G}(\mathbf{f}_1, \dots, \mathbf{f}_s)$ have positive determinants for $s = 1, \dots, n$.*

Proof. The previous proposition shows that if $\langle \, , \rangle$ is positive definite, then any basis will be good. Conversely, assume that $\mathbf{f}_1, \dots, \mathbf{f}_n$ is a basis such that $\mathcal{G}(\mathbf{f}_1, \dots, \mathbf{f}_s) > 0$ for $s = 1, \dots, n$. Then the recursive formula

$$\mathbf{e}_1 = \frac{\mathbf{f}_1}{\sqrt{\det \mathcal{G}(\mathbf{f}_1)}}, \qquad \mathbf{e}_s = \frac{\mathbf{f}_s - \sum_{i=1}^{s-1} \langle \mathbf{f}_s, \mathbf{e}_i \rangle \mathbf{e}_i}{\sqrt{\det \mathcal{G}(\mathbf{f}_1, \dots, \mathbf{f}_s) / \det \mathcal{G}(\mathbf{f}_1, \dots, \mathbf{f}_{s-1})}}$$

defines an orthonormal basis for $\langle \, , \rangle$, i.e. $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$. (Check this!) The existence of an orthonormal basis implies that $\langle \, , \rangle$ is positive definite, because if $\mathbf{v} = \sum_{i=1}^n v^i \mathbf{e}_i$, then $\langle \mathbf{v}, \mathbf{v} \rangle = \sum_{i=1}^n (v^i)^2 \geq 0$, and equality holds only when $\mathbf{v} = \mathbf{0}$. $\square$

The Principal Axis Theorem

Sometimes a Euclidean linear space $(V, \langle \, , \rangle)$ is equipped with a second symmetric bilinear function $\{ \, , \} \colon V \times V \to \mathbb{R}$. In general, a symmetric bilinear function on a linear space can be defined by its matrix with respect to a basis or by its quadratic form. Using the inner product of the Euclidean structure it can also be defined with the help of a self-adjoint map.

If $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a basis of the vector space V and $\{ \, , \}$ is a bilinear function on V then the $n \times n$ matrix $(b_{ij})_{1 \leq i,j \leq n}$ with entries $b_{ij} = \{\mathbf{x}_i, \mathbf{x}_j\}$ is called the *matrix representation* or simply the *matrix of* $\{ \, , \}$ *with respect to the basis* $(\mathbf{x}_1, \dots, \mathbf{x}_n)$. Fixing the basis we get a one to one correspondence between bilinear functions and $n \times n$ matrices. A bilinear form is symmetric if and only if its matrix with respect to a basis is symmetric.

Definition 1.2.61. The *quadratic form of a bilinear function* $\{\,,\}$ is the function defined by the equality $Q_{\{,\}}(\mathbf{x}) = \{\mathbf{x}, \mathbf{x}\}$. ※

Symmetric bilinear functions can be recovered from their quadratic forms with the help of the identity

$$\{\mathbf{x}, \mathbf{y}\} = \frac{1}{2}\big(Q_{\{,\}}(\mathbf{x}+\mathbf{y}) - Q_{\{,\}}(\mathbf{x}) - Q_{\{,\}}(\mathbf{y})\big). \tag{1.8}$$

The following proposition establishes a bijection between bilinear functions on a Euclidean linear space V and linear endomorphisms $L\colon V \to V$.

Proposition 1.2.62. *Let $(V, \langle\,,\rangle)$ be a Euclidean linear space.*

(i) *Then for any linear endomorphism $L\colon V \to V$, the map*

$$\{\,,\}_L \colon V \times V \to \mathbb{R}, \quad \{\mathbf{x}, \mathbf{y}\}_L = \langle L\mathbf{x}, \mathbf{y}\rangle$$

is a bilinear function on V.

(ii) *For any bilinear function $\{\,,\}$ on V, there is a unique linear map $L\colon V \to V$ such that $\{\,,\} = \{\,,\}_L$.*

(iii) *The bilinear function $\{\,,\}_L$ is symmetric if and only if L satisfies the identity $\langle L\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, L\mathbf{y}\rangle$.*

Proof. We prove only (ii), the rest is obvious. Choose an orthonormal basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ in V. Given a bilinear function $\{\,,\}$, the only possible choice for $L(\mathbf{x})$ is

$$L(\mathbf{x}) = \sum_{i=1}^{n} \langle L\mathbf{x}, \mathbf{e}_i\rangle \mathbf{e}_i = \sum_{i=1}^{n} \{\mathbf{x}, \mathbf{e}_i\}\mathbf{e}_i,$$

which proves uniqueness. On the other hand, if we define L by the last equality, then

$$\langle L\mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^{n} \{\mathbf{x}, \mathbf{e}_i\}\langle \mathbf{e}_i, \mathbf{y}\rangle = \Big\{\mathbf{x}, \sum_{i=1}^{n} \langle \mathbf{e}_i, \mathbf{y}\rangle \mathbf{e}_i\Big\} = \{\mathbf{x}, \mathbf{y}\},$$

so L is a good choice. □

Definition 1.2.63. A linear endomorphism L of a Euclidean linear space is said to be *self-adjoint (with respect to the Euclidean structure)* if it satisfies the identity $\langle L\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, L\mathbf{y}\rangle$. ※

Lemma 1.2.64. *All the eigenvalues of a self-adjoint map $L\colon V \to V$ are real.*

Proof. Let $x+iy \in \mathbb{C}$ be an eigenvalue of L, $\mathbf{v}+i\mathbf{w} \in \mathbb{C}\otimes V$ be a corresponding eigenvector. Then the real and imaginary part of the equation $L(\mathbf{v}+i\mathbf{w}) = (x+iy)(\mathbf{v}+i\mathbf{w})$ gives $L\mathbf{v} = x\mathbf{v} - y\mathbf{w}$ and $L\mathbf{w} = y\mathbf{v} + x\mathbf{w}$. Since L is self-adjoint, we have

$$x\langle \mathbf{v}, \mathbf{w}\rangle - y\|\mathbf{w}\|^2 = \langle L\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, L\mathbf{w}\rangle = x\langle \mathbf{v}, \mathbf{w}\rangle + y\|\mathbf{v}\|^2,$$

which yields $y(\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2) = 0$. However, $\mathbf{v}+i\mathbf{w} \neq \mathbf{0}$ because eigenvectors are non-zero vectors, therefore $y = 0$. □

Theorem 1.2.65 (Principal axis theorem)**.** *Let V be a finite dimensional Euclidean linear space and let $L\colon V \to V$ be a self-adjoint linear transformation on V. Then there exists an orthonormal basis of V consisting of eigenvectors of L.*

Proof. We prove by induction on the dimension n of V. The base case $n = 1$ is trivial. Assume that it is true for $n = k$. Suppose $n = k+1$. By Lemma 1.2.64, there exists a unit vector $\mathbf{v}_1$ in V which is an eigenvector of L. Let $W = \mathbf{v}_1^\perp = \{\mathbf{w} \in V \mid \mathbf{v}_1 \perp \mathbf{w}\}$. Then $L(W) \subset W$ since we have

$$\langle L\mathbf{w}, \mathbf{v}_1\rangle = \langle \mathbf{w}, L\mathbf{v}_1\rangle = \langle \mathbf{w}, \lambda_1\mathbf{v}_1\rangle = \lambda_1\langle \mathbf{w}, \mathbf{v}_1\rangle = 0$$

for any $\mathbf{w} \in W$, where λ_1 is the eigenvalue belonging to $\mathbf{v}_1$. Clearly $L\big|_W$ is self-adjoint. Since $\dim(W) = \dim(V) - 1 = k$, the induction assumption implies that there exists an orthonormal basis $(\mathbf{v}_2, \dots, \mathbf{v}_n)$ in W consisting of eigenvectors of $L\big|_W$. But each eigenvector of $L\big|_W$ is an eigenvector of L, so $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ is an orthonormal basis of V consisting of eigenvectors of L. □

Induced Euclidean Structures on Tensor Spaces

Let $(V, \langle\, , \rangle)$ be a finite dimensional Euclidean linear space. Our goal now is to extend the dot product in a natural way to tensor spaces and exterior power spaces constructed from V. Consider the map

$$\begin{aligned} K\colon V^{2k} = V^k \times V^k &\to \mathbb{R}, \\ ((\mathbf{v}_1, \dots, \mathbf{v}_k), (\mathbf{w}_1, \dots, \mathbf{w}_k)) &\mapsto \langle \mathbf{v}_1, \mathbf{w}_1\rangle \cdots \langle \mathbf{v}_k, \mathbf{w}_k\rangle. \end{aligned}$$

As this map is $2k$-linear, it induces a linear map $L\colon T^{(0,2k)}V \to \mathbb{R}$ such that $K = L \circ \otimes_{2k}$. Composing L with the tensor product operation $\otimes\colon T^{(0,k)} \times T^{(0,k)} \to T^{(0,2k)}$ we obtain a bilinear function on $T^{(0,k)}$, which we also denote by $\langle\, , \rangle$.

Proposition 1.2.66. *The induced bilinear function $\langle\, , \rangle$ on $T^{(0,k)}V$ is symmetric and positive definite. If $\mathbf{e}_1, \dots, \mathbf{e}_n$ is an orthonormal basis of V, then $\mathbf{e}_{i_1\dots i_k}$ $(1 \le i_1, \dots, i_k \le n)$ is an orthonormal basis of $T^{(0,k)}V$ with respect to $\langle\, , \rangle$.*

Proof. It is enough to show the second part of the proposition since orthonormal basis exists only for symmetric positive definite bilinear functions. This follows from

$$\begin{aligned}\langle \mathbf{e}_{i_1\dots i_k}, \mathbf{e}_{j_1\dots j_k}\rangle &= L(\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_k}\otimes\mathbf{e}_{j_1}\otimes\cdots\otimes\mathbf{e}_{j_k})\\ &= \langle \mathbf{e}_{i_1}, \mathbf{e}_{j_1}\rangle\cdots\langle \mathbf{e}_{i_k}, \mathbf{e}_{j_k}\rangle = \delta_{i_1,j_1}\cdots\delta_{i_k,j_k}.\end{aligned}$$ □

The linear space $A_k(V)$ of alternating tensors is a linear subspace of $T^{(0,k)}$, so the restriction of $\frac{1}{k!}\langle\,,\rangle$ defines a symmetric positive definite bilinear function on $A_k(V)$. Using the natural isomorphism α_k between $A_k(V)$ and $\Lambda^k(V)$ we obtain a positive definite symmetric bilinear function $\langle\,,\rangle$ also on $\Lambda^k(V)$ expressed by

$$\langle\omega_1,\omega_2\rangle = \frac{1}{k!}\langle\alpha_k(\omega_1),\alpha_k(\omega_2)\rangle \qquad \forall\omega_1,\omega_2\in\Lambda^k(V).$$

The reason why we divide by $k!$ is to make the following proposition true.

Proposition 1.2.67. *If $\mathbf{e}_1,\dots,\mathbf{e}_n$ is an orthonormal basis of V, then*

$$\{\mathbf{e}_{i_1}\wedge\cdots\wedge\mathbf{e}_{i_k} \mid 1\le i_1<\cdots<i_k\le n\}$$

is an orthonormal basis of $\Lambda^k(V)$.

Proof. If the increasing sequences $1\le i_1<\cdots<i_k\le n$ and $1\le j_1<\cdots<j_k\le n$ are not the same, then no permutation of them can coincide, so

$$\begin{aligned}&\langle \mathbf{e}_{i_1}\wedge\cdots\wedge\mathbf{e}_{i_k}, \mathbf{e}_{j_1}\wedge\cdots\wedge\mathbf{e}_{j_k}\rangle\\ &\quad= \frac{1}{k!}\left\langle \sum_{\pi\in\mathfrak{S}_k}\operatorname{sgn}\pi\cdot\mathbf{e}_{i_{\pi(1)}\dots i_{\pi(k)}}, \sum_{\sigma\in\mathfrak{S}_k}\operatorname{sgn}\sigma\cdot\mathbf{e}_{j_{\sigma(1)}\dots j_{\sigma(k)}}\right\rangle = 0.\end{aligned}$$

When the sequences are equal, then the permuted sequences $i_{\pi(1)},\dots,i_{\pi(k)}$ and $j_{\sigma(1)},\dots,j_{\sigma(k)}$ coincide if and only if $\pi=\sigma$, therefore

$$\begin{aligned}&\langle \mathbf{e}_{i_1}\wedge\cdots\wedge\mathbf{e}_{i_k}, \mathbf{e}_{i_1}\wedge\cdots\wedge\mathbf{e}_{i_k}\rangle\\ &\quad= \frac{1}{k!}\Big\langle \sum_{\pi\in\mathfrak{S}_k}\operatorname{sgn}\pi\cdot\mathbf{e}_{i_{\pi(1)}\dots i_{\pi(k)}}, \sum_{\sigma\in\mathfrak{S}_k}\operatorname{sgn}\sigma\cdot\mathbf{e}_{i_{\sigma(1)}\dots i_{\sigma(k)}}\Big\rangle\\ &\quad= \frac{1}{k!}\sum_{\pi\in\mathfrak{S}_k}\|\mathbf{e}_{i_{\pi(1)}\dots i_{\pi(k)}}\|^2 = \frac{k!}{k!} = 1.\end{aligned}$$ □

Proposition 1.2.68. *The following identity holds*

$$\langle \mathbf{v}_1\wedge\cdots\wedge\mathbf{v}_k, \mathbf{w}_1\wedge\cdots\wedge\mathbf{w}_k\rangle = \det\begin{pmatrix}\langle\mathbf{v}_1,\mathbf{w}_1\rangle & \cdots & \langle\mathbf{v}_1,\mathbf{w}_k\rangle\\ \vdots & \ddots & \vdots\\ \langle\mathbf{v}_k,\mathbf{w}_1\rangle & \cdots & \langle\mathbf{v}_k,\mathbf{w}_k\rangle\end{pmatrix}. \tag{1.9}$$

Proof. The left-hand side of (1.9) equals

$$\begin{aligned}
&\langle \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k, \mathbf{w}_1 \wedge \cdots \wedge \mathbf{w}_k \rangle \\
&= \frac{1}{k!} \Big\langle \sum_{\pi \in \mathfrak{S}_k} \operatorname{sgn} \pi \cdot \mathbf{v}_{\pi(1)} \otimes \cdots \otimes \mathbf{v}_{\pi(k)}, \sum_{\sigma \in \mathfrak{S}_k} \operatorname{sgn} \sigma \cdot \mathbf{w}_{\sigma(1)} \ldots \mathbf{w}_{\sigma(k)} \Big\rangle \\
&= \frac{1}{k!} \sum_{\pi \in \mathfrak{S}_k} \sum_{\sigma \in \mathfrak{S}_k} \operatorname{sgn} \pi \cdot \operatorname{sgn} \sigma \cdot \langle \mathbf{v}_{\pi(1)}, \mathbf{w}_{\sigma(1)} \rangle \cdots \langle \mathbf{v}_{\pi(k)}, \mathbf{w}_{\sigma(k)} \rangle \\
&= \frac{1}{k!} \sum_{\pi \in \mathfrak{S}_k} \operatorname{sgn} \pi \cdot \det \begin{pmatrix} \langle \mathbf{v}_{\pi(1)}, \mathbf{w}_1 \rangle & \cdots & \langle \mathbf{v}_{\pi(1)}, \mathbf{w}_k \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{v}_{\pi(k)}, \mathbf{w}_1 \rangle & \cdots & \langle \mathbf{v}_{\pi(k)}, \mathbf{w}_k \rangle \end{pmatrix} \\
&= \det \begin{pmatrix} \langle \mathbf{v}_1, \mathbf{w}_1 \rangle & \cdots & \langle \mathbf{v}_1, \mathbf{w}_k \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{v}_k, \mathbf{w}_1 \rangle & \cdots & \langle \mathbf{v}_k, \mathbf{w}_k \rangle \end{pmatrix}. \qquad \square
\end{aligned}$$

Let V^* be the dual space of V. The inner product is a bilinear function, therefore it defines a linear map $l\colon V \to V^*$ by $l(\mathbf{v})(\mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle$. The linear map l is injective, since if $l(\mathbf{v}) = 0$, then $l(\mathbf{v})(\mathbf{v}) = \|\mathbf{v}\|^2 = 0$, so $\mathbf{v} = \mathbf{0}$. This implies that l is a linear isomorphism between V and V^*. The linear isomorphism l can be used to identify V and its dual space, and gives rise to a Euclidean linear space structure $\langle \, , \rangle\colon V^* \times V^* \to \mathbb{R}$ on V^*, for which

$$\langle l(\mathbf{v}), l(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle.$$

Proposition 1.2.69. *If $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is an orthonormal basis of V, then its dual basis $(\mathbf{e}^1, \ldots, \mathbf{e}^n)$ is an orthonormal basis of V^*.*

Proof. Since the basis is orthonormal, we have

$$l(\mathbf{e}_i)(\mathbf{e}_j) = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij} = \mathbf{e}^i(\mathbf{e}_j),$$

thus, $l(\mathbf{e}_i) = \mathbf{e}^i$, and

$$\langle \mathbf{e}^i, \mathbf{e}^j \rangle = \langle l(\mathbf{e}_i), l(\mathbf{e}_j) \rangle = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}. \qquad \square$$

The Euclidean structure on V^* induces a Euclidean structure on the tensor spaces $T^{(k,0)}V$ and also on $\Lambda^k(V^*)$. Finally, since l identifies V with its dual space, tensors of type (k,l) can be identified with tensors of type $(k+l,0)$ and also with tensors of type $(0,k+l)$. Both identifications induces the same Euclidean structure on $T^{(k,l)}V$, with respect to which the basis $\mathbf{e}_{i_1 \ldots i_k}^{j_1 \ldots j_l}$ generated from any orthonormal basis of V will be an orthonormal basis in $T^{(k,l)}V$.

1.2.7 Hodge Star Operator

Let V be an oriented n-dimensional Euclidean linear space with inner product $\langle\,,\rangle$. Since $\Lambda^n V$ is an oriented 1-dimensional Euclidean linear space, it contains exactly one positively oriented unit vector $\omega \in \Lambda^n V$.

Definition 1.2.70. The *Hodge star operator* $*\colon \Lambda^k(V) \to \Lambda^{n-k}V$ is a linear operator. For $\eta \in \Lambda^k(V)$, $*\eta$ is defined as the unique element in $\Lambda^{n-k}V$ satisfying the identity

$$\zeta \wedge (*\eta) = \langle \zeta, \eta\rangle \omega \text{ for all } \zeta \in \Lambda^k V. \tag{1.10}$$

The following proposition says that the definition is correct.

Proposition 1.2.71. *For each $\eta \in \Lambda^k V$, there is exactly one element $*\eta \in \Lambda^{n-k}V$ which satisfies* (1.10).

Proof. The proof will give an explicit method to compute $*\eta$. Choose a positively oriented orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V. Then $\omega = \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n$. The system

$$\{\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_k} | 1 \leq i_1 < \cdots < i_k \leq n\} \tag{1.11}$$

is an orthonormal basis of $\Lambda^k V$, so we can write η and its unknown Hodge star $*\eta$ as

$$\eta = \sum_{1\leq i_1<\cdots<i_k\leq n} \eta^{i_1\dots i_k}\mathbf{e}_{i_1}\wedge\cdots\wedge\mathbf{e}_{i_k}, \quad *\eta = \sum_{1\leq j_1<\cdots<j_{n-k}\leq n} (*\eta)^{j_1\dots j_{n-k}}\mathbf{e}_{j_1}\wedge\cdots\wedge\mathbf{e}_{j_{n-k}}$$

Both sides of (1.10) are linear in ζ, so the equation holds for all $\zeta \in \Lambda^k(V)$ if and only if it is true for all vectors of a basis. Substituting the basis vector $\zeta = \mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_k}$ into (1.10) we obtain the equality

$$(*\eta)^{j_1\dots j_{n-k}}\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_k} \wedge \mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_{n-k}} = \eta^{i_1\dots i_k}\mathbf{e}_1 \wedge \dots \mathbf{e}_n,$$

where $j_1 < \cdots < j_{n-k}$ are the elements of the complementary set $\{1, 2, \dots, n\} \setminus \{i_1, \dots, i_k\}$ arranged in increasing order. Thus, denoting by $\epsilon(i_1, \dots, i_k)$ the sign of the permutation $(i_1, \dots, i_k, j_1, \dots, j_{n-k})$,

$$(*\eta)^{j_1\dots j_{n-k}} = \epsilon(i_1, \dots, i_k) \cdot \eta^{i_1\dots i_k}. \qquad \square$$

Corollary 1.2.72. *Since the Hodge operator maps the orthonormal basis* (1.11) *into an orthonormal basis of $\Lambda^{n-k}V$ it is an orthogonal transformation.*

Corollary 1.2.73. *As*

$$\operatorname{sgn}(i_1, \dots, i_k, j_1, \dots, j_{n-k}) = (-1)^{k(n-k)} \operatorname{sgn}(j_1, \dots, j_{n-k}, i_1, \dots, i_k),$$

we have $**|_{\Lambda^k(V)} = (-1)^{k(n-k)}\,\mathrm{id}_{\Lambda^k(V)}$.

Cross Product

Definition 1.2.74. Let V be a 3-dimensional oriented Euclidean vector field. Then the binary operation $\times\colon V\times V\to V$, $(\mathbf{a},\mathbf{b})\mapsto \mathbf{a}\times\mathbf{b}=*(\mathbf{a}\wedge\mathbf{b})$ is called the *cross product* operation. ※

Proposition 1.2.75. *The cross product has the following properties, which give a geometrical way to construct it.*

- $\mathbf{a}\times\mathbf{b}$ *is orthogonal to both* $\mathbf{a}$ *and* $\mathbf{b}$.
- *The length of* $\mathbf{a}\times\mathbf{b}$ *is equal to the area of the parallelogram spanned by* $\mathbf{a}$ *and* $\mathbf{b}$. *In particular,* $\mathbf{a}\times\mathbf{b}=\mathbf{0}$ *if and only if* $\mathbf{a}$ *and* $\mathbf{b}$ *are linearly dependent. If* $\mathbf{a}$ *and* $\mathbf{b}$ *are nonzero and the angle between them is* γ, *then* $\|\mathbf{a}\times\mathbf{b}\|=\|\mathbf{a}\|\cdot\|\mathbf{b}\|\cdot\sin(\gamma)$. *Geometrically,* $\|\mathbf{a}\|\cdot\|\mathbf{b}\|\cdot\sin(\gamma)$ *is the area of the parallelogram spanned by the vectors* $\mathbf{a}$ *and* $\mathbf{b}$.
- *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are linearly independent, then* $(\mathbf{a},\mathbf{b},\mathbf{a}\times\mathbf{b})$ *is a positively oriented basis of* V.

Proof. Let ω be the only positively oriented trivector of unit length in $\Lambda^3 V$. Then

$$0=\mathbf{a}\wedge(\mathbf{a}\wedge\mathbf{b})=\langle\mathbf{a},\mathbf{a}\times\mathbf{b}\rangle\omega \text{ and } 0=\mathbf{b}\wedge(\mathbf{a}\wedge\mathbf{b})=\langle\mathbf{b},\mathbf{a}\times\mathbf{b}\rangle\omega,$$

which implies that $\langle\mathbf{a},\mathbf{a}\times\mathbf{b}\rangle=\langle\mathbf{b},\mathbf{a}\times\mathbf{b}\rangle=0$.
Since the Hodge star operator preserves length,

$$\|\mathbf{a}\times\mathbf{b}\|^2=\|\mathbf{a}\wedge\mathbf{b}\|^2=\det\mathcal{G}(\mathbf{a},\mathbf{b})=\det\begin{pmatrix}\|\mathbf{a}\|^2 & \langle\mathbf{a},\mathbf{b}\rangle\\ \langle\mathbf{b},\mathbf{a}\rangle & |\mathbf{b}\|^2\end{pmatrix}.$$

The last determinant is 0 if any of the vectors $\mathbf{a}$ or $\mathbf{b}$ is $\mathbf{0}$. If none of them is $\mathbf{0}$ and the angle between them is γ, then $\langle\mathbf{a},\mathbf{b}\rangle=\|\mathbf{a}\|\cdot\|\mathbf{b}\|\cdot\cos(\gamma)$, and the value of the determinant is $\|\mathbf{a}\|^2\cdot\|\mathbf{b}\|^2\cdot(1-\cos^2(\gamma))=(\|\mathbf{a}\|\cdot\|\mathbf{b}\|\cdot\sin(\gamma))^2$. We also have

$$\mathbf{a}\wedge\mathbf{b}\wedge(\mathbf{a}\times\mathbf{b})=\langle\mathbf{a}\wedge\mathbf{b},*(\mathbf{a}\times\mathbf{b})\rangle\omega=\langle\mathbf{a}\wedge\mathbf{b},\mathbf{a}\wedge\mathbf{b}\rangle\omega=\|\mathbf{a}\wedge\mathbf{b}\|^2\omega.$$

If $\mathbf{a}$ and $\mathbf{b}$ are linearly independent, then this is a positive multiple of ω, which means that $(\mathbf{a},\mathbf{b},\mathbf{a}\times\mathbf{b})$ is a positively oriented basis of V. □

Proposition 1.2.76. *We have the following identities for the cross product:*

- $\mathbf{a}\times\mathbf{b}=-\mathbf{b}\times\mathbf{a}$ *(skew-commutativity);*
- $\begin{matrix}(\lambda\mathbf{a}_1+\mu\mathbf{a}_2)\times\mathbf{b}=\lambda\mathbf{a}_1\times\mathbf{b}+\mu\mathbf{a}_2\times\mathbf{b}\\ \mathbf{a}\times(\lambda\mathbf{b}_1+\mu\mathbf{b}_2)=\lambda\mathbf{a}\times\mathbf{b}_1+\mu\mathbf{a}\times\mathbf{b}_2\end{matrix}$ *(bilinearity);*

- $\langle \mathbf{a}, \mathbf{b}\times\mathbf{c}\rangle = \langle \mathbf{b}, \mathbf{c}\times\mathbf{a}\rangle = \langle \mathbf{c}, \mathbf{a}\times\mathbf{b}\rangle$ *(permutation rule for mixed product);*
- $\langle \mathbf{a}\times\mathbf{b}, \mathbf{c}\times\mathbf{d}\rangle = \det\begin{pmatrix}\langle \mathbf{a},\mathbf{c}\rangle & \langle \mathbf{a},\mathbf{d}\rangle\\ \langle \mathbf{b},\mathbf{c}\rangle & \langle \mathbf{b},\mathbf{d}\rangle\end{pmatrix}$ *(Lagrange identity);*
- $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \langle \mathbf{a},\mathbf{c}\rangle\mathbf{b} - \langle \mathbf{a},\mathbf{b}\rangle\mathbf{c}$ *(triple product expansion formula);*
- $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) + \mathbf{b}\times(\mathbf{c}\times\mathbf{a}) + \mathbf{c}\times(\mathbf{a}\times\mathbf{b}) = \mathbf{0}$ *(Jacobi identity).*

Proof. Skew-commutativity and bilinearity follow directly from analogous properties of the wedge product. The permutation rule is a consequence of the identity $\mathbf{a}\wedge(\mathbf{b}\wedge\mathbf{c}) = \langle \mathbf{a}, \mathbf{b}\times\mathbf{c}\rangle\omega$ and the alternating property $\mathbf{a}\wedge\mathbf{b}\wedge\mathbf{c} = \mathbf{b}\wedge\mathbf{c}\wedge\mathbf{a} = \mathbf{c}\wedge\mathbf{a}\wedge\mathbf{b}$ of the wedge product.
As the Hodge star operator preserves the inner product, $\langle \mathbf{a}\times\mathbf{b}, \mathbf{c}\times\mathbf{d}\rangle = \langle \mathbf{a}\wedge\mathbf{b}, \mathbf{c}\wedge\mathbf{d}\rangle$, and thus, Lagrange identity is a special case of equation (1.9).
Combining the permutation rule and the Lagrange identity, we obtain that

$$\langle \mathbf{a}\times(\mathbf{b}\times\mathbf{c}), \mathbf{d}\rangle = \langle \mathbf{d}\times\mathbf{a}, \mathbf{b}\times\mathbf{c}\rangle = \langle \mathbf{d},\mathbf{b}\rangle\langle \mathbf{a},\mathbf{c}\rangle - \langle \mathbf{d},\mathbf{c}\rangle\langle \mathbf{a},\mathbf{b}\rangle = \langle \langle \mathbf{a},\mathbf{c}\rangle\mathbf{b} - \langle \mathbf{a},\mathbf{b}\rangle\mathbf{c}, \mathbf{d}\rangle.$$

Since the first and last terms are equal for any $\mathbf{d}$, $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ must be equal to $\langle \mathbf{a},\mathbf{c}\rangle\mathbf{b} - \langle \mathbf{a},\mathbf{b}\rangle\mathbf{c}$.
Expanding the summands on the left-hand side of the Jacobi identity we obtain

$$(\langle \mathbf{a},\mathbf{c}\rangle\mathbf{b} - \langle \mathbf{a},\mathbf{b}\rangle\mathbf{c}) + (\langle \mathbf{b},\mathbf{a}\rangle\mathbf{c} - \langle \mathbf{b},\mathbf{c}\rangle\mathbf{a}) + (\langle \mathbf{c},\mathbf{b}\rangle\mathbf{a} - \langle \mathbf{c},\mathbf{a}\rangle\mathbf{b}) = \mathbf{0}. \qquad \square$$

Proposition 1.2.77. *Let $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ be a positively oriented orthonormal basis of V, and let (a^1, a^2, a^3) and (b^1, b^2, b^3) be the coordinates of the vectors $\mathbf{a}$ and $\mathbf{b}$ with respect to this basis. Then*

$$\mathbf{a}\times\mathbf{b} = \det\begin{pmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3\\ a^1 & a^2 & a^3\\ b^1 & b^2 & b^3\end{pmatrix}.$$

Proof. It is clear that $\mathbf{e}_1\times\mathbf{e}_2 = *(\mathbf{e}_1\wedge\mathbf{e}_2) = \mathbf{e}_3$, and similarly $\mathbf{e}_2\times\mathbf{e}_3 = \mathbf{e}_1$ and $\mathbf{e}_3\times\mathbf{e}_1 = \mathbf{e}_2$. Thus,

$$\begin{aligned}\mathbf{a}\times\mathbf{b} &= (a^1\mathbf{e}_1 + a^2\mathbf{e}_2 + a^3\mathbf{e}_3)\times(b^1\mathbf{e}_1 + b^2\mathbf{e}_2 + b^3\mathbf{e}_3)\\ &= (a^2b^3 - a^3b^2)\mathbf{e}_1 + (a^3b^1 - a^1b^3)\mathbf{e}_2 + (a^1b^2 - a^2b^1)\mathbf{e}_3. \qquad \square\end{aligned}$$

Sometimes we have to find a normal vector of an $(n-1)$-dimensional linear subspace W of an n-dimensional linear space V. If $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}$ is a basis of W, then the vector $*(\mathbf{a}_1\wedge\dots\wedge\mathbf{a}_{n-1})$ will be a normal vector of W. The $(n-1)$-variable operation $(\mathbf{a}_1, \dots, \mathbf{a}_{n-1}) \mapsto *(\mathbf{a}_1\wedge\dots\wedge\mathbf{a}_{n-1})$ is clearly a higher dimensional generalization of the cross product and it has analogous properties that can be proved in a similar way.

Proposition 1.2.78.

- *The vector* $\mathbf{c} = *(\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_{n-1})$ *is an alternating* $(n-1)$*-linear function of the vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$.
- $\mathbf{c}$ *is orthogonal to* $\mathbf{a}_i$ *for* $1 \leq i \leq n-1$.
- $\mathbf{c} = \mathbf{0}$ *if and only if the vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$ *are linearly dependent.*
- *If* $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$ *are linearly independent, then the length of* $\mathbf{c}$ *is the* $(n-1)$*-dimensional volume of the* $(n-1)$*-dimensional parallelepiped spanned by the vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$*, which is equal to* $\sqrt{\det \mathcal{G}(\mathbf{a}_1, \ldots, \mathbf{a}_{n-1})}$.
- $(\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}, \mathbf{c})$ *is a positively oriented basis of* V.
- *If the coordinates of* $\mathbf{a}_i$ *with respect to a positively oriented orthonormal basis* $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ *of* V *are* $(a_i^1, \ldots, a_i^n)$*, then*

$$\mathbf{c} = \det \begin{pmatrix} a_1^1 & \ldots & a_1^{n-1} & a_1^n \\ \vdots & \ddots & \vdots & \vdots \\ a_{n-1}^1 & \ldots & a_{n-1}^{n-1} & a_{n-1}^n \\ \mathbf{e}_1 & \ldots & \mathbf{e}_{n-1} & \mathbf{e}_n \end{pmatrix}.$$

The proof is left to the reader.

1.3 Geometry

1.3.1 Affine Geometry

Affine Spaces

In traditional axiomatic treatment of Euclidean geometry, vectors are defined as equivalence classes of ordered pairs of points (also called directed segments). If X is the set of points of the space, then an *ordered pair of points* is simply an element of the Cartesian product $X \times X$. If $(A, B) \in X \times X$ is an ordered pair of points, then A is called the *initial point*, and B is called the *endpoint of the ordered pair*. The ordered pairs of points (A, B) and (C, D) are said to be *equivalent* if the midpoint of the segment $[A, D]$ coincides with the midpoint of the segment $[B, C]$. It can be shown that this is indeed an equivalence relation. The equivalence classes are called *(free) vectors*. The set V of all vectors is equipped with a linear space structure. The sum of two vectors is constructed by the triangle or parallelogram rule, multiplication by scalars is defined in the usual way. The map $\Phi\colon X \times X \to V$, which assigns to each ordered pair of points (A, B) the vector $\Phi(A, B) = \overrightarrow{AB}$ represented by it, i.e., its equivalence class, satisfies the following properties.

(A1) For any $A \in X$ and $\mathbf{v} \in V$ there is a unique $B \in X$ such that $\Phi(A, B) = \mathbf{v}$.

(A2) (Triangle rule.) For any three points $A, B, C \in X$, we have $\Phi(A, B) + \Phi(B, C) = \Phi(A, C)$.

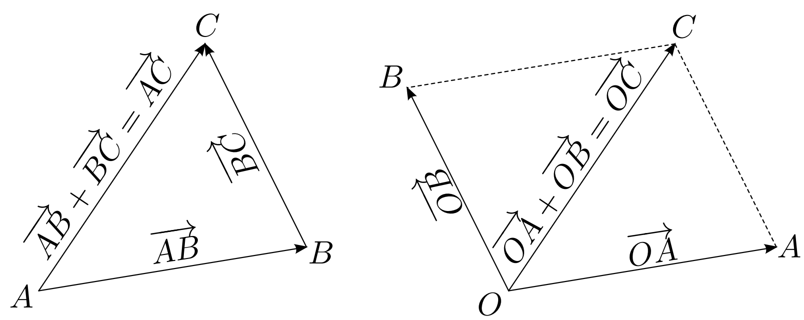

Figure 1.1: The construction of the sum of two vectors by the triangle and the parallelogram rule.

The notion of an affine space generalizes this picture.

Definition 1.3.1. An *affine space* is a triple $\mathbb{A} = (X, V, \Phi)$, where X is a set, the elements of which are called *points*, V is a linear space, the elements of which are called *vectors*, and $\Phi\colon X \times X \to V$ is a map satisfying conditions (A1) and (A2) above. The *dimension of the affine space* $\mathbb{A}$ is the (linear algebraic) dimension of the linear space V. ※

When it leads to no confusion, $\Phi(A, B)$ is also denoted by $\overrightarrow{AB}$.

Definition 1.3.2. Let $\mathbb{A} = (X, V, \Phi)$ and $\mathbb{B} = (Y, W, \Psi)$ be two affine spaces. Then an *affine transformation from* $\mathbb{A}$ *to* $\mathbb{B}$ consists of a map $T\colon X \to Y$ and a linear map $L\colon V \to W$ such that

$$L(\Phi(A, B)) = \Psi(T(A), T(B)) \text{ for all } A, B \in X. \tag{1.12}$$

Observe that since Φ is surjective, T determines L uniquely by (1.12). ※

Affine spaces form a category in which the morphisms are the affine transformations. An affine transformation is an isomorphism if and only if T is a bijection.

Example. Let V be an arbitrary linear space, $X = V$, and define $\Phi\colon X \times X \to V$ by $\Phi(\mathbf{p}, \mathbf{q}) = \mathbf{q} - \mathbf{p}$. Then $\mathbb{A}_V = (X, V, \Phi)$ is an affine space. Thus, every linear space carries an affine space structure. If V and W are linear spaces, then an affine transformation from $\mathbb{A}_V$ to $\mathbb{A}_W$ has the form $T(\mathbf{v}) = L(\mathbf{v}) + \mathbf{w}_0$, where $L\colon V \to W$ is a linear map, $\mathbf{w}_0 \in W$ is a fixed vector.

Linear spaces and affine spaces are very similar objects. The main difference is that in a linear space, we always have a distinguished point, the origin $\mathbf{0}$, whereas in an affine space, none of the points of X is distinguished. This is essentially the only difference, because if we choose any of the points of X for the origin, we can turn X into a linear space isomorphic to V.
Indeed, choose a point $O \in X$. By the first axiom (A1) of an affine space, the map $\Phi_O\colon X \to V$, $\Phi_O(P) = \Phi(O, P)$ is a bijection between X and V, and we have

$$\Phi(A, B) = \Phi_O(B) - \Phi_O(A)$$

by (A2). This identity means that Φ_O together with the linear map $L = \mathrm{id}_V$ is an isomorphism between the affine space $\mathbb{A} = (X, V, \Phi)$ and the affine space $\mathbb{A}_V$.
The vector $\Phi_O(P)$ is called the *position vector of A from the base point O.* Identification of points of an affine space with vectors with the help of a fixed base point O is called *vectorization of the affine space.*

Affine Subspaces

Definition 1.3.3. A nonempty subset $Y \subset X$ of the point set of an affine space $\mathbb{A} = (X, V, \Phi)$ is an *affine subspace of* $\mathbb{A}$ if the image $\Phi(Y \times Y) = W$ of $Y \times Y$ under Φ is a linear subspace of V and $(Y, W, \Phi|_{Y\times Y})$ is an affine space. W is called the *direction space* of the affine subspace, its elements are the *direction vectors of* Y, or the *vectors parallel to* Y. Since affine subspaces are affine spaces themselves, their *dimension* is properly defined. ※

The 0-dimensional affine subspaces of an affine space $\mathbb{A} = (X, V, \Phi)$ are the *points* of X. The 1-dimensional affine subspaces are called *straight lines.* The 2-dimensional affine subspaces are the ordinary *planes* of $\mathbb{A}$. In general, k-dimensional affine subspaces of an affine space will be called shortly *k-planes.* The $(n-1)$-planes of an n-dimensional affine space are called *hyperplanes.* The set of all k-dimensional affine subspaces of an affine space $\mathbb{A}$ is the *affine Grassmann manifold* $\mathrm{AGr}_k(\mathbb{A})$. At this moment we defined the affine Grassmann manifold just as a set, but this set can be endowed with a $(k+1)(n-k)$-dimensional manifold structure, which justifies its name.
The following proposition gives a more explicit description of affine subspaces in vectorized affine spaces.

Proposition 1.3.4. *A subset $Y \subset V$ of the affine space $\mathbb{A}_V$ is an affine subspace with direction space $W < V$ if and only if Y is a translate of W.*

Proof. If Y is an affine subspace with direction space W, then according to axiom (A1), for any $\mathbf{a} \in Y$, the map $Y \to W$, $\mathbf{b} \mapsto \mathbf{b} - \mathbf{a}$ is a bijection between Y and W, which means that Y is the translate of W by the vector $\mathbf{a}$.
Assume now that $Y = T_{\mathbf{a}}(W)$ is a translate of the linear subspace W of V, and show that Y is an affine subspace of $\mathbb{A}_V$ with direction space W. Two typical elements of Y have the form $\mathbf{y}_1 = \mathbf{a} + \mathbf{w}_1$ and $\mathbf{y}_2 = \mathbf{a} + \mathbf{w}_2$, where $\mathbf{w}_1$, $\mathbf{w}_2 \in W$. Then

$$\Phi(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{y}_2 - \mathbf{y}_1 = \mathbf{w}_2 - \mathbf{w}_1,$$

which shows that $\Phi(Y \times Y) = W$.
To check axiom (A1) for $(Y, W, \Phi|_{Y\times Y})$, observe that for $\mathbf{y}_0 = \mathbf{a} + \mathbf{w}_0 \in Y$, the map $Y \to W$, $\mathbf{y} \mapsto \Phi(\mathbf{y}_0, \mathbf{y}) = \mathbf{y} - \mathbf{y}_0$ is a translation by $-\mathbf{y}_0$ restricted onto Y. Since $T_{-\mathbf{y}_0} = T_{-\mathbf{w}_0} \circ T_{-\mathbf{a}}$ and $T_{-\mathbf{a}}$ maps Y onto W bijectively, and $T_{-\mathbf{w}_0}$ maps W onto itself bijectively, $T_{-\mathbf{y}_0}$ maps Y onto W bijectively.
Since the triangle rule (A2) is fulfilled for any triple in X, it is true for any three points in Y as well. □

Affine and 0-weight Linear Combinations

When we vectorize an affine space, points are identified with vectors and, therefore, we can take linear combinations of points. The result can be considered both a point and a vector. However, the resulting point, or vector may depend on the choice of the base point of the vectorization. From the viewpoint of affine geometry, linear combinations the result of which does not depend on the choice of the vectorization are of special importance.
Let us compare vectorizations with base points O and O'. If $\Phi(O, O') = \mathbf{a}$, and a point P corresponds to the vectors $\Phi_O(P) = \mathbf{p}$ and $\Phi_{O'}(P) = \mathbf{p}'$ under the vectorizations with these base points, then we have $\mathbf{p} = \mathbf{p}' + \mathbf{a}$ by the triangle rule (A2).
If the position vectors of the points $P_1, \ldots, P_k \in X$ from the base point O are $\mathbf{p}_1, \ldots, \mathbf{p}_k$, then the linear combination $\lambda_1\mathbf{p}_1 + \cdots + \lambda_k\mathbf{p}_k$ *as a vector in* V does not depend on the vectorization if and only if

$$\lambda_1\mathbf{p}_1 + \cdots + \lambda_k\mathbf{p}_k = \lambda_1(\mathbf{p}_1 - \mathbf{a}) + \cdots + \lambda_k(\mathbf{p}_k - \mathbf{a})$$

for any $\mathbf{a} \in V$, or, equivalently, if $\lambda_1 + \cdots + \lambda_k = 0$.

Definition 1.3.5. A linear combination $\lambda_1\mathbf{p}_1+\cdots+\lambda_k\mathbf{p}_k$ is a 0-*weight linear combination* if the sum $\lambda_1 + \cdots + \lambda_k$ of the coefficients is equal to 0. ※

According to the previous computation, we obtain the following statement.

Proposition 1.3.6. *For* 0*-weight linear combinations the vector* $\lambda_1\mathbf{p}_1+\cdots+\lambda_k\mathbf{p}_k \in V$ *does not depend on the vectorization, thus it can be denoted by* $\lambda_1 P_1+\cdots+\lambda_k P_k \in V$ *as well.*

We get a different condition on the coefficients if we want that the *point* with position vector $\lambda_1\mathbf{p}_1+\cdots+\lambda_k\mathbf{p}_k$ with base point O be the same as the point with position vector $\lambda_1\mathbf{p}'_1+\cdots+\lambda_k\mathbf{p}'_k$ with base point O'. For this we need that the equation

$$\lambda_1\mathbf{p}_1+\cdots+\lambda_k\mathbf{p}_k-\mathbf{a}=\lambda_1(\mathbf{p}_1-\mathbf{a})+\cdots+\lambda_k(\mathbf{p}_k-\mathbf{a})$$

should hold for any choice of $\mathbf{a}$. This condition is clearly equivalent to the condition that $\lambda_1+\cdots+\lambda_k=1$.

Definition 1.3.7. A linear combination $\lambda_1\mathbf{p}_1+\cdots+\lambda_k\mathbf{p}_k$ is an *affine combination* if the sum $\lambda_1+\cdots+\lambda_k$ of the coefficients is equal to 1. ※

The importance of affine combinations is summarized in the following proposition.

Proposition 1.3.8. *The point represented by an affine combination* $\lambda_1\mathbf{p}_1+\cdots+\lambda_k\mathbf{p}_k$ *of the position vectors of some points* $P_1,\dots,P_k$ *does not depend on the choice of the base point. This way, it makes sense to denote it by* $\lambda_1 P_1+\cdots+\lambda_k P_k \in X$.

Affine subspaces can be characterized in terms of affine combinations.

Proposition 1.3.9. *A subset* Y *of an affine space is an affine subspace if and only if it is nonempty and it contains all affine combinations of its points.*

Proof. Identify the space with $\mathbb{A}_V$ by vectorization.
If Y is an affine subspace, then it is a translation $T_{\mathbf{a}}W$ of its direction space. Since $\mathbf{0}\in W$, $\mathbf{a}\in Y$ is not empty. Furthermore, if $\mathbf{y}_i=\mathbf{a}+\mathbf{w}_i\in Y$, $(i=1,\dots,k)$ are some points, then an affine combination of the has the form

$$\begin{aligned}\lambda_1\mathbf{y}_1+\cdots+\lambda_k\mathbf{y}_k&=(\lambda_1+\cdots+\lambda_k)\mathbf{a}+(\lambda_1\mathbf{w}_1+\cdots+\lambda_k\mathbf{w}_k)\\&=\mathbf{a}+(\lambda_1\mathbf{w}_1+\cdots+\lambda_k\mathbf{w}_k).\end{aligned}$$

Since W is a linear space, $\mathbf{w}=\lambda_1\mathbf{w}_1+\cdots+\lambda_k\mathbf{w}_k\in W$, therefore, $\lambda_1\mathbf{y}_1+\cdots+\lambda_k\mathbf{y}_k=\mathbf{a}+\mathbf{w}\in Y$.
To show the other direction, assume now that Y is a nonempty subset of V containing all affine combinations of its points.
Consider the set $W$of all 0-weight linear combinations of elements of Y. Since the sum of two 0-weight linear combinations and any multiple of a 0-weight linear combination are 0-weight linear combinations, W *is a linear subspace of* V.

Choose a vector $\mathbf{a} \in Y$, and let us show that $Y = T_{\mathbf{a}}(W)$. Adding $\mathbf{a} = 1\cdot\mathbf{a}$ to a 0-weight linear combination of elements of Y we obtain an affine combination of some elements of Y, which belongs to Y by our assumptions on Y. Thus, $Y \supset T_{\mathbf{a}}(W)$. On the other hand, every $\mathbf{y}$ element of Y can be written as $\mathbf{y} = \mathbf{a} + (\mathbf{y} - \mathbf{a}) \in T_{\mathbf{a}}(W)$, which completes the proof. □

As a byproduct of the proof we obtain

Corollary 1.3.10. *The direction space of an affine subspace is the set of all* 0*-weight linear combinations of its elements.*

Corollary 1.3.11. *The intersection of affine subspaces of an affine space is either empty or it is also an affine subspace.*

Corollary 1.3.12. *For a nonempty subset $S \subset X$ there is a smallest affine subspace among the affine subspaces containing S.*

Definition 1.3.13. The smallest affine subspace containing the nonempty subset S is called the *affine subspace spanned by S* and it is denoted by $\mathrm{aff}[S]$. ※

Proposition 1.3.14. *The affine subspace spanned by the nonempty subset S consists of all affine combinations of the elements of S.*

Proof. Denote by Y_S the set of all affine combinations of elements of S. By Proposition 1.3.9, $\mathrm{aff}[S] \supseteq Y_S$. Since for any $\mathbf{s} \in S$, $1 \cdot \mathbf{s}$ is an affine combination of $\mathbf{s}$, therefore $S \subseteq Y_S$. It is easy to check that affine combinations of affine combinations of elements of S is again an affine combination of elements of S, so Y_S is an affine subspace containing S. However, among such affine subspaces $\mathrm{aff}[S]$ is the smallest one, so $\mathrm{aff}[S] \subseteq Y_S$. □

Affine Independence

A direction space of the affine subspace spanned by a system of $k+1$ points $P_0, \dots, P_k$ consists of 0-weight linear combinations of these points. This linear space is generated by the differences $P_1 - P_0, \dots, P_k - P_0$, therefore its dimension is at most k.

Definition 1.3.15. The points $P_0, \dots, P_k$ are called *affinely independent* if any of the following equivalent conditions is fulfilled.

- $P_0, \dots, P_k$ span a k-dimensional affine subspace.
- The vectors $P_1 - P_0, \dots, P_k - P_0$ are linearly independent.
- A 0-weight linear combination of $P_0, \dots, P_k$ equals $\mathbf{0}$ only if all the coefficients are equal to 0. ※

Affine Coordinate Systems

An *affine coordinate system* on an affine space $\mathbb{A} = (X, V, \Phi)$ is given by the following data:

- a point $O \in X$, which will be the origin of the coordinate system;
- and a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of V.

Given an affine coordinate system, the coordinates of a point P are the real numbers $(x^1, \ldots, x^n)$ for which

$$\Phi(O, P) = x^1\mathbf{e}_1 + \cdots + x^n\mathbf{e}_n.$$

Assigning to each point of X its coordinate vector gives an affine isomorphism between the affine space $\mathbb{A}$ and the affine space $\mathbb{A}_{\mathbb{R}^n}$.

Tangent Vectors at a Point, and the Tangent Bundle

A free vector $\mathbf{v} \in V$ of an affine space (X, V, Φ) has no given initial point or endpoint. It can be represented by any ordered pair of points (P, Q) for which $\Phi(P, Q) = \mathbf{v}$. Sometimes we have to consider vectors with a given base point instead of free vectors. For example, forces in Newtonian mechanics are vector like quantities. However, to describe a diagram of forces properly, it is not enough to know the directions and the magnitudes of the acting forces, we have to know also the (base) points at which the forces act on a given body. For instance, if we push a wardrobe by a horizontal force, sufficiently large to overcome friction, then the wardrobe will slide horizontally if we push it close to the floor, but it may fall over if it is pushed at the top. The reason of this fact is that the torque of a force depends on the point at which the force acts on a body.

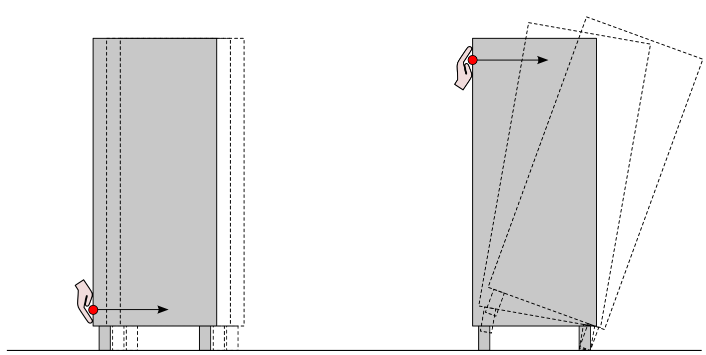

Figure 1.2: Pushing a body with equal forces at different base points.

Definition 1.3.16. Let $P \in X$ be a point of an affine space (X, V, Φ). A *vector based at* P or a *tangent vector at* P is a pair $(P, \mathbf{v})$, where $\mathbf{v} \in V$. The *tangent space* $T_P\mathbb{A}$ *(or* T_PX*) of the affine space* $\mathbb{A}$ *at* P is the set of all tangent vectors at P, i.e., $T_P\mathbb{A} = \{(P, \mathbf{v}) \mid \mathbf{v} \in V\}$. ※

Tangent vectors at P form a linear space with the operations

$$\lambda(P, \mathbf{v}) + \mu(P, \mathbf{w}) = (P, \lambda\mathbf{v} + \mu\mathbf{w}).$$

"Forgetting the base point" is a linear isomorphism $\iota_P \colon T_P\mathbb{A} \to V$, $\iota_P(P, \mathbf{v}) = \mathbf{v}$ between $T_P\mathbb{A}$ and V. If P and Q are two different points in X, then the linear isomorphism $\iota_Q^{-1} \circ \iota_P \colon T_P\mathbb{A} \to T_Q\mathbb{A}$ is called *parallel transport* between tangent spaces at P and Q.

Definition 1.3.17. The disjoint union of all tangent spaces of an affine space is called its tangent bundle $T\mathbb{A} = \bigcup_{P \in X} T_P\mathbb{A} = X \times V$. The map $\pi \colon T\mathbb{A} \to X$, which assigns to each tangent vector $(P, \mathbf{v})$ its base point P is called the *projection of the tangent bundle*. ※

1.3.2 Euclidean Spaces

Definition 1.3.18. An affine space (X, V, Φ) is a *Euclidean space* if the linear space V is endowed with positive definite symmetric bilinear function $\langle\,,\rangle$, making it a Euclidean linear space. ※

Since every linear subspace W of V inherits a Euclidean structure by restricting the inner product $\langle\,,\rangle$ onto $W \times W$, affine subspaces of a Euclidean space are Euclidean spaces as well.
The Euclidean structure enables us to introduce metric notions like distance, angle, area, volume etc.

Distance and the Isometry Group

Definition 1.3.19. The distance between two points P and Q of a Euclidean space (X, V, Φ) is $d(P, Q) = \|\Phi(P, Q)\|$. ※

Definition 1.3.20. A metric space is a pair (X, d), where X is a set, $d \colon X \times X \to \mathbb{R}$ is a functions, called *distance function*, satisfying the following axioms

- $d(P, Q) \geq 0$ for all $P, Q \in X$ and $d(P, Q) = 0$ if and only if $P = Q$.
- $d(P, Q) = d(Q, P)$ for all $P, Q \in X$ (symmetry).
- $d(P, Q) + d(Q, R) \geq d(P, R)$ for all $P, Q, R \in X$ (triangle inequality). ※

Proposition 1.3.21. *Every Euclidean space is a metric space with its distance function.*

Proof. The triangle inequality follows from Corollary 1.2.49

$$\begin{aligned} d(P,Q)+d(Q,R) &= \|\Phi(P,Q)\|+\|\Phi(Q,R)\| \geq \|\Phi(P,Q)+\Phi(Q,R)\| \\ &= \|\Phi(P,R)\| = d(P,R). \end{aligned}$$

The rest is trivial. □

Definition 1.3.22. A map $\varphi\colon X\to Y$ between the metric spaces (X,d_X) and (Y,d_Y) is said to be a *distance preserving map* or *isometry* if $d_Y(\varphi(P),\varphi(Q)) = d_X(P,Q)$ for all $P,Q\in X$. ※

An isometry is always injective but not necessarily surjective.

Exercise 1.3.23. Give an example of an isometry $\varphi\colon X\to X$ which maps a metric space (X,d) onto one of its proper subsets. ¤

Definition 1.3.24. Bijective isometries of a metric space form a group with respect to composition. The group of bijective isometries of the metric space X is called the *isometry group of* X, and is denoted by $\mathrm{Iso}(X)$. ※

Theorem 1.3.25. *Every isometry of a Euclidean space $\mathbb{E}^n=(X,V,\Phi)$ into itself is bijective. Isometries of $\mathbb{E}^n$ are affine transformations. An affine transformation $T\colon X\to X$ is an isometry if and only if the corresponding linear map $L\colon V\to V$, for which $L(\overrightarrow{AB})=\overrightarrow{T(A)T(B)}$, preserves the norm of vectors.*
A linear transformation of the Euclidean linear space V preserves the norm of vectors if and only if it preserves the dot product of vectors. If M is the matrix of L with respect to an orthonormal basis, then this property of L is also equivalent to the matrix equation

$$MM^T = M^TM = I. \tag{1.13}$$

Definition 1.3.26. Linear transformations of a Euclidean space preserving the dot product are called *orthogonal linear transformations.* Matrices satisfying equation (1.13) are called *orthogonal matrices.* ※

Definition 1.3.27. An *orientation preserving isometry* or *motion* of a Euclidean space is an isometry, the associated linear transformation of which has positive determinant. ※

Orientation preserving isometries form a normal subgroup $\mathrm{Iso}_+(\mathbb{E}^n)$ of index 2 in the isometry group $\mathrm{Iso}(\mathbb{E}^n)$ of the Euclidean space.

The Angle Between Affine Subspaces

Angle between affine subspaces of the same dimension will be defined as the angle between their direction spaces. Direction spaces are linear subspaces, so we first define the angle between linear subspaces.

Definition 1.3.28. The angle $\angle(e,f)$ between two 1-dimensional linear subspaces $e=\mathrm{lin}[\mathbf{v}]$ and $f=\mathrm{lin}[\mathbf{w}]$ of a Euclidean linear space V, is the smaller of α and $\pi-\alpha$, where α is the angle between $\mathbf{v}$ and $\mathbf{w}$. ※

Clearly, $\angle(e,f)$ is the unique angle in $[0,\pi/2]$ for which

$$\cos(\angle(e,f)) = \frac{|\langle \mathbf{v},\mathbf{w}\rangle|}{\|\mathbf{v}\|^2\cdot\|\mathbf{w}\|^2}.$$

Proposition 1.3.29. *$\angle$ is a metric on the projective space* $\mathrm{P}(V)$.

Proof. We prove only the triangle inequality, the rest is trivial. Let $e=\mathrm{lin}[\mathbf{v}]$, $f=\mathrm{lin}[\mathbf{w}]$, $g=\mathrm{lin}[\mathbf{z}]$ be three 1-dimensional subspaces and let us show that

$$\angle(e,f)+\angle(f,g)\geq\angle(e,g).$$

Assume that the vectors $\mathbf{v}$, $\mathbf{w}$ and $\mathbf{z}$ have unit length. Denote by α, β and γ the angles between $(\mathbf{v},\mathbf{w})$, $(\mathbf{w},\mathbf{z})$ and $(\mathbf{z},\mathbf{v})$ respectively. Changing the direction of the vectors $\mathbf{w}$ and $\mathbf{z}$ if necessary, we may assume that $\alpha,\beta\in[0,\pi/2]$. By Corollary 1.2.58, the Gram matrix of the vectors $\mathbf{v},\mathbf{w},\mathbf{z}$ has nonnegative determinant, thus

$$\begin{aligned}\det&\begin{pmatrix}1 & \cos(\alpha) & \cos(\gamma)\\ \cos(\alpha) & 1 & \cos(\beta)\\ \cos(\gamma) & \cos(\beta) & 1\end{pmatrix}\\ &= 1+2\cos(\alpha)\cos(\beta)\cos(\gamma)-\cos^2(\alpha)-\cos^2(\beta)-\cos^2(\gamma)\geq 0.\end{aligned}$$

This is a quadratic inequality for $\cos(\gamma)$, which is fulfilled if and only if $\cos(\gamma)$ is in the closed interval the endpoints of which are the roots of the polynomial

$$P(t)=(1-\cos^2(\alpha)-\cos^2(\beta))+2\cos(\alpha)\cos(\beta)t-t^2.$$

The roots of P are

$$\begin{aligned}&\cos(\alpha)\cos(\beta)\pm\sqrt{\cos^2(\alpha)\cos^2(\beta)+1-\cos^2(\alpha)-\cos^2(\beta)}=\\ &\qquad=\cos(\alpha)\cos(\beta)\pm\sqrt{(1-\cos^2(\alpha))(1-\cos^2(\beta))}=\cos(\alpha\pm\beta).\end{aligned}$$

Since the cosine function is strictly decreasing on the interval $[0,\pi]$, this implies

$$|\beta-\alpha|\leq\gamma\leq\alpha+\beta,$$

in particular,

$$\angle(e,f)+\angle(f,g)=\alpha+\beta\geq\gamma\geq\min\{\gamma,\pi-\gamma\}=\angle(e,g).$$ □

We define the angle between k-dimensional linear subspaces using the Plücker embedding $\mathrm{Gr}_k(V) \to \mathrm{P}(\Lambda^k V)$ (see Corollary 1.2.33).

Definition 1.3.30. The *angle between two k-dimensional linear subspaces* W_1, W_2 of a linear space V is the angle between the 1-dimensional linear subspaces $\mathrm{lin}[\mathbf{w}_1^1 \wedge \cdots \wedge \mathbf{w}_1^k]$ and $\mathrm{lin}[\mathbf{w}_2^1 \wedge \cdots \wedge \mathbf{w}_2^k]$, where $\mathbf{w}_1^1, \ldots, \mathbf{w}_1^k$ is a basis of W_1, $\mathbf{w}_2^1, \ldots, \mathbf{w}_2^k$ is a basis of W_2. ※

By Proposition 1.2.67, the angle α between W_1 and W_2 is the unique angle in the interval $[0, \pi/2]$ for which

$$\cos(\alpha) = \frac{\left| \det \begin{pmatrix} \langle \mathbf{w}_1^1, \mathbf{w}_2^1 \rangle & \ldots & \langle \mathbf{w}_1^1, \mathbf{w}_2^k \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{w}_1^k, \mathbf{w}_2^1 \rangle & \ldots & \langle \mathbf{w}_1^k, \mathbf{w}_2^k \rangle \end{pmatrix} \right|}{\sqrt{\det \mathcal{G}(\mathbf{w}_1^1, \ldots, \mathbf{w}_1^k)} \cdot \sqrt{\det \mathcal{G}(\mathbf{w}_2^1, \ldots, \mathbf{w}_2^k)}}.$$

According to Proposition 1.3.29, the *angle between subspaces is a metric on the Grassmann manifold* $\mathrm{Gr}_k(V)$.

Definition 1.3.31. The *angle between two affine subspaces of a Euclidean space* is the angle between their direction spaces. ※

Since two different affine subspaces can have the same direction space, the angle function is *not* a metric on the affine Grassmann manifold $\mathrm{AGr}_k(\mathbb{A})$.

Cartesian Coordinate Systems

Definition 1.3.32. An affine coordinate system given by the origin O and the basis vectors $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is a Cartesian coordinate system, if the basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is orthonormal. ※

A Cartesian coordinate system defines an isometric isomorphism between any n-dimensional Euclidean space and the *standard model* $\mathbb{A}_{\mathbb{R}^n}$ *of the n-dimensional Euclidean space*, where the inner product on $\mathbb{R}^n$ is the standard dot product.

Equations and Parameterizations of Affine Subspaces

Linear subspaces

Let us first deal with equations and parameterizations of linear subspaces of a linear space V. A k-dimensional linear subspace W of V can be given by k linearly independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ spanning W. Then a vector $\mathbf{x} \in V$ belongs to W if and only if $\mathbf{x}$ can be written as a linear combination of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$.

This way, the mapping

$$\mathbf{r}\colon \mathbb{R}^k \to \mathbf{V}, \qquad \mathbf{r}(x^1, \dots, x^k) = x^1\mathbf{a}_1 + \cdots + x^k\mathbf{a}_k$$

maps $\mathbb{R}^k$ bijectively onto W. The map $\mathbf{r}$ is called a *(linear) parameterization* of W. A linear subspace has many linear parameterizations, since it has many different bases.

We can also define a linear subspace as the kernel of a linear map, i.e., as the set of solutions of a system of linear equations.

For this purpose, observe that $\mathbf{x}$ is in W if and only if $\mathbf{a}_1, \dots, \mathbf{a}_k, \mathbf{x}$ are linearly dependent. By Proposition 1.2.32, linear dependence is equivalent to the equation

$$\mathbf{x} \wedge \mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_k = \mathbf{0}.$$

This is called the *equation of the linear subspace* W. This equation requires the vanishing of a $(k+1)$-vector. If $\dim V = n$, then the space of $\Lambda^{k+1}V$ has dimension $\binom{n}{k+1}$. If we fix a basis in V, then the equation of W becomes equivalent to a system of $\binom{n}{k+1}$ equations requiring the vanishing of all the coordinates of the $(k+1)$-vector on the left-hand side. Since the coordinates are linear functions of $\mathbf{x}$, all the equations in the system are linear. Obviously, these equations are not linearly independent in general. By the dimension formula (1.2) any maximal linearly independent subsystem of this system of equations contains exactly $(n-k)$ equations. Thus there are many different ways to convert the equation of a linear subspace to an independent system of $(n-k)$ linear equations.

In the special case $k = n-1$ of linear hyperplanes, however, we have only one linear equation on the coordinates, and it is uniquely determined by W up to a scalar multiplier. If we introduce an orientation and a Euclidean structure on V, then the equation of W can be written also as

$$\langle \mathbf{x}, *(\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_{n-1})\rangle = \mathbf{0}.$$

This way, W contains all vectors orthogonal to $*(\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_{n-1})$. This last vector is a *normal vector of* W.

Affine subspaces

Assume that we want to parameterize or write the equation of a k-dimensional affine subspace W spanned by $k+1$ affinely independent points $\mathbf{a}_0, \dots, \mathbf{a}_k$. It is not difficult to parameterize W, that is to obtain it as the image of a bijection $\mathbf{r}\colon \mathbb{R}^k \to W$. We know that a point belongs to W if and only if it is an affine combination of the points $\mathbf{a}_0, \dots, \mathbf{a}_k$, consequently, the mapping

$$\mathbf{r}(x^1, \dots, x^k) = (1 - (x^1 + \cdots + x^k))\mathbf{a}_0 + x^1\mathbf{a}_1 + \cdots + x^k\mathbf{a}_k$$

is a *parameterization of* W.

To write the equation of an affine subspace we use a trick rooted in projective geometry. Vectorize our affine space by choosing an origin. This identifies the space with $\mathbb{A}_V$, where V is a linear space. Consider the linear space $\langle V\rangle = \mathbb{R}\oplus V$ and embed V into $\langle V\rangle$ by the mapping $\mathbf{v}\mapsto\bar{\mathbf{v}}$, where $\bar{\mathbf{v}} = (1,\mathbf{v})$. The image $\bar{V}$ of V is the translate of the linear subspace $\{0\}\oplus V$ by the vector $(1,\mathbf{0})$, thus, it is an affine subspace of $\langle V\rangle$, just as the image $\bar{W}$ of W. Denote by $\langle W\rangle$ the linear subspace of $\langle V\rangle$ spanned by $\bar{W}$

$$\langle W\rangle = \mathrm{lin}[\bar{W}] = \mathrm{aff}[\bar{W}\cup\{(0,\mathbf{0})\}].$$

$\langle W\rangle$ is a $(k+1)$-dimensional linear subspace and uniquely determines W as the projection of $\bar{W} = \langle W\rangle\cap\bar{V}$ onto V. This way, *the map* $W\mapsto\langle W\rangle$ *gives an embedding of the affine Grassmann manifold* $\mathrm{AGr}_k(V)$ *into the linear Grassmann manifold* $\mathrm{Gr}_{k+1}(\langle V\rangle)$.

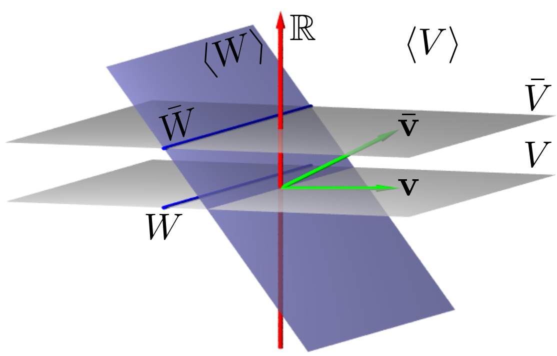

Figure 1.3: The correspondence between affine subspaces of V *and linear subspaces of* $\langle V\rangle$*.*

Using this picture, $\mathbf{x}$ belongs to the affine subspace W if and only if $\bar{\mathbf{x}}\in\langle W\rangle$. Since the linear subspace $\langle W\rangle$ is spanned by the vectors $\bar{\mathbf{a}}_0,\dots,\bar{\mathbf{a}}_k$, the *equation of the affine subspace* W is

$$\bar{\mathbf{x}}\wedge\bar{\mathbf{a}}_0\wedge\cdots\wedge\bar{\mathbf{a}}_k = \mathbf{0}.$$

Introducing a basis in V, this equation can be converted into a system of equations, in which each equation requires the vanishing of a linear combination of the coordinates of $\bar{\mathbf{x}}$. This means that every affine subspace can be defined by a system of inhomogeneous linear equations of the form

$$c_0 + c_1x^1 + \cdots + c_nx^n = 0. \tag{1.14}$$

The number of independent equations that define W is $(n+1)-(k+1) = n-k$. In the case of a hyperplane, we need only one equation of the form (1.14).

In a Cartesian coordinate system, the coefficients describe the hyperplane as follows. $\mathbf{N} = (c_1, \dots, c_n)$ is a normal vector of the hyperplane, and the hyperplane goes through the point $-c_0\mathbf{N}/\|\mathbf{N}\|^2$. The constant c_0 can be computed from any point $\mathbf{p}_0$ lying in the hyperplane as $c_0 = -\langle \mathbf{N}, \mathbf{p}_0 \rangle$.

Equations of Spheres

Definition 1.3.33. Let Σ be a $(k+1)$-dimensional subspace of an n-dimensional Euclidean space, $O \in \Sigma$ be a point, $r > 0$ be a positive number. Then the *k-dimensional sphere* or shortly the *k-sphere* in Σ centered at O with radius r is the set of those points in Σ the distance of which from O is r. A *hypersphere* is an $(n-1)$-dimensional sphere. The case $r = 0$ is considered to be a *degenerate case*, when the sphere degenerates to a point. ※

In the non-degenerate case, a k-sphere determines its $(k+1)$-plane, its center and its radius uniquely.
To write the equations of k-spheres introduce a Cartesian coordinate system on the space. This identifies our space with $\mathbb{R}^n$.
Consider first hyperspheres ($\Sigma = \mathbb{R}^n$). If $\mathbf{o}$ is the center of the sphere, then the equation of the sphere is

$$\|\mathbf{x} - \mathbf{o}\|^2 - r^2 = \|\mathbf{x}\|^2 + \langle -2\mathbf{o}, \mathbf{x} \rangle + (\|\mathbf{o}\|^2 - r^2) = 0.$$

This proves that every hypersphere can be defined by an equation of the form

$$a\|\mathbf{x}\|^2 + \langle \mathbf{b}, \mathbf{x} \rangle + c = 0, \tag{1.15}$$

where $a, c \in \mathbb{R}$, $\mathbf{b} \in \mathbb{R}^n$.

Definition 1.3.34. Equations of the form (1.15) will be called *hypersphere equations* or shortly *sphere equations.* ※

The solution set of a sphere equation is described in the following proposition.

Proposition 1.3.35. *Let $S = \{\mathbf{x} \mid a\|\mathbf{x}\|^2 + \langle \mathbf{b}, \mathbf{x} \rangle + c = 0\}$ be the set of solutions of the sphere equation* (1.15)*. Then we have the following cases.*

- *If the equation is* trivial*, i.e., $a = c = 0$, $\mathbf{b} = \mathbf{0}$, then the solution set is the whole space $\mathbb{R}^n$.*

- *If the equation is* contradictory*, i.e., $a = 0$, $\mathbf{b} = \mathbf{0}$, but $c \neq 0$, then $S = \emptyset$.*

- *If the equation is* inhomogeneous linear*, that is $a = 0$, but not trivial and not contradictory, that is $\mathbf{b} \neq \mathbf{0}$, then S is a hyperplane.*

- *If the equation is* quadratic, *i.e., $a \neq 0$, then S depends on the sign of the discriminant $d = \|\mathbf{b}\|^2 - 4ac$.*
  - *If $d > 0$, then S is a hypersphere centered at $-\mathbf{b}/(2a)$ with radius $\sqrt{d}/|2a|$.*
  - *If $d = 0$, then S is a degenerated hypersphere of radius 0 which contains a single point $-\mathbf{b}/(2a)$.*
  - *If $d < 0$, then $S = \emptyset$.*

Proof. The linear case is simple, the quadratic case follows from the equivalent rearrangement

$$\left\|\mathbf{x} - \left(-\frac{\mathbf{b}}{2a}\right)\right\|^2 = \frac{\|\mathbf{b}\|^2 - 4ac}{4a^2}$$

of equation (1.15). □

Proposition 1.3.36. *The intersection S of a hypersphere of radius r centered at O and a hyperplane Σ in an n-dimensional Euclidean space is described as follows. Let the orthogonal projection of O onto Σ be O' and denote by d the distance OO'. Then if $r > d$, then S is an $(n-2)$-sphere in Σ with radius $\sqrt{r^2 - d^2}$, and center at O'. S consists of the single point O' when $r = d$ and $S = \emptyset$ when $r > d$.*

The proof is a corollary of the Pythagorean theorem, since for any P in Σ, the triangle $POO'\triangle$ is a right triangle.

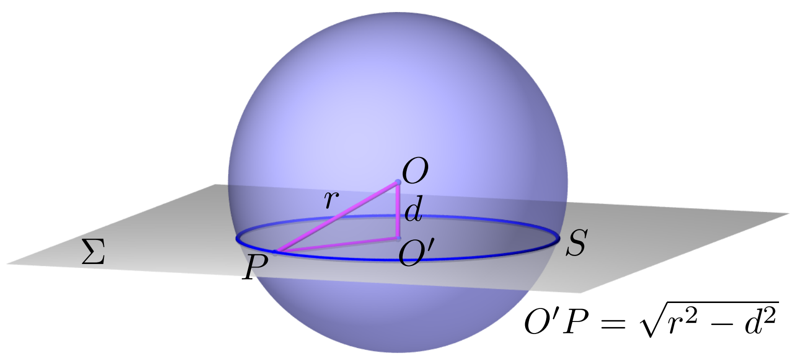

Figure 1.4: The intersection of a sphere and a hyperplane.

Corollary 1.3.37. *The intersection of a finite system of hyperspheres and hyperplanes is either empty or an affine subspace or a k-sphere.*

Proof. From an algebraic viewpoint, the intersection is the set of solutions of a finite number of sphere equations. If there is a quadratic sphere equation in the system, then subtracting a suitable multiple of it from the other equations, we can make all the other equations linear, so we may assume that there is at most one hypersphere in the family.
If there is no hypersphere at all, the statement follows from Corollary 1.3.11. If there is a hypersphere in the family, then we can prove the statement inductively, using the previous proposition. □

To write the equation of a k-sphere, we use an embedding of $\mathbb{R}^n$ into a larger linear, space analogous to the affine embedding $V \to \langle V \rangle = \mathbb{R} \oplus V$ that was used when we wrote equations of affine subspaces.
To simplify notation, denote the Euclidean linear space $\mathbb{R}^n$ by V. Set $\langle\langle V \rangle\rangle = \mathbb{R} \oplus V \oplus \mathbb{R}$. Embed V into $\langle\langle V \rangle\rangle$ with the map

$$\mathbf{v} \mapsto \breve{\mathbf{v}}, \text{ where } \breve{\mathbf{v}} = (1, \mathbf{v}, \|\mathbf{v}\|^2). \tag{1.16}$$

Because of the quadratic term in the last coordinate, this is not an affine embedding. Its image $\breve{V}$ is a paraboloid. The advantage of this embedding is that a linear equation on the coordinates of $\breve{\mathbf{v}}$ is a hypersphere equation on $\mathbf{v}$.

Proposition 1.3.38. *Let $\mathbf{a}_0, \ldots, \mathbf{a}_k$ be $(k+1) \geq 2$ affinely independent points. Then there is a unique $(k-1)$-sphere through these points and it can be defined by the equation*

$$\breve{\mathbf{x}} \wedge \breve{\mathbf{a}}_0 \wedge \cdots \wedge \breve{\mathbf{a}}_k = \mathbf{0}. \tag{1.17}$$

Proof. The $(k-1)$-sphere must be in the unique k-plane Σ spanned by the points. If the points were contained in two different $(k-1)$-spheres of Σ, then they would be contained in their intersection, which is a $(k-2)$-sphere. This would contradict affine independence of the points, since every $(k-2)$-sphere is contained in a $(k-1)$-plane. Thus, uniqueness is proved.
To complete the proof, we should check that the set given by the equation defines a $(k-1)$-sphere passing through the given points. Let S be the set of solutions of (1.17).
Since (1.17) is equivalent to a system of linear equations on the coordinates of $\breve{\mathbf{x}}$, which is a system of sphere equations on $\mathbf{x}$. Thus, by Corollary 1.3.37, S is either empty, or an affine subspace, or a sphere.
The set S cannot be empty, since it contains the points $\mathbf{a}_0, \ldots, \mathbf{a}_k$ as

$$\breve{\mathbf{a}}_i \wedge \breve{\mathbf{a}}_0 \wedge \cdots \wedge \breve{\mathbf{a}}_k = \mathbf{0}.$$

This also gives a lower bound on the dimension of S. If it is an m-sphere, then $m \geq k-1$, if it is an m-plane, then $m \geq k$.

If $\mathbf{x} \in S$, then $\breve{\mathbf{x}}$ is a linear combination of the vectors $\breve{\mathbf{a}}_0, \ldots, \breve{\mathbf{a}}_k$, hence

$$(1, \mathbf{x}, \|\mathbf{x}\|^2) = \alpha_0(1, \mathbf{a}_0, \|\mathbf{a}_0\|^2) + \cdots + \alpha_k(1, \mathbf{a}_k, \|\mathbf{a}_k\|^2)$$

for some $\alpha_0, \ldots, \alpha_k \in \mathbb{R}$. This equation splits into three components

$$\begin{aligned} 1 &= \alpha_0 + \cdots + \alpha_k, \\ \mathbf{x} &= \alpha_0\mathbf{a}_0 + \cdots + \alpha_k\mathbf{a}_k, \\ \|\mathbf{x}\|^2 &= \alpha_0\|\mathbf{a}_0\|^2 + \cdots + \alpha_k\|\mathbf{a}_k\|^2. \end{aligned}$$

The first two equations show that $\mathbf{x}$ must be an affine combination of the points $\mathbf{a}_0, \ldots, \mathbf{a}_k$. Thus, $S \subset \Sigma$, and we have only two possibilities left. Either S is a $(k-1)$-sphere, in which case we are done, or $S = \Sigma$. The latter case can be excluded by showing that the midpoint $(\mathbf{a}_0 + \mathbf{a}_1)/2$ is not in S. If it were in S, then by affine independence of the $\mathbf{a}_i$'s, the only possible choice for the α_i's would be $\alpha_0 = \alpha_1 = 1/2$ and $\alpha_i = 0$ for $2 \le i \le k$, however, the third equation is not fulfilled with these coefficients as

$$\left(\frac{1}{2}\|\mathbf{a}_0\|^2 + \frac{1}{2}\|\mathbf{a}_1\|^2\right) - \left\|\frac{\mathbf{a}_0 + \mathbf{a}_1}{2}\right\|^2 = \left\|\frac{\mathbf{a}_0 - \mathbf{a}_1}{2}\right\|^2 > 0.$$

This proves the proposition. □

Denote by $\mathrm{SPH}_k(V)$ the set of all k-spheres in the Euclidean linear space V. As a corollary of the previous proposition, we can construct an embedding $\mathrm{SPH}_k(V) \to \mathrm{Gr}_1(\Lambda^{k+1}\langle\!\langle V\rangle\!\rangle)$ in the following way. Given a k-sphere S, choose k affinely independent points $\mathbf{a}_0, \ldots, \mathbf{a}_k$ from it, and assign to S the 1-dimensional linear space spanned by the $(k+1)$-vector $\breve{\mathbf{a}}_0 \wedge \cdots \wedge \breve{\mathbf{a}}_k \in \Lambda^{k+1}\langle\!\langle V\rangle\!\rangle$.

Exercise 1.3.39. Prove that the center of the sphere defined by (1.17) is the point $\alpha_0\mathbf{a}_0 + \cdots + \alpha_k\mathbf{a}_k$, where the coefficients are obtained as the solution of the following system of linear equations

$$\begin{gathered} \alpha_0 + \cdots + \alpha_k = 1 \\ \alpha_0\langle \mathbf{a}_i - \mathbf{a}_0, \mathbf{a}_0\rangle + \cdots + \alpha_k\langle \mathbf{a}_i - \mathbf{a}_0, \mathbf{a}_k\rangle = \|\mathbf{a}_i\|^2 - \|\mathbf{a}_0\|^2 \text{ for } i = 1, \ldots, k. \end{gathered}$$ ¤

1.4 Topology

Definition 1.4.1. A pair (X, τ) is said to be a *topological space* if X is a set, τ is a collection of subsets of X, which we call the *open subsets* of X, such that

(i) the empty set $\emptyset$ and X are in τ;

(ii) the intersection of any two open subsets is also open;

(iii) the union of an arbitrary family of open subsets is open.

The family τ of open subsets is called the *topology on* X. ※

Definition 1.4.2. We say that a subset U of X is a *neighborhood of a point* $\mathbf{x} \in X$, if there is an open subset V such that $\mathbf{x} \in V \subset U$. ※

Definition 1.4.3. A subset Y of a topological space (X, τ) is said to be *closed* if $X \setminus Y$ is open. ※

Warning. Most of the subsets of a topological space are neither open nor closed.

Definition 1.4.4. If $A \subseteq X$ is a subset of a topological space (X, τ), then a point $\mathbf{p} \in X$ is an *interior point of* A if A is a neighborhood of $\mathbf{p}$. The point $\mathbf{p}$ is an *exterior point of* A if $X \setminus A$ is a neighborhood of $\mathbf{p}$. When all open sets containing $\mathbf{p}$ intersect both A and $X \setminus A$, $\mathbf{p}$ is called a *boundary point of* A. The sets of interior, exterior and boundary points of A are denoted by $\operatorname{int} A$, $\operatorname{ext} A$ and ∂A respectively. The *closure of the set* A is the union $\bar{A} = A \cup \partial A$. ※

Definition 1.4.5. A subset $A \subseteq X$ a topological space (X, τ) is said to be *dense in* X if its closure $\bar{A}$ is equal to the whole space X. The subset A is called *nowhere dense* if its closure has no interior points at all. ※

Examples.

- Let X be an arbitrary set. The *discrete topology on* X is the "maximal topology" on X, in which every subset is open.
- The *anti-discrete topology on* X is the "minimal topology" on X, in which only the empty set and X are open.

Metric Topology

The metric topology of a metric space (X, d) is introduced as follows.

Definition 1.4.6. The *open ball* in X with center $\mathbf{x} \in X$ and radius $\varepsilon > 0$ (or an *ε-ball centered at* $\mathbf{x}$) is the set $B_\varepsilon(\mathbf{x}) = \{\mathbf{y} \in X \mid d(\mathbf{x}, \mathbf{y}) < \varepsilon\}$. Similarly, the *closed ball* in X with center $\mathbf{x} \in X$ and radius $\varepsilon > 0$ (or a *closed ε-ball centered at* $\mathbf{x}$) is defined as the set $\bar{B}_\varepsilon(\mathbf{x}) = \{\mathbf{y} \in X \mid d(\mathbf{x}, \mathbf{y}) \leq \varepsilon\}$. ※

Definition 1.4.7. A subset U of a metric space X is called *open* if for each $\mathbf{x} \in U$ there is a positive ε such that the ball $B_\varepsilon(\mathbf{x})$ is contained in U. ※

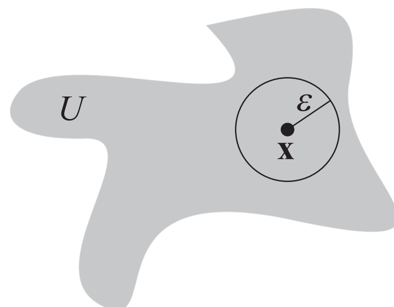

Figure 1.5: Open subsets of a metric space.

Proposition 1.4.8. *The family of open subsets is a topology on X.*

Proof. Obviously, the empty set and X are open. If U and V are open subsets, and $\mathbf{x}$ is a common point of them, then there exist positive numbers $\varepsilon_1, \varepsilon_2$ such that $B_{\varepsilon_1}(\mathbf{x}) \subset U$ and $B_{\varepsilon_2}(\mathbf{x}) \subset V$. Let ε be the smaller of ε_1 and ε_2. Then $B_\varepsilon(\mathbf{x}) \subset U \cap V$, showing that the intersection $U \cap V$ is open. Finally, let $\{U_i : i \in I\}$ be an arbitrary family of open sets, and $\mathbf{x}$ be an element of their union. Then we can find an index $j \in I$ and a positive ε such that $B_\varepsilon(\mathbf{x}) \subset U_j \subset \bigcup_{i \in I} U_i$, thus the union $\bigcup_{i \in I} U_i$ is open. □

This topology is referred to as the *metric topology of X*. A topology is called *metrizable*, if it can be derived from a metric. Characterization of metrizable topologies is a nontrivial problem of general topology, answered partially by Urysohn's metrization theorem and completely by the theorems of Nagata, Smirnov and Bing (see [9], Chapter 6).

Examples.

- Every affine space (over $\mathbb{R}$) can be turned into a Euclidean space by introducing an inner product on the linear space of its vectors. It can be checked that though the metric of the space depends on the choice of the inner product, the topology does not. Thus, every affine space has a *standard topology* metrizable by any of the Euclidean structures on it.

- Introducing a positive definite inner product on a linear space V, the angle between k-dimensional linear subspaces is a metric on the Grassmann manifold $\mathrm{Gr}_k(V)$. As in the previous example, the metric itself depends on the Euclidean structure on V, but the metric topology induced by it does not. This topology is the *standard topology on Grassmann manifolds.*

Subspace Topology

The topology of a topological space defines a topology on every of its subsets by the following construction.

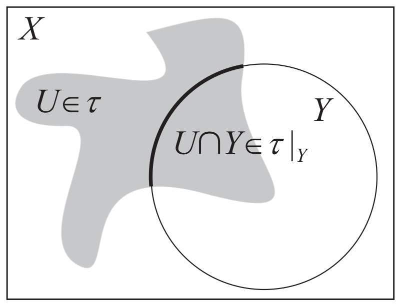

Figure 1.6: Open subsets in the subspace topology of $Y \subset X$.

Proposition 1.4.9. *Let Y be a subset of a topological space (X, τ). Then the family $\tau|_Y = \{U \cap Y | U \in \tau\}$ is a topology on Y.*

Proof. The proof of the proposition is straightforward from the following identities:

(i) $\emptyset \cap Y = \emptyset, \quad X \cap Y = Y$;

(ii) $(U \cap Y) \cap (V \cap Y) = (U \cap V) \cap Y$;

(iii) $\bigcup_{i \in I}(U_i \cap Y) = (\bigcup_{i \in I} U_i) \cap Y$. □

Definition 1.4.10. The topology $\tau|_Y$ is called the *subspace topology* or the *topology induced on Y by τ*. ※

Examples.

- As a special case of the subspace topology, all subsets of an affine space, in particular, all spheres of a Euclidean space inherit a subspace topology from the ambient space.

- At the end of Section 1.3.2 the affine Grassmann manifolds and the set of k-spheres of a Euclidean linear space were embedded into a projective space, which has a standard topology. Through these embeddings, affine Grassmann manifolds $\mathrm{AGr}_k(V)$ and the sets $\mathrm{SPH}_k(V)$ of k-spheres also inherit a subspace topology.

Factor Topology

The topology of the Grassmann manifolds of all k-dimensional affine or linear subspaces of an n-dimensional linear space can also be described with the help of the *factor space topology* construction.

Assume that a topological space (X,τ) is divided into a disjoint union of its subsets. Such a subdivision can always be thought of as a splitting of X into the equivalence classes of an equivalence relation $\sim$ on X. Denoting by $Y = X/_\sim$ the set of equivalence classes we have a natural mapping $\pi\colon X \to Y$ assigning to an element $\mathbf{x} \in X$ its equivalence class $[\mathbf{x}] \in Y$.

Proposition 1.4.11. *The set*

$$\tau' = \{U \subset Y | \pi^{-1}(U) \in \tau\}$$

is a topology on Y.

Proof. The proof follows from the following set theoretical identities:

(i) $\pi^{-1}(\emptyset) = \emptyset, \quad \pi^{-1}(Y) = X;$

(ii) $\pi^{-1}(U \cap V) = \pi^{-1}(U) \cap \pi^{-1}(V);$

(iii) $\pi^{-1}(\bigcup_{i\in I} U_i) = \bigcup_{i\in I} \pi^{-1}(U_i).$ □

Definition 1.4.12. The family τ' is called the *factor space topology* on Y. ※

As an example, consider the set

$$V(n,k) = \{(\mathbf{x}_1,\dots,\mathbf{x}_k) \in (\mathbb{R}^n)^k \,|\, \mathbf{x}_1,\dots,\mathbf{x}_k \text{ are linearly independent}\}.$$

$V(n,k)$ is an open subset in $(\mathbb{R}^n)^k = \mathbb{R}^{nk}$ hence it inherits a subspace topology from the standard topology of $\mathbb{R}^{nk}$. If we say that two elements of $V(n,k)$ are equivalent if they span the same k-dimensional linear subspace of $\mathbb{R}^n$, then the set $V(n,k)/_\sim$ of equivalence classes is essentially the same as the set $\mathrm{Gr}_k(\mathbb{R}^n)$ of all k-dimensional linear subspaces of $\mathbb{R}^n$. This set becomes a topological space with the factor space topology.

We can define a factor topology on the *affine Grassmann manifolds* similarly. We set

$$\tilde{V}(n,k) = \{(\mathbf{x}_0,\dots,\mathbf{x}_k) \in (\mathbb{R}^n)^{k+1} \mid \mathbf{x}_0,\dots,\mathbf{x}_k \text{ are not in a } (k-1)\text{-plane}\},$$

furnish $\tilde{V}(n,k)$ with the subspace topology inherited from $\mathbb{R}^{n(k+1)}$ and define an equivalence relation on $\tilde{V}(n,k)$ by $(\mathbf{x}_0,\dots,\mathbf{x}_k) \sim (\mathbf{y}_0,\dots,\mathbf{y}_k) \iff \mathbf{x}_0,\dots,\mathbf{x}_k$ and $\mathbf{y}_0,\dots,\mathbf{y}_k$ span the same k-plane.

$\tilde{V}(n,k)/_\sim$ is essentially the set of $\mathrm{AGr}_k(\mathbb{R}^n)$ affine k-dimensional subspaces $\mathbb{R}^n$ and it is equipped with the factor space topology.

Exercise 1.4.13. Show that the factor space topologies on $\mathrm{Gr}_k(\mathbb{R}^n)$ and $\mathrm{AGr}_k(\mathbb{R}^n)$ coincide with the metric topologies introduced above. ¤

Product Topology

Definition 1.4.14. If $(X_1, \tau_1), \ldots, (X_n, \tau_n)$ are topological spaces, then the *product topology* on the Cartesian product $X_1 \times \cdots \times X_n$ is defined as follows. A subset $U \subseteq X_1 \times \cdots \times X_n$ is open with respect to the product topology if and only if for each point $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in U$, we can find open subsets $U_1 \in \tau_1, \ldots, U_n \in \tau_n$ such that $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in U_1 \times \cdots \times U_n \subseteq U$. ※

Exercise 1.4.15. Show that the standard topology on $\mathbb{R}^n$ coincides with the product topology on n copies of $\mathbb{R}$. ¤

Convergence, Continuity

Definition 1.4.16. We say that a sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$ of points of a topological space (X, τ) converges to a point $\mathbf{x} \in X$, if for any neighborhood U of $\mathbf{x}$ there is a natural number N such that for $n > N$, $\mathbf{x}_n \in U$. ※

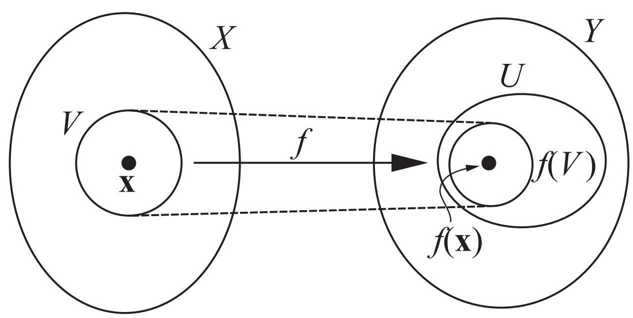

Figure 1.7: Continuity at a point.

Definition 1.4.17. Let (X, τ) and (Y, τ') be topological spaces. A mapping $f \colon X \to Y$ is said to be *continuous at the point* $\mathbf{x}$ *(with respect to the given topologies)* if for each neighborhood U of $f(\mathbf{x})$, we can find an open set $V \in \tau$ such that $\mathbf{x} \in V$ and $f(V) \subseteq U$. The mapping f is *continuous* if it is continuous at each point or, equivalently, if for each $U \in \tau'$, we have $f^{-1}(U) \in \tau$. ※

Exercise 1.4.18. Show that for $\mathbb{R}^n$ the above definition is equivalent to the "$\varepsilon - \delta$" definition of continuity at a point: $f \colon \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\mathbf{x} \in \mathbb{R}^n$ with respect to the standard topologies if and only if for any $\varepsilon > 0$ one can find a positive δ such that $|\mathbf{x} - \mathbf{x}'| < \delta$ implies $|f(\mathbf{x}) - f(\mathbf{x}')| < \varepsilon$. ¤

The map f is a *homeomorphism*, if it is a bijection such that both f and f^{-1} are continuous.

We say that two topological spaces are *homeomorphic* or *have the same topological type*, if there is a homeomorphism between them.
Homeomorphic topological spaces are considered to be the same from the viewpoint of topology.

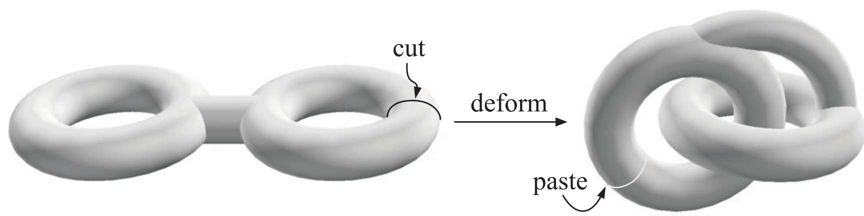

Figure 1.8: Intuitive description of a homeomorphism.

Intuitively, two spaces are homeomorphic if a rubber model of one of them can be deformed into that of the other. We are allowed to stretch and shrink the model but not allowed to cut the model or glue pieces together. More exactly, we may cut the model somewhere only if later on we glue together the parts we get in the same way as they were joined. Of course, this description of homeomorphism is applicable only for "nice spaces" such as surfaces, curves etc. and by no means substitutes the precise definition.
For example, the circle, the perimeter of a square, and the trefoil knot are homeomorphic, so are a solid disc and a solid square, however a circle is not homeomorphic to a solid disc. In most cases it is easy to show that two homeomorphic spaces are indeed homeomorphic: we only have to present a homeomorphism. However, to show that two topological spaces are not homeomorphic, we have to find a topological property, which is possessed only by one of the spaces. For example, the fact that $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}^m$ if $n \neq m$, is a non-trivial theorem of topology (the "dimension invariance theorem"), the proof of which uses techniques of algebraic topology or dimension theory.
Sometimes it is difficult to prove continuity of a map using the definition of continuity directly. However, there are some theorems, that make such proofs easier. For example, a mapping $h\colon X \to \mathbb{R}^n$ from a topological space (X, τ) into $\mathbb{R}^n$ is continuous if and only if all of its coordinate functions $h^i\colon X \to \mathbb{R}$ are continuous. If $f, g\colon X \to \mathbb{R}$ are real valued continuous functions on X then their real valued linear combinations $\alpha f + \beta g$ and their product fg are also continuous. The quotient f/g is also continuous everywhere, where g is not 0. Thus, for example, if $\mathbf{p} \in X \subset \mathbb{R}^m$ and $f\colon X \to \mathbb{R}^n$, then if $\mathbf{p}$ has a neighborhood U such that the coordinates of $f(\mathbf{q})$ can be expressed

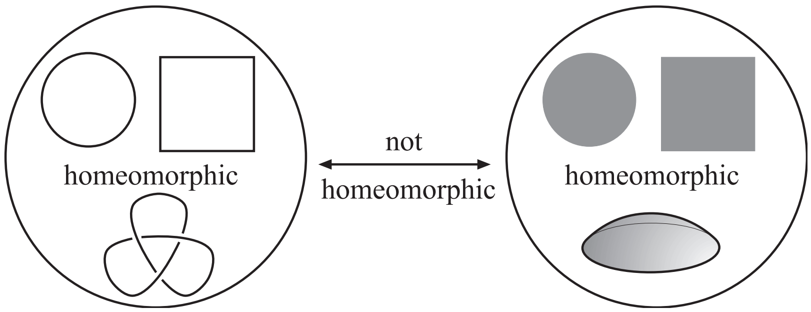

Figure 1.9: Examples of homeomorphic and non-homeomorphic topological spaces.

as rational functions (quotients of polynomials) of the coordinates of $\mathbf{q} \in U$, then f is surely continuous at $\mathbf{p}$.

Exercise 1.4.19. Show that $\mathbb{R}^n$ and the open balls in $\mathbb{R}^n$ (with the subspace topology) are homeomorphic. ¤

Exercise 1.4.20. Show that the "punctured sphere" $\mathbb{S}^n \setminus \{\mathbf{p}\}$, where

$$\mathbb{S}^n = \{\mathbf{x} \in \mathbb{R}^{n+1} \mid |\mathbf{x}| = 1\} \text{ and } \mathbf{p} \in \mathbb{S}^n,$$

is homeomorphic to the plane $\mathbb{R}^n$. ¤

Exercise 1.4.21. Show that the inverting map $\mathrm{GL}(V) \to \mathrm{GL}(V)$, $M \mapsto M^{-1}$ is a homeomorphisms. ¤

Exercise 1.4.22. Show that $\mathrm{Gr}_k(\mathbb{R}^n)$ is homeomorphic to $\mathrm{Gr}_{n-k}(\mathbb{R}^n)$. ¤

Exercise 1.4.23. A k-sphere in a Euclidean linear space V is uniquely determined by the affine subspace spanned by it, its center, and its radius. Thus, there is a bijection between $\mathrm{SPH}_k(V)$ and the set

$$\widetilde{\mathrm{SPH}}_k(V) = \{(\Sigma, O, r) \in \mathrm{AGr}_{k+1}(V) \times V \times \mathbb{R} \mid O \in \Sigma,\ r > 0\}.$$

Show that $\mathrm{SPH}_k(V)$ is homeomorphic to $\widetilde{\mathrm{SPH}}_k(V)$ endowed with the subspace topology inherited from the product topology of $\mathrm{AGr}_{k+1}(V) \times V \times \mathbb{R}$. ¤

1.4.1 Separation and Countability Axioms

Sometimes to prove substantial theorems on a topological space, we have to assume that it satisfies some simple properties that do not follow from the axioms of a topological space. Some of the frequently used properties can be formulated as axioms that can be added to the system of axioms of a topological space or as definitions of certain types of topological spaces that satisfy the additional axioms. We list here some of these axioms that will be used later. The axioms can be grouped according to their nature. First we start with the group of separation axioms.

Definition 1.4.24. A topological space (X, τ) is said to be T_0, if for any two distinct points $\mathbf{p} \neq \mathbf{q} \in X$, there is an open subset $U \in \tau$, which contains exactly one of the points. ※

Discrete topology on a set of at least two points is not T_0.

Definition 1.4.25. A topological space (X, τ) is said to be T_1, if for any two distinct points $\mathbf{p} \neq \mathbf{q} \in X$, there is an open subset $U \in \tau$, which contains $\mathbf{p}$ but does not contain $\mathbf{q}$. ※

If $X = \{\mathbf{p}, \mathbf{q}\}$ is a two point set, and the topology τ on X consists of the sets X, $\{\mathbf{q}\}$, and $\emptyset$, then (X, τ) is T_0 but not T_1. T_1-spaces are T_0.

Definition 1.4.26. A topological space (X, τ) is said to be T_2, or *Hausdorff*, if for any two distinct points $\mathbf{p} \neq \mathbf{q} \in X$, there exist open subsets $U, V \in \tau$ such that $\mathbf{p} \in U$, $\mathbf{q} \in V$ and $U \cap V = \emptyset$. ※

Every T_2-space is T_1. An example of a topological space which is T_1 but not T_2 is an infinite set X equipped with the *cofinite topology* τ. By definition, *tau* contains exactly the emptyset and those subsets of X, the complements of which are finite.

Definition 1.4.27. A topological space (X, τ) is *regular* if for any point $\mathbf{p} \in X$ and any closed set $F \subseteq X \setminus \mathbf{p}$, there are open subsets $U, V \in \tau$ such that $\mathbf{p} \in U$, $F \subseteq V$ and $U \cap V = \emptyset$. A regular T_0-space is called a T_3-space. ※

A set with at least two points is regular but not T_0. On the other hand, every T_3-space is T_2. A simple example of a Hausdorff space (X, τ) which is not regular was constructed by J.R. Munkres. In the example of Munkres, the set X is the set of real numbers $\mathbb{R}$. Elements of τ are sets of the form $U \setminus L$, where U is an open subset of $\mathbb{R}$ with respect to the standard (metric) topology, L is a subset of the set $K = \{1/n \mid 0 < n \in \mathbb{N}\}$. (X, τ) is not regular because the point 0 cannot be separated from the closed subset K with disjoint open subsets.

Definition 1.4.28. A topological space (X, τ) is called *normal* if for any two disjoint closed set F_1 and F_2, one can find disjoint open subsets U and V such that $U \supseteq F_1$ and $V \supseteq F_2$. A topological space is a T_4-space if it is normal and T_1. ※

The two point set $X = \{\mathbf{p}, \mathbf{q}\}$ with topology $\tau = \{X, \{\mathbf{q}\}, \emptyset\}$ is normal and T_0, but not T_1, hence not T_4. Every T_4 space is T_3. The *Sorgenfrey plane* is an example of a T_3 space which is not normal. The Sorgenfrey plane is obtained as the product space $(\mathbb{R}, \tau) \times (\mathbb{R}, \tau)$, where the (nonstandard) topology τ on $\mathbb{R}$ consists of all unions $\bigcup_{i \in I}[a_i, b_i)$ of left closed right open intervals. (See [9] Example 3. in §31, Ch. 4. for details.)
Countability axioms are of different character. They typically require the existence of a countable family of subsets with some properties. Recall that two sets have the same *cardinality* if and only if there is a bijection between them. There is an ordering on the family of cardinalities. The cardinality of a set A is *less then or equal to* the cardinality of the set B if there is an injective map from A into B. A set is called *countable* if it is either finite or has the same cardinality as the set of natural numbers. Equivalently, a set is countable if there is a sequence $\mathbf{p}_1, \mathbf{p}_2, \dots$ listing all elements of the set. Every infinite set contains a countable infinite set, therefore a countable infinite sets have the smallest cardinality among all infinite sets.

Definition 1.4.29. A topological space (X, τ) is *first-countable* if for any point $\mathbf{p} \in X$, there is a countable family of open sets $U_1, U_2, \dots$ such that for any $U \in \tau$ containing $\mathbf{p}$, there is an element U_i of the family for which $\mathbf{p} \in U_i \subseteq U$. ※

Metric topologies are always first-countable. Choosing for U_i the open ball of radius $1/i$ centered at $\mathbf{p}$, we obtain a countable family of neighborhoods of $\mathbf{p}$ satisfying the conditions.

Definition 1.4.30. A topological space (X, τ) is *second-countable* if there is a countable family of open sets $U_1, U_2, \dots$ such that every open set $U \in \tau$ is the union of those elements U_i that are covered by U, i.e., $U = \bigcup_{U_i \subseteq U} U_i$. ※

Every second-countable space is first-countable. The discrete topology on $\mathbb{R}$ is metrizable but not second-countable.

Definition 1.4.31. A topological space (X, τ) is said to be *separable* if contains a countable dense subset. ※

Proposition 1.4.32. *Every first-countable separable space is second-countable.*

1.4.2 Compactness

Definition 1.4.33. A topological space (X, τ) is said to be *compact* if from any open covering $X = \bigcup_{i \in I} U_i$, $U_i \in \tau$ of X, we can choose a finite subcovering $X = U_{i_1} \cup \dots \cup U_{i_k}$, where $i_1, \dots, i_k \in I$. ※

A related notion is sequential compactness.

Definition 1.4.34. A topological space (X, τ) is said to be *sequentially compact* if every sequence $x_1, x_2, \ldots$ in X has a convergent subsequence $x_{i_1}, x_{i_2}, \ldots$, $(1 \le i_1 < i_2 < \cdots)$. ※

We emphasize that the limit of the subsequence must be in X even if X is contained in a larger space as a subspace. Compactness and sequential compactness are independent properties, none of them implies the other in general. However, for metric spaces they are equivalent.

Proposition 1.4.35.

- *A first-countable compact space is sequentially compact.*
- *A second-countable sequentially compact space is compact.*
- *A sequentially compact metric space is separable, hence second-countable and compact.*

Compact metric spaces can also be characterized as follows.

Definition 1.4.36. A sequence of points $\mathbf{x}_1, \mathbf{x}_2, \ldots$ in a metric space (X, d) is a *Cauchy sequence* if for any $\varepsilon > 0$, there exists a number N such that we have $d(\mathbf{x}_m, \mathbf{x}_n) < \varepsilon$ whenever $i > N$ and $j > N$. A metric space (X, d) is said to be *complete* if all of its Cauchy sequences are convergent to a point of X. ※

Definition 1.4.37. A metric space (X, d) is *totally bounded* if and only if for any $\varepsilon > 0$, X can be covered by a finite collection of open balls of radius ε. ※

Proposition 1.4.38. *A metric space is compact if and only if it is complete and totally bounded.*

Corollary 1.4.39 (Heine–Borel theorem)**.** *A subset K of $\mathbb{R}^n$ with the subspace topology is compact if and only K is bounded and closed.*

Proposition 1.4.40. *The image $f(X)$ of a compact space X under a continuous map $f\colon X \to Y$ is a compact subspace of Y.*

In particular, if $f\colon X \to \mathbb{R}$ is a real function on a compact space, then $f(X)$ is bounded and closed in $\mathbb{R}$, therefore, it has a maximal and minimal element. This gives us the *extreme value theorem* due to Weierstrass.

Proposition 1.4.41. *A continuous function $f\colon X \to \mathbb{R}$ on a compact space (X, τ) is bounded and attains its maximum and minimum values.*

Definition 1.4.42. A function $f\colon X \to \mathbb{R}$ on a metric space (X, d) is *uniformly continuous*, if for any $\varepsilon > 0$, one can find a positive δ such that $x, y \in X$, $d(x, y) < \delta$ implies $|f(x) - f(y)| < \varepsilon$. ※

It is clear that every uniformly continuous function is continuous, but the converse is not true. For example, the function $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ is continuous, but not uniformly continuous, since no matter how small $|x-y|$ is, $|x^2-y^2| = |x-y||x+y|$ can be arbitrarily large. However, if the domain of the function is compact, then continuity implies uniform continuity.

Proposition 1.4.43. *A continuous function $f\colon X \to \mathbb{R}$ defined on a compact metric space (X,d) is uniformly continuous.*

Proof. Suppose to the contrary, that there is an $\varepsilon > 0$ for which we can not find a suitable δ. Then there exists a sequence of pairs of real numbers x_n, y_n such that $x_n, y_n \in X$, $d(x_n,y_n) < 1/n$, but $|f(x_n) - f(y_n)| > \varepsilon$. By compactness of X, we can select a convergent subsequence $x_{i_n} \to x$ of the sequence (x_n). Condition $d(x_n,y_n) < 1/n$ ensures that $y_{i_n} \to x$ as well, and so, by the continuity of f at x we have $|f(x_{i_n}) - f(y_{i_n})| \to 0$. But this contradicts the condition $|f(x_n) - f(y_n)| > \varepsilon$ for all n. □

1.4.3 Fundamental Group and Covering Spaces

Definition 1.4.44. A *pointed topological space* is a pair $(X, \mathbf{x}_0)$ consisting of a topological space X and a distinguished *base point* $\mathbf{x}_0 \in X$. ※

Pointed topological spaces form a category. A *morphism* $f\colon (X,\mathbf{x}_0) \to (Y,\mathbf{y}_0)$ of the category is a continuous map $f\colon X \to Y$ such that $f(\mathbf{x}_0) = \mathbf{y}_0$.

Definition 1.4.45. A *loop in the topological space X with base point* $\mathbf{x}_0 \in X$ is a continuous map $\gamma\colon [0,1] \to X$ such that $\gamma(0) = \gamma(1) = \mathbf{x}_0$. ※

Definition 1.4.46. The *concatenation of the loops γ and η* in X with a common base point $\mathbf{x}_0$ is the loop $\gamma * \eta$ defined by equipped with the group multiplication defined by

$$(\gamma * \eta)(t) = \begin{cases} \gamma(2t), & \text{if } 0 \le t \le \frac{1}{2}, \\ \eta(2t-1), & \text{if } \frac{1}{2} \le t \le 1. \end{cases}$$ ※

Definition 1.4.47. The loops γ and η in X with a common base point $\mathbf{x}_0 \in X$ are said to be *homotopic* if there is a continuous map $H\colon [0,1]\times[0,1] \to X$ such that $H(0,t) = \gamma(t)$, $H(1,t) = \eta(t)$, and $H(t,0) = H(t,1) = \mathbf{x}_0$ for all $t \in [0,1]$. ※

Being homotopic is an equivalence relation on the set of loops with a given base point. The equivalence class of a loop γ is also called its *homotopy class* and is denoted by $[\gamma]$.

Definition 1.4.48. The *fundamental group* $\pi_1(X,\mathbf{x}_0)$ *of a pointed topological space* $(X,\mathbf{x}_0)$ is the set of all homotopy classes of loops in X with base point $\mathbf{x}_0$ together with the multiplication $\cdot$ induced by concatenation as follows $[\gamma]\cdot[\eta] = [\gamma * \eta]$. ※

Proposition 1.4.49. *The product of the elements of the fundamental group is properly defined, and multiplication satisfies the group axioms.*

The fundamental group construction gives a functor from the category of pointed topological spaces to the category of groups. Indeed, every continuous map $f\colon (X, \mathbf{x}_0) \to (Y, \mathbf{y}_0)$ induces a group homomorphism $f_*\colon \pi_1(X, \mathbf{x}_0) \to \pi_1(Y, \mathbf{y}_0)$ by the formula $f_*([\gamma]) = [f \circ \gamma]$.

Definition 1.4.50. A topological space X is *connected* if X cannot be decomposed into the disjoint union of two nonempty open subsets. X is *path-connected* if for any two points $\mathbf{x}, \mathbf{y} \in X$, there is a continuous map $\gamma\colon [0, 1] \to X$ such that $\gamma(0) = \mathbf{x}$ and $\gamma(1) = \mathbf{y}$. A topological space is *locally path-connected* if for any point $\mathbf{x} \in X$ and any neighborhood U of $\mathbf{x}$, there is a path-connected neighborhood V of $\mathbf{x}$ which is contained in U. ※

Every path-connected space is connected. The union of the y-axis and the graph of the function $f\colon (0, 1) \to \mathbb{R}$, $f(x) = \sin(1/x)$ is connected but not path-connected with its subspace topology inherited from $\mathbb{R}^2$. However, a connected and locally path-connected topological space is path-connected. The set $\{(x, y) \in \mathbb{R} \mid y = 0 \text{ or } x \in \mathbb{Q}\}$ is path-connected but not locally path-connected.
The isomorphism class of the fundamental group of a path-connected space does not depend on the choice of the base point.

Definition 1.4.51. A topological space is said to be *simply connected* if it is path-connected and its fundamental group is trivial with respect to any base point. A topological space X is *semilocally simply-connected* if each point $\mathbf{x} \in X$ has a neighborhood for which the the embedding $\iota\colon U \to X$ induces a trivial homomorphism $\iota_*\colon \pi_1(U, \mathbf{x}) \to \pi_1(X, \mathbf{x})$. ※

The latter condition means geometrically that any loop in U based at $\mathbf{x}$ can be contracted to a point in X.
The fundamental group has many applications in topology. The main reason, why we introduced it here is its role played in the classification of covering spaces of a given space.

Definition 1.4.52. A continuous map $p\colon \tilde{X} \to X$ is called a *covering map* if each point $\mathbf{x} \in X$ has an open neighborhood $U \subseteq X$ such that $p^{-1}(U)$ can be decomposed as a disjoint union $p^{-1}(U) = \bigcup_{i \in I}^{*} U_i$ of open subsets U_i of X, such that the restriction $p|_{U_i}\colon U_i \to U$ is a homeomorphism for all $i \in I$. Neighborhoods U with this property are called *evenly-covered*, the sets U_i are called the *sheets/slices/layers over U*. ※

The cardinality of $p^{-1}(\mathbf{x})$ is locally constant on X. Thus, if X is connected, then it is constant. Whenever the cardinality of $p^{-1}(\mathbf{x})$ is constant on X, it

is called the *degree of the covering*. A *k-fold covering* is a covering of degree k.
Covering maps have the following lifting property for cubes.

Proposition 1.4.53. *Let $p\colon (\tilde{X}, \tilde{\mathbf{x}}_0) \to (X, \mathbf{x}_0)$ be a covering map between two pointed spaces, $\Phi\colon ([0,1]^k, \{\mathbf{0}\}) \to (X, \mathbf{x}_0)$ be a continuous map from a pointed k-dimensional cube to $(X, \mathbf{x}_0)$. Then there is a unique continuous map $\tilde{\Phi}\colon ([0,1]^k, \{\mathbf{0}\}) \to (X, \mathbf{x}_0)$ such that $p \circ \tilde{\Phi} = \Phi$.*

The map $\tilde{\Phi}$ is called the *lift of* Φ.

Corollary 1.4.54. *If $p\colon (\tilde{X}, \tilde{\mathbf{x}}_0) \to (X, \mathbf{x}_0)$ be a covering map between two pointed spaces,then the induced map $p_*\colon \pi_1(\tilde{X}, \tilde{\mathbf{x}}_0) \to \pi_1(X, \mathbf{x}_0)$ is injective.*

Thus, we can assign to each pointed covering map the subgroup $\operatorname{im} p_*$ of the fundamental group of $\pi_1(X, \mathbf{x}_0)$. It turns out that for "nice" spaces this subgroup determines the covering uniquely up to a natural notion of isomorphism.

Definition 1.4.55. Two covering maps $p_1\colon (\tilde{X}_1, \tilde{\mathbf{x}}_1) \to (X, \mathbf{x}_0)$ and $p_2\colon (\tilde{X}_2, \tilde{\mathbf{x}}_2) \to (X, \mathbf{x}_0)$ are *isomorphic* if there is a base point preserving homeomorphism $h\colon (\tilde{X}_1, \tilde{\mathbf{x}}_1) \to (\tilde{X}_2, \tilde{\mathbf{x}}_2)$ such that $p_2 \circ h = p_1$. We define *isomorphism of (non-pointed) coverings* in a similar way, ignoring the base points. ※

Theorem 1.4.56. *Let $(X, \mathbf{x}_0)$ be a path-connected, locally path-connected, and semilocally simply-connected topological space. Then there is a bijection between the set of base point preserving isomorphism classes of path-connected covering spaces $p\colon (\tilde{X}, \tilde{\mathbf{x}}_0) \to (X, \mathbf{x}_0)$ and the set of subgroups of $\pi_1(X, \mathbf{x}_0)$, obtained by assigning the subgroup $p_*(\pi_1(\tilde{X}, \tilde{\mathbf{x}}_0)) \leq \pi_1(X, \mathbf{x}_0)$ to the covering space $(\tilde{X}, \tilde{\mathbf{x}}_0)$.*
If base points are ignored, this correspondence gives a bijection between isomorphism classes of path-connected covering spaces $p\colon \tilde{X} \to X$ and conjugacy classes of subgroups of $\pi_1(X, \mathbf{x}_0)$.

In terms of this correspondence, the degree of the covering is the index $|\pi_1(X, \mathbf{x}_0) : p_*(\pi_1(\tilde{X}, \tilde{\mathbf{x}}_0))|$ of the subgroup $p_*(\pi_1(\tilde{X}, \tilde{\mathbf{x}}_0))$ in $\pi_1(X, \mathbf{x}_0)$. Spaces satisfying the conditions of the theorem have a unique simply connected covering space up to isomorphism. It corresponds to the trivial subgroup of the fundamental group of X. This covering space is the *universal covering space of X*.

1.5 Multivariable Calculus

Definition 1.5.1. A map $F\colon U \to \mathbb{R}^n$ defined on a neighborhood of $U \subset \mathbb{R}^m$ $\mathbf{x}_0 \in \mathbb{R}^m$ is *differentiable at* $\mathbf{x}_0$ and its *derivative at* $\mathbf{x}_0$ is the linear map

$A\colon \mathbb{R}^m \to \mathbb{R}^n$ if

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|F(\mathbf{x}) - F(\mathbf{x}_0) - A(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = 0.$$

The derivative of F at $\mathbf{x}_0$ and also its $n \times m$ matrix is denoted by $F'(\mathbf{x}_0)$. The derivative (matrix) $F'(\mathbf{x}_0)$ is also called the *Jacobian (matrix) of F at $\mathbf{x}_0$*. ※

Exercise 1.5.2. Prove that the derivative of a map at a point is uniquely determined by the condition defining it. ¤

Exercise 1.5.3. Show that if F is differentiable at a point $\mathbf{x}_0$, then it is also continuous at $\mathbf{x}_0$. ¤

The derivative map is the linear part of the affine transformation $\mathbf{x} \mapsto A(\mathbf{x}) + (F(\mathbf{x}_0) - A(\mathbf{x}_0))$ giving the best approximation of F around $\mathbf{x}_0$.

Proposition 1.5.4 (Chain Rule)**.** *If $U \subseteq \mathbb{R}^m$ is a neighborhood of $\mathbf{x}_0 \in U$, and the function $F\colon U \to \mathbb{R}^n$ is differentiable at $\mathbf{x}_0$, furthermore, the function $G\colon V \to \mathbb{R}^p$ is defined on a neighborhood of $F(\mathbf{x}_0)$ and it is also differentiable at $F(\mathbf{x}_0)$, then the composition $G \circ F\colon U \to \mathbb{R}^p$ is differentiable at $\mathbf{x}_0$ and*

$$(G \circ F)'(\mathbf{x}_0) = G'(F(\mathbf{x}_0)) \odot F'(\mathbf{x}_0).$$

The sign $\odot$ on the right-hand side denotes composition of linear maps or matrix multiplication depending on whether the derivative maps are thought of as linear maps or the matrices of these linear maps.

Definition 1.5.5. Let $F\colon U \to \mathbb{R}^n$ be differentiable at $\mathbf{x}_0 \in U$, where $U \subseteq \mathbb{R}^m$ is a neighborhood of $\mathbf{x}_0$. The *directional derivative of F along a tangent vector* $(\mathbf{x}_0, \mathbf{v}) \in T_{\mathbf{x}_0}\mathbb{R}^m$ is the vector

$$\partial_{\mathbf{v}} F(\mathbf{x}_0) = \frac{d}{dt} F(\mathbf{x}_0 + t\mathbf{v})|_{t=0} \in \mathbb{R}^n. \qquad ※$$

Recall that the tangent vector $(\mathbf{x}_0, \mathbf{v})$ consists of the vector $\mathbf{v}$, and a base point $\mathbf{x}_0$. There are several other notations used to denote directional derivatives, e.g., $\partial_{\mathbf{v}} F|_{\mathbf{x}_0}$, $\partial_{\mathbf{v}} F(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0}$, $(\partial_{\mathbf{v}}|_{\mathbf{x}_0})(F)$, sometimes ∂ is replaced by the letter D. We shall use usually the sign ∂, but otherwise any of these notations, always the one which seems to be most convenient typographically.
Remark also that though we defined $\partial_{\mathbf{v}} F(\mathbf{x}_0)$ as a vector, there is a natural choice of a base point for this vector, namely $F(\mathbf{x}_0)$, so it would have been also a possibility to define $\partial_{\mathbf{v}} F(\mathbf{x}_0)$ as a tangent vector of $\mathbb{R}^n$ at $F(\mathbf{x}_0)$.
Directional derivatives can be computed easily from the chain rule

$$\partial_{\mathbf{v}} F(\mathbf{x}_0) = F'(\mathbf{x}_0)(\mathbf{v}).$$

Definition 1.5.6. Let $F\colon U \to \mathbb{R}^n$ be as above. The *partial derivative* $\partial_i F(\mathbf{x}_0)$ *of* F *with respect to its* i*th variable at* $\mathbf{x}_0$ is its directional derivative

$$\partial_i F(\mathbf{x}_0) = \partial_{\mathbf{e}_i} F(\mathbf{x}_0)$$

along $(\mathbf{x}_0, \mathbf{e}_i)$, where $\mathbf{e}_i$ is the ith vector of the standard basis of $\mathbb{R}^m$. ※

Partial derivatives are also denoted in many different ways. In some cases, the variables of F are denoted by some fixed symbols like $(x_1, \dots, x_m)$. Then we may also use the notations $\partial_i F(\mathbf{x}_0) = \frac{\partial}{\partial x_i} F(\mathbf{x}_0) = \frac{\partial F}{\partial x_i}(\mathbf{x}_0) = \partial_{x_i} F(\mathbf{x}_0)$. Some of these notations can also be used when the variables do not carry numerical indices just denoted by different letters, like $(u, v, \dots)$. For example, in the latter case, $\partial_1 F(u_0, v_0, \dots)$ can also be denoted by $\partial_u F(u_0, v_0, \dots)$, $\frac{\partial}{\partial u} F(u_0, v_0, \dots)$, $\frac{\partial F}{\partial u}(u_0, v_0, \dots)$.

Warning. When the variables are denoted by fixed symbols, the variable symbols should not denote anything else, otherwise one can run into ambiguous expressions leading to confusion. For example, if $F(u, v) = uv$, then $\partial_u F(u, v) = v$. However, the expression $\partial_u F(u^2, v^2) = \partial_u u^2 v^2$ can mean both v^2 or $2uv^2$. The confusion is caused by the circumstance that instead of introducing two new variables, say (x, y) and making the substitution $(u, v) = (x^2, y^2)$, we denoted the new variables also by u and v, and made the strange substitution $(u, v) = (u^2, v^2)$. Thus, the different expressions $\partial_u F(x^2, y^2) = y^2$ and $\partial_x F(x^2, y^2) = 2xy^2$ collapsed.

We can express the matrix of the derivative of F with the help of its partial derivatives. Indeed, the columns of the matrix of $F'(\mathbf{x}_0)$ are the images of the standard basis vectors that is the partial derivatives $\partial_1 F(\mathbf{x}_0), \dots, \partial_m F(\mathbf{x}_0)$. To write the matrix more explicitly, suppose the the coordinate functions of F are $(F^1, \dots, F^n)$. The F^i's are real valued functions on U related to F by $F(\mathbf{x}) = (F^1(\mathbf{x}), \dots, F^n(\mathbf{x}))$. Then

$$F'(\mathbf{x}_0) = \begin{pmatrix} \partial_1 F^1(\mathbf{x}_0) & \dots & \partial_m F^1(\mathbf{x}_0) \\ \vdots & & \vdots \\ \partial_1 F^n(\mathbf{x}_0) & \dots & \partial_m F^n(\mathbf{x}_0) \end{pmatrix}$$

Definition 1.5.7. If the real valued function $F\colon U \to \mathbb{R}$ defined in a neighborhood U of $\mathbf{x}_0 \in \mathbb{R}^m$ is differentiable at $\mathbf{x}_0$, then its derivative matrix is a row vector

$$\operatorname{grad} F(\mathbf{x}_0) = (\partial_1 F(\mathbf{x}_0), \dots, \partial_m F(\mathbf{x}_0)),$$

which we call the *gradient vector of* F *at* $\mathbf{x}_0$. ※

According to this definition, for a vector valued function F, the rows of the derivative matrix of F are the gradient vectors of the coordinates functions of F.

Using the explicit form of the derivatives matrices and the multiplication rule for matrices, we can reformulate the chain rule equating the matrix coefficients.

$$\partial(F^i \circ G)$$

Proposition 1.5.8 (Chain Rule | a reformulation). *If $U \subseteq \mathbb{R}^m$ is a neighborhood of $\mathbf{x}_0 \in U$, and the function $F = (F^1, \dots, F^n)\colon U \to \mathbb{R}^n$ is differentiable at $\mathbf{x}_0$, furthermore, the function $G = (G^1, \dots, G^p)\colon V \to \mathbb{R}^p$ is defined on a neighborhood of $F(\mathbf{x}_0)$ and it is also differentiable at $F(\mathbf{x}_0)$, then*

$$\partial_i (G \circ F)^j(\mathbf{x}_0) = \partial_i (G^j \circ F)(\mathbf{x}_0) = \sum_{k=1}^{n} \partial_k G^j(F(\mathbf{x}_0)) \cdot \partial_i F^k(\mathbf{x}_0).$$

The existence of all directional derivatives of F at $\mathbf{x}_0$ is necessary for the differentiability of F at $\mathbf{x}_0$ but not sufficient. However we have the following theorem.

Theorem 1.5.9. *If $F\colon U \to \mathbb{R}^n$ is defined on an open subset U of $\mathbb{R}^m$ and all the partial derivatives $\partial_1 F(\mathbf{x}), \dots, \partial_m F(\mathbf{x})$ exist at each point $\mathbf{x} \in U$, and all depend continuously on $\mathbf{x}$, then F is differentiable at each point of U and the derivative of F is a continuous function of $\mathbf{x} \in U$.*

Definition 1.5.10. A function $F\colon U \to \mathbb{R}^n$ defined on an open subset U of $\mathbb{R}^m$ is said to be of *k times continuously differentiable* of a *function of class $\mathcal{C}^k$*, if all the k-th order partial derivatives $\partial_{i_1} \dots \partial_{i_k} F(\mathbf{x})$ exist and depend continuously on $\mathbf{x} \in U$.
Functions which are of class $\mathcal{C}^k$ for all k are said to be function of class $\mathcal{C}^\infty$ and called *infinitely many times differentiable* or *smooth functions.* ※

There is also a recursive definition of this notion. A function F is of class $\mathcal{C}^0$ if and only if F is continuous. For $k \geq 1$, F is of class $\mathcal{C}^k$ if and only if it is differentiable at each point of U and the map $F'\colon U \to \mathbb{R}^{n \times m}$ is of class $\mathcal{C}^{k-1}$.
There is important theorem which is referred to by many different names in the literature: “Young’s Theorem”, “Clairaut’s Theorem”, “Schwarz’s Theorem”, “Symmetry of Second Derivatives”, “Equality of Mixed Partials”.

Theorem 1.5.11 (Young’s Theorem). *If $F\colon U \to \mathbb{R}^n$ is defined on the open set $U \subseteq \mathbb{R}^m$ and is of class $\mathcal{C}^2$, then $\partial_i \partial_j F = \partial_j \partial_i F$ for any $1 \leq i, j \leq m$.*

Another fundamental theorem of multivariable calculus is the inverse function theorem

Theorem 1.5.12 (Inverse Function Theorem). *Suppose that U is an open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^n$ is a function of class $\mathcal{C}^1$ with invertible derivative $F'(\mathbf{x}_0)$ at $\mathbf{x}_0$. Then $\mathbf{x}_0$ has an open neighborhood $V \subseteq U$ such that*

$W = F(V)$ is an open neighborhood of $F(\mathbf{x}_0)$, $F|_V : V \to W$ is a bijection, and its inverse $(F|_V)^{-1} : W \to V$ is also of class $\mathcal{C}^1$.
If, in addition, F is of class $\mathcal{C}^k$, then $(F|_V)^{-1} : W \to V$ is also of class $\mathcal{C}^k$.

The derivative of the inverse map $G = (F|_V)^{-1}$ can be computed by differentiating the identity

$$(F \circ G)(\mathbf{x}) \equiv \mathbf{x}$$

by the chain rule. Rearranging we obtain

$$G'(\mathbf{x}) = (F'(G(\mathbf{x})))^{-1}.$$

Sometimes we want to speak about the differentiability of maps the domains of which is not an open subset (e.g. a closed interval). For this purpose, we say that a map $\mathbf{F} : A \to \mathbb{R}^n$ defined on an *arbitrary* set $A \subseteq \mathbb{R}^m$ *is of class* $\mathcal{C}^k$ or *k times differentiable* if there exists an open set $U \subseteq \mathbb{R}^m$ and a mapping $\tilde{F} : U \to \mathbb{R}^n$ of class $\mathcal{C}^k$ such that $A \subseteq U$ and $F = \tilde{F}|_A$.
We can define the partial derivatives $\partial_{i_1} \dots \partial_{i_r} F(\mathbf{x})$ of order $r \leq k$ of F as $\partial_{i_1} \dots \partial_{i_r} \tilde{F}(\mathbf{x})$, but it should be kept in mind that these derivatives are defined properly only on the closure of the interior of A. At other points, the derivatives usually depend on the choice of $\tilde{F}$.

Bump Functions

Proposition 1.5.13. *If $K \subseteq \mathbb{R}^n$ is a compact subset of $\mathbb{R}^n$ and $\varepsilon > 0$, then there exists a smooth function $h : \mathbb{R}^n \to [0, 1]$ such that $h(\mathbf{x})$ is equal to 1 if $\mathbf{x} \in K$ and $h(\mathbf{x}) = 0$ if $d(\mathbf{x}, K) \geq \varepsilon$.*

Proof. Define the function $h_1 : \mathbb{R} \to [0, 1]$ by the formula

$$h_1(t) = \begin{cases} e^{\frac{1}{t^2-1}}, & \text{if } t \in (-1, 1), \\ 0, & \text{if } t \notin (-1, 1). \end{cases}$$

Exercise 1.5.14. Prove that h_1 is a smooth function on $\mathbb{R}$. ¤

We can extend h_1 to $\mathbb{R}^n$ in a rotationally symmetric way by setting $h_n : \mathbb{R}^n \to \mathbb{R}$, $h_n(\mathbf{x}) = h_1(\|\mathbf{x}\|)$.

Exercise 1.5.15. Prove that h_n is a smooth function on $\mathbb{R}^n$. (Caution: The norm function $\mathbf{x} \to \|\mathbf{x}\|$ is not smooth at the origin.) ¤

The graph of h_n is a bell shaped hypersurface in $\mathbb{R}^{n+1}$. Let $c_n = \int_{\mathbb{R}^n} h_n(\mathbf{x}) d\mathbf{x}$ denote the volume of the domain under the bell.
Denote by χ the indicator function of the closed neighborhood of radius $\varepsilon/2$ of K, that is, $\chi : \mathbb{R} \to \mathbb{R}$ is the function defined by

$$\chi(\mathbf{x}) = \begin{cases} 1, & \text{if } d(\mathbf{x}, K) \leq \varepsilon/2, \\ 0, & \text{otherwise.} \end{cases}$$

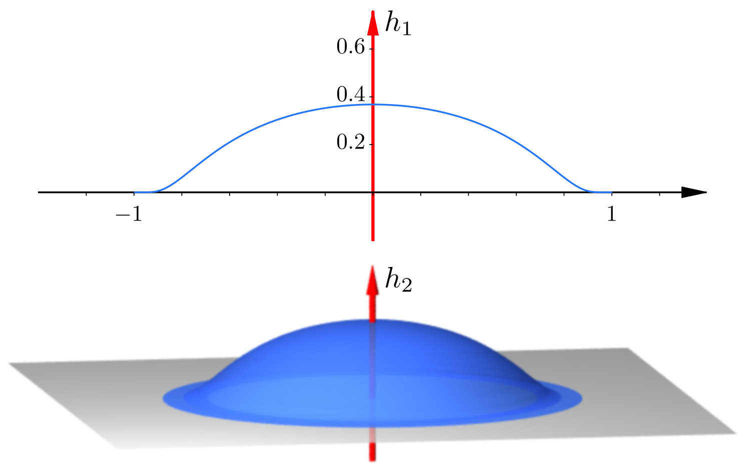

Figure 1.10: The graphs of the functions h_1 and h_2.

Then the function

$$h(\mathbf{x}) = \frac{2^n}{c_n \varepsilon^n} \int_{\mathbb{R}^n} \chi(\mathbf{y}) h_n \left(\frac{\mathbf{x} - \mathbf{y}}{\varepsilon/2} \right) d\mathbf{y}$$

has the desired properties. □

Exercise 1.5.16. Show that h is indeed smooth, $h|_K \equiv 1$, and that h is vanishing outside the ε-neighborhood of K. ¤

1.6 Measure and Integration

Extending the classical geometrical treatment of the notion of area and volume leads to the notion of Jordan content. Roughly speaking, assuming that we already know how to define the volume of a compact polyhedron, that is a finite union of n-dimensional simplices by dissecting it into non-overlapping simplices and adding the volumes of the simplices, a bounded subset A of $\mathbb{R}^n$ *is said to have Jordan content* if the supremum of the volumes of polytopes contained in A is equal to the infimum of the volumes of compact polytopes containing A. In this case, the common value of the supremum and the infimum is the *Jordan content* $J_n(A)$ *of* A.
Instead of Jordan content, we shall rather use Lebesgue measure as the "volume of sets". The reason for this is that whenever a set has Jordan content, it is also Lebesgue measurable and its Lebesgue measure equals its Jordan

content. However, “having Lebesgue measure” is a much less restrictive property then the property of having Jordan content. Lebesgue measure is also more convenient to work with due to its σ-additivity.

Measure Spaces

Definition 1.6.1. A *σ-algebra of subsets of a set X* is a family Σ of subsets of X satisfying the following axioms.

- $X \in \Sigma$.
- If $A \in \Sigma$, then $X \setminus A \in \Sigma$.
- If $A_1, A_2, \ldots$ is a sequence of elements of Σ, then $\bigcup_{i=1}^{\infty} A_i \in \Sigma$, i.e., Σ is closed under countable unions. ※

Definition 1.6.2. The *Borel algebra of a topological space* is the smallest σ-algebra containing its open subsets. *Borel sets are elements* of the Borel algebra. ※

Exercise 1.6.3. Show that closed subsets are Borel sets. ¤

Exercise 1.6.4. Show that if a topological space is Hausdorff, which means that for any two points $\mathbf{p} \neq \mathbf{q}$ there exist disjoint open set U and V such that $\mathbf{p} \in U$ and $\mathbf{q} \in V$, then compact subsets of the space are closed, hence Borel sets. ¤

Definition 1.6.5. If Σ is a σ-algebra of subsets of X, then a function $\mu \colon \Sigma \to [0, +\infty]$ is a *measure* if it satisfies the following axioms.

- $\mu(\emptyset) = 0$.
- If $A_1, A_2, \ldots$ is a sequence of pairwise disjoint elements of Σ, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$. ($\sigma$-additivity.)

The triple (X, Σ, μ) is called a *measure space*. Elements of Σ are the *μ-measurable sets* or shortly *measurable sets*. ※

Definition 1.6.6. A measure space (X, Σ, μ) is *complete* if any subset B which is contained in a measurable set A of measure $\mu(A) = 0$ is also measurable. ※

Exercise 1.6.7. Show that given a measure space (X, Σ, μ), there is a unique complete measure space $(X, \tilde{\Sigma}, \tilde{\mu})$ such that $\tilde{A} \in \tilde{\Sigma}$ if and only if there exist μ-measurable sets $A, B \in \Sigma$ such that $(A \setminus \tilde{A}) \cup (\tilde{A} \setminus A) \subseteq B$ and $\mu(B) = 0$, furthermore, if this is the case then $\tilde{\mu}(\tilde{A}) = \mu(A)$. The measure space $(X, \tilde{\Sigma}, \tilde{\mu})$ is called the *completion of the measure space* $(X, \tilde{\Sigma}, \tilde{\mu})$. ¤

Definition 1.6.8. A *Borel measure* μ on a topological space is a measure on its Borel sets. A Borel measure is called *outer regular* if for any Borel set B,

$$\mu(B) = \inf\{\mu(U) \mid B \subset U \text{ and } U \text{ is open}\}.$$

If compact sets of the space are Borel sets, for example, if the space is Hausdorff, then μ is said to be *inner regular* if for any Borel set B, we have

$$\mu(B) = \inf\{\mu(K) \mid K \supset B \text{ and } K \text{ is compact}\}.$$

The Borel measure μ is *regular* if it is both inner and outer regular. ※

Lebesgue Measure

There are several methods to construct the Lebesgue measure λ_n on $\mathbb{R}^n$. We can start with the construction of λ_1 as the completion of the unique regular Borel measure μ_1 the value of which on open intervals is given by $\mu_1((a,b)) = b - a$. Uniqueness of μ_1 is clear since any open subset of $\mathbb{R}$ is a countable union of disjoint open intervals, so the μ_1 measure of open subsets is uniquely prescribed by σ-additivity. Then the measure of other Borel sets is uniquely determined by outer regularity of μ_1. The proof of existence requires more work.
To proceed from λ_1 to λ_n, we can use the product measure construction.

Theorem 1.6.9. *Let (X_1, Σ_1, μ_1) and (X_2, Σ_2, μ_2) be two measure spaces. Denote by $\Sigma_1 \times \Sigma_2$ the smallest σ-algebras of subsets of $X_1 \times X_2$ containing the family of subsets $\{A_1 \times A_2 \mid A_1 \in \Sigma_1,\ A_2 \in \Sigma_2\}$. Then there exists a measure $\mu_1 \times \mu_2$ on $\Sigma \times \Sigma_2$ such that*

$$\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)$$

for all $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$. If both measure spaces are σ-finite, that is, if they can be presented as a countable union of sets of finite measure, then the measure $\mu_1 \times \mu_2$ is unique.

Definition 1.6.10. If μ_1 and μ_2 are σ-finite, then the unique measure $\mu_1 \times \mu_2$ is called the *product measure of μ_1 and μ_2*. ※

Exercise 1.6.11. Prove that if $X_1 = \mathbb{R}^k$, $X_2 = \mathbb{R}^l$ and Σ_i is the Borel algebra of X_i, then $\Sigma_1 \times \Sigma_2$ is the Borel algebra of $\mathbb{R}^{k+l}$. ¤

Starting from the Borel measure μ_1 on $\mathbb{R}$, the product measure construction gives a Borel measure $\mu_n = \mu_1 \times \mu_{n-1}$ on $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$ recursively for all n. Completion of the Borel measure μ_n is the *Lebesgue measure* λ_n.

Exercise 1.6.12. Show that $\mu_k \times \mu_l = \mu_{k+l}$. ¤

Integral of Measurable Functions

Definition 1.6.13. A function $f\colon X \to [-\infty, \infty]$ on a measure space (X, Σ, μ) is called *measurable* if the level sets $A_t = \{x \mid f(x) < t\}$ are in Σ for all $t \in \mathbb{R}$. ※

Exercise 1.6.14. Prove that for a measurable function f on the measure space (X, Σ, μ), the preimage $f^{-1}(B)$ of any Borel subset of $\mathbb{R}$ is also in Σ. ¤

The integral of measurable functions with respect to the measure μ is introduced in some steps.

Definition 1.6.15. The *indicator function* χ_A *of a subset* A if X is the function $\chi_A\colon X \to \mathbb{R}$ such that $\chi_A(x) = 1$ if and only if $x \in A$, otherwise $\chi_A(x) = 0$. A *step function* is a finite linear combination of indicator functions. ※

Exercise 1.6.16. Prove that a step function is measurable if and only if it can be written as a linear combination of indicator functions of measurable sets. If, in addition, f is non-negative, then it is possible to write it as a linear combination with non-negative coefficients. ¤

Definition 1.6.17. The integral of a non-negative measurable step function $f = \sum_{i=1}^{k} a_i \chi_{A_i}$, where $a_i \geq 0$ and $A_i \in \Sigma$ is

$$\int_X f d\mu = \sum_{i=1}^{k} a_i \mu(A_i).$$

As usual in integration theory, we use the convention $0 \cdot \infty = 0$. ※

Definition 1.6.18. The integral of a non-negative measurable function $f\colon X \to [0, +\infty]$ on the measure space (X, Σ, μ) is

$$\int_X f d\mu = \sup \left\{ \int_X s d\mu \mid 0 \leq s \leq f \text{ and } s \text{ is a measurable step function} \right\}.$$ ※

As for the integral of an arbitrary measurable function f decompose f into a non-negative and a non-positive part by setting

$$f_+(x) = \max\{f(x), 0\} \text{ and } f_-(x) = \max\{-f(x), 0\}.$$

Definition 1.6.19. We say that the *integral of the measurable function f exists* if at least one of the integrals $\int_X f_+ d\mu$ and $\int_X f_- d\mu$ is finite and in that case we define the integral as

$$\int_X f d\mu = \int_X f_+ d\mu - \int_X f_- d\mu.$$ ※

Sometimes we want to restrict the domain of integration to a measurable set D. This can be done easily with the help of indicator functions.

Definition 1.6.20. The integral of a measurable function f on the measurable set D is defined to be the integral

$$\int_D f d\mu = \int_X \chi_D \cdot f d\mu,$$

provided that the second integral exists. ※

The integral of a Lebesgue measurable function f on $\mathbb{R}^n$ will also be denoted by

$$\int_{\mathbb{R}^n} f(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^n} f d\lambda_n.$$

We have several important and useful tools for Lebesgue integrals. An important integration tool is Fubini's theorem.

Theorem 1.6.21 (Fubini's Theorem)**.** *If (X_1, Σ_1, μ_1) and (X_2, Σ_2, μ_2) are σ finite measure spaces and $f\colon X_1 \times X_2 \to [-\infty, \infty]$ is $\mu_1 \times \mu_2$ measurable, then $x_1 \mapsto f(x_1, x_2)$ is μ_1 measurable for any fixed $x_2 \in X_2$, the function $x_2 \mapsto \int_{X_1} f(x_1, x_2)d\mu_1(x_1)$ is μ_2 measurable, and*

$$\begin{aligned}\int_{X_2}\left(\int_{X_1} f(x_1, x_2)d\mu_1(x_1)\right)d\mu_2(x_2) &= \int_{X_1\times X_2} f d(\mu_1 \times \mu_2)\\ &= \int_{X_1}\left(\int_{X_2} f(x_1, x_2)d\mu_2(x_2)\right)d\mu_1(x_1).\end{aligned}$$

On $\mathbb{R}^{k+l} = \mathbb{R}^k \times \mathbb{R}^l$, Fubini's theorem gives

$$\int_{\mathbb{R}^l}\left(\int_{\mathbb{R}^k} f(\mathbf{x}, \mathbf{y})d\mathbf{x}\right)d\mathbf{y} = \int_{\mathbb{R}^{k+l}} f(\mathbf{z})d\mathbf{z} = \int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l} f(\mathbf{x}, \mathbf{y})d\mathbf{y}\right)d\mathbf{x},$$

where $\mathbf{z} = (\mathbf{x}, \mathbf{y})$. However, one has to modify slightly the statement for Lebesgue measures. For example, in the case of the Lebesgue measure, the partial integral $\int_{\mathbb{R}^k} f(\mathbf{x}, \mathbf{y})d\mathbf{x}$ may not exist for all $\mathbf{x}$. Nevertheless, the set of those points, for which the integral is not defined has Lebesgue measure 0. Since Lebesgue integrability and the Lebesgue integral of a function is not affected by the modification of the function on a set of measure zero, integration of functions whose values are not defined on a set of measure 0 makes sense.

Theorem 1.6.22 (Integration by Change of Variables)**.** *Let U and V be two open subsets of $\mathbb{R}^n$, $h\colon U \to V$ be a diffeomorphism of class $\mathcal{C}^1$, i.e., a* *bijection between U and V such that both h and h^{-1} are of class $\mathcal{C}^1$. Then for any Lebesgue measurable function $f\colon V \to [-\infty, \infty]$, the Lebesgue integral of f over V exists if and only if the Lebesgue integral of $(f \circ h) \cdot |\det h'|$ over U exists and if the integrals exist, then they are equal*

$$\int_V f(\mathbf{v})d\mathbf{v} = \int_U f(h(\mathbf{u})) \cdot |\det h'(\mathbf{u})|d\mathbf{u}.$$

Theorem 1.6.23 (Sard's Lemma)**.** *Let $U \subseteq \mathbb{R}^m$ be an open subset. For a $\mathcal{C}^1$-map $f\colon U \to \mathbb{R}^n$, define the set Σ_f of singular points of f by*

$$\Sigma_f = \{\mathbf{x} \in U \mid \operatorname{rk}(f'(\mathbf{x})) = \dim(\operatorname{im} f'(\mathbf{x})) < n\}.$$

Then the set $f(\Sigma_f)$ of singular values has Lebesgue measure 0, i.e., $\lambda_n(f(\Sigma_f)) = 0$.

Combining the last two theorems we obtain a generalization of the first one.

Theorem 1.6.24 (Integration by Non-Bijective Change of Variables)**.** *Suppose that U and V are open subsets of $\mathbb{R}^n$ and $h\colon U \to V$ is a map of class $\mathcal{C}^1$. For a point $\mathbf{v} \in V$, denote by $\#h^{-1}(\mathbf{v})$ the number of h-preimages of $\mathbf{v}$. Then for any Lebesgue measurable function $f\colon V \to [-\infty, \infty]$ we have*

$$\int_U f(h(\mathbf{u}))|\det h'(\mathbf{u})|d\mathbf{u} = \int_V f(\mathbf{v})\#h^{-1}(\mathbf{v})d\mathbf{v},$$

provided that both integrals exist. The integrals may not exist, but if any of them exists the other exists as well.

Measure and Integration on the Sphere

One can introduce a measure μ on the unit sphere $\mathbb{S}^{n-1} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}$ using the Lebesgue measure λ on $\mathbb{R}^n$. For a set $A \subseteq \mathbb{S}^{n-1}$, define the spherical cone $c(A)$ over A as the set $c(A) = \{t\mathbf{u} \mid \mathbf{u} \in \mathbb{S}^{n-1}, t \in (0, 1]\}$. The set A will be μ-measurable if and only if its cone $c(A)$ is Lebesgue measurable, and if this happens to be the case, then we define the μ-measure of A as $\mu(A) = n\lambda(c(A))$.

The factor n is motivated by the heuristic picture that any set can be split into very small ones. When A is very small, A is almost flat, and the spherical cone over it is almost like a pyramid over A of height 1, and in $\mathbb{R}^n$, the volume of a pyramid is $1/n$ times the volume of the base times the height.

If $f\colon \mathbb{S}^{n-1} \to \mathbb{R}$ is a μ-measurable function, then its integral can also be expressed as an integral with respect to the Lebesgue measure. Extend the function f to a function f^c on the unit ball $B^n = c(\mathbb{S}^{n-1})$ by

$$f^c(\mathbf{x}) = \begin{cases} f\left(\dfrac{\mathbf{x}}{\|\mathbf{x}\|}\right), & \text{if } \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, & \text{if } \mathbf{x} = \mathbf{0}. \end{cases}$$

Then

$$\int_{\mathbb{S}^{n-1}} f d\mu = n \int_{B^n} f^c d\lambda. \tag{1.18}$$

Equation (1.18) can be checked easily for step functions. The general case can be proved by approximating measurable functions with step functions. A Lebesgue integral over a ball, can be obtained by integrating the function over concentric spheres with respect to a suitably scaled measure and then integrating these spherical integrals with respect to the radius.

Theorem 1.6.25 (Integration in Spherical Coordinates)**.** *Let $f\colon B^n_R \to \mathbb{R}$ be a Lebesgue measurable function on the n-dimensional ball B^n_R of radius R centered at the origin. Then*

$$\int_{B^n_R} f d\lambda = \int_0^R r^{n-1} \int_{\mathbb{S}^{n-1}} f(r\mathbf{u}) d\mathbf{u} dr,$$

where integrals over $\mathbb{S}^{n-1}$ are taken with respect to the measure μ, all the other integrals are computed with respect to the Lebesgue measure.

Proof. Using equation (1.18), the right-hand side can be transferred to an integral on $B^n \times [1, R]$

$$\int_0^R r^{n-1} \int_{\mathbb{S}^{n-1}} f(r\mathbf{u}) d\mathbf{u} dr = n \int_0^R \int_{B^n \setminus \{\mathbf{0}\}} r^{n-1} f\left(r\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) d\mathbf{x} dr.$$

Compute the last integral by the substitution

$$h\colon (B^n_R \setminus \{\mathbf{0}\}) \times (0, 1] \to (B^n \setminus \{\mathbf{0}\}) \times (0, R], \qquad h(\mathbf{y}, s) = \Big(s\frac{\mathbf{y}}{\|\mathbf{y}\|}, \|\mathbf{y}\|\Big).$$

The integrand composed with h is

$$\|\mathbf{y}\|^{n-1} f\Big(\|\mathbf{y}\| \frac{\mathbf{y}/\|\mathbf{y}\|}{\big\|\mathbf{y}/\|\mathbf{y}\|\big\|}\Big) = \|\mathbf{y}\|^{n-1} f(\mathbf{y}).$$

Thinking of $\mathbf{y}$ as a column vector, the determinant of the derivative matrix of h is

$$\begin{aligned} \det(h'(\mathbf{y}, s)) &= \det \left(\begin{array}{c|c} \dfrac{sI_n}{\|\mathbf{y}\|} - \dfrac{s\mathbf{y}\mathbf{y}^T}{\|\mathbf{y}\|^3} & \dfrac{\mathbf{y}}{\|\mathbf{y}\|} \\ \hline \mathbf{y}^T/\|\mathbf{y}\| & 0 \end{array}\right) \\ &= \det \left(\begin{array}{c|c} \dfrac{sI_n}{\|\mathbf{y}\|} & \dfrac{\mathbf{y}}{\|\mathbf{y}\|} \\ \hline \mathbf{y}^T/\|\mathbf{y}\| & 0 \end{array}\right), \end{aligned}$$

where I_n is the $n \times n$ unit matrix. The second equality holds because subtracting suitable multiples of the last column from the preceding ones we can eliminate the term $s\mathbf{y}\mathbf{y}^T/\|\mathbf{y}\|^3$ from the upper left corner.

Exercise 1.6.26. Show that for any column vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, we have

$$\det \left(\begin{array}{c|c} I_n & \mathbf{a} \\ \hline \mathbf{b}^T & 0 \end{array} \right) = -\langle \mathbf{a}, \mathbf{b} \rangle. \qquad ¤$$

Applying the result of the exercise,

$$\det \left(\begin{array}{c|c} \dfrac{sI_n}{\|\mathbf{y}\|} & \dfrac{\mathbf{y}}{\|\mathbf{y}\|} \\ \hline \mathbf{y}^T/\|\mathbf{y}\| & 0 \end{array} \right)$$
$$= \left(\frac{s}{\|\mathbf{y}\|} \right)^{n+1} \det \left(\begin{array}{c|c} I_n & \dfrac{\mathbf{y}}{s} \\ \hline \mathbf{y}^T/s & 0 \end{array} \right) = -\left(\frac{s}{\|\mathbf{y}\|} \right)^{n-1}.$$

Therefore, the substitution gives

$$\begin{aligned} n\int_0^R \int_{B^n \setminus \{\mathbf{0}\}} r^{n-1} f\left(r\frac{\mathbf{x}}{\|\mathbf{x}\|} \right) d\mathbf{x}dr &= n\int_0^1 \int_{B_R^n} \|\mathbf{y}\|^{n-1} f(\mathbf{y}) \left(\frac{s}{\|\mathbf{y}\|} \right)^{n-1} d\mathbf{y}ds \\ &= n\int_0^1 s^{n-1}ds \int_{B_R^n} f(\mathbf{y})d\mathbf{y} = \int_{B_R^n} f(\mathbf{y})d\mathbf{y}, \end{aligned}$$

so the theorem is proved. □

As a corollary we obtain other ways to express spherical integral with Lebesgue integral.

Corollary 1.6.27. *Let $f\colon \mathbb{S}^{n-1} \to \mathbb{R}$ be a μ-measurable function, $g\colon [0,1] \to \mathbb{R}$ be a continuous function for which $m_g = \int_0^1 g(r)r^{n-1}dr \neq 0$. Then*

$$\int_{\mathbb{S}^{n-1}} f d\mu = \frac{1}{m_g} \int_{B^n} g(\|\mathbf{x}\|) f\left(\frac{\mathbf{x}}{\|\mathbf{x}\|} \right) d\mathbf{x}.$$

For example, choosing $g \equiv 1$ we obtain (1.18), $g(r) = r$ yields $m_g = 1/(n+1)$ and

$$\int_{\mathbb{S}^{n-1}} f d\mu = (n+1) \int_{B^n} \|\mathbf{x}\| f\left(\frac{\mathbf{x}}{\|\mathbf{x}\|} \right) d\mathbf{x}. \tag{1.19}$$

1.7 Ordinary Differential Equations

Definition 1.7.1. Let $U \subseteq \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$. A *vector field over* U is a map $\mathbf{F}\colon U \to T\mathbb{R}^n$ such that $\mathbf{F}(\mathbf{p}) \in T_{\mathbf{p}}\mathbb{R}^n$ for all $\mathbf{p} \in U$. ※

A tangent vector based at $\mathbf{p}$ is a pair $(\mathbf{p}, \mathbf{v})$, where $\mathbf{v} \in \mathbb{R}^n$, therefore, $\mathbf{F}(\mathbf{p})$ must have the form $(\mathbf{p}, F(\mathbf{p}))$. As the function $F\colon U \to \mathbb{R}^n$ obtained from $\mathbf{F}$ by ignoring the base points determines $\mathbf{F}$ uniquely, we can define a vector field uniquely by a mapping from U to $\mathbb{R}^n$.

Definition 1.7.2. Let $U \subseteq \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$, $F\colon U \to \mathbb{R}^n$ be a vector field on U. A *(first order autonomous vector valued) ordinary differential equation with right-hand side* F is the problem of finding differentiable parameterized curves, i.e., differentiable maps $\gamma\colon I \to U$ satisfying the equation

$$\gamma'(t) = F(\gamma(t)) \text{ for all } t \in I,$$

where I is a finite or infinite interval in $\mathbb{R}$. Solutions of the problem are called the *integral curves of the differential equation* or that of the vector field F. F is also called the *right-hand side* of the differential equation. ※

More generally a kth order vector valued ordinary differential equation is given by a map $F\colon U \times (\mathbb{R}^n)^{k-1} \to \mathbb{R}^n$ and is posing the problem to find k-times differentiable curves $\gamma\colon I \to U$ satisfying

$$\gamma^{(k)}(t) = F(\gamma(t), \gamma'(t), \dots, \gamma^{(k-1)}(t)) \text{ for all } t \in I.$$

Despite its more general form, every kth order differential equation is equivalent to a first order one. The equivalent problem is to find a curve $(\gamma, \eta_1, \dots, \eta_{k-1})\colon I \to U \times (\mathbb{R}^n)^{k-1}$ which satisfies the first order differential equation

$$(\gamma, \eta_1, \dots, \eta_{k-1})'(t) = (\eta_1(t), \dots, \eta_{k-1}(t), F(\gamma(t), \eta_1(t), \dots, \eta_{k-1}(t))).$$

The adjective *autonomous* refers to the circumstance that the right-hand side F depends only on the position $\gamma(t)$ but not on other parameters, say t. For example, a time dependent non-autonomous differential equation has the form

$$\gamma'(t) = F(t, \gamma(t)), \tag{1.20}$$

where $F\colon \mathbb{R} \times U \to \mathbb{R}^n$ is the time dependent right-hand side. However, such a differential equation can also be rephrased as an autonomous differential equation. Indeed, it is equivalent to finding curves $(\tau, \gamma)\colon I \to \mathbb{R} \times U$ satisfying

$$(\tau, \gamma)'(t) = (1, F(\tau(t), \gamma(t))).$$

For these reasons, we shall summarize here the fundamental theorems of ordinary differential equations only for autonomous first order systems.

Definition 1.7.3. A *maximal integral curve* of a differential equation is a solution which cannot be extended to a larger interval as a solution of the differential equation. ※

Theorem 1.7.4 (Existence and Uniqueness of Solutions)**.** *If the right-hand side $F\colon U \to \mathbb{R}^n$ of a first order differential equation is continuously differentiable, then for any $\mathbf{p} \in U$, there is a unique maximal integral curve $\gamma_{\mathbf{p}}\colon (a_{\mathbf{p}}, b_{\mathbf{p}}) \to U$ such that $-\infty \leq a_{\mathbf{p}} < 0 < b_{\mathbf{p}} \leq +\infty$ and $\gamma_{\mathbf{p}}(0) = \mathbf{p}$.*

Remark that for continuous right-hand side only the existence part of the theorem is true, uniqueness may fail.

Theorem 1.7.5 (Smooth Dependence on the Initial Condition)**.** *Suppose that the right-hand side of a first order ordinary differential equation is smooth. Then the set $W = \{(\mathbf{p}, t) \mid \mathbf{p} \in U, t \in (a_{\mathbf{p}}, b_{\mathbf{p}})$ is an open subset of $U \times \mathbb{R}$, and the map $\Phi\colon W \to U$, $\Phi(\mathbf{p}, t) = \gamma_{\mathbf{p}}(t)$ is a smooth map.*

In other words, the point at which we arrive at after traveling for time t along an integral curve starting at $\mathbf{p}$ depends smoothly on $\mathbf{p}$ and t. As a corollary, for any $t \in \mathbb{R}$, the set $W_t = \{\mathbf{p} \mid (\mathbf{p}, t) \in W\}$ is an open subset of U, and the map $\Phi_t\colon W_t \to U$, $\Phi_t(\mathbf{p}) = \Phi(\mathbf{p}, t)$ is smooth. If $(\mathbf{p}, t_0) \in W$ and $\mathbf{q} = \gamma_{\mathbf{p}}(t_0)$, then the map $(a_{\mathbf{p}} - t_0, b_{\mathbf{p}} - t_0) \to U$, $t \mapsto \gamma_{\mathbf{p}}(t + t_0)$ is a maximal integral curve starting at $\mathbf{q}$, therefore, $\gamma_{\mathbf{q}}(t) = \gamma_{\mathbf{p}}(t+t_0)$, $\mathbf{q} \in W_{-t_0}$ and $\Phi_{-t_0}(\mathbf{q}) = \mathbf{p}$. In conclusion, $\Phi_t\colon W_t \to W_{-t}$ is a diffeomorphism for all $t \in \mathbb{R}$.

Definition 1.7.6. The family $\{\Phi_t\}_{t\in\mathbb{R}}$ is called the *flow* or *one-parameter family of diffeomorphisms* of the ordinary differential equation or the vector field F. ※

The flow satisfies the following group property. If $(\mathbf{p}, t) \in W$ and $(\Phi_t(\mathbf{p}), s) \in W$, then $(\mathbf{p}, t + s \in W)$ and

$$\Phi_{t+s}(\mathbf{p}) = \Phi_s(\Phi_t(\mathbf{p})).$$

The typical reason why an integral curve cannot be extended to $[0, \pm\infty)$ is that it is running out to infinity or to the boundary of U within a finite time.

Theorem 1.7.7 (Unboundedness of Maximal Solutions in Time or Space)**.** *If for a maximal integral curve $\gamma_{\mathbf{p}}\colon (a_{\mathbf{p}}, b_{\mathbf{p}}) \to U$, $a_{\mathbf{p}} \neq -\infty$ (or $b_{\mathbf{p}} \neq +\infty$), then the trace $\gamma((a_{\mathbf{p}}, 0])$ (or $\gamma([0, b_{\mathbf{p}}))$, respectively,) cannot be covered by a bounded closed subset of U.*

Linear Differential Equations

Denote by $\mathbb{R}^{n\times m}$ the linear space of $n \times m$ matrices and identify $\mathbb{R}^n$ with the space $\mathbb{R}^{n\times 1}$ of column vectors. A *linear differential equation* is an equation of the form

$$\mathbf{x}' = A \cdot \mathbf{x}, \tag{1.21}$$

where $\mathbf{x}\colon I \to \mathbb{R}^n$ is an unknown column vector valued function defined on the interval I, $A\colon I \to \mathbb{R}^{n\times n}$ is a given matrix valued function, $\cdot$ denotes matrix multiplication. As the matrix A typically depends on t, linear differential equations are usually non-autonomous.
Of course, the fundamental theorems on ordinary differential equations are true also for linear differential equations, but linearity has some additional consequences not true in the general case. These are sum up in the following theorem.

Theorem 1.7.8.

- *If $A\colon I \to \mathbb{R}^{n\times n}$ is smooth, $t_0 \in I$ is a given initial point and $\mathbf{x}_0 \in \mathbb{R}^n$ is an arbitrary initial value, then there is a unique solution $\mathbf{x}\colon I \to \mathbb{R}^n$ of* (1.21) *for which $\mathbf{x}(t_0) = \mathbf{x}_0$.*

  Recall that for a general differential equation, solution exist only in a certain, maybe small neighborhood of the initial point t_0. In the linear case, however, solution exists on the whole interval I.

- *Solutions of the linear differential equation* (1.21) *form an n-dimensional linear space with respect to pointwise addition and multiplication by real numbers. This property is a characterization of linearity of a differential equation.*

- *In the special case when A is a constant matrix, solution of* (1.21) *with initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$ can be written explicitly as*

$$\mathbf{x}(t) = e^{A(t-t_0)}\mathbf{x}_0,$$

  where the exponential of a matrix M is defined as $e^M = \sum_{k=0}^{\infty} \frac{M^k}{k!}$.

Systems of Total Differential Equations

Systems of total differential equations are multivariable generalizations of the non-autonomous differential equation (1.20).

Definition 1.7.9. A *system of total differential equations* is an equation of the form

$$G'(\mathbf{u}) = F(\mathbf{u}, G(\mathbf{u})) \tag{1.22}$$

for an unknown multivariable function G, where $F\colon \Omega \to \mathbb{R}^{n\times m}$ is a given matrix valued function on an open subset of $\Omega \subset \mathbb{R}^m \times \mathbb{R}^n$. $G\colon U \to \mathbb{R}^n$ is a *solution* of the system, if it is defined on an open subset U of $\mathbb{R}^m$, its graph $\{(\mathbf{u}, G(\mathbf{u}) \mid \mathbf{u} \in U\}$ is contained in Ω, and (1.22) is satisfied by all $\mathbf{u} \in U$. A system of total differential equations is *integrable* if for any $(\mathbf{u}_0, \mathbf{v}_0) \in \Omega$, there is a solution G of the system satisfying the initial condition $G(\mathbf{u}_0) = \mathbf{v}_0$.
※

If G is a solution with given initial value $G(\mathbf{u}_0) = \mathbf{v}_0$, $(\mathbf{e}_1, \dots, \mathbf{e}_m)$ is the standard basis of $\mathbb{R}^m$, then for any $1 \le i \le m$, the curve $\gamma_i(t) = G(\mathbf{u}_0 + t\mathbf{e}_i)$ satisfies the ordinary differential equation

$$\gamma_i'(t) = F(\mathbf{u}_0 + t\mathbf{e}_i, \gamma_i(t)) \cdot \mathbf{e}_i$$

with the initial condition $\gamma_i(0) = \mathbf{u}_0$. (In this equation, vectors are column vectors, "$\cdot$" denotes matrix multiplication.) Solving these differential equations we can compute the values of G in a neighborhood of $\mathbf{u}_0$ along segments through $\mathbf{u}_0$ parallel to one of the coordinate axes. As any point of $\mathbb{R}^m$ can be connected to $\mathbf{u}_0$ by a broken line the segments of which are parallel to one of the coordinate axes, iterating this process we can compute the values of G in a small neighborhood of $\mathbf{u}_0$. This proves that the solution of a system of total differential equations with a given initial condition is unique in a neighborhood of the initial point. However, the solution may not exist. When $m \ge 2$, there are many ways to choose the broken line connecting $\mathbf{u}_0$ to a point $\mathbf{u}_1$ nearby, and it can happen that computing $G(\mathbf{u}_1)$ with the help of different broken lines we get different values. In such a way the system has no solution with the given initial value and the system is not integrable.
Frobenius' Theorem gives a necessary and sufficient condition for the integrability of a system of total differential equations. The easiest way to paraphrase the condition is that the system is integrable if and only if it does not contradict to Young's Theorem (Theorem 1.5.11). Compute what this means in terms of formulae. Let $F^i_j(\mathbf{u}, \mathbf{v})$ be the matrix element of $F(\mathbf{u}, \mathbf{v})$ in the ith row and jth column. If $G(\mathbf{u}) = (G^1(\mathbf{u}), \dots, G^n(\mathbf{u}))^\top$ is a solution of the system, then

$$\partial_j G^i(\mathbf{u}) = F^i_j(\mathbf{u}, G(\mathbf{u})) \qquad \text{for all } 1 \le i \le n,\ 1 \le j \le m.$$

Differentiating this equality with respect to the kth variable we obtain

$$\begin{aligned}\partial_k\partial_j G^i(\mathbf{u}) &= \partial_k F^i_j(\mathbf{u}, G(\mathbf{u})) + \sum_{s=1}^n \partial_{m+s} F^i_j(\mathbf{u}, G(\mathbf{u}))\partial_k G^s(\mathbf{u}) \\ &= \partial_k F^i_j(\mathbf{u}, G(\mathbf{u})) + \sum_{s=1}^n \partial_{m+s} F^i_j(\mathbf{u}, G(\mathbf{u})) F^s_k(\mathbf{u}, G(\mathbf{u})).\end{aligned}$$

By Young's Theorem, the right hand side of this equality should not change if we flip the role of j and k.

Theorem 1.7.10 (Frobenius' Theorem). *The system of total differential equations* (1.22) *is integrable if and only if*

$$\partial_k F^i_j + \sum_{s=1}^n \partial_{m+s} F^i_j F^s_k = \partial_j F^i_k + \sum_{s=1}^n \partial_{m+s} F^i_k F^s_j$$

holds on Ω *for all* $1 \leq i \leq n$, *and* $1 \leq j, k \leq m$.

A geometrical version of Frobenius' Theorem will be proved in Section 4.3.2.

Chapter 2

Curves in $\mathbb{E}^n$

2.1 The Notion of a Curve

In elementary geometry, one meets a lot of examples of curves: straight lines, circles, conic sections, cubic curves, graphs of functions defined on an interval or the whole real line, intersections of surfaces etc. Based on these examples everyone gets the feeling of what a curve is, however, it is not easy to give an exact definition of a curve which is satisfactory in all respect. To illustrate this, we give some commonly used definitions of certain classes of curves.

Definition 2.1.1. A *simple arc* in a topological space is a subset Γ homeomorphic to a closed interval $[a, b]$ of $\mathbb{R}$. A *parameterization of a simple arc* is a homeomorphism $\gamma\colon [a, b] \to \Gamma$. ※

Probably anyone agrees that simple arcs are curves, but since open segments, straight lines, conic sections and many other important examples of curves are not simple arcs, this class of curves is too narrow.

We could define curves as *finite unions of simple arcs*. This wider class includes circles, ellipses, but still excludes non-compact examples like straight lines, hyperbolae, parabolas. Non-compact examples would be included if we considered *countable unions of simple arcs*. This class of curves seems to be wide enough, but maybe too wide. For example, it contains the set of all those points in $\mathbb{R}^n$ which have at least one rational coordinate and it is questionable whether we could call this set a curve.

Definition 2.1.2. A 1-*dimensional topological manifold with boundary* is a second countable Hausdorff topological space, in which each point has an open neighborhood homeomorphic either to an open interval or to a left-closed, right-open interval of $\mathbb{R}$. ※

This is a technical definition, but there is a simple description of 1-dimensional topological manifolds with boundary. They have a finite or countable number of open connected components and each connected component is homeomorphic either to an open, closed, or half-closed interval, or to a circle.

The class of 1-dimensional manifolds with boundary is wider than the class of simple arcs, it includes much more important examples of curves, but as it fixes the local structure of a curve quite strictly, it *excludes examples of curves having certain kind of singularities*. For example, *figure-eight shaped curves*, like Bernoulli's lemniscate shown in Figure 2.13, are not a topological manifolds, because the self-intersection point in the middle does not have a neighborhood with the required property.

Definition 2.1.3. An *algebraic plane curve in* $\mathbb{R}^2$ is the set of solutions of a polynomial equation $P(x, y) = 0$, where $P \neq 0$ is a polynomial in two variables with real coefficients. ※

Algebraic plane curves may have a finite number of singular points, for example self intersections, so they are not necessarily 1-dimensional manifolds, but removing the singular points, the remaining set is a 1-dimensional manifold, maybe empty. On the other hand, algebraic curves are very specific curves. For example, if a straight line intersects an algebraic curve in an infinite number of points, then it is contained in the curve. In particular the graphs periodic non-constant functions (like the sine function) are not an algebraic curves.

One can also define curves as 1-*dimensional topological or metric spaces.* For such a definition one must have a proper notion of dimension. Possible definitions of dimension for a topological or metric space are discussed in a branch of topology called *dimension theory*, and within the framework of *geometric measure theory.* These theories are out of the main focus of this textbook.

All the above definitions define curves as topological spaces or subsets of topological spaces having a certain property satisfied by a sufficiently large family of known examples of curves.

The second approach, which will be more suitable for our purposes, derives curves from the motion of a point. This view is reflected in the definition of a continuous curve:

Definition 2.1.4. A *continuous parameterized curve* in a topological space is a continuous map of an interval I into the space. The interval I can have any of the forms (a, b), $(a, b]$, $[a, b)$, $[a, b]$, $(-\infty, b)$, $(-\infty, b]$, $(a, +\infty)$, $[a, +\infty)$, $(-\infty, +\infty) = \mathbb{R}$, where $a, b \in \mathbb{R}$. If $I = [a, b]$, then the images of a and b are the *initial* and *terminal points* of the curve respectively. The path is said to *connect* the initial point to the terminal point. ※

We stress that according to this definition, a *parameterized curve* is a *map* and not a set of points as in the earlier definitions. However, we can associate to any parameterized curve a subset of the ambient space, the set of points traced out by the moving point.

Definition 2.1.5. The *trace* or *trajectory of a parameterized curve* $\gamma\colon I \to X$ is the image of the map γ. We say that γ *is a parameterization of the subset* A of X when A is the trace of γ. ※

In many cases, the trace of a continuous parameterized curve is a curve in one of the above sense, but there are examples, when the image is not a curve at all. The Italian mathematician Giuseppe Peano (1858-1932) constructed a continuous parameterized curve that passes through each point of a square. Such a pathology can not occur if we restrict ourselves to smooth curves.

Definition 2.1.6. A *smooth parameterized curve in the Euclidean space* $\mathbb{E}^n$ is a smooth map $\gamma\colon I \to \mathbb{E}^n$ from an interval I into $\mathbb{E}^n$. ※

Definition 2.1.7. We say that the continuous curve $\gamma_1\colon I_1 \to \mathbb{R}^n$ is obtained from the curve $\gamma_2\colon I_2 \to \mathbb{R}^n$ by a *reparameterization* if there is a homeomorphism $\phi\colon I_1 \to I_2$ such that $\gamma_1 = \gamma_2 \circ \phi$. We say that the reparameterization *preserves orientation* if ϕ is increasing. A reparameterization is called *regular* if ϕ is smooth and $\phi'(t) \neq 0$ for all $t \in I_1$. ※

Intuitively, orientation preserving reparameterizations describe motions along the same route with different timings.

2.2 The Length of a Curve

Definition 2.2.1. The *length* of a continuous curve $\gamma\colon [a,b] \to \mathbb{R}^n$ is the limit of the lengths of inscribed broken lines with consecutive vertices $\gamma(t_0), \gamma(t_1), \dots, \gamma(t_N)$, where $a = t_0 < t_1 < \cdots < t_N = b$ and the limit is taken as $\max_{1\leq i\leq N} |t_i - t_{i-1}|$ tends to zero. Provided that this limit is finite, the curve is called *rectifiable.* ※

Exercise 2.2.2. Show that the limit of the lengths of the inscribed broken lines always exists and it is equal to the supremum of the lengths. Construct a continuous curve $\gamma\colon [a,b] \to \mathbb{R}^2$ having infinite length. ¤

The following theorem yields a formula that can be used in practice to compute the length of curves.

Theorem 2.2.3. *A smooth curve $\gamma\colon [a,b] \to \mathbb{R}^n$ is always rectifiable and its length is equal to the integral*

$$l(\gamma) = \int_a^b |\gamma'(t)| dt.$$

Proof. Denoting by $x_1, x_2, \ldots, x_n$ the coordinate functions of $\mathbf{x}$ the length of the broken line considered in Definition 2.2.1 is equal to

$$\lambda = \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{n} (x_j(t_i) - x_j(t_{i-1}))^2}.$$

By the Lagrange mean value theorem we can find real numbers ξ_{ij} such that

$$x_j(t_i) - x_j(t_{i-1}) = x_j'(\xi_{ij})(t_i - t_{i-1}), \qquad t_{i-1} < \xi_{ij} < t_i.$$

Using these equalities we get

$$\lambda = \sum_{i=1}^{N} (t_i - t_{i-1}) \sqrt{\sum_{j=1}^{n} x_j'(\xi_{ij})^2}\ .$$

Fix a positive ε. By Proposition 1.4.43, for each $\varepsilon > 0$, we can find a positive δ such that $t, t^* \in [a, b]$ and $|t - t^*| < \delta$ imply $|x_j'(t) - x_j'(t^*)| < \varepsilon$ for all $1 \leq j \leq n$.
Suppose that the approximating broken line is fine enough in the sense that $|t_i - t_{i-1}| < \delta$ for all $1 \leq i \leq N$. Then we have by the triangle inequality

$$\left| \sqrt{\sum_{j=1}^{n} x_j'(\xi_{ij})^2} - \sqrt{\sum_{j=1}^{n} x_j'(t_i)^2} \right| \leq \sqrt{\sum_{j=1}^{n} (x_j'(\xi_{ij}) - x_j'(t_i))^2} \leq \varepsilon\sqrt{n}.$$

Making use of this estimation we see that

$$\left| \lambda - \sum_{i=1}^{N} \left((t_i - t_{i-1}) \sqrt{\sum_{j=1}^{n} x_j'(t_i)^2} \right) \right| \leq \varepsilon\sqrt{n}(b - a) \tag{2.1}$$

In this formula

$$\sum_{i=1}^{N} \left((t_i - t_{i-1}) \sqrt{\sum_{j=1}^{n} x_j'(t_i)^2} \right) = \sum_{i=1}^{N} (t_i - t_{i-1}) |\gamma'(t_i)|$$

is just an integral sum which converges to the integral $\int_a^b |\gamma'(t)| dt$ when $\max_i |t_i - t_{i-1}|$ tends to zero. Taking into account inequality (2.1) we see that in this case the length λ of the inscribed broken lines also tends to this integral. □

Let $\gamma\colon I \to \mathbb{R}^n$ be a smooth curve, $a \in I$ be a given point. Consider the function $s\colon I \to \mathbb{R}$

$$s(t) = \int_a^t |\gamma'(\tau)| d\tau.$$

$s(t)$ is the signed length of the arc of the curve between $\gamma(a)$ and $\gamma(t)$. It is a monotone but not necessarily a strictly monotone function of t in general. This fact motivates the following definition.

Definition 2.2.4. A smooth curve is said to be *regular* if $\gamma'(t) \neq \mathbf{0}$ for all $t \in I$. ※

If γ is a regular curve, then s defines a regular reparameterization of γ. The map $\gamma \circ s^{-1}\colon s(I) \to \mathbb{R}^n$ is referred to as a *natural* or *unit speed parameterization* of the curve γ or as a *parameterization of* γ *by arc length*. The second name is justified by the fact that the speed vector

$$(\gamma \circ s^{-1})'(t) = \gamma'(s^{-1}(t)) \cdot (s^{-1})'(t) = \gamma'(s^{-1}(t)) \frac{1}{s'(s^{-1}(t))} = \frac{\gamma'(s^{-1}(t))}{|\gamma'(s^{-1}(t))|}$$

of this parameterization has unit length at each point.

Exercise 2.2.5. The curve *cycloid* is the trajectory of a peripheral point of a circle that rolls along a straight line. Find a parameterization of the cycloid and compute the length of one of its arcs. ¤

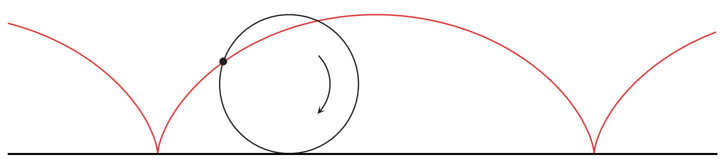

Figure 2.1: Generation of the cycloid.

Sometimes it is more convenient to use the polar coordinate system in the plane.

Definition 2.2.6. The *polar coordinates* (r, ϕ) of a point $(x, y) \in \mathbb{R}^2$ in the plane are the distance $r = \sqrt{x^2 + y^2}$ of the point from the origin and the direction angle ϕ of the vector (x, y). The direction angle is not defined at the origin $(0, 0)$, and it is defined only modulo 2π at other points. There is no continuous choice of ϕ for the whole punctured plane $\mathbb{R}^2 \setminus \{(0, 0)\}$. However, one can choose ϕ continuously on the complement of any closed half-line starting at the origin. E.g., on the complement of the half-line $\{(x, y) \mid x \leq 0,\ y = 0\}$, ϕ can be defined continuously by the formula $\phi = 2\arctan\left(\frac{y}{x+\sqrt{x^2+y^2}}\right) \in (-\pi, \pi)$. ※

Cartesian coordinates can be expressed in terms of the polar coordinates as

$$x = r\cos(\phi), \quad y = r\sin(\phi).$$

Exercise 2.2.7. Let $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{\mathbf{0}\}$ be a smooth curve and denote by $(r(t), \phi(t))$ the polar coordinates of $\gamma(t)$, where $\phi(t)$ is chosen to be a smooth function of t. Prove that the length of γ is equal to the integral

$$\int_a^b \sqrt{r^2(t)\phi'^2(t) + r'^2(t)}\, dt.$$ ¤

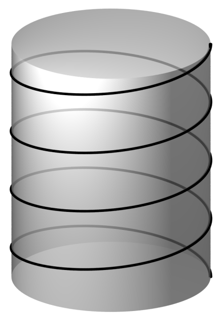

Figure 2.2: A helix.

Exercise 2.2.8. Find a natural reparameterization of the *helix* $\gamma(t) = (a\cos t, a\sin t, bt)$. ¤

2.3 Crofton's Formula

There are more sophisticated integral formulas for the length of a curve. In this section, we discuss some of them.

Topology and measure on the set of straight lines

Let $\mathcal{E} = \mathrm{AGr}_1(\mathbb{R}^2)$ be the set of straight lines in the Euclidean plane $\mathbb{R}^2$. For $(\theta, p) \in \mathbb{R}^2$, denote by $e_{\theta,p} \in \mathcal{E}$ the straight line defined by the equation $x\cos(\theta) + y\sin(\theta) = p$. As every straight line can be defined by such an equation, the map $\rho\colon \mathbb{R}^2 \to \mathcal{E}$, $\rho(\theta, p) = e_{\theta,p}$ is surjective. However, ρ is not injective since $\rho(\theta, p) = \rho(\tilde{\theta}, \tilde{p})$ if and only if $\theta - \tilde{\theta} = k\pi$ and $p = (-1)^k\tilde{p}$ for some integer $k \in \mathbb{Z}$. In particular, the band $[0, \pi) \times \mathbb{R}$, closed from the left and open from the right, is mapped onto $\mathcal{E}$ bijectively, and a boundary point (π, p) on its right side corresponds to the same line as the point $(0, -p)$. This implies easily that if we equip the set $\mathcal{E}$ with the factor topology induced by

the surjective map ρ, then $\mathcal{E}$ will be a Möbius band without its boundary circle, and ρ becomes a covering map, the universal covering map of $\mathcal{E}$.
The Lebesgue measure λ on $\mathbb{R}^2$ defines a measure ν on $\mathcal{E}$ with the help of the covering map ρ as follows. Let a subset $A \subset \mathcal{E}$ be ν-measurable if and only if $\rho^{-1}(A) \cap ([0,\pi] \times \mathbb{R})$ Lebesgue measurable and then let $\nu(A) = \lambda(\rho^{-1}(A) \cap ([0,\pi] \times \mathbb{R}))$.

Proposition 2.3.1. *The measure ν is invariant under the isometry group of the plane, that is, if $\Phi \in Iso(\mathbb{R}^2)$ is an arbitrary isometry, $A \subset \mathcal{E}$ is a ν-measurable set of straight lines, then $\Phi(A)$ is also ν-measurable, and $\nu(A) = \nu(\Phi(A))$.*

Proof. The isometry group of the plane is generated by the following three types of transformations:

- Rotations R_α by angle α about the origin;
- Translations T_a by a vector $(a, 0)$ parallel to the x axis;
- Reflection M in the x axis.

Hence it is enough to check the invariance of the measure under these transformations.
It is clear that $R_\alpha(e_{\theta,p}) = e_{\theta+\alpha,p}$. Since the translation $(\theta, p) \mapsto (\theta + \alpha, p)$ preserves Lebesgue measure, the statement is true for the rotations R_α.
The action of the translation T_a on the straight line parameters is not so simple as $T_a(e_{\theta,p}) = e_{\theta,p+a\cos(\theta)}$, nevertheless, as the determinant of the derivative of the transformation $(\theta, p) \mapsto (\theta, p + a\cos(\theta))$ is 1, this transformation also preserves Lebesgue measure.
Finally, the action of M on the line parameters is given by the map $(\theta, p) \mapsto (-\theta, p)$. The latter transformation is a reflection, therefore preserves the Lebesgue measure. □

The Planar Crofton Formula

Theorem 2.3.2. *Let $\gamma\colon [a, b] \to \mathbb{R}^2$ is a $\mathcal{C}^1$ curve, $m\colon \mathcal{E} \to \mathbb{N} \cup \{\infty\}$ is defined by $m(e) = \#\{t \in [a, b] \mid \gamma(t) \in e\}$. In other words, $m(e)$ is the number of intersection points of the curve γ and the straight line e counted with multiplicity. Then the length of γ is*

$$l_\gamma = \frac{1}{2}\int_{\mathcal{E}} m\, d\nu.$$

Proof. Consider the $\mathcal{C}^1$-map $h\colon [0,\pi] \times [a, b] \to [0,\pi] \times \mathbb{R}$ defined by

$$h(\theta, t) = (\theta, x(t)\cos\theta + y(t)\sin\theta),$$

where $x(t)$ and $y(t)$ are the coordinates of $\gamma(t)$, and apply Theorem 1.6.24 to h and the constant 1 function on $[0,\pi] \times \mathbb{R}$. The number of preimages of a point $(\theta, p) \in [0,\pi] \times \mathbb{R}$ is the number $m(e_{\theta,p})$ of intersection points of the straight line $e_{\theta,p}$ with the curve γ counted with multiplicities. The determinant of the derivative matrix of h at (θ, t) is

$$\det \begin{pmatrix} 1 & 0 \\ -x(t)\sin\theta + y(t)\cos\theta & x'(t)\cos\theta + y'(t)\sin\theta \end{pmatrix} = x'(t)\cos\theta + y'(t)\sin\theta$$

Thus, we get

$$\int_{\mathcal{E}} m\, d\nu = \int_{[0,\pi]\times\mathbb{R}} m(e_{\theta,p}) d\theta dp = \int_a^b \int_0^\pi |x'(t)\cos\theta + y'(t)\sin\theta| d\theta dt.$$

To compute the integral $\int_0^\pi |x'(t)\cos\theta + y'(t)\sin\theta| d\theta$, fix the value of t and write the speed vector $\gamma'(t) = (x'(t), y'(t))$ as $v(t)(\cos\phi, \sin\phi)$, where $v(t) = \|\gamma'(t)\|$, ϕ is a direction angle of the speed vector or any angle if $v(t) = 0$. Then

$$\int_0^\pi |x'(t)\cos\theta + y'(t)\sin\theta| d\theta = v(t) \int_0^\pi |(\cos(\theta - \phi)| d\theta.$$

Since the absolute value of the cosine function is periodic with period π, the last integral does not depend on ϕ and its value is

$$\int_0^\pi |\cos(\theta - \phi)| d\theta = \int_0^\pi |\cos(\theta)| d\theta = 2\int_0^{\pi/2} \cos(\theta) d\theta = 2.$$

Combining these equations we get

$$\int_{\mathcal{E}} m\, d\nu = \int_a^b \int_0^\pi |x'(t)\cos\theta + y'(t)\sin\theta| d\theta dt = \int_a^b 2v(t) dt = 2l_\gamma,$$

as we wanted to show. □

Topology and Measure on the Set of Hyperplanes

One can generalize the planar Crofton's formula for curves in $\mathbb{R}^n$. In the higher dimensional version, we have to count the number of intersection points of the curve with hyperplanes. Let $\mathcal{H} = \mathrm{AGr}_{n-1}(\mathbb{R}^n)$ denote the set of hyperplanes in $\mathbb{R}^n$. Let $B^n = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \le 1\}$ be the unit ball centered at the origin, $\mathbb{S}^{n-1} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}$ be its boundary sphere. Assign to each pair $(\mathbf{u}, p) \in \mathbb{S}^{n-1} \times \mathbb{R}$ the hyperplane $H_{\mathbf{u},p} = \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{u}, \mathbf{x}\rangle = p\}$.
The map $\rho\colon \mathbb{S}^{n-1} \times \mathbb{R} \to \mathcal{H}$, $(\mathbf{u}, p) \mapsto H_{\mathbf{u},p}$ is surjective and since $H_{\mathbf{u},p} = H_{\tilde{\mathbf{u}},\tilde{p}}$ if and only if $(\mathbf{u}, p) = (\tilde{\mathbf{u}}, \tilde{p})$ or $(\mathbf{u}, p) = (-\tilde{\mathbf{u}}, -\tilde{p})$, ρ is a double covering.
The map ρ induces a factor topology and also a measure ν on $\mathcal{H}$. A subset $A \subset \mathcal{H}$ is ν-measurable if and only if $\rho^{-1}(A)$ is $(\mu \times \lambda_1)$-measurable, where μ is the surface measure on $\mathbb{S}^{n-1}$, λ_1 is the Lebesgue measure on the real line, and if A is ν-measurable, then its measure is $\nu(A) = \frac{1}{2}(\mu \times \lambda)_1(\rho^{-1}(A))$.

Crofton Formula in $\mathbb{R}^n$

Theorem 2.3.3. *Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a $\mathcal{C}^1$ curve, and $m\colon \mathcal{H} \to \mathbb{N} \cup \{\infty\}$ be the map assigning to a hyperplane H the number $m(H) = \#\{t \in [a,b] \mid \gamma(t) \in H\}$ of intersection points of the curve γ and the hyperplane H counted with multiplicities. Then the length of γ is*

$$l_\gamma = \frac{1}{\omega_{n-1}} \int_{\mathcal{H}} m d\nu.$$

where ω_{n-1} is the volume of the $(n-1)$-dimensional unit ball.

Proof. The proof is analogous to the planar case. Consider the $\mathcal{C}^1$-map

$$h\colon B^n \times [a,b] \to B^n \times \mathbb{R}, \qquad h(\mathbf{x},t) = (\mathbf{x}, \langle \mathbf{x}, \gamma(t)\rangle),$$

and apply Theorem 1.6.24 to it and the constant 1 function on $B^n \times \mathbb{R}$, For $\mathbf{x} \neq \mathbf{0}$, the number of h-preimages of $(\mathbf{x},p)$ is the number of intersection points of the hyperplane $H_{\frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{p}{\|\mathbf{x}\|}}$.
The number of h-preimages of $(\mathbf{0},0)$ is ∞ while for $p \neq 0$ $h^{-1}(\mathbf{0},p) = \emptyset$. However, these values can be ignored since the set $\{\mathbf{0}\} \times \mathbb{R}$ has measure 0 in $B^n \times \mathbb{R}$.
The determinant of the derivative matrix of h at $(\mathbf{x},t)$ is

$$\det(h'(\mathbf{x},t)) = \det \left(\begin{array}{cccc|c} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ \hline & & \gamma(t) & & \langle \mathbf{x}, \gamma'(t)\rangle \end{array}\right) = \langle \mathbf{x}, \gamma'(t)\rangle.$$

Thus, Theorem 1.6.24 yields

$$\int_{B^n \times [a,b]} |\langle \mathbf{x}, \gamma'(t)\rangle| d\mathbf{x} dt = \int_{B^n \times \mathbb{R}} m(H_{\frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{p}{\|\mathbf{x}\|}}) d\mathbf{x} dp$$

Substituting $p = \|\mathbf{x}\|\bar{p}$ in the second integral we see that

$$\begin{aligned}\int_{B^n \times \mathbb{R}} m\big(H_{\frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{p}{\|\mathbf{x}\|}}\big) d\mathbf{x} dp &= \int_{B^n \times \mathbb{R}} \|\mathbf{x}\| m\big(H_{\frac{\mathbf{x}}{\|\mathbf{x}\|}, \bar{p}}\big) d\mathbf{x} d\bar{p} \\ &= \frac{1}{n+1} \int_{\mathbb{R}} \int_{\mathbb{S}^{n-1}} m(H_{\mathbf{u},\bar{p}}) d\mathbf{u} d\bar{p} = \frac{2}{n+1} \int_{\mathcal{H}} m d\nu.\end{aligned}$$

To compute the integral $\int_{B^n} |\langle \mathbf{x}, \gamma'(t)\rangle| d\mathbf{x}$ for a fixed t, write $\gamma'(t)$ as $v(t)\mathbf{u}$, where $v(t) = \|\gamma'(t)\|$, $\mathbf{u}$ is a unit vector. Slice the ball with the hyperplanes

$H_{\mathbf{u},\tau}$ orthogonal to $\mathbf{u}$. If $\tau \in [-1,1]$, then $H_{\mathbf{u},\tau} \cap B^n$ is an $(n-1)$-dimensional ball of radius $\sqrt{1-\tau^2}$. Since the function $\langle \mathbf{x}, \gamma'(t)\rangle$ is equal to the constant $v(t)\tau$ on this ball,

$$\int_{H_{\mathbf{u},\tau}\cap B^n} |\langle \mathbf{x}, \gamma'(t)\rangle| d\mathbf{x} = \omega_{n-1} v(t) |\tau| (1-\tau^2)^{(n-1)/2}.$$

Integrating the integrals over the slices

$$\begin{aligned}\int_{B^n} |\langle \mathbf{x}, \gamma'(t)\rangle| d\mathbf{x} &= \omega_{n-1} v(t) \int_{-1}^{1} |\tau| (1-\tau^2)^{(n-1)/2} d\tau \\ &= 2\omega_{n-1} v(t) \int_{0}^{1} \tau (1-\tau^2)^{(n-1)/2} d\tau \\ &= 2\omega_{n-1} v(t) \left[\frac{-1}{n+1} (1-\tau^2)^{(n+1)/2} \right]_0^1 = \frac{2\omega_{n-1}}{n+1} v(t)\end{aligned}$$

by Fubini's theorem. Integrating with respect to t we get

$$\int_{B^n \times [a,b]} |\langle \mathbf{x}, \gamma'(t)\rangle| d\mathbf{x} dt = \int_a^b \frac{2\omega_{n-1}}{n+1} v(t) dt = \frac{2\omega_{n-1}}{n+1} l_\gamma \tag{2.2}$$

which completes the proof. □

Crofton Formula for Spherical Curves

To finish this section, we compute yet another version of the Crofton formula for curves lying on the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$. This formula will express the length of a curve using only the number of intersection points with hyperplanes passing through the origin. The scheme of producing the formula is the same as before.

Denote by $\mathcal{H}_0 = \mathrm{Gr}_{n-1}(\mathbb{R}^n)$ the set of all hyperplanes passing through the origin. For $\mathbf{u} \in \mathbb{S}^{n-1}$, let $H_{\mathbf{u}} \in \mathcal{H}_0$ be the hyperplane orthogonal to $\mathbf{u}$. The map $\rho\colon \mathbb{S}^{n-1} \to \mathcal{H}_0$, $\mathbf{u} \mapsto H_{\mathbf{u}}$ is a double cover of $\mathcal{H}_0$. The spherical measure μ on $\mathbb{S}^{n-1}$ induces a measure ν on $\mathcal{H}_0$ for which $A \subset \mathcal{H}_0$ is ν-measurable if and only if $\rho^{-1}(A)$ is μ-measurable and if this is the case, then $\nu(A) = \frac{1}{2}\mu(A)$.

Theorem 2.3.4. *Let $\gamma\colon [a,b] \to \mathbb{S}^{n-1}$ be a spherical curve. For $H \in \mathcal{H}_0$, denote by $m(H)$ the number of intersection points of H with γ counted with multiplicity, i.e.,*

$$m(H) = \#\{t \in [a,b] \mid \gamma(t) \in H\}.$$

Then the length l_γ of γ can be obtained as the integral

$$l_\gamma = \frac{2\pi}{n\omega_n} \int_{\mathcal{H}_0} m d\nu = \frac{\pi}{n\omega_n} \int_{S^{n-1}} m(H_{\mathbf{u}}) d\mathbf{u},$$

where ω_n is the volume of the n-dimensional unit ball.

Proof. For a fixed $t \in [a,b]$, any vector $\mathbf{w} \in \mathbb{R}^n$ can be decomposed uniquely into the sum of two vectors so that the first vector is parallel to $\gamma(t)$, the second one is orthogonal to it. The decomposition is

$$\mathbf{w} = \langle \mathbf{w}, \gamma(t)\rangle\gamma(t) + (\mathbf{w} - \langle \mathbf{w}, \gamma(t)\rangle\gamma(t)).$$

Consider the map

$$h\colon B^n \times [a,b] \to \mathbb{R}^n \times \mathbb{R}, \qquad (\mathbf{w}, t) \to (\mathbf{w} - \langle \mathbf{w}, \gamma(t)\rangle\gamma(t), \langle \mathbf{w}, \gamma(t)\rangle)$$

encoding the two components of this decomposition.
If $(\mathbf{x}, s) \in \mathbb{R}^n \times \mathbb{R}$ and $\mathbf{x} \neq \mathbf{0}$, then $(\mathbf{w}, t) \in h^{-1}(\mathbf{x}, s)$ if and only if $t \in [a,b]$ is a point such that $\gamma(t) \perp \mathbf{w}$ and $\mathbf{w} = \mathbf{x} + s\gamma(t)$. Since $\|\mathbf{w}\|^2 = \|\mathbf{x}\|^2 + s^2$, $\mathbf{w} \in B^n$ if and only if $(\mathbf{x}, s) \in B^{n+1}$. Thus, the image of h is the unit ball $B^{n+1} \subset \mathbb{R}^n \times \mathbb{R}$, and the number of h-preimages of $(\mathbf{x}, s) \in B^{n+1}$ is $m\big(H_{\frac{\mathbf{x}}{\|\mathbf{x}\|}}\big)$.
The determinant of the derivative matrix of h at $(\mathbf{w}, t)$ is

$$\det(h'(\mathbf{w}, t)) = \det\left(\begin{array}{c|c} I_n - \gamma(t)\gamma^T(t) & \gamma(t) \\ \hline -\langle \mathbf{w}, \gamma'(t)\rangle\gamma^T(t) - \langle \mathbf{w}, \gamma(t)\rangle{\gamma'}^T(t) & \langle \mathbf{w}, \gamma'(t)\rangle \end{array}\right).$$

Adding suitable multiples of the last column to the previous ones we can eliminate the $\gamma(t)\gamma^T(t)$ term from the upper left corner. Then applying Exercise 1.6.26 we get

$$\begin{aligned}\det(h'(\mathbf{w}, t)) &= \det\left(\begin{array}{c|c} I_n & \gamma(t) \\ \hline -\langle \mathbf{w}, \gamma(t)\rangle\gamma'(t) & \langle \mathbf{w}, \gamma'(t)\rangle \end{array}\right) \\ &= \langle \mathbf{w}, \gamma'(t)\rangle + \langle \mathbf{w}, \gamma(t)\rangle\langle \gamma(t), \gamma'(t)\rangle.\end{aligned}$$

This reduces to $\det(h'(\mathbf{w}, t)) = \langle \mathbf{w}, \gamma'(t)\rangle$ as γ is on the unit sphere, therefore $(\|\gamma\|^2)' = 2\langle \gamma, \gamma'\rangle \equiv 0$.
Applying Theorem 1.6.24 for h and the constant 1 function on B^{n+1} yields

$$\begin{aligned}\int_a^b \int_{B^n} |\langle \mathbf{w}, \gamma'(t)\rangle| d\mathbf{w} dt &= \int_{(\mathbf{x},s)\in B^{n+1}} m(H_{\frac{\mathbf{x}}{\|\mathbf{x}\|}}) d\mathbf{x} ds \\ &= \int_{B^n} 2\sqrt{1 - \|\mathbf{x}\|^2} m(H_{\frac{\mathbf{x}}{\|\mathbf{x}\|}}) d\mathbf{x}.\end{aligned}$$

The second equation comes from Fubini's theorem by slicing the ball B^{n+1} by straight lines orthogonal to the hyperplane $\mathbb{R}^n \times \{0\}$. By equation (2.2), we have

$$\int_a^b \int_{B^n} |\langle \mathbf{w}, \gamma'(t)\rangle| d\mathbf{w} dt = \frac{2\omega_{n-1}}{n+1} l_\gamma.$$

According to Corollary 1.6.27, if we denote the integral $\int_0^1 r^{n-1}\sqrt{1-r^2}dr$ by a_n, then

$$\int_{B^n} 2m(H_{\frac{\mathbf{x}}{\|\mathbf{x}\|}})\sqrt{1-\|\mathbf{x}\|^2}d\mathbf{x} = 2a_n \int_{\mathbb{S}^{n-1}} m(H_{\mathbf{u}})\, d\mathbf{u}.$$

The last three equations show that

$$l_\gamma = c_n \int_{S^{n-1}} m(H_{\mathbf{u}})\, d\mathbf{u}, \tag{2.3}$$

where $c_n = \frac{(n+1)a_n}{\omega_{n-1}}$. The value of c_n can be obtained by computing a_n, but it can also be obtained by evaluating (2.3) for a great circle. The length of a great circle on $\mathbb{S}^{n-1}$ is 2π and almost all hyperplanes through the origin cut a great circle in exactly 2 points, that is the set of exceptional hyperplanes, that contain the great circle has measure 0. This way, $\int_{\mathbb{S}^{n-1}} m(H_{\mathbf{u}})d\mathbf{u}$ for a great circle is twice the surface measure $\mu(\mathbb{S}^{n-1}) = n\omega_n$ of the sphere. This gives that $c_n = \pi/(n\omega_n)$. □

Exercise 2.3.5. Find a direct proof of the equations

$$a_n = \int_0^1 r^{n-1}\sqrt{1-r^2}dr = \frac{\omega_{n-1}\pi}{n(n+1)\omega_n} = \frac{\omega_{n+1}}{2n\omega_n}$$
$$= \begin{cases} \frac{2^{k-1}(k-1)!}{(2k+1)!!}, & \text{if } n = 2k, \\ \frac{(2k-1)!!}{(2k+2)!!}\cdot\frac{\pi}{2}, & \text{if } n = 2k+1. \end{cases}$$ ¤

2.4 The Osculating k-planes

One of the most important tools of analysis is linearization, or more generally, the approximation of general objects with more easily treatable ones. E.g. the derivative of a function is the best linear approximation, Taylor polynomials are the best polynomial approximations of the function around a point. Adapting this idea to the theory of curves, the following questions arise naturally. Given a curve $\gamma\colon [a,b] \to \mathbb{R}^n$ and a point $\gamma(\bar{t})$ on it, how can we find the straight line, circle, conic or polynomial curve of degree less than or equal to n etc. that approximates the curve around $\gamma(\bar{t})$ the best or how can we find the k-plane, k-sphere, quadric surface etc. that is tangent to the curve at $\gamma(\bar{t})$ with the highest possible order?

We shall deal now with the problem of finding the k-plane which is tangent to a curve at a given point with the highest possible order. The classical approach to this problem is the following. A k-plane is determined uniquely by $k+1$ points of it not lying in a $(k-1)$-plane. Let us take $k+1$ points $\gamma(t_0), \gamma(t_1), \ldots, \gamma(t_k)$ on the curve. If γ is a curve of "general type" then these points span a unique k-plane, which will be denoted by $A(t_0, \ldots, t_k)$. The k-plane we look for is the limit position of the k-planes $A(t_0, \ldots, t_k)$ as $t_0, \ldots, t_k$ tend to $\bar{t}$.
To properly understand the last sentence, we need a definition of convergence of k-planes. In section 1.4, we defined a topology on the affine Grassmann manifold $AGr_k(\mathbb{R}^n)$ in different but equivalent ways. Constructing the topology as a factor topology gives the following notion of convergence of k-planes.

Proposition 2.4.1. *The sequence of k-planes $X_1, X_2, \ldots$ tends to the k-plane X with respect to the topology of the affine Grassmann manifold* $\mathrm{AGr}_k(\mathbb{R}^n)$ *if and only if one can find points $\mathbf{p}_j \in X_j$, $\mathbf{p} \in X$ and linearly independent direction vectors $\mathbf{v}_1^j, \ldots, \mathbf{v}_k^j$ of X_j and $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of X such that* $\lim_{j\to\infty} \mathbf{p}_j = \mathbf{p}$ *and* $\lim_{j\to\infty} \mathbf{v}_i^j = \mathbf{v}_i$ *for $i = 1, \ldots, k$.*

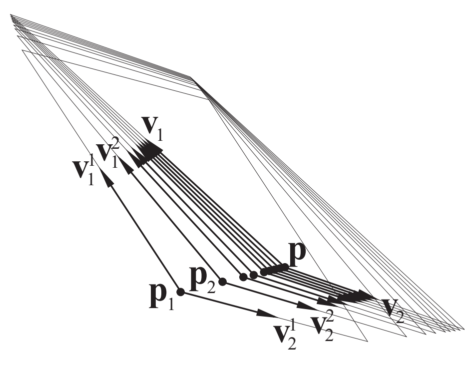

Figure 2.3: Convergence of k-planes

Exercise 2.4.2. Show that convergence of k-planes as defined above is the same as convergence with respect to the topology constructed in section 1.4. ¤

The topology of each Grassmann manifold is metrizable, hence Hausdorff, which implies that a convergent sequence in a Grassman manifold can have at most one limit point. We give a direct proof of this fact, relying only on Proposition 2.4.1.

Proposition 2.4.3. *A sequence of k-planes can have at most one limit.*

Proof. Suppose that the sequence $X_1, X_2, \dots$ has two limits, say X and Y. Then by the definition, one can find points $\mathbf{p}_j$, $\mathbf{q}_j \in X_j$, $\mathbf{p} \in X$, $\mathbf{q} \in Y$ and linearly independent direction vectors $\{\mathbf{v}_1^j, \dots, \mathbf{v}_k^j\}$ and $\{\mathbf{w}_1^j, \dots, \mathbf{w}_k^j\}$ of X_j, $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ of X and $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ of Y, such that $\lim \mathbf{p}_j = \mathbf{p}$, $\lim \mathbf{q}_j = \mathbf{q}$, $\lim \mathbf{v}_i^j = \mathbf{v}_i$ and $\lim \mathbf{w}_i^j = \mathbf{w}_i$ for $i = 1, \dots, k$ as j tends to infinity. Since $\{\mathbf{v}_1^j, \dots, \mathbf{v}_k^j\}$ and $\{\mathbf{w}_1^j, \dots, \mathbf{w}_k^j\}$ span the same linear space, which contains the direction vector $\mathbf{p}_j - \mathbf{q}_j$, there exist a unique $k \times k$ matrix $(a_{rs}^j)_{1 \le r,s \le k}$ and a vector $(b_1^j, \dots, b_k^j)$ such that

$$\begin{aligned} \mathbf{v}_i^j &= \sum_{s=1}^k a_{is}^j \mathbf{w}_s^j \qquad \text{for } i = 1, \dots, k; \\ \mathbf{p}_j - \mathbf{q}_j &= \sum_{s=1}^k b_s^j \mathbf{w}_s^j \end{aligned} \tag{2.4}$$

The components a_{rs}^j of this matrix and the numbers b_s^j can be determined by solving the system (2.4) of linear equations, thus by Cramer's rule, they are rational functions (quotients of polynomials) of the components of the vectors $\mathbf{v}_i^j$ and $\mathbf{w}_i^j$. Using the fact that rational functions are continuous at each point where they are defined one can show that the limits $\lim a_{rs}^j = a_{rs}$, $\lim b_r^j = b_r$ exist as $j \to \infty$. Taking $j \to \infty$ in (2.4) we obtain

$$\begin{aligned} \mathbf{v}_i &= \sum_{s=1}^k a_{is} \mathbf{w}_s \qquad \text{for } i = 1, \dots, k; \\ \mathbf{p} - \mathbf{q} &= \sum_{s=1}^k b_s \mathbf{w}_s; \end{aligned}$$

from which follows that the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ and $\mathbf{w}_1, \dots, \mathbf{w}_k$ span the same linear space, i.e. the k-planes X and Y must be parallel and that the point $\mathbf{q}$ is a common point of them, consequently $X = Y$.

□

Definition 2.4.4. Let $\gamma\colon [a, b] \to \mathbb{R}^n$ be a curve, $\bar{t} \in (a, b)$, $1 \le k \le n$. If the k-planes $A(t_0, \dots, t_k)$ are defined for parameters close enough to $\bar{t}$ and their limit exists as $t_0, \dots, t_k \to \bar{t}$, then the limit is called the *osculating k-plane* of the curve γ at $\bar{t}$. The osculating 1-plane of a curve is just the *tangent* of the curve. (The word "osculate" comes from the Latin "osculatio", which means kissing or kiss). ※

The definition of the osculating k-plane is justified by intuition. It is the k-plane that passes through $k + 1$ points "infinitely close" to a given point. However, it is usually not convenient to determine the osculating k-plane using its definition directly. Fortunately, we have the following theorem.

Theorem 2.4.5. *Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a smooth curve, $\bar{t} \in (a,b)$, $1 \le k \le n$. If the derivatives $\gamma'(\bar{t}), \gamma''(\bar{t}), \dots, \gamma^{(k)}(\bar{t})$ are linearly independent, then the osculating k-plane of γ is defined at $\bar{t}$ and it is the k-plane that passes through $\gamma(\bar{t})$ with direction vectors $\gamma'(\bar{t}), \gamma''(\bar{t}), \dots, \gamma^{(k)}(\bar{t})$.*

To prove the theorem we need some preparation.

Definition 2.4.6. Let $f\colon [a,b] \to \mathbb{R}^n$ be a vector valued function, $t_0, t_1, \dots . \in [a,b]$ are different numbers. The *higher order divided differences* or *difference quotients* are defined recursively

$$\begin{aligned} f_0(t_0) &:= f(t_0), \\ f_1(t_0,t_1) &:= \frac{f(t_1)-f(t_0)}{t_1-t_0}, \\ &\vdots \\ f_k(t_0,t_1,\dots,t_k) &:= \frac{f_{k-1}(t_1,\dots,t_k)-f_{k-1}(t_0,\dots,t_{k-1})}{t_k-t_0}. \end{aligned}$$

※

Exercise 2.4.7. Show that the k-th order divided difference is a symmetric function of the variables $t_0, \dots, t_k$ and has the following explicit form

$$f_k(t_0,t_1,\dots,t_k) = \sum_{i=0}^{k} f(t_i)\frac{1}{\omega'(t_i)},$$

where $\omega(t)$ is the polynomial $(t-t_0)(t-t_1)\dots.(t-t_k)$, hence

$$\omega'(t_i) = (t_i-t_0)(t_i-t_1)\dots(t_i-t_{i-1})(t_i-t_{i+1})\dots(t_i-t_k).$$

¤

Lemma 2.4.8. *If $f\colon [a,b] \to \mathbb{R}$ is a smooth function, then there exists a number $\xi \in (a,b)$ such that*

$$f_k(t_0,t_1,\dots,t_k) = \frac{f^{(k)}(\xi)}{k!}.$$

Proof. Let $P(t)$ be the polynomial of degree $\le k$ for which $f(t_i) = P(t_i)$ for $i = 0,1,\dots,k$. Such a polynomial exists and is unique. P is unique, since if Q is also a polynomial of degree $\le k$ such that $f(t_i) = P(t_i) = Q(t_i)$ for $i = 0,\dots,k$, then the polynomial $P-Q$ is of degree $\le k$, and $P-Q$ has $k+1$ roots, which is possible only in the case when $P-Q=0$, i.e., $P=Q$. We show the existence of P by an explicit construction. Set

$$P_i(t) = \frac{(t-t_0)(t-t_1)\dots(t-t_{i-1})(t-t_{i+1})\dots(t-t_k)}{(t_i-t_0)(t_i-t_1)\dots(t_i-t_{i-1})(t_i-t_{i+1})\dots(t_i-t_k)}.$$

Obviously, P_i is a polynomial of degree k such that $P_i(t_j) = \delta_{ij}$. (The *Kronecker δ symbol* denotes $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.) Thus, the polynomial

$$P(t) = \sum_{i=0}^{k} f(t_i)P_i(t)$$

is good for our purposes.

The difference $f - P$ has $k+1$ roots. Since by the mean value theorem the interval between two zeros of a smooth function contains an interior point at which the derivative vanishes, the derivative $f' - P'$ has at least k roots. Similarly, $f'' - P''$ has at least $k-1$ roots, etc. $f^{(k)} - P^{(k)}$ vanishes at a certain point $\xi \in [a, b]$. The k-th derivative of a polynomial of degree k is $k!$ times the coefficient of the highest power, from which

$$f^{(k)}(\xi) = P^{(k)}(\xi) = k! \sum_{i=0}^{k} f(t_i)\frac{1}{\omega'(t_i)} = k! f_k(t_0, t_1, \dots, t_k). \qquad \square$$

Corollary 2.4.9. *Since ξ can be chosen from the interval spanned by the points $t_0, \dots, t_k$, if these points tend to $\bar{t} \in [a, b]$, then ξ also tends to $\bar{t}$, consequently $f_k(t_0, t_1, \dots, t_k) = \dfrac{f^{(k)}(\xi)}{k!}$ tends to $\dfrac{f^{(k)}(\bar{t})}{k!}$.*

Corollary 2.4.10. *Applying the previous corollary to the components of a vector valued function $f\colon [a, b] \to \mathbb{R}^n$, we obtain that $f_k(t_0, t_1, \dots, t_k)$ tends to $\dfrac{f^{(k)}(\bar{t})}{k!}$ as $t_0, \dots, t_k$ tend to $\bar{t}$.*

Proof of Theorem 2.4.5. Let us recall that if $\mathbf{p}_0, \dots, \mathbf{p}_k$ are position vectors of $k+1$ points in $\mathbb{R}^n$, then the affine plane spanned by them consists of linear combinations the coefficients in which have sum equal to 1

$$A(\mathbf{p}_0, \dots, \mathbf{p}_k) = \{\alpha_0\mathbf{p}_0 + \dots + \alpha_k\mathbf{p}_k \mid \alpha_0 + \dots + \alpha_k = 1\}.$$

The direction vectors of this affine plane are linear combinations $\alpha_0\mathbf{p}_0 + \dots + \alpha_k\mathbf{p}_k$ such that $\alpha_0 + \dots + \alpha_k = 0$.

We claim that if $\gamma\colon [a, b] \to \mathbb{R}^n$ is a curve in $\mathbb{R}^n$, then $\gamma_k(t_0, \dots, t_k)$, $(k \geq 1)$ is a direction vector of the affine subspace spanned by the points $\gamma(t_0), \dots \gamma(t_k)$. To see this, it suffices to show that $\sum_{i=0}^{k} 1/\omega'(t_i) = 0$. Consider the function $f = 1$ and construct the polynomial P of degree $\leq k$ which coincides with f at $t_0, \dots, t_k$ using the general formulae. By the above proposition, there exists a number ξ such that

$$0 = f^{(k)}(\xi) = P^{(k)}(\xi) = \sum_{i=0}^{k} \frac{1}{\omega'(t_i)},$$

as we wanted to show.

This way, $\gamma(t_0)$ is a point and $\gamma_1(t_1, t_0), \dots, \gamma_k(t_k, \dots, t_0)$ are direction vectors of the affine subspace spanned by the points $\gamma(t_0), \dots, \gamma(t_k)$. If γ is smooth, then $\gamma(t_0)$ tends to $\gamma(\bar{t})$, $\gamma_1(t_1, t_0)$ tends to $\gamma'(\bar{t})$, and so on, $\gamma_k(t_k, \dots, t_0)$ tends to $\gamma^{(k)}(\bar{t})/k!$ as the points $t_0, \dots, t_k$ tend to $\bar{t} \in [a, b]$. Since by our assumption the first k derivatives of γ are linearly independent at $\bar{t}$, so are the vectors $\gamma_1(t_1, t_0), \dots, \gamma_k(t_k, \dots, t_0)$ if $t_0, \dots, t_k$ are in a sufficiently small neighborhood of $\bar{t}$, and in this case the k-plane $A(t_0, \dots, t_k)$ tends to the k-plane that passes through $\gamma(\bar{t})$ with direction vectors $\gamma'(\bar{t}), \gamma''(\bar{t})/2, \dots, \gamma^{(k)}(\bar{t})/k!$. □

Exercise 2.4.11. Prove by induction on k that $\gamma_k(t_0, \dots, t_k)$, $(k \geq 1)$ is a direction vector of the affine subspace spanned by the points $\gamma(t_0), \dots, \gamma(t_k)$. ¤

Definition 2.4.12. Let V be an n-dimensional vector space. A *flag* in V is a sequence of linear subspaces $\{\mathbf{0}\} = V_0 \subset V_1 \subset \dots \subset V_n = V$ such that $\dim V_i = i$. An *affine flag* is a sequence of affine subspaces $A_0 \subset A_1 \subset \dots \subset A_n = V$ such that $\dim A_i = i$. ※

Definition 2.4.13. A curve $\gamma \colon I \to \mathbb{R}^n$ is called a *curve of general type in* $\mathbb{R}^n$ if the first $n-1$ derivatives $\gamma'(t), \gamma''(t), \dots, \gamma^{(n-1)}(t)$ are linearly independent for all $t \in I$. ※

Definition 2.4.14. The *osculating flag* of a curve of general type at a given point is the affine flag consisting of the osculating k-planes for $k = 0, 1, \dots, n-1$ and the whole space. ※

2.5 Frenet Frames and Curvatures, the Fundamental Theorem of Curve Theory

Our plan is the following. A curve of general type in $\mathbb{R}^n$ is not contained in any affine subspace of dimension $k < n-1$ (prove this!), so we may pose the question how far it is from being contained in a k-plane. In other words, we want to measure the deviation of the curve from its osculating k-plane. One way to do this is that we measure how quickly the osculating flag rotates as we travel along the curve. Since the faster we travel along the curve the faster change we observe, it is natural to consider the speed of rotation of the osculating flag with respect to the unit speed parameterization of the curve. This will lead us to quantities that describe the way a curve is winding in space. These quantities will be called the curvatures of the curve. There is one question of technical character left: how can we measure the speed of rotation of an affine subspace? This problem can be solved by introducing an orthonormal basis at each point in such a way that the first k basis vectors span the osculating k-plane at the point in question, then measuring the speed of change of this basis.

Definition 2.5.1. A (smooth) *vector field along a curve* $\gamma\colon I \to \mathbb{R}^n$ is a smooth mapping $\mathbf{v}\colon I \to \mathbb{R}^n$. ※

Remark. There is no formal difference between a curve and a vector field along a curve. The difference is only in the interpretation. When we think of a map $\mathbf{v}\colon I \to \mathbb{R}^n$ as a vector field along the curve $\gamma\colon I \to \mathbb{R}^n$ we represent (depict) $\mathbf{v}(t)$ by a directed segment starting from $\gamma(t)$.

Definition 2.5.2. A *moving (orthonormal) frame along a curve* $\gamma\colon I \to \mathbb{R}^n$ is a collection of n vector fields $\mathbf{t}_1, \dots, \mathbf{t}_n$ along γ such that $\langle \mathbf{t}_i(t), \mathbf{t}_j(t)\rangle = \delta_{ij}$ for all $t \in I$. ※

There are many moving frames along a curve and most of them have nothing to do with the geometry of the curve. This is not the case for Frenet frames.

Definition 2.5.3. A moving frame $\mathbf{t}_1, \dots, \mathbf{t}_n$ along a curve γ is called a *Frenet frame* if for all k, $1 \le k \le n$, $\gamma^{(k)}(t)$ is contained in the linear span of $\mathbf{t}_1(t), \dots, \mathbf{t}_k(t)$. ※

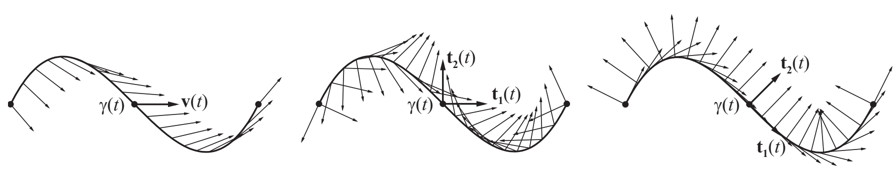

Figure 2.4: A vector field, a moving frame and the distinguished Frenet frame along a curve.

Exercise 2.5.4. Construct a curve which has no Frenet frame and one with infinitely many Frenet frames. Show that a curve of general type in $\mathbb{R}^n$ has exactly 2^n Frenet frames. ¤

According to the exercise, a Frenet frame along a curve of general type is almost unique. To select a distinguished Frenet frame from among all of them, we use orientation.

Definition 2.5.5. A Frenet frame $\mathbf{t}_1, \dots, \mathbf{t}_n$ of a curve γ of general type in $\mathbb{R}^n$ is called a *distinguished Frenet frame* if for all k, $1 \le k \le n-1$, the vectors $\mathbf{t}_1(t), \dots, \mathbf{t}_k(t)$ have the same orientation in their linear span as the vectors $\gamma'(t), \dots, \gamma^{(k)}(t)$, and the basis $\mathbf{t}_1(t), \dots, \mathbf{t}_n(t)$ is positively oriented with respect to the standard orientation of $\mathbb{R}^n$. ※

Proposition 2.5.6. *A curve of general type possesses a unique distinguished Frenet frame.*

Proof. We can determine the first $n-1$ vector fields of the distinguished Frenet frame by application of the *Gram–Schmidt orthogonalization process* pointwise to the first $n-1$ derivatives of γ. According to this recursive procedure, explained in Theorem 1.2.55, we start with setting

$$\mathbf{t}_1 = \frac{\gamma'}{\|\gamma'\|}.$$

If $\mathbf{t}_1, \dots, \mathbf{t}_{k-1}$ have already been defined, where $k \leq n-1$, then we compute the vector

$$\mathbf{f}_k = \gamma^{(k)} - (\langle \gamma^{(k)}, \mathbf{t}_1 \rangle \mathbf{t}_1 + \dots + \langle \gamma^{(k)}, \mathbf{t}_{k-1} \rangle \mathbf{t}_{k-1}),$$

and then set

$$\mathbf{t}_k = \frac{\mathbf{f}_k}{\|\mathbf{f}_k\|} = \frac{\gamma^{(k)} - (\langle \gamma^{(k)}, \mathbf{t}_1 \rangle \mathbf{t}_1 + \dots + \langle \gamma^{(k)}, \mathbf{t}_{k-1} \rangle \mathbf{t}_{k-1})}{\|\gamma^{(k)} - (\langle \gamma^{(k)}, \mathbf{t}_1 \rangle \mathbf{t}_1 + \dots + \langle \gamma^{(k)}, \mathbf{t}_{k-1} \rangle \mathbf{t}_{k-1})\|}.$$

To finish the proof, we have to show that given $n-1$ mutually orthogonal unit vectors $\mathbf{t}_1, \dots, \mathbf{t}_{n-1}$ in $\mathbb{R}^n$, there is a unique vector $\mathbf{t}_n$ for which the vectors $\mathbf{t}_1, \dots, \mathbf{t}_n$ form a positively oriented orthonormal basis of $\mathbb{R}^n$. The condition that a vector is perpendicular to $\mathbf{t}_1, \dots, \mathbf{t}_{n-1}$ is equivalent to a system of $n-1$ linearly independent linear equation, the solutions of which form a 1-dimensional linear subspace (a straight line). There are exactly two opposite unit vectors parallel to a given straight line, and exactly one of them will fulfill the orientation condition. (Replacing a vector of an ordered basis by its opposite changes the orientation.) □

Exercise 2.5.7. Show that if $\mathbf{e}_1, \dots, \mathbf{e}_n$ is the standard basis of $\mathbb{R}^n$ and

$$\mathbf{t}_i = \alpha_{i1}\mathbf{e}_1 + \dots + \alpha_{in}\mathbf{e}_n \quad \text{for} \quad i = 1, \dots, n-1,$$

then $\mathbf{t}_n$ can be obtained as the formal determinant of the matrix

$$\begin{pmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{(n-1)1} & \dots & \alpha_{(n-1)n} \\ \mathbf{e}_1 & \dots & \mathbf{e}_n \end{pmatrix}.$$

¤

Proposition 2.5.8. *Let γ be a curve of general type in $\mathbb{R}^n$. Denote by $\mathbf{t}_1, \dots, \mathbf{t}_n$ its distinguished Frenet frame and set $v = \|\gamma'\|$. Let the matrix $(\alpha_{ij})_{1 \leq i,j \leq n}$ be such that*

$$\frac{1}{v}\mathbf{t}_i' = \sum_{j=1}^{n} \alpha_{ij} \mathbf{t}_i. \tag{2.5}$$

Then

(i) $\alpha_{ij} = 0$ *provided that* $j > i+1$, *and*

(ii) *the matrix* $(\alpha_{ij})_{1\le i,j\le n}$ *is skew-symmetric, i.e.* $\alpha_{ij} = -\alpha_{ji}$.

Proof. (i) Since $\mathbf{t}_i$, $1 \le i \le n-1$, is a linear combination of the vectors $\gamma', \dots, \gamma^{(i)}$, $\mathbf{t}_i'$ is a linear combination of the vectors $\gamma', \dots, \gamma^{(i+1)}$. As the latter vectors are linear combinations of the vectors $\mathbf{t}_1, \dots, \mathbf{t}_{(i+1)}$, the first statement is proved.
(ii) Since $\langle \mathbf{t}_i, \mathbf{t}_j \rangle \equiv \delta_{ij}$ is a constant function, we get

$$\alpha_{ij} + \alpha_{ji} = \frac{1}{v}(\langle \mathbf{t}_i', \mathbf{t}_j \rangle + \langle \mathbf{t}_i, \mathbf{t}_j' \rangle) = 0$$

by differentiation. The proposition is proved. □

According to Propositions 2.5.8, only the entries $\alpha_{i,i+1} = -\alpha_{i+1,i}$ of the matrix (α_{ij}) may differ from zero. Setting

$$\kappa_1 = \alpha_{12},\ \kappa_2 = \alpha_{23},\ \dots, \kappa_{n-1} = \alpha_{n-1,n},$$

we see that equations (2.5) collapse to the following form

$$\begin{aligned}
\frac{1}{v}\mathbf{t}_1' &= \kappa_1 \mathbf{t}_2 \\
\frac{1}{v}\mathbf{t}_2' &= -\kappa_1 \mathbf{t}_1 + \kappa_2 \mathbf{t}_3 \\
&\ \ \vdots \\
\frac{1}{v}\mathbf{t}_{n-1}' &= -\kappa_{n-2} \mathbf{t}_{n-2} + \kappa_{n-1} \mathbf{t}_n \\
\frac{1}{v}\mathbf{t}_n' &= -\kappa_{n-1} \mathbf{t}_{n-1}.
\end{aligned}$$

These formulae are called the *Frenet formulae for a curve of general type in* $\mathbb{R}^n$. The functions $\kappa_1, \dots, \kappa_{n-1}$ are called the *curvature functions of the curve.*
We formulate two invariance theorems concerning the curvatures of a curve. They are intuitively clear and their proof is straightforward.

Proposition 2.5.9 (Invariance under isometries)**.** *Let* γ *be a curve of general type in* $\mathbb{R}^n$, $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ *be an isometry (distance preserving bijection). Then the curvature functions* $\kappa_1, \dots, \kappa_{n-2}$ *of the curves* γ *and* $\Phi \circ \gamma$ *are the same. The last curvatures* κ_{n-1} *of these curves coincide if* Φ *is orientation preserving and they differ (only) in sign if* Φ *is orientation reversing.*

Proposition 2.5.10 (Invariance under reparameterization)**.** *If $\tilde{\gamma}$ is a regular reparameterization of the curve γ i.e. $\tilde{\gamma} = \gamma \circ h$ for some smooth function h with property $h' > 0$ or $h' < 0$, then the curvature functions of $\tilde{\gamma}$ and γ are related to one another by $\tilde{\kappa}_i = \kappa_i \circ h$ for $1 \leq i \leq n-2$ and $\tilde{\kappa}_{n-1} = (\mathrm{sgn}(h'))^{\frac{n(n+1)}{2}} \kappa_{n-1} \circ h$.*

Exercise 2.5.11. Assume a curve γ of general type in $\mathbb{R}^n$ lies in $\mathbb{R}^{n-1} \subset \mathbb{R}^n$. Then we can compute the curvatures $\kappa_1, \ldots, \kappa_{n-1}$ of this curve considering γ a curve in $\mathbb{R}^n$ and also we may compute the curvatures $\tilde{\kappa}_1, \ldots, \tilde{\kappa}_{n-2}$ of this curve considering γ a curve in $\mathbb{R}^{n-1}$. What is the relationship between these two sets of numbers? ¤

Computation of the Curvature Functions

Our goal now is to express the curvatures of a curve γ of general type in $\mathbb{R}^n$ in terms of the derivatives of γ. For this purpose write the derivatives of γ as linear combinations of the distinguished Frenet frame. By the definition of the Frenet frame, the kth derivative must be the linear combination of the first k Frenet vector fields

$$\begin{aligned}
\gamma' &= a_1^1 \mathbf{t}_1, \\
\gamma'' &= a_2^1 \mathbf{t}_1 + a_2^2 \mathbf{t}_2, \\
&\ \,\vdots \\
\gamma^{(k)} &= a_k^1 \mathbf{t}_1 + \cdots + a_k^k \mathbf{t}_k, \\
&\ \,\vdots \\
\gamma^{(n)} &= a_n^1 \mathbf{t}_1 + a_n^2 \mathbf{t}_2 + \cdots + a_n^n \mathbf{t}_n.
\end{aligned} \tag{2.6}$$

The coefficients a_i^j can be expressed recursively by the speed length function $v = \|\gamma'\|$ and the curvature functions. To illustrate this, let us compute the first three decompositions assuming $n \geq 3$.
The speed vector is simply

$$\gamma' = v\mathbf{t}_1. \tag{2.7}$$

Differentiating and applying the first Frenet equation

$$\gamma'' = v'\mathbf{t}_1 + v\mathbf{t}_1' = v'\mathbf{t}_1 + v^2\kappa_1\mathbf{t}_2. \tag{2.8}$$

Differentiating again and using the first two Frenet equations we get

$$\begin{aligned}
\gamma''' &= v''\mathbf{t}_1 + v'\mathbf{t}_1' + (v^2\kappa_1)'\mathbf{t}_2 + v^2\kappa_1\mathbf{t}_2' \\
&= v''\mathbf{t}_1 + v'v\kappa_1\mathbf{t}_2 + (v^2\kappa_1)'\mathbf{t}_2 + v^2\kappa_1 v(\kappa_2\mathbf{t}_3 - \kappa_1\mathbf{t}_1) \\
&= (v'' - v^3\kappa_1^2)\mathbf{t}_1 + (v'v\kappa_1 + (v^2\kappa_1)')\mathbf{t}_2 + v^3\kappa_1\kappa_2\mathbf{t}_3.
\end{aligned} \tag{2.9}$$

In principle, iterating this method we can compute all the coefficients a_i^j but already the third derivative shows that the coefficients become complicated and do not show any pattern except for the last coefficients.

Proposition 2.5.12.

$$a_k^k = v^k \kappa_1 \cdots \kappa_{k-1} \text{ for } 1 \le k \le n.$$

Proof. We prove the equation by induction on k. The base cases are verified above for $k \le 3$. Assume now that the statement holds for $k-1 < n$. Then differentiating

$$\gamma^{(k-1)} = a_{k-1}^1 \mathbf{t}_1 + \cdots + a_{k-1}^{k-2} \mathbf{t}_{k-2} + a_{k-1}^{k-1} \mathbf{t}_{k-1}.$$

and applying the Frenet formulae we obtain

$$\begin{aligned}\gamma^{(k)} &= \sum_{j=1}^{k-1} ((a_{k-1}^j)' \mathbf{t}_j + a_{k-1}^j \mathbf{t}_j') \\ &= ((a_{k-1}^1)' \mathbf{t}_1 + a_{k-1}^1 v \kappa_1 \mathbf{t}_2) + \sum_{j=2}^{k-1} ((a_{k-1}^j)' \mathbf{t}_j + a_{k-1}^j v(\kappa_j \mathbf{t}_{j+1} - \kappa_{j-1} \mathbf{t}_{j-1})).\end{aligned}$$

We see that $\mathbf{t}_k$ appears only in the last summand, when $j = k-1$ and its coefficient is

$$a_k^k = a_{k-1}^{k-1} v \kappa_{k-1} = (v^{k-1} \kappa_1 \cdots \kappa_{k-2}) v \kappa_{k-1} = v^k \kappa_1 \cdots \kappa_{k-1}. \qquad \square$$

Proposition 2.5.13. *The curvature functions $\kappa_1, \dots, \kappa_{n-2}$ of a curve of general type in $\mathbb{R}^n$ are positive. (However, there is not any restriction on the sign of κ_{n-1}.)*

Proof. By the orientation condition on the distinguished Frenet frame, for $k \le n-1$, $\gamma', \dots, \gamma^{(k)}$ has the same orientation as $\mathbf{t}_1, \dots, \mathbf{t}_k$ in the linear space they span. According to the definition of "having the same orientation, this means that the determinant

$$\Delta_k = \det \begin{pmatrix} a_1^1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_i^1 & \dots & a_i^i & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_k^1 & \dots & a_k^i & \dots & a_k^k \end{pmatrix} = a_1^1 \cdots a_k^k \tag{2.10}$$

of the lower triangular matrix of coefficients is positive. Since the quotients of positive numbers is positive, we obtain also that $a_1^1 = \Delta_1$ and $a_k^k = \Delta_k / \Delta_{k-1}$

for $2 \leq k \leq n-1$ are positive. However, then the quotients $a_k^k/a_{k-1}^{k-1} = v\kappa_{k-1}$ are also positive for $2 \leq k \leq n-1$. Since $v = \|\gamma'\| > 0$,this means that all the curvatures $\kappa_1, \dots, \kappa_{n-2}$ are positive.
The above computation shows also that

$$\operatorname{sgn}(\kappa_{n-1}) = \operatorname{sgn}(a_n^n/a_{n-1}^{n-1}) = \operatorname{sgn}(a_n^n) = \operatorname{sgn}(\Delta_n/\Delta_{n-1}) = \operatorname{sgn}(\Delta_n).$$

Consequently, the sign of the last curvature coincides with the the sign of Δ_n, which is positive (or negative) if and only if $(\gamma', \dots, \gamma^{(n)})$ is a positively (or negatively) oriented basis of $\mathbb{R}^n$. It is 0 if and only if $\gamma', \dots, \gamma^{(n)}$ are linearly dependent. The remark at the end of the proposition follows from the Fundamental Theorem of Curve Theory, which will be proved below. This theorem that for any smooth function κ_{n-1} on an interval, there is a curve of general type in $\mathbb{R}^n$ the last curvature of which is κ_{n-1}. □

As a byproduct of the proof, we obtain the following expressions for the curvatures

$$\kappa_1 = \frac{a_2^2}{va_1^1} = \frac{\Delta_2/\Delta_1}{v\Delta_1} = \frac{\Delta_2}{v^3}$$

and

$$\kappa_k = \frac{a_{k+1}^{k+1}}{va_k^k} = \frac{\Delta_{k+1}/\Delta_k}{v\Delta_k/\Delta_{k-1}} = \frac{\Delta_{k+1}\Delta_{k-1}}{v\Delta_k^2}$$

for $2 \leq k \leq n-1$. Thus, to find formulae for the curvatures it is enough to express the determinants Δ_k or the diagonal elements a_k^k in terms of the derivatives of γ.
If we take the wedge product of the first k equations (2.6) and apply Proposition 1.2.31, we get

$$\gamma' \wedge \cdots \wedge \gamma^{(k)} = \Delta_k \mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k. \tag{2.11}$$

As $\mathbf{t}_1, (t) \dots, \mathbf{t}_n(t)$ an orthonormal basis for each parameter t,

$$\|\gamma' \wedge \cdots \wedge \gamma^{(k)}\| = |\Delta_k|. \tag{2.12}$$

Since the determinants Δ_k are positive for $k \leq n-2$, the absolute value can be omitted for all but the last k. The length of the k-vector $\gamma' \wedge \cdots \wedge \gamma^{(k)}$ can be expressed using (1.9) with the help of the Gram matrix $\mathcal{G}(\gamma', \dots, \gamma^{(k)})$. Thus for $k \leq n-1$,

$$\begin{aligned}\Delta_k = \|\gamma' \wedge \cdots \wedge \gamma^{(k)}\| &= \sqrt{\mathcal{G}(\gamma', \dots, \gamma^{(k)})} \\ &= \sqrt{\det \begin{pmatrix} \langle \gamma', \gamma' \rangle & \dots & \langle \gamma', \gamma^{(k)} \rangle \\ \vdots & \ddots & \vdots \\ \langle \gamma^{(k)}, \gamma' \rangle & \dots & \langle \gamma^{(k)}, \gamma^{(k)} \rangle \end{pmatrix}}\end{aligned}$$

We lost information on the sign of Δ_n when we took the absolute value of the sides of (2.11), so to obtain Δ_n together with its sign, return to equation (2.11) and consider it for $k = n$. As the Frenet frame assigns a positively oriented orthonormal basis to each parameter, denoting by $\mathbf{e}_1, \dots, \mathbf{e}_n$ the standard basis of $\mathbb{R}^n$

$$\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_n = \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n,$$

consequently

$$\gamma' \wedge \cdots \wedge \gamma^{(n)} = \Delta_n \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n.$$

By the last equation we can obtain Δ_n by writing the derivatives of γ as linear combinations of the standard basis and computing the determinant of the matrix of coefficients. If we decompose γ as a linear combination

$$\gamma = (\gamma_1, \dots, \gamma_n) = \gamma_1 \mathbf{e}_1 + \cdots + \gamma_n \mathbf{e}_n$$

of the standard basis, then

$$\gamma^{(k)} = (\gamma_1^{(k)}, \dots, \gamma_n^{(k)}) = \gamma_1^{(k)} \mathbf{e}_1 + \cdots + \gamma_n^{(k)} \mathbf{e}_n,$$

and therefore

$$\Delta_n = \det \begin{pmatrix} \gamma_1' & \dots & \gamma_n' \\ \gamma_1'' & \dots & \gamma_n'' \\ \vdots & \vdots & \vdots \\ \gamma_1^{(n)} & \dots & \gamma_n^{(n)} \end{pmatrix} = \det \begin{pmatrix} \gamma' \\ \gamma'' \\ \vdots \\ \gamma^{(n)} \end{pmatrix}$$

We summarize our computation in the following theorem.

Theorem 2.5.14. *Let $\gamma \colon I \to \mathbb{R}^n$ be a curve of general type in $\mathbb{R}^n$, $1 \le k \le n-1$. Then the curvature functions of γ can be computed by the equations*

$$\kappa_1 = \frac{\Delta_2}{v^3} \quad \text{and} \quad \kappa_k = \frac{\Delta_{k+1}\Delta_{k-1}}{v\Delta_k^2} \text{ for } k \ge 2,$$

where $v = \|\gamma'\|$, and the numbers Δ_k are given by

$$\Delta_k = \sqrt{\det \begin{pmatrix} \langle \gamma', \gamma' \rangle & \dots & \langle \gamma', \gamma^{(k)} \rangle \\ \vdots & \ddots & \vdots \\ \langle \gamma^{(k)}, \gamma' \rangle & \dots & \langle \gamma^{(k)}, \gamma^{(k)} \rangle \end{pmatrix}}$$

$$\text{for } k < n \text{ and } \Delta_n = \det \begin{pmatrix} \gamma' \\ \gamma'' \\ \vdots \\ \gamma^{(n)} \end{pmatrix}.$$

When we want to compute the curvatures of a curve directly, without computing the Frenet frame, the quickest method is probably the evaluation of the formulae in the theorem. However, if we need to compute the Frenet frame as well, and we compute them following the proof of Proposition 2.5.6, then curvatures pop up simply during the Gram–Schmidt orthogonalization process. Indeed, in the kth step of the procedure we compute the vector

$$\mathbf{f}_k = \gamma^{(k)} - \sum_{i=1}^{k-1} \langle \gamma^{(k)}, \mathbf{t}_i \rangle \mathbf{t}_i.$$

Taking the decompositions (2.6) into account, we see that

$$\langle \gamma^{(k)}, \mathbf{t}_i \rangle = a_k^i,$$

and

$$\mathbf{f}_k = \gamma^{(k)} - \sum_{i=1}^{k-1} a_k^i \mathbf{t}_i = a_k^k \mathbf{t}_k.$$

This way, the length of $\mathbf{f}_k$, which we also have to compute for evaluating the equation

$$\mathbf{t}_k = \frac{\mathbf{f}_k}{\|\mathbf{f}_k\|},$$

is equal to

$$\|\mathbf{f}_k\| = |a_k^k| = v^k |\kappa_1 \cdots \kappa_{k-1}|,$$

and therefore

$$|\kappa_k| = \frac{\|\mathbf{f}_{k+1}\|}{v\|\mathbf{f}_k\|}.$$

This formula allows us to compute also $|\kappa_{n-1}|$, but to determine the sign of κ_{n-1}, one needs further consideration of the orientation of the first n derivatives of γ.

Exercise 2.5.15. Compute the curvatures of the *moment curve* $\gamma(t) = (t, t^2, \ldots, t^n)$ at $t = 0$. ¤

Fundamental Theorem of Curve Theory

Theorem 2.5.16 (Fundamental Theorem of Curve Theory)**.** *Given $n-1$ positive smooth functions $v, \kappa_1, \ldots, \kappa_{n-2}$ and a smooth function κ_{n-1} on an interval I, there exists a curve $\gamma\colon I \to \mathbb{R}^n$ of general type in $\mathbb{R}^n$ such that $\|\gamma'\| = v$ and the curvatures of γ are the prescribed functions $\kappa_1, \ldots, \kappa_{n-1}$. This curve is unique up to orientation preserving isometries of the space, that is, if $\tilde{\gamma}$ is another curve with the same properties, then there is an orientation preserving isometry $\Phi \in \mathrm{Iso}_+(\mathbb{R}^n)$ such that $\tilde{\gamma} = \Phi \circ \gamma$.*

Proof. We can eliminate the freedom given by orientation preserving isometries if we fix a parameter $t_0 \in I$ and restrict our attention to curves $\gamma\colon I \to \mathbb{R}^n$ such that $\gamma(t_0) = \mathbf{0}$ and the Frenet basis of γ at t_0 coincides with the standard basis of $\mathbb{R}^n$. We are to show that within this family of curves, each collection of allowed curvature functions corresponds to a unique curve. The proof below describes a way to reconstruct the curve from its curvatures by solving some differential equations.

Let $\mathbf{T}$ denote the $n \times n$ matrix, whose rows are the Frenet vector fields of the unknown curve $\gamma\colon I \to \mathbb{R}^n$. If $C = (c_{ij})\colon I \to \mathbb{R}^{n\times n}$ is the matrix valued function the only non-zero entries of which are $c_{i,i+1} = -c_{i+1,i} = v\kappa_i$ for $1 \le i \le n-1$, then $\mathbf{T}$ must satisfy the *linear* differential equation

$$\mathbf{T}' = C \cdot \mathbf{T} \tag{2.13}$$

Thus, prescribing the Frenet basis at a given point $\mathbf{T}(t_0)$, we can obtain the whole moving Frenet frame as a unique solution of (2.13). For each $t \in I$, $\mathbf{T}(t)$ must be an *orthogonal matrix*, that is a matrix the rows of which form an orthonormal basis of $\mathbb{R}^n$. Orthogonality is equivalent to the equation $\mathbf{T}\cdot\mathbf{T}^T = \mathbf{T}^T\cdot\mathbf{T} = I_n$, where $\mathbf{T}^T$ is the transposition of $\mathbf{T}$, I_n is the $n\times n$ unit matrix. Furthermore, since the Frenet frame must be a positively oriented basis of $\mathbb{R}^n$, $\mathbf{T}(t)$ must have positive determinant for all $t \in I$.

Recall that orthogonal matrices with positive determinant are called *special orthogonal matrices*. The set (in fact group) of $n \times n$ special orthogonal matrices is usually denoted by $SO(n)$.

We claim that *if* $\mathbf{T}$ *is a solution of* (2.13) *and there is a number* t_0*, for which* $\mathbf{T}(t_0) \in SO(n)$*, then* $\mathbf{T}(t) \in SO(n)$ *for all* $t \in I$. This means that if we take care of the restrictions on T when we choose the initial matrix $\mathbf{T}_0$, then we do not have to worry about the other values of $\mathbf{T}$.

To show the statement, set $\mathbf{M} = \mathbf{T}^T \cdot \mathbf{T}$. Then

$$\mathbf{M}' = (\mathbf{T}')^T\cdot\mathbf{T} + \mathbf{T}^T\cdot\mathbf{T}' = (C\cdot\mathbf{T})^T\cdot\mathbf{T} + \mathbf{T}^T\cdot C\cdot\mathbf{T} = \mathbf{T}^T\cdot(C^T + C)\cdot\mathbf{T} = \mathbf{0}, \tag{2.14}$$

as C is skew symmetric. (2.14) shows that $\mathbf{M}$ is constant, therefore $\mathbf{M} \equiv \mathbf{M}(t_0) = I_n$. This proves that $\mathbf{T}(t)$ is an orthogonal matrix for all $t \in I$.

As for the determinant of $\mathbf{T}(t)$, observe that the determinant of an orthogonal matrix $A \in \mathbb{R}^{n\times n}$ satisfies

$$(\det A)^2 = \det A \cdot \det A^T = \det(A \cdot A^T) = \det I_n = 1,$$

hence $\det A = \pm 1$. Since $\det(\mathbf{T}(t))$ is a continuous function of t, which takes only the values ± 1, it must be constant. We have assumed that $\mathbf{T}(t_0) > 0$, thus $\mathbf{T}(t) = 1 > 0$ for all $t \in I$.

The derivative of γ is $v\mathbf{t}_1$, thus we can obtain γ from the Frenet frame by integrating $v\mathbf{t}_1$

$$\gamma(t) = \int_{t_0}^{t} v(\tau)\mathbf{t}_1(\tau)d\tau. \tag{2.15}$$

Solving (2.13) we get the only possible Frenet frame, then formula (2.15) defines the only curve which can satisfy the conditions. This proves uniqueness. To show the existence part of the theorem we have to check that the curve we have obtained has the prescribed curvature and length of speed functions. The lengthy but straightforward verification of this fact is left to the reader. □

Exercise 2.5.17. Plot the curve, called *astroid*, given by the parameterization $\gamma\colon [0, 2\pi] \to \mathbb{R}^2$, $\gamma(t) = (\cos^3(t), \sin^3(t))$. Is the astroid a smooth curve? Is it regular? Is it a curve of general type? If the answer is no for a property, characterize those arcs of the astroid which have the property. Compute the length of the astroid. Show that the segment of a tangent lying between the axis intercepts has the same length for all tangents. ¤

Exercise 2.5.18. Find the distinguished Frenet frame and the equation of the osculating 2-plane of the *elliptical helix* $t \mapsto (a\cos t, b\sin t, ct)$ at the point $(a, 0, 0)$ (a, b and c are given positive numbers). ¤

Exercise 2.5.19. Show that a curve of general type in $\mathbb{R}^n$ is contained in an $(n-1)$-dimensional affine subspace if and only if $\kappa_{n-1} \equiv 0$. ¤

Exercise 2.5.20. Prove that if a curve of general type in $\mathbb{R}^3$ has constant curvatures then it is either a circle or a helix. ¤

Exercise 2.5.21. Describe those curves of general type in $\mathbb{R}^n$ which have constant curvatures. ¤

2.6 Obtaining Curvatures as Limits of Some Geometrical Quantities

In this section, we compute the limit of some geometrical quantities depending on a collection of points on the curve as the points tend simultaneously to a given curve point. The limits will be expressed by the curvatures at the given point, thus, each of these results gives rise to a geometrical interpretation of curvatures.

The angle between osculating k-planes

Let $\gamma\colon I \to \mathbb{R}^n$ be a curve of general type in $\mathbb{R}^n$ *parameterized by arc length.* Fix a point $t_0 \in I$ and for $t_0 + s \in I$, denote by $\alpha_k(s)$ the angle between the osculating k-planes of γ at t_0 and at $t_0 + s$ for every $k < n$.

Proposition 2.6.1. *As the parameter $t_0 + s \in I$ tends to t_0, the quotient $\alpha_k(s)/|s|$ tends to $|\kappa_k(t_0)|$.*

Proof. The direction space of the osculating k-plane at $t \in I$ is spanned by the first k Frenet vectors $\mathbf{t}_1(t), \dots, \mathbf{t}_k(t)$. Thus, the angle $\alpha_k(s)$ is the angle between the unit k-vectors $(\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k)(t_0)$ and $(\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k)(t_0 + s)$. By elementary geometry, if an isosceles triangle has two equal sides of unit length enclosing an angle α, then the length of the third side is $2\sin(\alpha/2)$. Applying this theorem gives

$$\frac{\alpha_k(s)}{|s|} = \frac{\alpha_k(s)/2}{\sin(\alpha_k(s)/2)} \left\| \frac{(\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k)(t_0 + s) - (\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k)(t_0)}{s} \right\|.$$

When s tends to 0, the angle $\alpha_k(s)$ also tends to 0, so

$$\lim_{s \to 0} \frac{\alpha_k(s)/2}{\sin(\alpha_k(s)/2)} = 1.$$

As a corollary, we get

$$\begin{aligned} \lim_{s \to 0} \frac{\alpha_k(s)}{|s|} &= \lim_{s \to 0} \left\| \frac{(\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k)(t_0 + s) - (\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k)(t_0)}{s} \right\| \\ &= \|(\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k)'(t_0)\|. \end{aligned}$$

Computing the derivative with the Leibniz rule and applying the Frenet formulae

$$\begin{aligned} (\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k)' &= \sum_{i=1}^{k} \mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_i' \wedge \cdots \wedge \mathbf{t}_k \\ &= (\kappa_1 \mathbf{t}_2 \wedge \mathbf{t}_2 \wedge \cdots \wedge \mathbf{t}_k + \sum_{i=2}^{k} \mathbf{t}_1 \wedge \cdots \wedge (\kappa_i \mathbf{t}_{i+1} - \kappa_{i-1} \mathbf{t}_{i-1}) \wedge \cdots \wedge \mathbf{t}_k. \end{aligned}$$

As the wedge product of vectors vanishes if two of the vectors are equal, the only nonzero summand on the right-hand side occurs at $i = k$

$$(\mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_k)' = \mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_{k-1} \wedge (\kappa_k \mathbf{t}_{k+1}).$$

Combining all we have

$$\begin{aligned}\lim_{s\to 0}\frac{\alpha_k(s)}{|s|} &= \|(\mathbf{t}_1\wedge\cdots\wedge\mathbf{t}_k)'(t_0)\| \\ &= |\kappa_k(t_0)|\cdot\|(\mathbf{t}_1\wedge\cdots\wedge\mathbf{t}_{k-1}\wedge\mathbf{t}_{k+1})(t_0)\| = |\kappa_k(t_0)|,\end{aligned}$$

as it was to be proven. □

Volume and Dihedral Angles of Inscribed Simplices

Let $\gamma\colon I\to\mathbb{R}^n$ be a curve of general type in $\mathbb{R}^n$, $k\le n$. Fix a point $\bar t\in I$ and choose $k+1$ pairwise distinct parameters $t_0,\dots,t_k\in I$. Then the curve points $\gamma(t_0),\dots,\gamma(t_k)$ span a k-dimensional simplex, the k-dimensional volume of which we denote by $V_k(t_0,\dots,t_k)$. It is clear that if all the parameters $t_0,\dots,t_k$ tend to $\bar t$ simultaneously, then this volume tends to 0.

Proposition 2.6.2. *The limit of the volume $V_k(t_0,\dots,t_k)$ divided by the product of the edge lengths of the simplex spanned by the points $\gamma(t_0),\dots,\gamma(t_k)$ as all the parameters t_i tend to $\bar t$, is equal to*

$$\lim_{\forall i:t_i\to\bar t}\frac{V_k(t_0,\dots,t_k)}{\prod_{0\le i<j\le k}\|\gamma(t_j)-\gamma(t_i)\|}=\frac{1}{k!}\cdot\left(\prod_{m=1}^{k}\frac{1}{m!}\right)|\kappa_1^{k-1}(\bar t)\cdot\kappa_2^{k-2}(\bar t)\cdots\kappa_{k-1}(\bar t)|.$$

Proof. The volume of a simplex can be obtained from the wedge product of the edge vectors running out from one of the vertices

$$V_k(t_0,\dots,t_k)=\frac{1}{k!}\|(\gamma(t_1)-\gamma(t_0))\wedge\cdots\wedge(\gamma(t_k)-\gamma(t_0))\|.$$

We know that the mth order divided difference of γ has the following explicit form

$$\gamma_m(t_0,\dots,t_m)=\sum_{i=0}^{m}\frac{1}{\prod_{\substack{j=0\\ j\ne i}}^{m}(t_i-t_j)}\gamma(t_i).$$

During the proof of Theorem 2.4.5 we showed that right-hand side is a 0 weight linear combination of the points $\gamma(t_i)$. Consequently, equation

$$\gamma_m(t_0,\dots,t_m)=\sum_{i=1}^{m}\frac{1}{\prod_{\substack{j=0\\ j\ne i}}^{m}(t_i-t_j)}(\gamma(t_i)-\gamma(t_0)) \tag{2.16}$$

also holds. Consider equation (2.16) for $m=1,\dots,k$ and take the wedge product of these k equations. Applying Proposition 1.2.31 we get

$$\begin{aligned}&\gamma_1(t_0,t_1)\wedge\cdots\wedge\gamma_k(t_0,\dots,t_k)\\ &\qquad=\frac{1}{\prod_{m=1}^{k}\prod_{j=0}^{m-1}(t_m-t_j)}(\gamma(t_1)-\gamma(t_0))\wedge\cdots\wedge(\gamma(t_k)-\gamma(t_0)).\end{aligned} \tag{2.17}$$

Taking the length of both sides and rearranging the equation gives

$$\frac{V_k(t_0,\dots,t_k)}{\prod_{0\le i<j\le k}\|\gamma(t_j)-\gamma(t_i)\|}=\frac{1}{k!}\cdot\frac{\|\gamma_1(t_0,t_1)\wedge\cdots\wedge\gamma_k(t_0,\dots,t_k)\|}{\prod_{0\le i<j\le k}\|\gamma_1(t_i,t_j)\|}.$$

As we know the limits of the higher order divided differences as all the t_i's tend to $\bar{t}$, we can compute the limit of the right-hand side easily

$$\lim_{\forall i: t_i\to\bar{t}}\frac{V_k(t_0,\dots,t_k)}{\prod_{0\le i<j\le k}\|\gamma(t_j)-\gamma(t_i)\|}=\frac{1}{k!}\cdot\left(\prod_{m=1}^{k}\frac{1}{m!}\right)\cdot\frac{\|(\gamma'\wedge\cdots\wedge\gamma^{(k)})(\bar{t})\|}{\|\gamma'(\bar{t})\|^{(k+1)k/2}}.$$

Making use of (2.12), (2.10) and Proposition 2.5.12, we obtain

$$\|\gamma'\wedge\cdots\wedge\gamma^{(k)}\|=|\Delta_k|=|a_1^1\cdots a_k^k|=v^{(k+1)k/2}|\kappa_1^{k-1}\cdot\kappa_2^{k-2}\cdots\kappa_{k-1}|,$$

which completes the proof. □

By elementary geometry, if the area of a triangle is A and its side length are a, b, c, then the radius R of the circumcircle of the triangle is expressed by the formula

$$R=\frac{abc}{4A}.$$

Thus, our theorem for $k=2$ has the following corollary.

Corollary 2.6.3. *For a curve $\gamma\colon I\to\mathbb{R}^n$ of general type in $\mathbb{R}^n$, and for any three distinct points t_0, t_1, t_2 in I, denote by $R(t_0,t_1,t_2)\in[0,\infty]$ the circumradius of the triangle $\gamma(t_0)\gamma(t_1)\gamma(t_2)$. Then*

$$\lim_{t_0,t_1,t_2\to\bar{t}}R(t_0,t_1,t_2)=\frac{1}{|\kappa_1(\bar{t})|}.$$

Though for $k>2$ the limit in Proposition 2.6.2 involves several curvatures, it is possible to construct expressions of the volumes of faces of various dimensions of a simplex whose limit gives one single curvature. An example of such a formula is Darboux's formula for the second curvature.

Corollary 2.6.4 (Darboux)**.** *Using the notations of Proposition 2.6.2,*

$$\lim_{t_0,t_1,t_2,t_3\to\bar{t}}\frac{9}{2}\frac{V_3(t_0,t_1,t_2,t_3)}{\sqrt{V_2(t_1,t_2,t_3)V_2(t_0,t_2,t_3)V_2(t_0,t_1,t_3)V_2(t_0,t_1,t_2)}}=|\kappa_2(\bar{t})|.$$

Consider again a curve $\gamma\colon I\to\mathbb{R}^n$ of general type in $\mathbb{R}^n$, and let $2\le k\le n$ be an integer. Fix a point $\bar{t}\in I$ and if $k=n$, then suppose that $\kappa_{n-1}(\bar{t})\neq 0$. By Proposition 2.6.2, when the points $t_0,\dots,t_k\in I$ are sufficiently close to $\bar{t}$, the simplex S spanned by $\gamma(t_0),\dots,\gamma(t_k)$ will be non-degenerate. Consider

two facets of S, say the facet F spanned by the points $\gamma(t_0), \dots, \gamma(t_{k-1})$ and $\hat{F}$ spanned by $\gamma(t_1), \dots, \gamma(t_k)$. These two facets intersect along a $(k-2)$-dimensional face $L = F \cap \hat{F}$. Let $\alpha = \alpha(t_0, \dots, t_k)$ be the angle of the $(k-1)$-planes spanned by the facets F and $\hat{F}$.
We remark that if β is the angle between the outer unit normals of the facets F and $\hat{F}$, then the dihedral angle of the simplex at the $(k-2)$-dimensional face L is defined to be the complementary angle $\pi - \beta$. The angle α of the $(k-1)$-planes is the smaller of β and the dihedral angle $\phi - \beta$, that is, $\alpha = \min\{\beta, \pi - \beta\} \in [0, \pi/2]$.

Proposition 2.6.5. *If the distinct parameters $t_0, \dots, t_k \in I$ tend to $\bar{t}$ simultaneously, then*

$$\lim_{\forall i: t_i \to \bar{t}} \frac{\alpha(t_0, \dots, t_k)}{\|\gamma(t_k) - \gamma(t_0)\|} = \frac{\kappa_{k-1}(\bar{t})}{k}.$$

Proof. Denote by P the vertex $\gamma(t_k)$, by P' is orthogonal projection onto the $(k-1)$-plane containing F, and by P'' its orthogonal projection onto the $(k-2)$-plane spanned by L. The triangle $PP'P''\triangle$ is a right triangle with right angle at P' and angle α at P''. For this reason, $\sin(\alpha) = \frac{PP'}{PP''}$. PP' is the height of the original simplex corresponding to the facet F, PP'' is the height of the simplex $\hat{F}$ corresponding to its face L. Thus, we can compute the lengths of these segments from the volume formula for simplices:

$$PP' = k \cdot \frac{V_k(t_0, \dots, t_k)}{V_{k-1}(t_0, \dots, t_{k-1})}, \text{ and } PP'' = (k-1) \cdot \frac{V_{k-1}(t_1, \dots, t_k)}{V_{k-2}(t_1, \dots, t_{k-1})}.$$

For simplicity, denote by E_S, E_F, $E_{\hat{F}}$, and E_L the product of the edge lengths of the simplices S, F, $\hat{F}$, and L respectively. It is clear that $E_S \cdot E_L = E_F \cdot E_{\hat{F}} \cdot \|\gamma(t_k) - \gamma(t_0)\|$. Consequently

$$\begin{aligned} \frac{\sin(\alpha)}{\|\gamma(t_k) - \gamma(t_0)\|} &= \frac{PP' \cdot E_F \cdot E_{\hat{F}}}{PP'' \cdot E_S \cdot E_{\hat{L}}} \\ &= \frac{k}{k-1} \cdot \frac{\dfrac{V_k(t_0, \dots, t_k)}{E_S} \cdot \dfrac{V_{k-2}(t_1, \dots, t_{k-1})}{E_L}}{\dfrac{V_{k-1}(t_0, \dots, t_{k-1})}{E_F} \cdot \dfrac{V_{k-1}(t_1, \dots, t_k)}{E_{\hat{F}}}}. \end{aligned}$$

Computing the limit of the right-hand side applying Proposition 2.6.2 we obtain

$$\lim_{\forall i: t_i \to \bar{t}} \frac{\sin(\alpha(t_0,\dots,t_k))}{\|\gamma(t_k)-\gamma(t_0)\|} = \frac{k}{k-1} \frac{\left|\dfrac{\kappa_1^{k-1}(\bar{t})\cdots\kappa_{k-1}(\bar{t})}{1!2!\cdots k!\cdot k!}\right| \cdot \left|\dfrac{\kappa_1^{k-3}(\bar{t})\cdots\kappa_{k-3}(\bar{t})}{1!2!\cdots(k-2)!\cdot(k-2)!}\right|}{\left|\dfrac{\kappa_1^{k-2}(\bar{t})\cdots\kappa_{k-2}(\bar{t})}{1!2!\cdots(k-1)!\cdot(k-1)!}\right|^2} = \frac{\kappa_{k-1}(\bar{t})}{k}.$$

The above equation shows that α tends to 0 as all the t_i's tend to $\bar{t}$, therefore $\alpha/\sin(\alpha)$ tends to 1, so the equation implies the statement. □

2.7 Osculating Spheres

According to Corollary 2.6.3, the circumradius of a triangle $\gamma(t_0)\gamma(t_1)\gamma(t_2)\triangle$ inscribed in a curve of general type γ tends to $1/|\kappa_1(\bar{t})|$ as the parameters t_0, t_1, t_2 tend to $\bar{t}$. In this section we show that under some conditions not only the radii, but also the circles themselves have a limit position. In fact we prove such a statement for the circumspheres of inscribed simplices of any given dimension. Then we investigate how to compute centers and radii of the limit spheres.

The set $\mathrm{SPH}(\mathbb{R}^n)$ of all k-spheres has a natural topology induced from the embedding into $\mathrm{Gr}_1(\langle\langle V\rangle\rangle)$. This embedding was described in the subsection "Equations of Spheres" of Section 1.3.2. According to Exercise 1.4.23, convergence of a sequence of k-spheres can be split into three conditions. Namely, a sequence $S_1, S_2, \dots$ of k-spheres tends to a $k$sphere S if and only if tending with i to infinity the affine subspace $\mathrm{aff}[S_i]$ tends to the affine subspace $\mathrm{aff}[S]$, the center of S_i tends to the center of S and the radius of S_i tends to the radius of S.

Theorem 2.7.1. *Assume that for a smooth parameterized curve $\gamma\colon I \to \mathbb{R}^n$ and a parameter $\bar{t} \in I$, the first $(k+1)$ derivatives of γ at $\bar{t}$, $\gamma'(\bar{t}), \dots, \gamma^{(k+1)}(\bar{t})$ are linearly independent. Then $\bar{t}$ has a neighborhood U such that for any $k+2$ distinct parameters $t_0, \dots, t_{k+1}$ from $U \cap I$, the points $\gamma(t_0), \dots, \gamma(t_{k+1})$ are affinely independent, therefore define a unique k-sphere $S(t_0, \dots, t_{k+1})$. Furthermore, the limit of the k-sphere $S(t_0, \dots, t_{k+1})$ exists as the parameters $t_0, \dots, t_{k+1}$ tend to $\bar{t}$.*

Proof. The existence of the open neighborhood U was proven as part of Theorem 2.4.5.

To show convergence, consider the embedding $V \to \langle\langle V\rangle\rangle$, $\mathbf{v} \mapsto \check{\mathbf{v}}$ defined by equation (1.16). Then by Proposition 1.3.38, the equation of $S(t_0, \dots, t_{k+1})$

is

$$\breve{\mathbf{x}}\wedge\breve{\gamma}(t_0)\wedge\cdots\wedge\breve{\gamma}(t_{k+1}) = \breve{\mathbf{x}}\wedge\breve{\gamma}(t_0)\wedge(\breve{\gamma}(t_1)-\breve{\gamma}(t_0))\wedge\cdots\wedge(\breve{\gamma}(t_{k+1})-\breve{\gamma}(t_0)) = \mathbf{0}.$$

If we apply formula (2.17) for $\breve{\gamma}$, the equation of the sphere can be rewritten as

$$\breve{\mathbf{x}} \wedge \breve{\gamma}(t_0) \wedge \breve{\gamma}_1(t_0, t_1) \wedge \cdots \wedge \breve{\gamma}_{k+1}(t_0, \ldots, t_{k+1}) = \mathbf{0}.$$

As the limit of higher rth order divided difference of $\breve{\gamma}$ as all the parameters tend to $\bar{t}$ exists and equals $\breve{\gamma}^{(r)}(\bar{t})/r!$, we have

$$\begin{aligned}\lim_{\forall i: t_i \to \bar{t}} & \breve{\gamma}(t_0) \wedge \breve{\gamma}_1(t_0, t_1) \wedge \cdots \wedge \breve{\gamma}_{k+1}(t_0, \ldots, t_{k+1}) \\ &= \frac{1}{1! \cdots (k+1)!} \breve{\gamma}(\bar{t}) \wedge \breve{\gamma}'(\bar{t}) \wedge \cdots \wedge \breve{\gamma}^{(k+1)}(\bar{t}) \neq \mathbf{0}.\end{aligned}$$

Hence the limit of the k-sphere $S(t_0, \ldots, t_{k+1})$ exists as the parameters $t_0, \ldots,$ t_{k+1} tend to $\bar{t}$, and the equation of the limit sphere is

$$\breve{\mathbf{x}} \wedge \breve{\gamma}(\bar{t}) \wedge \breve{\gamma}'(\bar{t}) \wedge \cdots \wedge \breve{\gamma}^{(k+1)}(\bar{t}) = \mathbf{0}. \tag{2.18}$$

□

Definition 2.7.2. The limit of the k-spheres $S(t_0, \ldots, t_{k+1})$ as each parameter t_i tends to $\bar{t}$ is called the *osculating k-sphere of γ at $\bar{t}$.* ※

The equation of the osculating k-sphere at $\bar{t}$ is (2.18). The position of the center is not seen immediately from this equation, but it can be obtained by solving some linear equations (see Exercise 1.3.39), the coefficients in which are combinations of the dot products of the derivatives of γ. Thus, if γ is smooth, then the center $\mathbf{o}_k(\bar{t})$ of the osculating k-sphere at $\bar{t}$ is a smooth function of $\bar{t}$ on intervals, where the first $k+1$ derivatives are linearly independent.

Assume that $\gamma\colon I \to \mathbb{R}^n$ is a curve of general type in $\mathbb{R}^n$. For $k < n-1$, denote by $\mathbf{o}_k\colon I \to \mathbb{R}^n$ the center of the osculating k-sphere at t, as a smooth function of t, and by $R_k\colon I \to \mathbb{R}$ the radius function of the osculating k-sphere. We define $\mathbf{o}_{n-1}$ and R_{n-1} similarly, but these maps are defined only on the open subset J of I, where the last curvature function κ_{n-1} is not 0. Our goal now is to find some recursive formulae to compute these functions. We start with some simple geometrical facts.

Since the osculating spheres at t go through the curve point $\gamma(t)$,

$$R_k^2 = \|\mathbf{o}_k - \gamma\|^2. \tag{2.19}$$

Proposition 2.7.3. *The difference $\mathbf{o}_k - \mathbf{o}_{k-1}$ is parallel to $\mathbf{t}_{k+1}$.*

Proof. Let $\bar{t} \in I$ be a fixed point and take $k+2$ distinct points $t_0, \dots, t_{k+1}$ in I. For $m = k-1$ and k, denote by $S_m = S_m(t_0, \dots, t_{m+1})$ the m-sphere through the curve points $\gamma(t_0), \dots, \gamma(t_{m+1})$, and denote by $O_m = O_m(t_0, \dots, t_{m+1})$ its center. S_{k-1} is a subsphere of S_k, therefore, its center O_{k-1} is the orthogonal projection of the center of S_k onto the affine subspace spanned by S_{k-1}. This orthogonality must be preserved when we take the limit as all parameters tend to $\bar{t}$. So in the limit we obtain that $\mathbf{o}_k - \mathbf{o}_{k-1}$ must be orthogonal to the osculating k-plane. On the other hand, the difference vector $\mathbf{o}_k - \mathbf{o}_{k-1}$ is a direction vector of the osculating $(k+1)$-plane. However, among direction vectors of the osculating $(k+1)$-plane only vectors parallel to $\mathbf{t}_{k+1}$ are orthogonal to the osculating k-plane. □

Corollary 2.7.4. *There exists smooth functions $\sigma_k \colon I \to \mathbb{R}$ for $k \le n-1$ and $\sigma_n \colon J \to \mathbb{R}$ such that*

$$\mathbf{o}_k = \gamma + \sigma_1 \mathbf{t}_1 + \cdots + \sigma_{k+1} \mathbf{t}_{k+1} \quad \textit{for} \quad k = 0, \dots, n-1. \tag{2.20}$$

The corollary is proved by a trivial induction on k, starting from the base case $\mathbf{o}_0 = \gamma$. Using this form of $\mathbf{o}_k$, and equation (2.19) we can express the radius R_k with the functions σ_i as

$$R_k^2 = \sigma_1^2 + \cdots + \sigma_{k+1}^2 \tag{2.21}$$

Proposition 2.7.5. *The derivative $\mathbf{o}_k'$ is orthogonal to the osculating k-plane.*

Proof. Use the same notation as in the proof of the previous proposition. Keeping the parameters $t_0, \dots, t_{k+1}$ fixed for a moment, consider the function $F \colon I \to \mathbb{R}$ given by

$$F(t) = \|\gamma(t) - O_k(t_0, \dots, t_{k+1})\|^2 - R_k(t_0, \dots, t_{k+1})^2.$$

F is a smooth function on I vanishing at the $k+2$ distinct points $t_0, \dots, t_{k+1}$. Since by the Lagrange mean value theorem, the derivative of a smooth function on an interval has a root strictly between any two roots of the function, F' has at least $k+1$ roots inside the interval $T = [\min_{0 \le i \le (k+1)} t_i, \max_{0 \le i \le (k+1)} t_i]$. Going on inductively, the jth derivative $F^{(j)}$ has at least $k+2-j$ roots inside the interval T for $j \le k+1$. Choose a root $\xi_j \in T$ of $F^{(j)}$ for $j = 1, \dots, k+1$. Then

$$\begin{aligned} F^{(j)}(\xi_j) = {} & 2\langle \gamma^{(j)}(\xi_j), \gamma(\xi_j) - O_k(t_0, \dots, t_{k+1}) \rangle \\ & + \sum_{s=1}^{j-1} \binom{j}{s} \langle \gamma^{(s)}(\xi_j), \gamma^{(j-s)}(\xi_j) \rangle = 0. \end{aligned}$$

Take the limit of these equations as the parameters $t_0, \dots, t_{k+1}$ tend to $\bar{t}$. Then T shrinks onto $\bar{t}$ and therefore ξ_j tends to $\bar{t}$ as well, so we get

$$2\langle \gamma^{(j)}(\bar{t}), \gamma(\bar{t}) - \mathbf{o}_k(\bar{t})\rangle + \sum_{s=1}^{j-1} \binom{j}{s} \langle \gamma^{(s)}(\bar{t}), \gamma^{(j-s)}(\bar{t})\rangle = 0.$$

This equation is valid for any $\bar{t}$. Rearranging the equation we can bring it to a simpler form

$$\langle \gamma^{(j)}, \mathbf{o}_k\rangle = \frac{1}{2}(\|\gamma\|^2)^{(j)} \qquad \text{for } j = 1, \dots, k+1.$$

Assume that $j \le k$ end differentiate this equation. We obtain

$$\langle \gamma^{(j+1)}, \mathbf{o}_k\rangle + \langle \gamma^{(j)}, \mathbf{o}_k'\rangle = \frac{1}{2}(\|\gamma\|^2)^{(j+1)}.$$

On the other hand, since $j+1$ is still less than or equal to $k+1$, we have

$$\langle \gamma^{(j+1)}, \mathbf{o}_k\rangle = \frac{1}{2}(\|\gamma\|^2)^{(j+1)}.$$

The difference of the last two equations gives that $\mathbf{o}_k'$ is orthogonal to the first k derivatives of γ, consequently, it is perpendicular to the osculating k-plane. □

Proposition 2.7.6. *The coefficients σ_i can be computed recursively by the formulae*

$$\begin{aligned} \sigma_1 &= 0, \\ \sigma_2 &= \frac{1}{\kappa_1}, \\ \sigma_{k+1} &= \frac{1}{\kappa_k}\left(\frac{\sigma_k'}{v} + \sigma_{k-1}\kappa_{k-1}\right). \end{aligned}$$

Proof. A 0-sphere is a pair of points on a straight line, and the center is at the midpoint. The limit of the center of the 0-sphere $\{\gamma(t_0), \gamma(t_1)\}$ as $t_0, t_1 \to \bar{t}$ is $\gamma(\bar{t})$. Thus

$$\mathbf{o}_0 = \gamma + \sigma_1 \mathbf{t}_1 = \gamma,$$

showing $\sigma_1 = 0$.
Differentiating (2.20) for $k \ge 1$ we obtain

$$\mathbf{o}_k' = v\mathbf{t}_1 + \sum_{j=2}^{k+1}(\sigma_j'\mathbf{t}_j + v \cdot \sigma_j \cdot (\kappa_j \mathbf{t}_{j+1} - \kappa_{j-1}\mathbf{t}_{j-1})).$$

By the previous proposition, $\mathbf{o}_k'$ is orthogonal to $\mathbf{t}_k$, consequently, the coefficient of $\mathbf{t}_k$ on the right-hand side must vanish. When $k = 1$, the coefficient of $\mathbf{t}_1$ is

$$v - v\sigma_2\kappa_1 = 0,$$

from which $\sigma_2 = 1/\kappa_1$.
If $1 < k \leq n-1$ then the coefficient of $\mathbf{t}_k$ in the decomposition of $\mathbf{o}_k'$ is

$$\sigma_k' + v \cdot (\sigma_{k-1} \cdot \kappa_{k-1} - \sigma_{k+1} \cdot \kappa_k) = 0. \tag{2.22}$$

Solving this equation for σ_{k+1} we obtain the recursive formula of the proposition. □

Remark. Observe that equation (2.22) can be rearranged into a form resembling the Frenet equations

$$\frac{1}{v}\sigma_k' = \kappa_k\sigma_{k+1} - \kappa_{k-1}\sigma_k.$$

With the help of the proposition, we can compute the location of the centers of the osculating k-spheres recursively. The initial cases are

$$\begin{aligned}
\mathbf{o}_1 &= \gamma + \frac{1}{\kappa_1}\mathbf{t}_2, \\
\mathbf{o}_2 &= \gamma + \frac{1}{\kappa_1}\mathbf{t}_2 - \frac{\kappa_1'}{v \cdot \kappa_2 \cdot \kappa_1^2}\mathbf{t}_3 \\
\mathbf{o}_3 &= \gamma + \frac{1}{\kappa_1}\mathbf{t}_2 - \frac{\kappa_1'}{v \cdot \kappa_2 \cdot \kappa_1^2}\mathbf{t}_3 + \left(\frac{\kappa_2}{\kappa_1 \cdot \kappa_3} - \frac{1}{v \cdot \kappa_3} \cdot \left(\frac{\kappa_1'}{v \cdot \kappa_2 \cdot \kappa_1^2}\right)'\right)\mathbf{t}_4. \\
&\vdots
\end{aligned}$$

Proposition 2.7.7. *The radius functions R_k of the osculating k-spheres obey the following recursive rules*

$$\begin{aligned}
R_0^2 &= 0, \\
R_1^2 &= \frac{1}{\kappa_1^2}, \\
R_{k+1}^2 &= R_k^2 + \frac{R_k^2}{R_k^2 - R_{k-1}^2}\left(\frac{R_k'}{v \cdot \kappa_{k+1}}\right)^2.
\end{aligned}$$

Proof. The initial cases are obvious. For $1 \leq k \leq n-2$, differentiate equation (2.21). Omitting $\sigma_1 = 0$ and dividing by 2 we obtain

$$R_k R_k' = \sum_{j=2}^{k+1} \sigma_j \sigma_j'.$$

Express σ_j' from the recursive equation (2.22). This provides

$$\begin{aligned}R_kR_k' &= \sum_{j=2}^{k+1} v\sigma_j(\sigma_{j+1}\kappa_j - \sigma_{j-1}\kappa_{j-1})\\ &= \sum_{j=2}^{k+1} v\sigma_{j+1}\sigma_j\kappa_j - v\sigma_j\sigma_{j-1}\kappa_{j-1} = v\sigma_{k+2}\sigma_{k+1}\kappa_{k+1}.\end{aligned}$$

If we square this equation and use $\sigma_{s+1}^2 = R_s^2 - R_{s-1}^2$, then we get

$$(R_kR_k')^2 = v^2(R_{k+1}^2 - R_k^2)(R_k^2 - R_{k-1}^2)\kappa_{k+1}^2.$$

Solving this equation for R_{k+1}^2 results the recursive formula of the proposition. □

2.8 Plane Curves

This section we deduce some facts on plane curves from the general theory of curves. A plane curve $\gamma\colon I \to \mathbb{R}^2$ is given by two coordinate functions.

$$\gamma(t) = (x(t), y(t)) \qquad t \in I.$$

The curve γ is of general type if the vector γ' is a linearly independent "system of vectors". Since a single vector is linearly independent if and only if it is non-zero, curves of general type in the plane are the same as regular curves. From this point on we assume that γ is regular.
The Frenet vector fields $\mathbf{t}_1, \mathbf{t}_2$ are denoted by $\mathbf{t}$ and $\mathbf{n}$ in classical differential geometry and they are called the *(unit) tangent* and the *(unit) normal vector fields* of the curve. There is only one curvature function of a plane curve $\kappa = \kappa_1$. The Frenet formulae have the form

$$\begin{aligned}\mathbf{t}' &= v\kappa\mathbf{n},\\ \mathbf{n}' &= -v\kappa\mathbf{t},\end{aligned}$$

where $v = |\gamma'|$.
Let us find explicit formulae for $\mathbf{t}$, $\mathbf{n}$ and κ. Obviously,

$$\mathbf{t} = \frac{1}{v}(x', y') = \frac{1}{\sqrt{x'^2 + y'^2}}(x', y').$$

The normal vector $\mathbf{n}$ is the last vector of the Frenet basis so it is determined by the condition that $(\mathbf{t}, \mathbf{n})$ is a positively oriented orthonormal basis, that is, in our case, $\mathbf{n}$ is obtained from $\mathbf{t}$ by a 90 degree rotation in positive direction.

The right angled rotation in the positive direction takes the vector (a, b) to the vector $(-b, a)$, thus

$$\mathbf{n} = \frac{1}{v}(-y', x') = \frac{1}{\sqrt{x'^2 + y'^2}}(-y', x').$$

To express κ, let us start from the equation

$$\gamma' = v\mathbf{t}.$$

Differentiating and using the first Frenet formula,

$$\gamma'' = v'\mathbf{t} + v\mathbf{t}' = v'\mathbf{t} + v^2\kappa\mathbf{n}.$$

Taking dot product with $\mathbf{n}$ and using $\langle \mathbf{t}, \mathbf{n} \rangle = 0$ we get

$$\langle \gamma'', \mathbf{n} \rangle = v^2\kappa,$$

which gives

$$\kappa = \frac{\langle \gamma'', \mathbf{n} \rangle}{v^2} = \frac{-x''y' + y''x'}{v^3} = \frac{\det\begin{pmatrix} x' & y' \\ x'' & y'' \end{pmatrix}}{(x'^2 + y'^2)^{3/2}},$$

in accordance with Theorem 2.5.14.

2.8.1 Evolute, Involute, Parallel Curves

Since a 1-sphere in a plane is simply a circle, osculating 1-spheres of curve are called its *osculating circles*. The osculating circle of a regular plane curve γ exists at a point t if and only if $\kappa(t) \neq 0$. Then the center $\mathbf{o}_1(t)$ of the radius $R_1(t)$ of the osculating circle are given by

$$\mathbf{o}_1(t) = \gamma(t) + \frac{1}{\kappa(t)}\mathbf{n}(t), \qquad R_1(t) = \frac{1}{|\kappa(t)|}.$$

Definition 2.8.1. The center of the osculating circle is called the *center of curvature*, the radius of the osculating circle is called the *radius of curvature* of the given curve at the given point. ※

The center of the osculating circle can be obtained also as the limit of the intersection point of two normal lines of γ.

Theorem 2.8.2. *Assume that $\gamma\colon I \to \mathbb{R}^2$ is a smooth regular plane curve, $\bar{t} \in I$ is such that $\kappa(\bar{t}) \neq 0$. Denote by $M(t_0, t_1)$ the intersection point of the normals of γ taken at t_0 and $t_1 \neq t_0$. Then $M(t_0, t_1)$ exists if t_0 and t_1 are sufficiently close to $\bar{t}$ and*

$$\lim_{t_0, t_1 \to \bar{t}} M(t_0, t_1) = \mathbf{o}_1(\bar{t}).$$

Proof. The intersection point of the normals exist if the corresponding tangents are not parallel. Since $\mathbf{t}'(t) = \kappa(t)\mathbf{n}(t) \neq \mathbf{0}$, the unit tangent $\mathbf{t}$ rotates with nonzero angular speed at $\bar{t}$, hence no two tangents are parallel in a small neighborhood of $\bar{t}$.
Assume t_0, t_1 are close to $\bar{t}$ and write $M(t_0, t_1)$ as

$$M(t_0, t_1) = \gamma(t_0) + m(t_0, t_1)\mathbf{n}(t_0)$$

using the fact that $M(t_0, t_1)$ is on the normal at t_0. Since $M(t_0, t_1)$ is also on the normal at t_1,

$$\langle M(t_0, t_1) - \gamma(t_1), \mathbf{t}(t_1)\rangle = \langle \gamma(t_0) + m(t_0, t_1)\mathbf{n}(t_0) - \gamma(t_1), \mathbf{t}(t_1)\rangle = 0.$$

Solving this equation for $m(t_0, t_1)$ gives

$$m(t_0, t_1) = \frac{\langle \gamma(t_1) - \gamma(t_0), \mathbf{t}(t_1)\rangle}{\langle \mathbf{n}(t_0), \mathbf{t}(t_1)\rangle} = \frac{\left\langle \dfrac{\gamma(t_1) - \gamma(t_0)}{t_1 - t_0}, \mathbf{t}(t_1)\right\rangle}{\left\langle \dfrac{\mathbf{n}(t_0) - \mathbf{n}(t_1)}{t_1 - t_0}, \mathbf{t}(t_1)\right\rangle}.$$

It is easy to calculate the limit of the right-hand side

$$\lim_{t_0, t_1 \to \bar{t}} m(t_0, t_1) = \frac{\langle \gamma'(\bar{t}), \mathbf{t}(\bar{t})\rangle}{-\langle \mathbf{n}'(\bar{t}), \mathbf{t}(\bar{t})\rangle} = \frac{\langle v(\bar{t})\mathbf{t}(\bar{t}), \mathbf{t}(\bar{t})\rangle}{\langle v(\bar{t})\kappa(\bar{t})\mathbf{t}(\bar{t}), \mathbf{t}(\bar{t})\rangle} = \frac{1}{\kappa(\bar{t})}.$$

In conclusion,

$$\lim_{t_0, t_1 \to \bar{t}} M(t_0, t_1) = \lim_{t_0, t_1 \to \bar{t}} \gamma(t_0) + m(t_0, t_1)\mathbf{n}(t_0) = \gamma(\bar{t}) + \frac{1}{\kappa(\bar{t})}\mathbf{n}(\bar{t}) = \mathbf{o}_1(\bar{t}). \quad \square$$

Exercise 2.8.3. Let us illuminate a curve by rays of light, parallel to the normal line of γ at $\gamma(\bar{t})$. Denote by $F(t_1, t_2)$ the intersection point of the rays reflected from $\gamma(t_1)$ and $\gamma(t_2)$. Show that $\lim_{t_1, t_2 \to \bar{t}} F(t_1, t_2) \to \gamma(\bar{t}) + \dfrac{1}{2\kappa(\bar{t})}\mathbf{t}_2(\bar{t})$. ¤

Definition 2.8.4. The locus of the centers of curvature of a curve is the *evolute* of the curve. The evolute is defined for arcs along which the curvature is not zero. ※

If $\gamma\colon I \to \mathbb{R}^2$ is a curve with a nowhere zero curvature function κ and unit normal vector field $\mathbf{n}$, then the evolute can be parameterized by the mapping $\mathbf{o}_1\colon I \to \mathbb{R}^2$, $\mathbf{o}_1 = \gamma + (1/\kappa)\mathbf{n}$.

Exercise 2.8.5. Show that the evolute of the ellipse $\gamma(t) = (a\cos t, b\sin t)$ is the "affine astroid" $\mathbf{o}_1(t) = \left(\frac{a^2-b^2}{a}\cos^3 t, \frac{b^2-a^2}{b}\sin^3 t\right)$. ¤

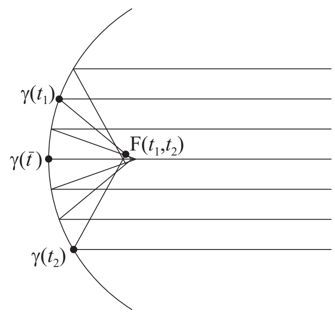

Figure 2.5: Parallel rays reflected from a curve crossing one of the rays orthogonally.

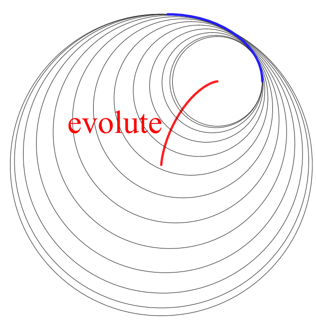

Figure 2.6: Osculating circles and the evolute of a curve

The evolute of a curve was introduced by Christiaan Huygens (1629-1695). He used the fact that the evolute of a cycloid is a congruent cycloid to modify the pendulum used in pendulum clocks to increase the accuracy of the clock. He could also be led to the notion of the evolute in connection with his investigations on the propagation of wave fronts. If we generate a curvilinear wave on the surface of calm water (e.g. we drop a wire into it), the wave starts moving. Geometrically, the shape of the wave front at consecutive moments of time is described by the parallel curves of the original curve.

Definition 2.8.6. Let γ be a regular plane curve with normal vector field $\mathbf{n}$. A *parallel curve of* γ is a curve of the form $\gamma_d = \gamma + d\mathbf{n}$, where $d \in \mathbb{R}$ is a fixed real number. ※

Experiments on wave fronts show that even if the initial curve is smooth and regular, singularities may appear on its parallel curves. As the distance d varies, singularities on γ_d are usually born and disappear in pairs.

Definition 2.8.7. Let γ be a smooth parameterized curve, t a point of its parameter interval. We say that $\gamma(t)$ (or t) is a *singular point*(or *singular parameter*) of the curve γ if $\gamma'(t) = 0$. ※

Exercise 2.8.8. Study the singularities of the parallel curves of an ellipse. ¤

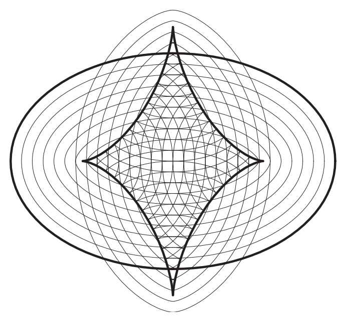

Figure 2.7: Parallel curves of an ellipse and the evolute swept out by the singular points of them.

Proposition 2.8.9. *Singular points of the parallel curves of a regular curve γ trace out the evolute of γ.*

Proof. Since $\gamma_d' = \gamma' + d\mathbf{n}' = v\mathbf{t} - vd\kappa\mathbf{t} = (1 - d\kappa)v\mathbf{t}$, singular parameters are characterized by $1 - d\kappa(t) = 0$. Then the corresponding singular points $\gamma_d(t) = \gamma(t) + (1/\kappa(t))\mathbf{n}(t)$ lie on the evolute of the curve. It is also easy to show that any evolute point is a singular point of a suitable parallel curve. □

Proposition 2.8.10. *Let $\gamma\colon [a, b] \to \mathbb{R}^2$ be a regular plane curve with monotone non-vanishing curvature κ, and evolute $\mathbf{o}_1 = \gamma + (1/\kappa)\mathbf{n}$.*

- *If $\kappa'(t) \neq 0$, then $\mathbf{o}_1$ is regular at t and its tangent line at t coincides with the normal line of γ at t.*
- *The length of the arc of the evolute between $\mathbf{o}_1(t_1)$ and $\mathbf{o}_1(t_2)$, $t_1 < t_2$, is the difference of the radii of curvatures $|1/\kappa(t_1) - 1/\kappa(t_2)|$.*

Proof. The speed vector of $\mathbf{o}_1$ is

$$\mathbf{o}_1{}' = \gamma' + (1/\kappa)'\mathbf{n} + (1/\kappa)\mathbf{n}' = v\mathbf{t} + (1/\kappa)'\mathbf{n} - (1/\kappa)v\kappa\mathbf{t} = (1/\kappa)'\mathbf{n}.$$

This equation shows that if $\kappa'(t) \neq 0$, then $\mathbf{o}_1$ is regular at t and its tangent at t is parallel to the normal $\mathbf{n}(t)$ of γ. This proves the first part of the proposition. As for the length of the evolute, it is equal to the integral

$$\int_{t_1}^{t_2} |\mathbf{o}_1{}'(\tau)|d\tau = \int_{t_1}^{t_2} |(1/\kappa)'(\tau)\mathbf{n}(\tau)|d\tau = \int_{t_1}^{t_2} |(1/\kappa)'(\tau)|d\tau$$
$$= \left|\int_{t_1}^{t_2} (1/\kappa)'(\tau)d\tau\right| = |1/\kappa(t_1) - 1/\kappa(t_2)|.$$

At the third equality we used that $1/\kappa$ is monotone, thus either $(1/\kappa)' \geq 0$ or $(1/\kappa)' \leq 0$ everywhere. □

The above proposition gives a method to reconstruct the curve γ from its evolute. Suppose for simplicity that $\kappa > 0$ and $\kappa' \geq 0$. Take a thread of length $1/\kappa(a)$ and fix one of its ends to $\mathbf{o}_1(a)$. Then pulling the other end of the thread from the initial position $\gamma(a)$ wrap the thread tightly onto the curve $\mathbf{o}_1$. By Proposition 2.8.10, the moving end of the thread will draw γ.

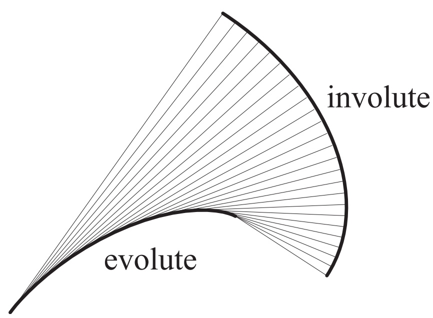

Figure 2.8: Relation of the evolute and involute constructions.

The thread construction gives involutes of a curve.

Definition 2.8.11. Let $\gamma: I \to \mathbb{R}^2$ be a regular curve with unit tangent vector field $\mathbf{t}$, $a \in I$. An *involute* of the curve γ is a curve $\hat{\gamma}$ of the form $\hat{\gamma} = \gamma + (l - s)\mathbf{t}$, where l is a given real number, $s(t) = \int_a^t |\gamma'(\tau)|d\tau$ is the signed length of the arc of γ between $\gamma(a)$ and $\gamma(t)$. ※

A curve has many involutes corresponding to the different choices of the initial point $a \in I$ and the length l of the thread.

Corollary 2.8.12. *A curve satisfying the conditions of Proposition 2.8.10 is an involute of its evolute.*

Proposition 2.8.13. *Let γ be a unit speed curve with $\kappa > 0$, $\hat{\gamma}(s) = \gamma(s) + (l - s)\mathbf{t}(s)$ an involute of it such that l is greater than the length of γ. Then the evolute of $\hat{\gamma}$ is γ.*

Proof. We have

$$\hat{\gamma}'(s) = \mathbf{t}(s) - \mathbf{t}(s) + (l-s)\kappa(s)\mathbf{n}(s) = (l-s)\kappa(s)\mathbf{n}(s),$$
$$\hat{\gamma}''(s) = [(l-s)\kappa(s)]'\mathbf{n}(s) - (l-s)\kappa^2(s)\mathbf{t}(s).$$

The first equation implies that the Frenet frame $\hat{\mathbf{t}}, \hat{\mathbf{n}}$ of $\hat{\gamma}$ is related to that of γ by $\hat{\mathbf{t}} = \mathbf{n}$, $\hat{\mathbf{n}} = -\mathbf{t}$. Computing the curvature $\hat{\kappa}$ of $\hat{\gamma}$,

$$\hat{\kappa}(s) = \frac{\langle \hat{\gamma}''(s), \hat{\mathbf{n}}(s)\rangle}{|\hat{\gamma}'(s)|^3} = \frac{(l-s)^2\kappa^3(s)}{(l-s)^3\kappa^3(s)} = \frac{1}{(l-s)}.$$

Thus, the evolute of $\hat{\gamma}$ is $\hat{\gamma} + (1/\hat{\kappa})\hat{\mathbf{n}} = \gamma + (l-s)\mathbf{t} - (l-s)\mathbf{t} = \gamma$. □

We formulate some further results on the evolute and involute as exercises.

Exercise 2.8.14. (a) Suppose that the regular curves γ_1 and γ_2 have regular evolutes. Show that γ_1 and γ_2 are parallel if and only if their evolutes are the same.
(b) Show that if two involutes of a regular curve are regular, then they are parallel. ¤

Exercise 2.8.15. The curve *cardioid* is the trajectory of a peripheral point of a circle rolling about a fixed circle of the same radius.
– Find a smooth parameterization of the cardioid.
– Compute its length.
– Show that its evolute is also a cardioid. ¤

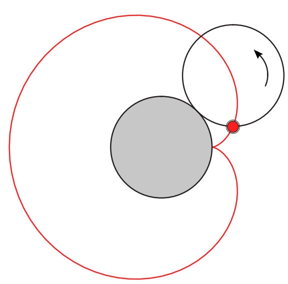

Figure 2.9: Generation of the cardioid.

Exercise 2.8.16. The *chain curve* is the graph of the hyperbolic cosine function

$$\cosh(x) = \frac{e^x + e^{-x}}{2}.$$

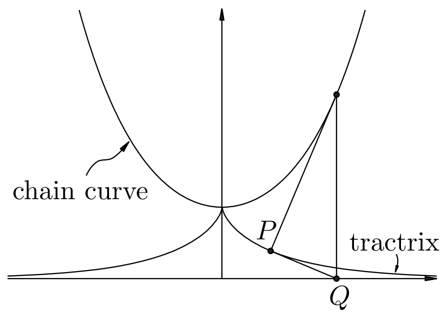

Figure 2.10: Construction of the tractrix as an involute of the chain curve.

– Determine the involute of the chain curve through the point $(0, 1)$. (This curve is called "tractrix".)
– Let the tangent of the tractrix at P intersect the x-axis at Q. Show that the segment PQ has unit length. ¤

Exercise 2.8.17. Let γ be a regular plane curve for which the curvature function and its derivative are positive. Show that for any $t_1 < t_2$ from the parameter domain of γ the osculating circle of γ at $\gamma(t_1)$ contains the osculating circle at $\gamma(t_2)$. ¤

2.8.2 The Rotation Number Theorem

Let $\gamma\colon [a, b] \to \mathbb{R}^2$ be a *unit speed* curve. The direction angle $\alpha(s)$ of the tangent $\mathbf{t}(s)$ is determined only up to an integer multiple of 2π, however one can see easily the existence of a *differentiable* function $\alpha\colon [a, b] \to \mathbb{R}$ such that $\alpha(s)$ is a direction angle of $\mathbf{t}(s)$ for all $s \in [a, b]$ (see Proposition 2.8.27 below). Then

$$\mathbf{t}(s) = (\cos\alpha(s), \sin\alpha(s)).$$

Differentiating with respect to s gives $\mathbf{t}' = \alpha'(-\sin\alpha, \cos\alpha)$. If we compare this equality with Frenet equations, we see immediately that

$$\kappa = \alpha',$$

i.e. the curvature is the derivative of the direction angle (angular speed) of the tangent vector with respect to the arc length.

Definition 2.8.18. The *total curvature* of a curve is the integral of its curvature function with respect to arc length

$$\int_a^b \kappa(s)ds.$$ ※

By the relation $\kappa = \alpha'$, the total curvature of a curve is $\alpha(b) - \alpha(a)$, thus it measures the rotation made by the tangent vector during the motion along the curve.

Definition 2.8.19. A curve $\gamma\colon [a,b] \to \mathbb{R}^n$ is a *smooth closed curve*, if there exists a smooth mapping $\tilde{\gamma}\colon \mathbb{R} \to \mathbb{R}^n$ such that $\tilde{\gamma}\big|_{[a,b]} = \gamma$, and $\tilde{\gamma}$ is periodic with period $b-a$, that is $\tilde{\gamma}(t+b-a) \equiv \tilde{\gamma}(t)$. ※

In other words, a smooth curve is smoothly closed if it returns to its initial point and the periodic repetition of the same motion crosses the initial point smoothly. Since the tangent vector of a smooth closed planar curve is the same at the endpoints $\gamma(a)$ and $\gamma(b)$, the direction angles $\alpha(b)$ and $\alpha(a)$ differ in an integer multiple of 2π.

Definition 2.8.20. The integer $(\alpha(b) - \alpha(a))/(2\pi)$ is called the *rotation number* of the smooth closed planar curve γ. ※

Exercise 2.8.21. Construct a smooth closed curve with an arbitrarily given rotation number $k \in \mathbb{Z}$. ¤

Solving the exercise we can observe that all our efforts to construct a curve with rotation number neither 1 nor -1 having no self-intersection are in vain. The reason for this is the famous "Umlaufsatz" (rotation number theorem).

Definition 2.8.22. A curve $\gamma\colon [a,b] \to \mathbb{R}^n$ is a *simple closed curve* if it is closed and $\gamma(t) = \gamma(t')$ holds if and only if $t = t'$ or $\{t, t'\} = \{a, b\}$. ※

Theorem 2.8.23 (Umlaufsatz)**.** *The rotation number of a planar smooth simple closed curve in the plane is equal to ± 1, or equivalently, the total curvature of such a curve is $\pm 2\pi$.*

The proof of the theorem will use some concepts borrowed from topology.

Definition 2.8.24. Let S^1 denote the unit circle $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\}$. The mapping $\pi\colon \mathbb{R} \to S^1$, $\pi(t) = (\cos t, \sin t)$ is called the *universal covering map of S^1*. Given a continuous mapping $\phi\colon X \to S^1$ from a topological space X into S^1, we say that ϕ has a *lift* to $\mathbb{R}$ if there exists a continuous mapping $\bar{\phi}\colon X \to \mathbb{R}$ such that $\phi = \pi \circ \bar{\phi}$. ※

Exercise 2.8.25. Construct a continuous mapping from a topological space into the circle which has no lift to $\mathbb{R}$. ¤

Lemma 2.8.26. *Suppose that the image of the mapping $\Phi\colon X \to S^1$ does not cover the point $(\cos\alpha, \sin\alpha) \in S^1$ and that the restriction ϕ of Φ onto a subspace $Y \subset X$ has got a lift $\bar{\phi}$ such that $\bar{\phi}(Y)$ is contained in an interval of the form $(\alpha + 2k\pi, \alpha + 2(k+1)\pi)$, $k \in \mathbb{Z}$. Then $\bar{\phi}$ can be extended to a lift $\bar{\Phi}$ of Φ. If furthermore Y is non-empty, X is path connected (i.e. any two points of X can be connected by a continuous curve lying in X) then the lift $\bar{\Phi}$ is unique, and maps X into the interval $(\alpha + 2k\pi, \alpha + 2(k+1)\pi)$.*

Proof. The restriction of π onto $(\alpha+2k\pi, \alpha+2(k+1)\pi)$ is a homeomorphism between $(\alpha+2k\pi, \alpha+2(k+1)\pi)$ and $S^1 \setminus \{(\cos\alpha, \sin\alpha)\}$. Thus $\bar{\Phi}$ can be defined as

$$\bar{\Phi} = (\pi|_{(\alpha+2k\pi,\alpha+2(k+1)\pi)})^{-1} \circ \Phi.$$

If X is path connected, then so is its image under a continuous lift $\bar{\Phi}$. Consequently $\bar{\Phi}(X)$ must be contained in one of the intervals $(\alpha+2k\pi, \alpha+2(k+1)\pi)$. If Y is non-empty then k is uniquely determined, hence $\bar{\Phi}$ must have the form $(\pi|_{(\alpha+2k\pi,\alpha+2(k+1)\pi)})^{-1} \circ \Phi$. □

Proposition 2.8.27. *Any continuous mapping $\phi\colon [a,b] \to S^1$ from an interval into the circle has got a lift.*

Proof. Choose a partition $a = t_0 < t_1 < \dots < t_k = b$ of the interval $[a,b]$ fine enough to insure that the restriction of ϕ onto $[t_i, t_{i+1}]$ does not cover the whole circle. This is possible because of the uniform continuity of ϕ. Then using Lemma 2.8.26 we can define a lift $\bar{\phi}$ of ϕ step by step, extending $\bar{\phi}$ recursively to $[a, t_i]$, $i = 1, 2, \dots, k$. □

Remark. If ϕ is differentiable, then so is its lift.

Proposition 2.8.28. *Any continuous mapping $\phi\colon T \to S^1$ from a rectangle $T = [a,b] \times [c,d]$ into the circle has got a lift to $\mathbb{R}$.*

Proof. Divide the rectangle into $n \times n$ small rectangles so that the image of any of the small rectangles does not cover the circle. Then we may define a lift $\bar{\phi}$ of ϕ recursively, applying at each step the lemma above. $\bar{\phi}$ can be defined first on the small rectangles of the first row going from left to right then on the rectangles of the second row, etc.

1	2	3	...	n
n+1	n+2	n+3	...	2n

□

Proof of the "Umlaufsatz". Let us choose a point $\mathbf{p}$ on the regular simple closed curve γ in such a way that the curve is lying on one side of the tangent at $\mathbf{p}$ and parameterize the curve by arc length starting from $\mathbf{p}$. Denoting this parameterization also by $\gamma\colon [0,l] \to \mathbb{R}^2$, we define a mapping $\phi\colon [0,l]\times[0,l] \to S^1$ by

$$\phi(t_1,t_2) = \begin{cases} -\gamma'(0), & \text{if } \{t_1,t_2\} = \{l,0\}, \\ \dfrac{\gamma(t_1)-\gamma(t_2)}{|\gamma(t_1)-\gamma(t_2)|}, & \text{if } t_1 > t_2 \text{ and } (t_1,t_2) \neq (l,0), \\ \gamma'(t_1), & \text{if } t_1 = t_2, \\ \dfrac{\gamma(t_2)-\gamma(t_1)}{|\gamma(t_2)-\gamma(t_1)|}, & \text{if } t_2 > t_1 \text{ and } (t_1,t_2) \neq (0,l). \end{cases}$$

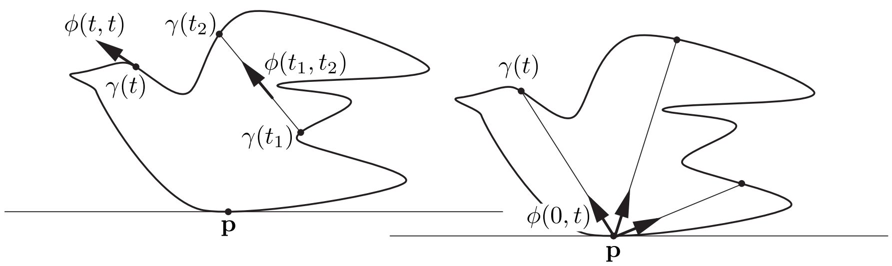

Figure 2.11: The geometrical meaning of ϕ.

It is easy to see that ϕ is continuous, so it has a continuous lift $\bar{\phi}$. If the function $\alpha\colon [0,l] \to \mathbb{R}$ is defined by $\alpha(t) = \bar{\phi}(t,t)$, then $\alpha(t)$ is a direction angle of the speed vector $\mathbf{t}(t)$ of γ, thus $\alpha(l) - \alpha(0)$ is 2π times the rotation number of γ. Consider the functions $\xi(t) = \bar{\phi}(0,t)$ and $\vartheta(t) = \bar{\phi}(t,l)$. $\xi(t)$ is a direction angle of the unit vector $(\gamma(t)-\mathbf{p})/|\gamma(t)-\mathbf{p}|$, $\vartheta(t)$ is a direction angle of its opposite. Thus ξ and ϑ differ only in a constant of the form $(2k+1)\pi$. Since the vectors $(\gamma(t) - \mathbf{p})/|\gamma(t) - \mathbf{p}|$ point in a half-plane bounded by the tangent of γ at $\mathbf{p}$, the image of ξ is contained in an open interval of length 2π (see Lemma 2.8.26). Thus $\xi(l)-\xi(0)$, which has obviously the form $(2m+1)\pi$ for some $m \in \mathbb{Z}$, must be equal to $\pm\pi$. Hence, we conclude that

$$\begin{aligned}\alpha(l) - \alpha(0) &= \bar{\phi}(l,l) - \bar{\phi}(0,0) = (\bar{\phi}(l,l) - \bar{\phi}(0,l)) + (\bar{\phi}(0,l) - \bar{\phi}(0,0))\\ &= \vartheta(l) - \vartheta(0) + \xi(l) - \xi(0) = 2(\xi(l) - \xi(0)) = \pm 2\pi. \qquad \square\end{aligned}$$

Remark. With some more work but using essentially the same idea, one can generalize the "Umlaufsatz" for piecewise smooth closed simple curves. The generalization says that for a simple closed polygon with smooth curvilinear edges, the sum of oriented external angles plus the sum of the total curvatures of the edges equals $\pm 2\pi$.

Exercise 2.8.29. Let γ be a simple regular closed curve of length l with curvature function κ. Choose a real number d such that $1 \geq \kappa \cdot d$. How long is the parallel curve $\gamma_d = \gamma + d\mathbf{n}$? ¤

2.8.3 Convex Curves

Definition 2.8.30. A simple closed plane curve γ is *convex*, if for any point $\mathbf{p} = \gamma(\bar{t})$, the curve lies on one side of the tangent line of γ at $\mathbf{p}$. In other words the function $\langle \gamma(t) - \mathbf{p}, \mathbf{n}(\bar{t}) \rangle$ must be ≥ 0 or ≤ 0 for all t. ※

Exercise 2.8.31. Show that a simple closed curve is convex if and only if every arc of the curve lies on one side of the straight line through the endpoints of the arc. ¤

Convex curves can be characterized with the help of the curvature function.

Theorem 2.8.32. *A simple closed curve is convex if and only if* $\kappa \geq 0$ *or* $\kappa \leq 0$ *everywhere along the curve.*

Proof. Assume first that γ is a naturally parameterized convex curve. Let $\alpha(t)$ be a continuous direction angle for the tangent $\mathbf{t}(t)$. As we know, $\alpha' = \kappa$, thus, it suffices to show that α is a weakly monotone function. This follows if we show that if α takes the same value at two different parameters t_1, t_2, then α is constant on the interval $[t_1, t_2]$. The rotation number of a simple curve is ± 1, hence the image of $\mathbf{t}$ covers the whole unit circle. As a consequence, we can find a point at which

$$\mathbf{t}(t_3) = -\mathbf{t}(t_1) = -\mathbf{t}(t_2).$$

If the tangent lines at t_1, t_2, t_3 were different, then one of them would be between the others and this tangent would have points of the curve on both sides. This contradicts convexity, hence two of these tangents say the tangents at $\mathbf{p} = \gamma(t_i)$ and $\mathbf{q} = \gamma(t_j)$ coincide.

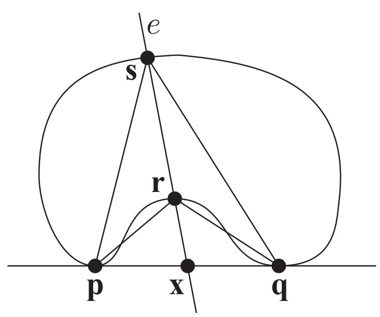

Figure 2.12

We claim that the segment $\overline{\mathbf{pq}}$ is an arc of γ. It is enough to prove that this segment is in the image of γ. Assume to the contrary that a point $\mathbf{x} \in \overline{\mathbf{pq}}$ is not covered by γ. Drawing a line $e \neq \mathbf{pq}$ through $\mathbf{x}$, we can find at least two intersection points $\mathbf{r}$ and $\mathbf{s}$ of e and the curve, since e separates $\mathbf{p}$ and $\mathbf{q}$ and γ has two essentially disjoint arcs connecting $\mathbf{p}$ to $\mathbf{q}$. Since $\mathbf{pq}$ is a tangent of γ, the points $\mathbf{r}$ and $\mathbf{s}$ must lie on the same side of it. As a consequence, we get that one of the triangles $\mathbf{pqr}$ and $\mathbf{pqs}$, say the first one is inside the other. However, this leads to a contradiction, since for such a configuration the tangent through $\mathbf{r}$ necessarily separates two vertices of the triangle $\mathbf{pqs}$, which lie on the curve.

If γ is defined on the interval $[a, b]$, then $\gamma(a) = \gamma(b)$ is either on the segment $\overline{\mathbf{pq}}$ or not. The first case is not possible, because then α would be constant on the intervals $[a, t_1]$ and $[t_2, b]$, yielding

$$\alpha(a) = \alpha(t_1) = \alpha(t_2) = \alpha(b)$$

and

$$\text{rotation number } = (\alpha(b) - \alpha(a))/2\pi = 0.$$

In the second case α is constant on the interval $[t_1, t_2]$, as we wanted to show. Now we prove the converse. Assume that γ is a simple closed curve with $\kappa \geq 0$ everywhere and assume to the contrary that γ is not convex (the case $\kappa \leq 0$ can be treated analogously). Then we can find a point $\mathbf{p} = \gamma(t_1)$, such that the tangent at $\mathbf{p}$ has curve points on both of its sides. Let us find on each side a curve point, say $\mathbf{q} = \gamma(t_2)$ and $\mathbf{r} = \gamma(t_3)$ respectively, lying at maximal distance from the tangent at $\mathbf{p}$. Then the tangents at $\mathbf{p}, \mathbf{q}$ and $\mathbf{r}$ are different and parallel. Since the unit tangent vectors $\mathbf{t}(t_i)$, $i = 1, 2, 3$ have parallel directions, two of them, say $\mathbf{t}(t_i)$ and $\mathbf{t}(t_j)$ must be equal. The points $\mathbf{a} = \gamma(t_i)$ and $\mathbf{b} = \gamma(t_j)$ divide the curve into two arcs. Denoting by K_1 and K_2 the total curvatures of these arcs, we deduce that these total curvatures have the form $K_1 = 2k_1\pi, K_2 = 2k_2\pi$, where $k_1, k_2 \in \mathbf{Z}$, since the unit tangents at the ends of the arcs are equal. On the other hand, we have $k_1 + k_2 = 1$ by the Umlaufsatz and $k_1 \geq 0$, $k_2 \geq 0$ by the assumption $\kappa \geq 0$. This is possible only if one of the total curvatures K_1 or K_2 is equal to zero. Since $\kappa \geq 0$, this means that $\kappa = 0$ along one of the arcs between $\mathbf{a}$ and $\mathbf{b}$. But then this arc would be a straight line segment, implying that the tangents at $\mathbf{a}$ and $\mathbf{b}$ coincide. The contradiction proves the theorem. □

2.8.4 The Four Vertex Theorem

Definition 2.8.33. A point $\gamma(t)$ of a regular plane curve γ is called a *vertex* if $\kappa'(t) = 0$. ※

Vertices of a curve correspond to the singular points of the evolute. By compactness, the curvature function of a closed curve attains its maximum and minimum somewhere, hence every closed curve has at least two vertices.

Exercise 2.8.34. Find a parameterization of *Bernoulli's lemniscate*

$$\{P \in \mathbb{R}^2 \mid \overline{PA} \cdot \overline{PB} = 1/4(\overline{AB})^2\},$$

where $A \neq B$ are given points in the plane and show that it is a closed curve with exactly two vertices. Determine the rotation number of the lemniscate. ¤

Exercise 2.8.35. Find the points on the ellipse $\gamma(t) = (a\cos t, b\sin t)$ at which the curvature is minimal or maximal $(a > b > 0)$. ¤

Theorem 2.8.36 (Four Vertex Theorem)**.** *A convex closed curve has at least 4 vertices.*

This result is sharp. For example, an ellipse has exactly four vertices.

Proof. Local maxima and minima of the curvature function yield vertices. One can always find a local minimum on an arc bounded by two local maxima, hence if we have two local maxima or minima of the curvature then we must have at least four vertices. Thus we have to exclude the case when the curvature function has one absolute maximum at A and one absolute minimum at B, and strictly monotone on the arcs bounded by A and B. In such a case, choose a coordinate system with origin at A and x-axis AB.
The arcs of the curve bounded by A and B do not cut the x-axis at points other than A and B. Indeed, if there were a further intersection point C, then the curve would be split into three arcs by A,B and C in such a way that on each arc we could find a point at which the tangent to the curve is parallel to the straight line ABC. If the three tangents at these points were different then the one in the middle would cut the curve apart contradicting convexity, if two of the tangents coincided then we could find a straight line segment contained in the curve yielding an infinite number of vertices.
If the two arcs bounded by A and B lied on the same side of AB then the line AB would be a common tangent of the curve at A and B. In this case the segment $\overline{AB}$ would be contained in the curve, yielding an infinite number of vertices as before.
We conclude that for a suitable orientation of the y-axis, $y(t)\kappa'(t) \geq 0$ for every $t \in [a, b]$, where $\gamma(t) = (x(t), y(t))$, $t \in [a, b]$ is a unit speed parameterization of the curve. Hence we get

$$\int_a^b y(t)\kappa'(t)dt > 0.$$

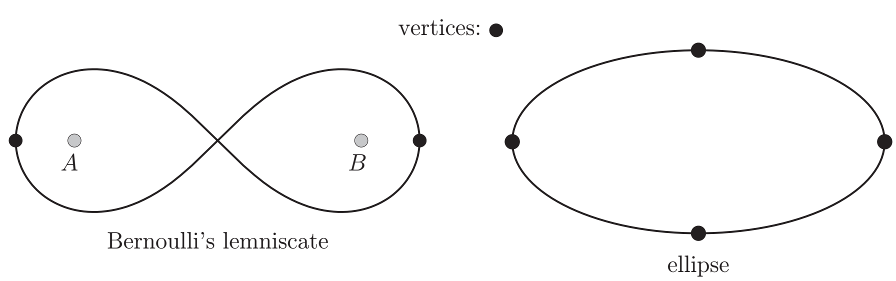

Figure 2.13: Vertices of Bernoulli's lemniscate and that of an ellipse.

Integrating by parts,

$$\int_a^b y(t)\kappa'(t)dt = [y(t)\kappa(t)]_a^b - \int_a^b y'(t)\kappa(t)dt = -\int_a^b y'(t)\kappa(t)dt.$$

The unit tangent vector field of the curve is $\mathbf{t} = (x', y')$, the unit normal vector field is $\mathbf{n} = (-y', x')$, hence by the first Frenet formula,

$$x'' = -\kappa y'.$$

Integrating,

$$-\int_a^b y'(t)\kappa(t)dt = \int_a^b x''(t)dt = [x'(t)]_a^b = 0.$$

This is a contradiction since a positive number can not be equal to 0. □

2.9 Curves in $\mathbb{R}^3$

A 3-dimensional curve is a curve of general type if its first two derivatives are not parallel. From now on we shall suppose that the curves under consideration are all of general type.

The distinguished Frenet frame vector fields $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_3$ of a 3-dimensional curve are denoted in classical differential geometry by $\mathbf{t}$, $\mathbf{n}$ and $\mathbf{b}$ and they are called the *(unit) tangent*, the *principal normal* and the *binormal* vector fields of the curve respectively. These vector fields define a coordinate system at each point of the curve. The coordinate planes of this coordinate system are given the following names. We are already familiar with the plane that goes through a given curve point and spanned by the directions of the tangent and the principal normal. It is the *osculating plane* of the curve. The plane that is spanned by the principal normal and the binormal is the plane that contains all straight lines orthogonally intersecting the curve at the given point.

For this obvious reason, this plane is called the *normal plane* of the curve. The third coordinate plane, that is the plane spanned by the tangent and the binormal directions is the *rectifying plane* of the curve. The reason for this naming will become clear later. As we know from the general theory, a 3-dimensional curve of general type has two curvature functions κ_1, which is always positive and κ_2, which may have any sign. The first curvature κ_1 is denoted by the classics simply by κ and it is referred to as the *curvature* of the curve while the second curvature κ_2 is called the *torsion* of the curve and is denoted by τ.

Using the classical notation, Frenet's formulae for a space curve can be written as follows.

$$\begin{aligned}\mathbf{t}' &= v\kappa\mathbf{n}\\ \mathbf{n}' &= -v\kappa\mathbf{t} + v\tau\mathbf{b}\\ \mathbf{b}' &= -v\tau\mathbf{n}.\end{aligned}$$

Let us find explicit formulae for the computation of these vectors and curvatures in an economical way. Assume that $\gamma\colon [a,b] \to \mathbb{R}^3$ is a curve of general type. The unit tangent vector field $\mathbf{t}$ can be obtained by normalizing the speed vector γ'

$$\mathbf{t} = \frac{\gamma'}{\|\gamma'\|}.$$

To obtain the principal normal $\mathbf{n}$ we can use the general method based on the Gram–Schmidt orthogonalization process

$$\mathbf{n} = \frac{\gamma'' - \langle \gamma'', \mathbf{t}\rangle\mathbf{t}}{\|\gamma'' - \langle \gamma'', \mathbf{t}\rangle\mathbf{t}\|} = \frac{\|\gamma'\|^2\gamma'' - \langle \gamma'', \gamma'\rangle\gamma'}{\big\|\|\gamma'\|^2\gamma'' - \langle \gamma'', \gamma'\rangle\gamma'\big\|},$$

and after this we can compute the binormal as a cross product of $\mathbf{t}$ and $\mathbf{n}$

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}.$$

In certain cases, it can be convenient to calculate the binormal first. The binormal vector is the unit normal vector of the osculating plane, for which $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ is positively oriented. The osculating plane is spanned by the first two derivatives γ', γ'' of γ, furthermore the pair $(\mathbf{t}, \mathbf{n})$ defines the same orientation of the osculating plane as the pair (γ', γ''), thus the basis $(\gamma', \gamma'', \mathbf{b})$ is positively oriented. Hence,

$$\mathbf{b} = \frac{\gamma' \times \gamma''}{\|\gamma' \times \gamma''\|}.$$

Having computed $\mathbf{b}$, $\mathbf{n}$ can be obtained as

$$\mathbf{n} = \mathbf{b} \times \mathbf{t} = \frac{(\gamma' \times \gamma'') \times \gamma'}{\|\gamma' \times \gamma'')\| \cdot \|\gamma'\|}.$$

The formulae obtained for the curvature and the torsion in Theorem 2.5.14 can be rewritten using standard vector operations. The first curvature is

$$\kappa = \frac{\|\gamma' \wedge \gamma''\|}{v^3} = \frac{\|\gamma' \times \gamma''\|}{v^3},$$

while the torsion is

$$\tau = \frac{\det\begin{pmatrix}\gamma' \\ \gamma'' \\ \gamma'''\end{pmatrix}\|\gamma'\|}{v\|\gamma' \wedge \gamma''\|^2} = \frac{\det\begin{pmatrix}\gamma' \\ \gamma'' \\ \gamma'''\end{pmatrix}}{\|\gamma' \times \gamma''\|^2} = \frac{\langle \gamma' \times \gamma'', \gamma''' \rangle}{\|\gamma' \times \gamma''\|^2}.$$

Recall that the numerator of this fraction is the determinant of the matrix the rows of which are $\gamma', \gamma'', \gamma'''$ and geometrically it is the signed volume of the parallelepiped spanned by $\gamma', \gamma'', \gamma'''$, where the sign is positive if and only if $(\gamma', \gamma'', \gamma''')$ is a positively oriented basis.

2.9.1 Orthogonal Projections onto Planes Spanned by Frenet Vectors

Study the shape of the orthogonal projections of a curve onto the planes spanned by the vectors of the distinguished Frenet frame. For simplicity, suppose that the curve γ is parameterized by arc length and examine the curve around $\gamma(0)$. Since $v \equiv 1$, equations (2.7)–(2.9) can be reduced to the form

$$\begin{aligned}
\gamma' &= \mathbf{t} \\
\gamma'' &= \mathbf{t}' = \kappa \mathbf{n} \\
\gamma''' &= \kappa' \mathbf{n} + \kappa \mathbf{n}' = \kappa' \mathbf{n} + \kappa(-\kappa \mathbf{t} + \tau \mathbf{b}) = -\kappa^2 \mathbf{t} + \kappa' \mathbf{n} + \kappa\tau \mathbf{b}
\end{aligned}$$

We can approximate the curve γ around $\gamma(0)$ by its Taylor expansion.

$$\gamma(t) = \gamma(0) + \gamma'(0)t + \frac{\gamma''(0)}{2}t^2 + \frac{\gamma'''(0)}{6}t^3 + o(t^3).$$

Recall that the "little oh" notation $o(t^3)$ is used in the following sense. If f, g, and h are functions defined around 0, then we write

$$f(t) = g(t) + o(h(t))$$

if $\dfrac{f(t) - g(t)}{h(t)}$ tends to zero as t tends to 0. Expressing the derivatives of γ with the help of Frenet vectors we get

$$\begin{aligned}
\gamma(t) - \gamma(0) &= \\
&= \left(t - \kappa^2(0)\frac{t^3}{6}\right)\mathbf{t}(0) + \left(\kappa(0)\frac{t^2}{2} + \kappa'(0)\frac{t^3}{6}\right)\mathbf{n}(0) \\
&+ \left(\kappa(0)\tau(0)\frac{t^3}{6}\right)\mathbf{b}(0) + o(t^3).
\end{aligned}$$

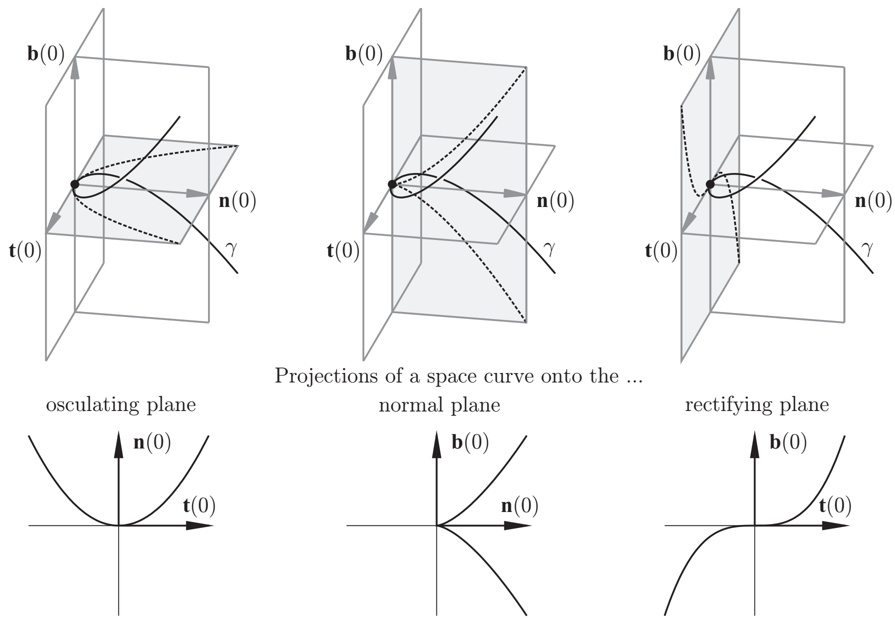

Figure 2.14

Looking at this expansion we conclude that the projection of the curve on the osculating plane is well approximated by the parabola $t\mathbf{t}(0)+(\kappa(0)t^2/2)\mathbf{n}(0)$. Observe that the curvature of this parabola at $t=0$ is $\kappa(0)$.
The projection onto the normal plane has approximately the same shape as the semicubical parabola $(\kappa(0)t^2/2)\mathbf{n}(0) + (\kappa(0)\tau(0)t^3/6)\mathbf{n}(0)$, in particular, it has a so called cusp singularity at $t=0$.
Finally, the projection onto the rectifying plane has the Taylor expansion

$$\gamma(0) + \big(t + o(t)\big)\mathbf{t}(0) + \Big(\kappa(0)\tau(0)\frac{t^3}{6} + o(t^3)\Big)\mathbf{b}(0).$$

so its shape is like the graph of a cubic function. It is easy to see, that the curvature of this projection is 0 at $t=0$, thus it is almost straight around the origin. That is the reason why the rectifying plane was given just this name: projection of the curve onto the rectifying plane straightens the curve and "rectifying" means straightening.

Exercise 2.9.1. Given a unit speed curve of general type in $\mathbb{R}^3$ with distinguished Frenet frame $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$, find a vector field ω along the curve such that $\mathbf{t}_i' = \omega \times \mathbf{t}_i$ holds for $i = 1, 2, 3$. (ω is called the *Darboux vector field* of the curve.) ¤

Exercise 2.9.2. Suppose that the osculating planes of a curve of general type in $\mathbb{R}^3$ have a point in common. Show that the curve is a plane curve. (In particular, if a point is moving in a central force field, its motion is planar.) ¤

Exercise 2.9.3. Suppose that the normal planes of a regular curve in $\mathbb{R}^3$ go through a fixed point O. Show that the curve lies on a sphere centered at O. ¤

Exercise 2.9.4. Show that the rectifying planes of a curve of general type in $\mathbb{R}^3$ can never share a point in common. ¤

2.9.2 Fenchel's Theorem

Though a regular space curve is not necessarily a curve of general type, the usual formulae allow us to define the unit tangent vector field and the (first) curvature function for a regular space curve as well.

Definition 2.9.5. The *unit tangent vector field* of a regular space curve $\gamma\colon I \to \mathbb{R}^3$ is the vector field $\mathbf{t} = \gamma'/v$, where $v = \|\gamma'\|$. The *curvature of* γ is the function

$$\kappa = \frac{\|\mathbf{t}'\|}{v} = \frac{\|\gamma' \times \gamma''\|}{v^3}.$$

The *total curvature of* γ is the integral

$$\int_I v(t)\kappa(t)dt.$$ ※

We saw that the total curvature of a smooth closed plane curve can take only isolated values, the integer multiples of 2π. This is related to the fact that the curvature of a planar curve is a *signed* function. When we try to increase the total curvature of a plane curve making it wavier by continuously deforming it within the class of closed regular plane curves, we shall not succeed, because the total curvature of arcs where the curvature is negative keeps equilibrium with the total curvature of the arcs where the curvature is negative.

In contrast to the planar case, curvature of a regular space curve is nonnegative even if the curve is contained in a plane. In the latter case, the (spatial) curvature of the curve is the absolute value of the signed planar curvature. For this reason, the total curvature of a regular space curve can be increased continuously and can be made arbitrarily large by making the curve wavier. Hence for regular closed smooth space curves, the interesting question is, how small the total curvature can be. This question is answered by a theorem of Moritz Werner Fenchel (1905-1988).

Theorem 2.9.6 (Fenchel)**.** *The total curvature of a regular smooth closed space curve is at least 2π. Equality holds if and only if the curve is a convex plane curve.*

Remark. We do not assume that γ is simple.

The proof of the theorem is based on the following Lemma from spherical geometry.

Lemma 2.9.7. *If a closed curve on the unit sphere $\mathbb{S}^2$ centered at the origin $\mathbf{0}$ has length $l \leq 2\pi$, then the curve is contained in a closed hemisphere. If the length is strictly less than 2π, then the curve is contained in an open hemisphere.*

Proof. Recall that the spherical distance of two points on the sphere $\mathbb{S}^2$ is defined to be the angle between their position vectors from the center. It is a corollary of the spherical triangle inequality, that the length of any spherical curve connecting the points P and Q is greater than or equal to the spherical distance of P and Q, and equality holds if and only if the curve is a shorter arc of a great circle passing through P and Q. This arc is unique if P and Q are not antipodal, but when P and Q are antipodal, there are infinitely many semicircles of minimal length connect them.
Consider a closed spherical curve Γ of length $l \leq 2\pi$. Choose two points P and Q on Γ, which divide Γ into two arcs Γ_1, Γ_2 of equal length $l/2 \leq \pi$.
If P and Q are antipodal points, then their spherical distance is π, so the two arcs of Γ connecting them must be two semicircles. In this case the length of Γ is exactly 2π and Γ is contained in one of the closed hemispheres bounded by the great circle containing the semicircle Γ_1. We infer that the Lemma holds in this case.

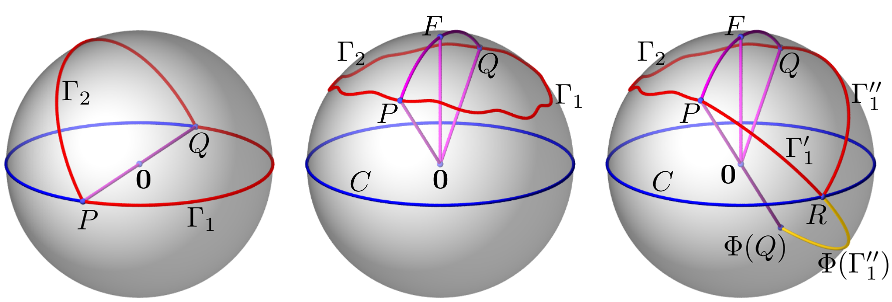

Figure 2.15

Assume now that P and Q are not antipodal. Then let F be the midpoint of the shortest great circle arc connecting them. $\mathbf{0}F$ is the bisector of the angle $P\mathbf{0}Q$ and reflecting P in the line $\mathbf{0}F$ gives Q.

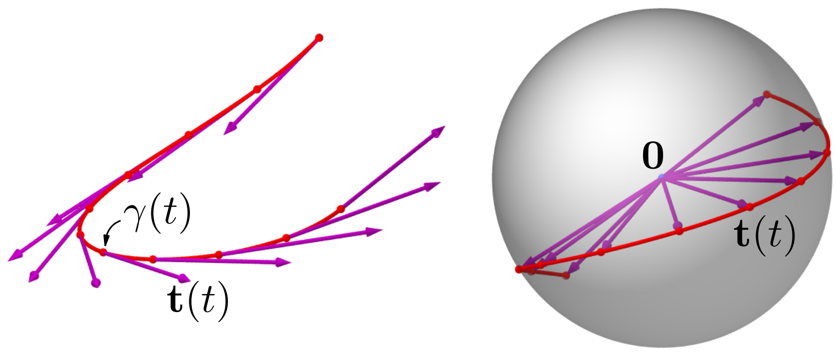

Figure 2.16: The tantrix of a parameterized curve γ.

Denote by C the great circle, the plane of which is orthogonal to $\mathbf{0}F$.
We prove that if $l < 2\pi$, then Γ does not intersect C. Assume to the contrary, that there is an intersection point R. We can assume without loss of generality, that R is on the arc Γ_1 and it cuts Γ_1 into an arc Γ_1' connecting P to R, and an arc Γ_1'' connecting R to Q. Let Φ denote the reflection in the plane aff$[C]$. It is easy to see that $\Phi(Q)$ is antipodal to P, and concatenation of the arcs Γ_1' and $\Phi(\Gamma_1'')$ connect P to $\Phi(Q)$. However, this is a contradiction, since the total length of these two arcs is $l/2 < \pi$, but the spherical distance of P and $\Phi(Q)$ is π. The contradiction proves that if $l < 2\pi$, then Γ is contained in one of the open hemispheres bounded by C.
If we allow $l = 2\pi$, then the arc Γ_1 and Γ_2 may intersect C. Nevertheless, if, say Γ_1 intersects C in a point R, as above, then Γ_1' and $\Phi(\Gamma_1'')$ must be two arcs of a semicircle connecting P and $\Phi(Q)$. This means that even in this case the arc Γ_1 is in the closed hemisphere bounded by C and containing P. The same is true for Γ_2. □

Proof of Fenchel's Theorem. Let $\gamma\colon [a,b] \to \mathbb{R}^3$ be a regular smooth closed curve, $\mathbf{t}\colon [a,b] \to \mathbb{S}^2$ be its unit tangent vector field. $\mathbf{t}$ parameterizes a closed spherical curve Γ. Γ is called the *tangent indicatrix* or shortly the *tantrix* of γ. The length of Γ equals

$$\int_a^b \|\mathbf{t}'(t)\| dt = \int_a^b v(t)\kappa(t) dt,$$

that is the total curvature of γ.
Apply the Lemma to the curve Γ. The parameterization $\mathbf{t}$ is not necessarily injective, but this is not important as the proof of the Lemma works also for curves with a non-injective parameterization.
If we assume to the contrary that the total curvature of γ is less than 2π, then Γ is a closed spherical curve of length $< 2\pi$, so it can be covered by an

open hemisphere. Let $\mathbf{u} \in \mathbb{S}^2$ be the spherical center of the open hemisphere containing Γ. Then $\mathbf{u}$ encloses an acute angle with all the vectors $\mathbf{t}(t)$, thus, $\langle \mathbf{u}, \mathbf{t}(t)\rangle > 0$ for all $t \in [a, b]$. From this we obtain the inequality

$$\langle \mathbf{u}, \gamma\rangle' = v \cdot \langle \mathbf{u}, \mathbf{t}\rangle > 0,$$

which implies that $\langle \mathbf{u}, \gamma\rangle$ is a strictly increasing function. However, this contradicts the assumption that γ is closed, as this leads to

$$\langle \mathbf{u}, \gamma(a)\rangle = \langle \mathbf{u}, \gamma(b)\rangle.$$

The contradiction proves that the total curvature of γ is at least 2π.
Consider the case of equality. Applying the lemma for the case when γ has total curvature 2π, we obtain a unit vector $\mathbf{u}$ satisfying $\mathbf{u}, \mathbf{t}\rangle \geq 0$. Then the above argument provides that the function $\langle \mathbf{u}, \gamma\rangle$ is weakly increasing and takes the same values at the endpoints of the interval $[a, b]$. Thus, $\langle \mathbf{u}, \gamma\rangle$ is constant, which means that γ lies in a plane orthogonal to $\mathbf{u}$.
We can also see that γ is simple. Indeed, if γ had a self-intersection point $\gamma(t_1) = \gamma(t_2)$, $a < t_1 < t_2 \leq b$ then we could split γ into two arcs $\gamma_1 = \gamma|_{[t_1,t_2]}$ and $\gamma_2 = \tilde{\gamma}|_{[t_2,t_1+(b-a)]}$, where $\tilde{\gamma}\colon \mathbb{R} \to \mathbb{R}^3$ is the periodic extension of γ. Then the points $P = \mathbf{t}(t_1)$ and $Q = \mathbf{t}(t_2)$ would cut Γ into two corresponding arcs Γ_1 and Γ_2. We show that each of these arcs must have length strictly larger than π. This would contradict the fact that Γ has total length 2π.
Assume to the contrary that, say Γ_1 has length $\leq \pi$. Then the great circle C that covers Γ_1 has a closed semicircle with midpoint $\mathbf{v} \in C$ which also covers Γ_1. Following the above line of reasoning, this implies that $\langle \mathbf{v}, \gamma\rangle$ must be weakly increasing on $[t_1, t_2]$ and take the same values at t_1 and t_2, therefore, $\langle \mathbf{v}, \gamma\rangle$ must be constant on $[t_1, t_2]$. Differentiating this equation we infer that $\mathbf{t}$ must stay orthogonal to $\mathbf{v}$ on $[t_1, t_2]$. But $\mathbf{t}$ is also orthogonal to the unit vector $\mathbf{u} \perp \mathbf{v}$, consequently, $\mathbf{t}$ must be constant on $[t_1, t_2]$, equal to either $\mathbf{u} \times \mathbf{v}$ or $-\mathbf{u} \times \mathbf{v}$. However, in this case, $\langle \gamma, \mathbf{u} \times \mathbf{v}\rangle$ is strictly increasing or decreasing on $[t_1, t_2]$, which contradict the assumption $\gamma(t_1) = \gamma(t_2)$.
Let us orient the plane of γ and introduce the signed planar curvature of γ, which we denote by $\kappa^{\pm}$ to distinguish it from $\kappa = |\kappa^{\pm}|$. Since γ is simple, the Umlaufsatz gives

$$\int_a^b v(t)|\kappa^{\pm}(t)|dt = \int_a^b v(t)\kappa(t)dt = 2\pi = \left|\int_a^b v(t)\kappa^{\pm}(t)dt\right|.$$

This equality can be true only if $\kappa^{\pm} \geq 0$ or $\kappa^{\pm} \leq 0$ everywhere, and this is equivalent to the convexity of the curve. □

2.9.3 The Fáry-Milnor Theorem

Fenchel proved his theorem in 1929. In 1947 K. Borsuk generalized Fenchel's theorem for curves in the n-dimensional space, and formulated the conjecture that the total curvature of a nontrivial knot in the three dimensional space must exceed 4π. The conjecture with a weak inequality $\geq 4\pi$ was proved by I. Fáry in 1949. Independently, in the same year J. W. Milnor published an analogous, but a bit more technical proof in which the strict inequality was verified.

A simple closed curve gives an injective continuous map from the circle into the space. Such a map is also called a *knot*. A knot is said to be *trivial*, if the map from the circle extends to a continuous injective map from the disc bounded by the circle into $\mathbb{R}^3$.

Theorem 2.9.8 (I. Fáry, J.W. Milnor)**.** *The total curvature of a nontrivial knot in $\mathbb{R}^3$ is greater than 4π.*

Below we prove only the week inequality $\geq 4\pi$.

Proof of the weak form. Assume that the total curvature of a simple regular closed smooth curve is less than 4π. We want to show that the curve is a trivial knot. The tantrix Γ of γ is a closed spherical curve of length $< 4\pi$. Let us apply the spherical Crofton formula for Γ. For a unit vector $\mathbf{u} \in \mathbb{S}^2$, let $m(\mathbf{u})$ denote the number of intersection points of Γ with the great circle contained in the plane orthogonal to $\mathbf{u}$, counted with multiplicity

$$m(\mathbf{u}) = \#\{t \in [a,b] \mid \langle \mathbf{u}, \mathbf{t}(t) \rangle = 0\}.$$

Then by the Crofton formula, the length of Γ can be given as an integral

$$l_\Gamma = \frac{1}{4}\int_{\mathbb{S}^2} m(\mathbf{u}) d\mathbf{u}.$$

The surface area of $\mathbb{S}^2$ is 4π, and the value of the integral on the right-hand side must be less than 16π, the subset of those points of $\mathbb{S}^2$ at which the value of m is at most 3 must have positive measure. In particular, we can find a vector $\mathbf{u} \in \mathbb{S}^2$ such that the dot product $\langle \mathbf{u}, \mathbf{t}(t) \rangle$ vanishes at no more than 3 points.

We can think of $\mathbf{u}$ as a vertical vector pointing upward, and $\langle \mathbf{u}, \gamma(t) \rangle$ as the height of the point $\gamma(t)$. The height of $\gamma(t)$ increases strictly on intervals where $\langle \mathbf{u}, \mathbf{t}(t) \rangle > 0$ and decreases strictly where $\langle \mathbf{u}, \mathbf{t}(t) \rangle < 0$. The roots of $\langle \mathbf{u}, \mathbf{t} \rangle$ divide the image of γ into at most 3 simple arcs along which the height changes in a strictly monotonous way. There must exist both an arc along which the height increases and another one where it decreases, otherwise the curve could not return to its initial point. If there are three arcs along which the height is a strictly monotone function then two of the arcs can be joined so that the height will strictly monotone also on the union. We can summarize the information we got as follows. The curve traced out by γ can be split into two simple arcs joining the unique highest and the unique lowest point of the trace. As we move along any of the two arcs the height changes in a strictly monotonous way.

Take any horizontal plane between the highest and lowest points of the trace of γ. Such a plane intersects both arcs at a unique point. Connect these two points by a segment. The union of all such horizontal segments connecting two points of $\operatorname{im}\gamma$ fill a topologically embedded disc in $\mathbb{R}^3$, the boundary of which is the trace of γ. We conclude that γ is a trivial knot. $\square$

Chapter 3

Hypersurfaces in $\mathbb{R}^n$

3.1 General Theory

3.1.1 Definition of a Parameterized Hypersurface

Definition 3.1.1. A *smooth parameterized hypersurface* in $\mathbb{R}^n$ is a differentiable mapping $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ from an open domain Ω of $\mathbb{R}^{n-1}$ into the n-dimensional space. ※

Smooth curves on a parameterized hypersurface are curves of the form $\gamma(t) = \mathbf{r}(\mathbf{u}(t))$, where the mapping $t \mapsto \mathbf{u}(t)$ is a smooth curve lying in the parameter domain Ω. Curves of the form $t \mapsto \mathbf{r}(u_1, \dots, u_{i-1}, t, u_{i+1}, \dots, u_{n-1})$, where $u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_{n-1}$ are fixed numbers are called the *parameter lines* or *coordinate lines* on the hypersurface. The speed vectors of the parameter lines $t \mapsto \mathbf{r}(u_1, \dots, u_{i-1}, t, u_{i+1}, \dots, u_{n-1})$, which are just the partial derivatives of the mapping $\mathbf{r}$ with respect to the i-th variable, will be denoted by $\mathbf{r}_i(u_1, \dots, u_{i-1}, t, u_{i+1}, \dots, u_{n-1})$.
Since we shall often work with formulae containing partial derivatives of a function, it is convenient to introduce the shorthand convention that we shall denote the partial derivative of a multivariable function F with respect to its i-th variable by F_i. In general, the higher order partial derivative $\dfrac{\partial^k F}{\partial u_{i_1} \dots \partial u_{i_k}}$ of F will be denoted by $F_{i_1 \dots i_k}$. If there is a danger of confusion with ordinary lower indices, the lower indices of the function will be separated from the indices of variables with respect to which we have to take the partial derivative by a comma. Thus, $\dfrac{\partial^k F_{j_1 \dots j_r}}{\partial u_{i_1} \dots \partial u_{i_k}}$ will be denoted by $F_{j_1 \dots j_r, i_1 \dots i_k}$. Similar conventions will be applied to functions the variables of which are denoted by fixed alphabetical symbols. For example, the partial derivative

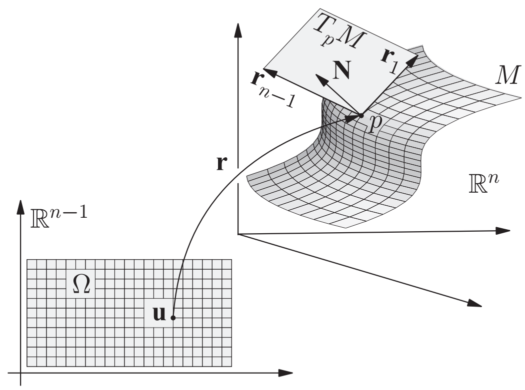

Figure 3.1

of a function F of the variables $u, v, w, \ldots$ with respect to the variable u will be denoted shortly by F_u. We also use this abbreviation for higher order partial derivatives. E.g., F_{uv} is a short form of $\partial_u \partial_v F$. For conventions about denoting partial derivatives see also the remarks following Definition 1.5.6.

Definition 3.1.2. A parameterized hypersurface is *regular* if the vectors $\mathbf{r}_1(\mathbf{u}), \ldots, \mathbf{r}_{n-1}(\mathbf{u})$ are linearly independent for any $\mathbf{u} \in \Omega$. In this case we also say that $\mathbf{r}$ is an *immersion* of the domain Ω into $\mathbb{R}^n$. ※

Regularity does not imply that $\mathbf{r}$ is injective. However, if $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ is a regular parameterized hypersurface, then each point $\mathbf{u} \in \Omega$ has a neighborhood $U \subset \Omega$ such that $\mathbf{r}\big|_U$ maps U bijectively onto $\mathbf{r}(U)$. Thus, studying a regular parameterized hypersurface locally (i.e. in a small neighborhood of a point) we may assume without loss of generality that $\mathbf{r}$ is injective. We shall usually denote the image of $\mathbf{r}$ by M.

Definition 3.1.3. The *affine tangent space* of a regular parameterized hypersurface at the point $p = \mathbf{r}(\mathbf{u}) \in M$ is the hyperplane through $\mathbf{r}(\mathbf{u})$ spanned by the direction vectors $\mathbf{r}_1(\mathbf{u}), \ldots, \mathbf{r}_{n-1}(\mathbf{u})$. The *(linear) tangent space* of M at p is the linear space T_pM of the direction vectors of the affine tangent space. The *unit normal vector of the hypersurface at the point* $\mathbf{r}(\mathbf{u})$ is defined to be the unit normal vector $\mathbf{N}(\mathbf{u})$ of the tangent plane, for which $\mathbf{r}_1(\mathbf{u}), \ldots, \mathbf{r}_{n-1}(\mathbf{u}), \mathbf{N}(\mathbf{u})$ is a positively oriented basis of $\mathbb{R}^n$. ※

For parameterized surfaces in $\mathbb{R}^3$, the unit normal vector field can be calculated with the help of cross product

$$\mathbf{N}(u_1, u_2) = \frac{\mathbf{r}_1(u_1, u_2) \times \mathbf{r}_2(u_1, u_2)}{|\mathbf{r}_1(u_1, u_2) \times \mathbf{r}_2(u_1, u_2)|}.$$

To get a similar formula in higher dimensions, we need a suitable generalization of the cross product.
Let $\mathbf{r}_i = (r_i^1, \dots, r_i^n) \in \mathbb{R}^n$, $i = 1, 2, \dots, n-1$, be n-dimensional vectors, $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ be the standard basis of $\mathbb{R}^n$. Then according to Proposition 1.2.78, the *generalized cross product* $*(\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1})$ of the vectors $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$ can be computed by the equality

$$*(\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}) = \det \begin{pmatrix} r_1^1 & \dots & r_1^n \\ \dots & \dots & \dots \\ r_{n-1}^1 & \dots & r_{n-1}^n \\ \mathbf{e}_1 & \dots & \mathbf{e}_n \end{pmatrix}.$$

The vector $*\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}$ is orthogonal to $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$, it is different from $\mathbf{0}$ if and only if $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$ are linearly independent, and finally, $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$, $(-1)^{n-1}\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}$ is a positively oriented basis of $\mathbb{R}^n$.
Consequently, for regular hypersurfaces, we have

$$\mathbf{N}(u) = \frac{*(\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1})}{\|\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\|}.$$

3.1.2 Curvature of Curves on a Hypersurface

Now we shall study a problem which connects curve theory to surface theory. If a curve lies on a surface, curvedness of the surface forces the curve to bend. Thus, curvedness of a surface can be detected by the curvatures of the curves lying on the surface. It is clear heuristically that the curvature of a curve on a given surface should be the same as the curvature of the intersection curve of the surface and the osculating plane of the curve provided that the osculating plane is not tangent to the surface. This is indeed true and thus we may pose the question how to compute the curvature of the curve using only information on the surface and the position of the osculating plane of the curve. The existence of a formula that answers this question will prove our heuristics.
Consider the curve $\gamma(t) = \mathbf{r}(\mathbf{u}(t))$ lying on the regular parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$, where $\mathbf{u} = (u_1, \dots, u_{n-1})$ is a curve in Ω. Express the first two derivatives of γ using the Frenet formulae on one hand and the special form of γ on the other. By the chain rule, we get

$$v\mathbf{t}_1 = \gamma' = \sum_{i=1}^{n-1} u_i' \mathbf{r}_i(\mathbf{u})$$

and

$$v'\mathbf{t}_1 + v^2 \kappa_1 \mathbf{t}_2 = \gamma'' = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} u_i' u_j' \mathbf{r}_{ij}(\mathbf{u}) + \sum_{i=1}^{n-1} u_i'' \mathbf{r}_i(\mathbf{u}).$$

Multiplying the last equation by the normal vector of the hypersurface and using the fact that it is orthogonal to the tangent vectors $\mathbf{t}_1, \mathbf{r}_1, \dots, \mathbf{r}_{n-1}$, we obtain

$$v^2\kappa_1\langle \mathbf{N}(\mathbf{u}), \mathbf{t}_2\rangle = \langle \mathbf{N}(\mathbf{u}), \gamma''\rangle = \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\langle \mathbf{N}(\mathbf{u}), \mathbf{r}_{ij}(\mathbf{u})\rangle u_i' u_j',$$

from which

$$\kappa_1 = \frac{1}{\langle \mathbf{N}(\mathbf{u}), \mathbf{t}_2\rangle} \cdot \frac{\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\langle \mathbf{N}(\mathbf{u}), \mathbf{r}_{ij}(\mathbf{u})\rangle u_i' u_j'}{v^2}. \tag{3.1}$$

Let us study this expression. We claim that the right hand side is determined by the osculating plane of the curve and the hypersurface provided that the osculating plane is not tangent to the surface.
Let us start with the expression

$$k(\gamma') = \frac{\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\langle \mathbf{N}(\mathbf{u}), \mathbf{r}_{ij}(\mathbf{u})\rangle u_i' u_j'}{v^2}.$$

Since the quantities $\langle \mathbf{N}(\mathbf{u}), \mathbf{r}_{ij}(\mathbf{u})\rangle$ are determined by the parameterization of the hypersurface, the functions $u_1', \dots, u_{n-1}'$ are the components of the speed vector γ' of the curve with respect to the basis $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$ of the tangent space, v is the length of the speed vector γ', $k(\gamma')$ depends only on the speed vector γ' of the curve (that justifies the notation $k(\gamma')$).

Definition 3.1.4. Let $\mathbf{v} \neq \mathbf{0}$ be an arbitrary tangent vector of the regular parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ at $\mathbf{r}(\mathbf{u})$. The intersection curve of the hypersurface and the plane through $\mathbf{r}(\mathbf{u})$ spanned by the direction vectors $\mathbf{N}(u)$ and $\mathbf{v}$ is called the *normal section of the hypersurface in the direction* $\mathbf{v}$. Assume that the normal section is parameterized in such a way that the speed vector of the parameterization at $\mathbf{r}(\mathbf{u})$ is $\mathbf{v}$. (Such a parameterization always exists in a neighborhood of $\mathbf{r}(u)$.) Orienting the cutting normal plane by the ordered basis $(\mathbf{v}, \mathbf{N}(\mathbf{u}))$, we may consider the signed curvature of the normal section, which will be called the *normal curvature of the hypersurface in the direction* $\mathbf{v}$ and will be denoted by $k(\mathbf{v})$. ※

Applying (3.1) for normal sections one may see easily that the normal curvature of a parameterized hypersurface in the direction $\mathbf{v} = v_1\mathbf{r}_1(\mathbf{u}) + \dots + v_{n-1}\mathbf{r}_{n-1}(\mathbf{u})$ is just

$$k(\mathbf{v}) = \frac{\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\langle \mathbf{N}(\mathbf{u}), \mathbf{r}_{ij}(\mathbf{u})\rangle v_i v_j}{v^2}, \tag{3.2}$$

where $v = \|\mathbf{v}\|$. Since $k(\lambda\mathbf{v}) = k(\mathbf{v})$ for any $\lambda \neq 0$, the normal curvature depends only on the direction of $\mathbf{v}$.

Returning to the curve γ we see that $k(\gamma')$ is determined by the tangent line of γ at the given point which is the intersection of the osculating plane of γ and the tangent space of the hypersurface.
Since the osculating plane and the tangent line determine the second Frenet vector $\mathbf{t}_2$ uniquely up to sign, we conclude that the curvature $\kappa_1 = (1/\langle \mathbf{N}(\mathbf{u}), \mathbf{t}_2\rangle)k(\gamma')$ of the curve is determined by the osculating plane up to sign and since the curvature κ_1 is positive, if $n \geq 3$, both $\mathbf{t}_2$ and κ_1 are determined uniquely (and not only up to sign) by the osculating plane.
To finish this unit with, we formulate an obvious consequence of the formula expressing the curvature of a curve lying on a hypersurface.

Corollary 3.1.5 (Meusnier's theorem)**.** *If the osculating plane of a curve γ lying on a hypersurface is not contained in the tangent space of the hypersurface at a given point $\gamma(t) = \mathbf{r}(\mathbf{u}(t))$, then the curvature of the curve and the normal curvature of the surface in the direction $\gamma'(t)$ are related to one another by the equation $\kappa_1(t) = \frac{1}{\cos\alpha}k(\gamma'(t))$, where α is the angle between the normal vector $\mathbf{N}(\mathbf{u}(t))$ of the hypersurface and the second Frenet vector $\mathbf{t}_2(t)$ of the curve.*

3.1.3 The Weingarten Map and the Fundamental Forms

Definition 3.1.6. Let $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ be a parameterized hypersurface. A *vector field along the hypersurface* is a mapping $X\colon \Omega \to \mathbb{R}^n$ from the domain of parameters into $\mathbb{R}^n$. X is a *tangential vector field*, if $X(\mathbf{u})$ is tangent to the hypersurface at $\mathbf{r}(\mathbf{u})$ for all $\mathbf{u} \in \Omega$. ※

For example, the mappings $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$ and $\mathbf{N}$ are vector fields along the hypersurface, and the first $n-1$ of them are tangential.
Given a vector field along a hypersurface, we would like to express the speed of change of the vector field as we move along the surface, in terms of the speed of our motion. This is achieved by the following.

Definition 3.1.7. Let $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ be a parameterized hypersurface, $X\colon \Omega \to \mathbb{R}^n$ be a vector field along it, $\mathbf{u}_0 \in \Omega$, $\mathbf{v}$ a tangent vector of the hypersurface at $\mathbf{r}(\mathbf{u}_0)$. We define the *derivative $\partial_{\mathbf{v}}X$ of the vector field X with respect to the tangent vector* $\mathbf{v}$ as $\partial_{\mathbf{v}}X = (X \circ \mathbf{u})'(0)$, where $\mathbf{u}\colon [-1,1] \to \Omega$ is a curve in the parameter domain such that $\mathbf{u}(0) = \mathbf{u}_0$ and $(r \circ \mathbf{u})'(0) = \mathbf{v}$. ※

Since by the chain rule

$$(X \circ \mathbf{u})'(0) = \sum_{i=1}^{n-1} u_i'(0)X_i(\mathbf{u}(0)),$$

where $(u_1'(0), \dots, u_{n-1}'(0))$ are the components of $\mathbf{v}$ in the basis $\mathbf{r}_1(\mathbf{u}_0), \dots, \mathbf{r}_{n-1}(\mathbf{u}_0)$ of the tangent space at $\mathbf{r}(\mathbf{u}_0)$, we have the following formula

$$\partial_{\mathbf{v}} X = \sum_{i=1}^{n-1} v_i X_i(\mathbf{u}_0),$$

where $v_1, \dots, v_{n-1}$ are the components of the vector $\mathbf{v}$ in the basis $\mathbf{r}_1(\mathbf{u}_0), \dots, \mathbf{r}_{n-1}(\mathbf{u}_0)$. This formula shows that the definition of $\partial_{\mathbf{v}} X$ is correct, i.e. independent of the choice of the curve $\mathbf{u}(t)$.
The way in which a hypersurface curves around in $\mathbb{R}^n$ is closely related to the way the normal direction changes as we move from point to point.

Lemma 3.1.8. *The derivative $\partial_{\mathbf{v}}\mathbf{N}$ of the unit normal vector field of a hypersurface with respect to a tangent vector $\mathbf{v}$ at $p = \mathbf{r}(\mathbf{u})$ is tangent to the hypersurface at $\mathbf{r}(\mathbf{u})$.*

Proof. We have to show that $\partial_{\mathbf{v}}\mathbf{N}$ is orthogonal to $\mathbf{N}(\mathbf{u})$. Indeed, differentiating the relation $1 \equiv \langle \mathbf{N}, \mathbf{N} \rangle$, we get

$$0 = \langle \partial_{\mathbf{v}}\mathbf{N}, \mathbf{N} \rangle + \langle \mathbf{N}, \partial_{\mathbf{v}}\mathbf{N} \rangle = 2\langle \partial_{\mathbf{v}}\mathbf{N}, \mathbf{N} \rangle. \qquad \square$$

Definition 3.1.9. Let us denote by M the parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ and by T_pM the linear space of its tangent vectors at $p = \mathbf{r}(\mathbf{u}_0)$. The linear map

$$L_p\colon T_pM \to T_pM, \qquad L_p(\mathbf{v}) = -\partial_{\mathbf{v}}\mathbf{N}$$

is called the *Weingarten map* or *shape operator* of M at p. ※

We define two bilinear forms on each tangent space of the hypersurface.

Definition 3.1.10. Let M be a parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$, $\mathbf{u}_0 \in \Omega$, T_pM the linear space of tangent vectors of M at $p = \mathbf{r}(\mathbf{u})$, $L_p\colon T_pM \to T_pM$ the Weingarten map. The *first fundamental form* of the hypersurface is the bilinear function I_p on T_pM obtained by restriction of the dot product onto T_pM

$$I_p(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle \qquad \text{for } \mathbf{v}, \mathbf{w} \in T_pM.$$

The *second fundamental form* of the hypersurface is the bilinear function II_p on T_pM defined by the equality

$$II_p(\mathbf{v}, \mathbf{w}) = \langle L_p\mathbf{v}, \mathbf{w} \rangle \qquad \text{for } \mathbf{v}, \mathbf{w} \in T_pM.$$ ※

The first fundamental form is obviously a positive definite symmetric bilinear function on the tangent space. Its matrix representation with respect to the basis $(\mathbf{r}_1(\mathbf{u}_0), \dots, \mathbf{r}_{n-1}(\mathbf{u}_0))$ has entries $\langle \mathbf{r}_i(\mathbf{u}_0), \mathbf{r}_j(\mathbf{u}_0) \rangle$.
An important property of the Weingarten map and the second fundamental form is stated in the following theorem.

Theorem 3.1.11. *The second fundamental form is symmetric, i.e.*

$$\langle L_p \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, L_p \mathbf{w} \rangle \qquad \textit{for } \mathbf{v}, \mathbf{w} \in T_p M,$$

or in other words, the Weingarten map is self-adjoint *(with respect to the first fundamental form).*

Lemma 3.1.12.

$$\mathit{II}_p(\mathbf{r}_i(\mathbf{u}_0), \mathbf{r}_j(\mathbf{u}_0)) = \langle \mathbf{r}_{ij}(\mathbf{u}_0), \mathbf{N}(\mathbf{u}_0) \rangle. \tag{3.3}$$

Proof. We know that the normal vector field $\mathbf{N}$ is perpendicular to any tangential vector field, thus

$$\langle \mathbf{N}, \mathbf{r}_j \rangle \equiv 0.$$

Differentiating this identity with respect to the i-th parameter we get

$$\langle \partial_{\mathbf{r}_i} \mathbf{N}, \mathbf{r}_j \rangle + \langle \mathbf{N}, \mathbf{r}_{ji} \rangle \equiv 0,$$

from which

$$\langle \mathbf{N}, \mathbf{r}_{ji} \rangle \equiv \langle -\partial_{\mathbf{r}_i} \mathbf{N}, \mathbf{r}_j \rangle = \langle L_{\mathbf{r}} \mathbf{r}_i, \mathbf{r}_j \rangle. \qquad \square$$

Proof of Theorem 3.1.11. It is enough to prove that the matrix of the second fundamental form with respect to the basis $\mathbf{r}_1(\mathbf{u}_0), \dots, \mathbf{r}_{n-1}(\mathbf{u}_0)$ is symmetric. However, this follows from the lemma, since by Young's theorem $\mathbf{r}_{ij} = \mathbf{r}_{ji}$. □

Comparing the identity (3.3) with (3.2) we see that the normal curvature is the quotient of the quadratic forms of the second and first fundamental forms

$$k(\mathbf{v}) = \frac{\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \langle \mathbf{N}(\mathbf{u}_0), \mathbf{r}_{ij}(\mathbf{u}_0) \rangle v_i v_j}{|\mathbf{v}|^2} = \frac{\mathit{II}_p(\mathbf{v}, \mathbf{v})}{I_p(\mathbf{v}, \mathbf{v})},$$

where $\mathbf{v} = \sum_{i=1}^{n-1} v_i \mathbf{r}_i(\mathbf{u}_0)$ is a tangent vector of the hypersurface at $p = \mathbf{r}(\mathbf{u}_0)$. The expression

$$k(\mathbf{v}) = \frac{\mathit{II}_p(\mathbf{v}, \mathbf{v})}{I_p(\mathbf{v}, \mathbf{v})},$$

gives rise to a linear algebraic investigation of the normal curvature. It is natural to ask at which directions the normal curvature attains its extrema. Since $k(\lambda \mathbf{v}) = k(\mathbf{v})$ for any $\lambda \neq 0$, it is enough to consider this question for the restriction of k onto the unit sphere S in the tangent space. The unit sphere of a Euclidean space is a compact (i.e., closed and bounded) subset, thus, by Weierstrass' theorem, any continuous function defined on it attains its maximum and minimum.

Definition 3.1.13. Let f be a differentiable function defined on the unit sphere S of a Euclidean vector space. We say that the vector $\mathbf{v} \in S$ is a *critical point* of f if for any curve $\gamma\colon [-1,1] \to S$ on the sphere such that $\gamma(0) = \mathbf{v}$ the derivative of the composite function $f \circ \gamma$ vanishes at 0. ※

Clearly, points at which a function is locally minimal or maximal are critical, but the converse is not true. The following proposition gives a characterization of critical points for the restriction of the normal curvature onto the unit sphere of the tangent space.

Proposition 3.1.14. *Let V be a finite dimensional vector space with a positive definite symmetric bilinear function $\langle , \rangle$ and let $L\colon V \to V$ be a self-adjoint linear transformation on V. Set $S = \{\mathbf{x} \in V | \langle \mathbf{x}, \mathbf{x}\rangle = 1\}$ and define $f\colon S \to \mathbb{R}$ by $f(\mathbf{x}) = \langle L\mathbf{x}, \mathbf{x}\rangle$. Then $\mathbf{v} \in S$ is a critical point of f if and only if $\mathbf{v}$ is an eigenvector of L.*

Proof. For any curve $\gamma\colon [-1,1] \to S$ such that $\gamma(0) = \mathbf{v}$, we have

$$\begin{aligned}\frac{d}{dt}\langle L(\gamma(t)), \gamma(t)\rangle\big|_{t=0} &= \langle L(\gamma'(0)), \gamma(0)\rangle + \langle L(\gamma(0)), \gamma'(0)\rangle \\ &= \langle L\gamma'(0), \mathbf{v}\rangle + \langle L\mathbf{v}, \gamma'(0)\rangle = 2\langle L\mathbf{v}, \gamma'(0)\rangle.\end{aligned}$$

This means that $\mathbf{v}$ is a critical point of f if and only if $L\mathbf{v}$ is orthogonal to all vectors of the form $\gamma'(0)$. Since the speed vectors $\gamma'(0)$ of spherical curves through $\mathbf{v} = \gamma(0)$ range over the tangent space of the sphere S, $\mathbf{v}$ is a critical point of f if and only if $L\mathbf{v}$ is orthogonal to the tangent space of S at $\mathbf{v}$. However, since the normal vector of this tangent space is $\mathbf{v}$, the latter condition is satisfied if and only if $L\mathbf{v}$ is a scalar multiple of $\mathbf{v}$, i.e. $\mathbf{v}$ is an eigenvector of L. □

Applying Proposition 3.1.14 to the Weingarten map L_p we obtain that the critical vectors of the normal curvature function k_p at p are the eigenvectors of L_p. According to the Principal Axis Theorem 1.2.65, there is an orthonormal basis of the tangent space T_pM consisting of eigenvectors of the Weingarten map.

Definition 3.1.15. For a hypersurface M in $\mathbb{R}^n$ parameterized by $\mathbf{r}$, $\mathbf{r}(\mathbf{u}_0) = p \in M$, the eigenvalues $\kappa_1(p), \dots, \kappa_{n-1}(p)$ of the Weingarten map $L_p\colon T_pM \to T_pM$ are called the *principal curvatures* of M at p, the unit eigenvectors of L_p are called *principal curvature directions*. ※

If the principal curvatures are ordered so that $\kappa_1(p) \le \kappa_2(p) \le \dots \le \kappa_{n-1}(p)$, the discussion above shows that $\kappa_{n-1}(p)$ is the maximal, $\kappa_1(p)$ is the minimal value of the normal curvature $k(\mathbf{v})$.

Theorem 3.1.16 (Euler's formula)**.** *Let* $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$ *be an orthonormal basis of* T_pM *consisting of principal curvature directions,* $\kappa_1(p), \dots, \kappa_{n-1}(p)$ *be the corresponding principal curvatures. Then the normal curvature* $k(\mathbf{v})$ *in the direction* $\mathbf{v} \in T_pM$*,* $|\mathbf{v}| = 1$*, is given by*

$$k(\mathbf{v}) = \sum_{i=1}^{n-1} \kappa_i(p)\langle \mathbf{v}, \mathbf{v}_i\rangle^2 = \sum_{i=1}^{n-1} \kappa_i(p)\cos^2(\theta_i),$$

where $\theta_i = \arccos(\langle \mathbf{v}, \mathbf{v}_i\rangle)$ *is the angle between* $\mathbf{v}$ *and* $\mathbf{v}_i$*.*

Proof. Since $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ is an orthonormal basis, $\mathbf{v}$ can be expressed as

$$\mathbf{v} = \sum_{i=1}^{n-1} \langle \mathbf{v}, \mathbf{v}_i\rangle \mathbf{v}_i.$$

Making use of this formula, we obtain

$$\begin{aligned} k(\mathbf{v}) = \langle L_p(\mathbf{v}), \mathbf{v}\rangle &= \Big\langle L_p\Big(\sum_{i=1}^{n-1} \langle \mathbf{v}, \mathbf{v}_i\rangle \mathbf{v}_i\Big), \sum_{i=1}^{n-1} \langle \mathbf{v}, \mathbf{v}_i\rangle \mathbf{v}_i\Big\rangle \\ &= \Big\langle \sum_{i=1}^{n-1} \langle \mathbf{v}, \mathbf{v}_i\rangle \kappa_i(p)\mathbf{v}_i, \sum_{i=1}^{n-1} \langle \mathbf{v}, \mathbf{v}_i\rangle \mathbf{v}_i\Big\rangle = \sum_{i=1}^{n-1} \kappa_i(p)\langle \mathbf{v}, \mathbf{v}_i\rangle^2. \end{aligned}$$ □

The determinant and trace of the Weingarten map, that is the product and sum of the principal curvatures are of particular importance in differential geometry.

Definition 3.1.17. For M a hypersurface, $p \in M$, the determinant $K(p)$ of the Weingarten map L_p is called the *Gaussian* or *Gauss-Kronecker curvature* of M at p, $H(p) = \operatorname{tr}(L_p)/(n-1)$ is called the *mean curvature* or *Minkowski curvature.* ※

When we want to compute the principal curvatures and directions of a hypersurface at a point we generally work with a matrix representation of the Weingarten map. Dealing with a parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$, it is natural to take matrix representations with respect to the basis $\mathbf{r}_1(\mathbf{u}), \dots, \mathbf{r}_{n-1}(\mathbf{u})$ of the tangent space at $\mathbf{r}(\mathbf{u})$. Let us denote by $\mathcal{G} = (g_{ij})_{1\le i,j\le n-1}$, $\mathcal{B} = (b_{ij})_{1\le i,j\le n-1}$ and $\mathcal{L} = (l^i_j)_{1\le i,j\le n-1}$ the matrix representations of the first and second fundamental forms and the Weingarten map respectively, with respect to this basis (g_{ij}, b_{ij} and l^i_j are functions on the parameter domain, i is the row index, j is the column index). Components of $\mathcal{G}$ and $\mathcal{B}$ can be calculated according to the equations

$$\begin{aligned} g_{ij} &= \langle \mathbf{r}_i, \mathbf{r}_j\rangle, \\ b_{ij} &= \langle \mathbf{N}, \mathbf{r}_{ij}\rangle \qquad \text{(cf. Lemma 3.1.12).} \end{aligned}$$

By definition, the coefficients l_j^i are obtained from the equations

$$L_{\mathbf{r}}(\mathbf{r}_j) = \sum_{i=1}^{n-1} l_j^i \mathbf{r}_i \qquad i = 1, 2, \dots, n-1.$$

The relationship between the matrices $\mathcal{G}$, $\mathcal{B}$, and $\mathcal{L}$ follows from the following equalities

$$b_{ji} = b_{ij} = \langle L_{\mathbf{r}}\mathbf{r}_i, \mathbf{r}_j \rangle = \left\langle \sum_{k=1}^{n-1} l_i^k \mathbf{r}_k, \mathbf{r}_j \right\rangle = \sum_{k=1}^{n-1} l_i^k \langle \mathbf{r}_k, \mathbf{r}_j \rangle = \sum_{k=1}^{n-1} g_{jk} l_i^k$$

expressing that $\mathcal{B} = \mathcal{G}\mathcal{L}$. As $\mathcal{G}$ is the matrix of a positive definite bilinear function, it is invertible (its determinant is positive). Multiplying the equation $\mathcal{B} = \mathcal{L}\mathcal{G}$ with the inverse of $\mathcal{G}$ we get the expression

$$\mathcal{L} = \mathcal{G}^{-1}\mathcal{B}$$

for the matrix of the Weingarten operator.

Corollary 3.1.18. *The Gaussian curvature of a hypersurface is equal to*

$$K = \frac{\det \mathcal{B}}{\det \mathcal{G}}.$$

Recall from linear algebra that in order to determine the eigenvalues of a linear mapping with matrix representation L one has to find the roots of the characteristic polynomial $p_L(\lambda) = \det(L - \lambda I)$, where I denotes the identity matrix.
Having determined the eigenvalues of the linear mapping, components of eigenvectors with respect to the fixed basis are obtained as non-zero solutions of the linear system of equations $L\mathbf{v} = \lambda\mathbf{v}$, where λ is an eigenvalue of L.

Exercise 3.1.19. Determine the Weingarten map for a sphere of radius r at one of its points. ¤

Exercise 3.1.20. Find the normal curvature $k(\mathbf{v})$ for each tangent direction $\mathbf{v}$, the principal curvatures and the principal curvature directions, and compute the Gaussian and mean curvatures of the following surfaces *at the given point p*.

(a) $(x_1^2/a^2) + (x_2^2/b^2) + (x_3^2/c^2) = 1$, $p = (a, 0, 0)$ (ellipsoid);

(b) $(x_1^2/a^2) + (x_2^2/b^2) - (x_3^2/c^2) = 1$, $p = (a, 0, 0)$ (one-sheeted hyperboloid);

(c) $x_1^2 + \left(\sqrt{x_2^2 + x_3^2} - 2\right)^2 = 1$, $p = (0, 3, 0)$ or $p = (0, 1, 0)$ (torus). ¤

Exercise 3.1.21. Suppose that the principal curvatures of a connected parameterized surface in $\mathbb{R}^3$ vanish. Show that the surface is a part of a plane. ¤

Exercise 3.1.22. Find the Gaussian curvature $K\colon M \to \mathbb{R}$ for the following surfaces

(a) $x_1^2 + x_2^2 - x_3^2 = 0$, $x_3 > 0$ (cone);

(b) $(x_1^2/a^2) + (x_2^2/b^2) - (x_3^2/c^2) = 1$ (hyperboloid);

(c) $(x_1^2/a^2) + (x_2^2/b^2) - x_3 = 0$ (elliptic paraboloid);

(d) $(x_1^2/a^2) - (x_2^2/b^2) - x_3 = 0$ (hyperbolic paraboloid). ¤

Exercise 3.1.23. Let M be a (hyper)surface in $\mathbb{R}^3$, $p \in M$. Show that for each $\mathbf{v}, \mathbf{w} \in T_pM$,

$$L_p(\mathbf{v}) \times L_p(\mathbf{w}) = K(p)\mathbf{v} \times \mathbf{w}.$$ ¤

3.1.4 Umbilical Points

Definition 3.1.24. A point $p = \mathbf{r}(\mathbf{u})$ of a regular parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ is called an *umbilical point* or *umbilic* if the principal curvatures at p are equal. A umbilical point p is said to be *spherical* if the common value $\kappa(\mathbf{u})$ of the principal curvatures at p is not equal to zero, and *planar* if $\kappa(\mathbf{u}) = 0$. ※

At an umbilical point p, the Weingarten map is a constant multiple of the identity map of T_pM, where the multiplier is $\kappa(\mathbf{u})$. To compute umbilical points in practice, it is more convenient to consider the matrix equation $\mathcal{B}(\mathbf{u}) = \kappa\mathcal{G}(\mathbf{u})$ instead of the equation $\mathcal{G}^{-1}(\mathbf{u})\mathcal{B}(\mathbf{u}) = \kappa I_{n-1}$. By the symmetry of the fundamental forms, this matrix equation gives $n(n+1)/2$ ordinary equations for the unknown vector $\mathbf{u}$ and the scalar κ.
The following theorem gives a characterization of those hypersurfaces which have only umbilical points.

Theorem 3.1.25. *Suppose that the parameter domain Ω of a regular parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ is connected and that all the points of the hypersurface are umbilical. If $n \geq 3$, then the trace $M = \mathbf{r}(\Omega)$ of $\mathbf{r}$ is contained either in a hypersphere or in a hyperplane.*

Proof. Denote by $\kappa(\mathbf{u})$ the common value of the principal curvatures at $\mathbf{r}(\mathbf{u})$. First we show that κ is constant. For any pair of indices $1 \leq i < j \leq n-1$, we have

$$\mathbf{N}_i = -\kappa\mathbf{r}_i \quad \text{and} \quad \mathbf{N}_j = -\kappa\mathbf{r}_j,$$

since $\mathbf{r}_i$ and $\mathbf{r}_j$ are principal directions as any tangent vector is. Differentiating the first equation with respect to the jth variable, and the second with respect to the ith, we get

$$\mathbf{N}_{ij} = -\kappa_j \mathbf{r}_i - \kappa \mathbf{r}_{ij} \quad \text{and} \quad \mathbf{N}_{ji} = -\kappa_i \mathbf{r}_j - \kappa \mathbf{r}_{ji},$$

from which $\kappa_j \mathbf{r}_i = \kappa_i \mathbf{r}_j$. Since $\mathbf{r}_i$ and $\mathbf{r}_j$ are linearly independent, the last equation can hold only if $\kappa_i = \kappa_v = j$, i.e., if κ is constant.
Case 1: $\kappa \equiv 0$. In this case $\mathbf{N}_i \mathbf{0}$ for all $1 \leq i \leq n-1$, therefore the normal vector is constant along the hypersurface. The derivative of the function $\langle \mathbf{N}, \mathbf{r} \rangle$ with respect to the ith variable is $\langle \mathbf{N}, \mathbf{r}_i \rangle = 0$ because $\mathbf{N}$ is perpendicular to the tangent vectors $\mathbf{r}_i$, hence $\langle \mathbf{N}, \mathbf{r} \rangle$ is constant and $M = \mathbf{r}(\Omega)$ is contained in a hyperplane with equation $\langle \mathbf{N}, \mathbf{x} \rangle = \mathit{const}$.
Case 2: $\kappa \neq 0$. We claim that in this case M is contained in a hypersphere. The facts we have so far suggest that if the claim is true, then the center of the hypersphere should be $\mathbf{r} + (1/\kappa)\mathbf{N}$. Setting $\mathbf{p}(\mathbf{u}) = \mathbf{r}(\mathbf{u}) + (1/\kappa)\mathbf{N}(\mathbf{u})$, we have to make sure first that $\mathbf{p}(\mathbf{u})$ does not depend on $(\mathbf{u})$. Indeed, using the fact that κ is constant, differentiation with respect to the ith variable results

$$\mathbf{p}_i = \mathbf{r}_i + (1/\kappa)\mathbf{N}_i = \mathbf{r}_i - (1/\kappa)\kappa \mathbf{r}_i = \mathbf{0}$$

for any i.
Now to show that the surface lies on a hypersphere centered at $\mathbf{p}$ we have to prove that the function $\|\mathbf{r} - \mathbf{p}\|$ is constant. This is clear as

$$\|\mathbf{r}(\mathbf{u}) - \mathbf{p}\| = \left\| \frac{1}{\kappa} \mathbf{N}(\mathbf{u}) \right\| = \left| \frac{1}{\kappa} \right|$$

is constant. The theorem is proved. □

3.1.5 The Fundamental Equations of Hypersurface Theory

Now we derive some formulae for hypersurfaces. Consider a regular parameterized hypersurface $\mathbf{r} \colon \Omega \to \mathbb{R}^{n-1}$. The partial derivatives $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$ define a basis of the tangent space of the hypersurface at each point. If we add to these vectors the normal vector of the hypersurface, we get a basis of $\mathbb{R}^n$ at each point of the hypersurface. The system of the vector fields $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}, \mathbf{N}$ along $\mathbf{r}$ is called the *Gauss frame* of the hypersurface. Gauss frame plays similar role in the theory of hypersurfaces as Frenet frame does in curve theory. Similarity is not complete however, since a Gauss frame is much more dependent on the parameterization. Nevertheless, in the same way as for Frenet frames, it is important to know how the derivatives of the frame vector fields with respect to the parameters can be expressed as a linear combination of the frame vectors. For this we have to determine the coefficients $\Gamma_{ij}^k, \alpha_{ij}, \beta_j^k, \gamma_j$ in the expressions

$$\mathbf{r}_{ij} = \sum_{k=1}^{n-1} \Gamma_{ij}^k \mathbf{r}_k + \alpha_{ij}\mathbf{N}, \qquad \mathbf{N}_j = \sum_{k=1}^{n-1} \beta_j^k \mathbf{r}_k + \gamma_j \mathbf{N}. \tag{3.4}$$

Let us begin with the simple observation that since $\mathbf{N}_j$ is known to be tangential, and $\mathbf{N}_j = -L_{\mathbf{r}}(\mathbf{r}_j)$, where $L_{\mathbf{r}}$ is the Weingarten map,

$$\gamma_j = 0 \text{ for all } j,$$

and $(-\beta_j^k)_{j,k=1}^{n-1}$ is the matrix $\mathcal{L} = \mathcal{G}^{-1}\mathcal{B}$ of the Weingarten map with respect to the basis $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$. Denote by g_{ij} and b_{ij} the entries of the first and second fundamental forms as usual, and denote by g^{ij} the components of the inverse matrix of the matrix of the first fundamental form. (Attention! Entries of $\mathcal{G}$ and $\mathcal{G}^{-1}$ are distinguished by the position of indices.) Then

$$\beta_j^k = -l_j^k = -\sum_{i=1}^{n-1} g^{ki} b_{ij}.$$

Taking the dot product of the first equation of (3.4) with $\mathbf{N}$ we gain the equality $\langle \mathbf{r}_{ij}, \mathbf{N}\rangle = \alpha_{ij}$ and since $\langle \mathbf{r}_{ij}, \mathbf{N}\rangle = b_{ij}$,

$$\alpha_{ij} = b_{ij} \text{ for all } i, j.$$

There is only one question left: What are the coefficients Γ_{ij}^k equal to? Let us take the dot product of the first equation of (3.4) with $\mathbf{r}_l$

$$\langle \mathbf{r}_{ij}, \mathbf{r}_l\rangle = \sum_{k=1}^{n-1} \Gamma_{ij}^k \langle \mathbf{r}_k, \mathbf{r}_l\rangle = \sum_{k=1}^{n-1} \Gamma_{ij}^k g_{kl},$$

or denoting the dot product $\langle \mathbf{r}_{ij}, \mathbf{r}_l\rangle$ shortly by Γ_{ijl},

$$\Gamma_{ijl} = \sum_{k=1}^{n-1} \Gamma_{ij}^k g_{kl}.$$

The coefficients Γ_{ij}^k and Γ_{ijk} are called the *Christoffel symbols of the first and second type* respectively. The last equation shows how to express Christoffel symbols of the second type with the help of Christoffel symbols of the first type. It can also be used to express Christoffel symbols of the first type in terms of secondary Christoffel symbols. Indeed, multiplying the equation with g^{ls}, summing up for l and using $\sum_{l=1}^{n-1} g_{kl} g^{ls} = \delta_k^s$, we get

$$\sum_{l=1}^{n-1} \Gamma_{ijl} g^{ls} = \sum_{l=1}^{n-1}\sum_{k=1}^{n-1} \Gamma_{ij}^k g_{kl} g^{ls} = \sum_{k=1}^{n-1} \Gamma_{ij}^k \delta_k^s = \Gamma_{ij}^s.$$

Now let us determine the Christoffel symbols of the second type. Differentiating the equality $g_{ij} = \langle \mathbf{r}_i, \mathbf{r}_j \rangle$ with respect to the k-th variable and then permuting the role of the indices i, j, k we get the equalities

$$\begin{aligned} g_{ij,k} &= \langle \mathbf{r}_{ik}, \mathbf{r}_j \rangle + \langle \mathbf{r}_i, \mathbf{r}_{jk} \rangle, \\ g_{jk,i} &= \langle \mathbf{r}_{ji}, \mathbf{r}_k \rangle + \langle \mathbf{r}_j, \mathbf{r}_{ki} \rangle, \\ g_{ki,j} &= \langle \mathbf{r}_{kj}, \mathbf{r}_i \rangle + \langle \mathbf{r}_k, \mathbf{r}_{ij} \rangle. \end{aligned}$$

Solving this linear system of equations for the secondary Christoffel symbols standing on the right-hand side, we obtain

$$\Gamma_{ijk} = \langle \mathbf{r}_{ij}, \mathbf{r}_k \rangle = \frac{1}{2}(g_{ik,j} + g_{jk,i} - g_{ij,k})$$

and

$$\Gamma_{ij}^k = \sum_{l=1}^{n-1} \Gamma_{ijl} g^{lk} = \sum_{l=1}^{n-1} \frac{1}{2}(g_{il,j} + g_{jl,i} - g_{ij,l}) g^{lk}. \tag{3.5}$$

Observe that the Christoffel symbols *depend only on the first fundamental form* of the hypersurface.

Now we ask the following question. Suppose we are given $2(n-1)^2$ smooth functions g_{ij}, b_{ij} $i, j = 1, 2, \ldots, n-1$ on an open domain Ω of $\mathbb{R}^{n-1}$. When can we find a regular parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ the fundamental forms of which are represented by the matrices $\mathcal{G} = (g_{ij})$ and $\mathcal{B} = (b_{ij})$. We have some obvious restrictions on the functions g_{ij} and b_{ij}. First, $g_{ij} = g_{ji}$, $b_{ij} = b_{ji}$, and since $\mathcal{G}$ is the matrix of a positive definite bilinear form, the determinants of the corner submatrices $(g_{ij})_{i,j=1}^k$ must be positive for $k = 1, \ldots, n-1$. However, several examples show that these conditions are not enough to guarantee the existence of a hypersurface. For example, if $\mathcal{G}$ is the identity matrix everywhere, while $\mathcal{B} = f\mathcal{G}$ for some function on Ω, then the hypersurface (if exists) consists of umbilics. We know however that if a surface consists of umbilics, then the principal curvatures are constant, so although our choice of $\mathcal{B}$ and $\mathcal{G}$ satisfies all the conditions we have listed so far, it does not correspond to a hypersurface unless f is constant. So there must be some further relations between the components of $\mathcal{B}$ and $\mathcal{G}$. Our plan to find some of these correlations is the following. Let us express $\mathbf{r}_{ijk}$ and $\mathbf{r}_{ikj}$ as a linear combination of the Gauss frame vectors. The coefficients we get are functions of the entries of the first and second fundamental forms. Since $\mathbf{r}_{ijk} = \mathbf{r}_{ikj}$, the corresponding coefficients in the expressions for these vectors must be equal and it can be hoped that this way we arrive at further non-trivial relations between $\mathcal{G}$ and $\mathcal{B}$. This was the philosophy, and now let us perform the computation.

$$\begin{aligned}\mathbf{r}_{ijk} &= \left(\sum_{l=1}^{n-1}\Gamma_{ij}^{l}\mathbf{r}_l + b_{ij}\mathbf{N}\right)_{,k} = \sum_{l=1}^{n-1}(\Gamma_{ij,k}^{l}\mathbf{r}_l + \Gamma_{ij}^{l}\mathbf{r}_{lk}) + b_{ij,k}\mathbf{N} + b_{ij}\mathbf{N}_k \\ &= \sum_{l=1}^{n-1}\left(\Gamma_{ij,k}^{l}\mathbf{r}_l + \Gamma_{ij}^{l}\Big(\sum_{s=1}^{n-1}\Gamma_{lk}^{s}\mathbf{r}_s + b_{lk}\mathbf{N}\Big)\right) + b_{ij,k}\mathbf{N} - b_{ij}\sum_{l=1}^{n-1}\sum_{s=1}^{n-1} b_{ks}g^{sl}\mathbf{r}_l \\ &= \sum_{l=1}^{n-1}(\Gamma_{ij,k}^{l} + \sum_{s=1}^{n-1}\Gamma_{ij}^{s}\Gamma_{sk}^{l} - b_{ij}\sum_{s=1}^{n-1} b_{ks}g^{sl})\mathbf{r}_l + (b_{ij,k} + \sum_{l=1}^{n-1}\Gamma_{ij}^{l}b_{lk})\mathbf{N}.\end{aligned}$$

Comparing the coefficient of $\mathbf{r}_l$ in $\mathbf{r}_{ijk}$ and $\mathbf{r}_{ikj}$, we obtain

$$\Gamma_{ij,k}^{l} - \Gamma_{ik,j}^{l} + \sum_{s=1}^{n-1}(\Gamma_{ij}^{s}\Gamma_{sk}^{l} - \Gamma_{ik}^{s}\Gamma_{sj}^{l}) = \sum_{s=1}^{n-1}(b_{ij}b_{ks} - b_{ik}b_{js})g^{sl},$$

while comparison of the coefficient of $\mathbf{N}$ gives

$$b_{ij,k} - b_{ik,j} = \sum_{l=1}^{n-1}(\Gamma_{ik}^{l}b_{lj} - \Gamma_{ij}^{l}b_{lk}).$$

The first $(n-1)^4$ equations (we have an equation for all i, j, k, l), are the *Gauss equations* for the hypersurface. The second family of $(n-1)^3$ equations are the *Codazzi–Mainardi equations*.

Exercise 3.1.26. Express the second order derivatives $\mathbf{N}_{ij}$ and $\mathbf{N}_{ji}$ as a linear combination of the Gauss frame vectors. Compare the corresponding coefficients and prove that their equality follows from the Gauss and Codazzi–Mainardi equations. ¤

The exercise points out that a similar try to derive new relations between $\mathcal{G}$ and $\mathcal{B}$ does not lead to really new results. This is no wonder, since the Gauss and Codazzi–Mainardi equations together with the previously listed obvious conditions on $\mathcal{G}$ and $\mathcal{B}$ form a complete system of necessary and sufficient conditions for the existence of a hypersurface with fundamental forms $\mathcal{G}$ and $\mathcal{B}$.

Theorem 3.1.27 (Fundamental theorem of hypersurfaces)**.** *Let $\Omega \subset \mathbb{R}^n$ be an open connected and simply connected subset of $\mathbb{R}^{n-1}$ (e.g. an open ball or cube), and suppose that we are given two smooth $(n-1)\times(n-1)$ matrix valued functions $\mathcal{G}$ and $\mathcal{B}$ on Ω such that $\mathcal{G} = (g_{ij})$ and $\mathcal{B} = (b_{ij})$ assign to every point a symmetric matrix, $\mathcal{G}$ gives the matrix of a positive definite bilinear form. In this case, if the functions Γ_{ij}^{k} derived from the components of $\mathcal{G}$ according to* *the formulae (3.5) satisfy the Gauss and Codazzi–Mainardi equations, then there exists a regular parameterized hypersurface* $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ *for which the matrix representations of the first and second fundamental forms are* $\mathcal{G}$ *and* $\mathcal{B}$ *respectively. Furthermore, this hypersurface is unique up to rigid motions of the whole space. In other words, if* $\mathbf{r}_1$ *and* $\mathbf{r}_2$ *are two such hypersurfaces, then there exists an orientation preserving isometry* $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ *for which* $\mathbf{r}_2 = \Phi \circ \mathbf{r}_1$.

Let us denote the expressions standing on the left-hand sides of the Gauss equations by

$$R^l_{kji} := \Gamma^l_{ij,k} - \Gamma^l_{ik,j} + \sum_{s=1}^{n-1}(\Gamma^s_{ij}\Gamma^l_{sk} - \Gamma^s_{ik}\Gamma^l_{sj}).$$

Then Gauss equations can be abbreviated writing

$$R^l_{kji} = \sum_{s=1}^{n-1}(b_{ij}b_{ks} - b_{ik}b_{js})g^{sl}.$$

Let us multiply this equation by g_{lm} and take a summation for l

$$\begin{aligned}\sum_{l=1}^{n-1} R^l_{kji}g_{lm} &= \sum_{l=1}^{n-1}\sum_{s=1}^{n-1}(b_{ij}b_{ks} - b_{ik}b_{js})g^{sl}g_{lm} \\ &= \sum_{s=1}^{n-1}(b_{ij}b_{ks} - b_{ik}b_{js})\delta^s_m = (b_{ij}b_{km} - b_{ik}b_{jm}).\end{aligned}$$

Introducing the functions $R_{kjim} := \sum_{l=1}^{n-1} R^l_{kji}g_{lm}$, we may write

$$R_{kjim} = (b_{ij}b_{km} - b_{ik}b_{jm}).$$

Let us observe, that the functions R_{imjk} can be expressed in terms of the first fundamental form $\mathcal{G}$.

Corollary 3.1.28 (Theorema Egregium)**.** *The Gaussian curvature of a regular parameterized surface in* $\mathbb{R}^3$ *can be expressed in terms of the first fundamental form as follows*

$$K = \frac{R_{1221}}{\det \mathcal{G}}.$$

Theorema Egregium (meaning Remarkable Theorem in Latin) is one of those theorems of Gauss he was very proud of. The surprising fact is not the actual form of this formula but the mere existence of a formula that expresses the Gaussian curvature in terms of the first fundamental form. The geometrical

meaning of the existence of such a formula is that the *Gaussian curvature does not change when we bend the surface.*
Bending of a hypersurface is a deformation of it which preserves the lengths of curves drawn onto the hypersurface. For example, when a flat sheet of paper is bent into a cylinder, or a cone, it is a bending in the mathematical sense as well, but blowing up a balloon is not a bending, because it increases the lengths of curves drawn on the balloon.
If we choose an injective regular parameterization $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ of a hypersurface $M = \mathbf{r}(\Omega)$ and the mapping $\Phi\colon M \to \mathbb{R}^n$ describes where a point of M goes when we deform the surface, then $\tilde{\mathbf{r}} = \Phi \circ \mathbf{r}\colon \Omega \to \mathbb{R}^n$ is a parameterization of the image hypersurface $\Phi(M)$. The deformation $\Phi = \tilde{\mathbf{r}} \circ \mathbf{r}^{-1}$ is uniquely determined by the parameterizations $\mathbf{r}$ and $\tilde{\mathbf{r}}$. This motivates the following formal definition.

Definition 3.1.29. We say that the regular parameterized hypersurfaces $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ and $\tilde{\mathbf{r}}\colon \Omega \to \mathbb{R}^n$ are *obtained from one another by a bending* if for any smooth curve $\mathbf{u}\colon [a,b] \to \Omega$, the lengths of the curves $\gamma = \mathbf{r} \circ \mathbf{u}$ and $\tilde{\gamma} = \tilde{\mathbf{r}} \circ \mathbf{u}$ are equal. ※

The connection between bendings and the first fundamental form is given in the following simple proposition.

Proposition 3.1.30. *The parameterized hypersurfaces $r\colon \Omega \to \mathbb{R}^n$ and $\tilde{\mathbf{r}}\colon \Omega \to \mathbb{R}^n$ are obtained from one another by a bending if and only if the matrix $\mathcal{G}$ of the first fundamental form of $\mathbf{r}$ with respect to the basis $(\mathbf{r}_1, \dots, \mathbf{r}_{n-1})$ is the same as the matrix $\tilde{\mathcal{G}}$ of the first fundamental form of $\tilde{\mathbf{r}}$ with respect to the basis $(\tilde{\mathbf{r}}_1, \dots, \tilde{\mathbf{r}}_{n-1})$.*

Proof. Let $\mathbf{u}\colon [a,b] \to \Omega$, $\mathbf{u}(t) = (u^1(t), \dots, u^{n-1}(t))$ be a smooth curve. Then the length of the curve $\gamma = \mathbf{r} \circ \mathbf{u}$ is equal to

$$\begin{aligned} l_\gamma &= \int_a^b \|\gamma'(t)\| dt = \int_a^b \sqrt{\Big\langle \sum_{i=1}^{n-1} \mathbf{r}_i(\mathbf{u}(t)) u^{i\prime}(t), \sum_{j=1}^{n-1} \mathbf{r}_j(\mathbf{u}(t)) u^{j\prime}(t) \Big\rangle}\, dt \\ &= \int_a^b \sqrt{\sum_{i,j=1}^{n-1} u^{i\prime}(t) u^{j\prime}(t) g_{ij}(\mathbf{u}(t))}\, dt, \end{aligned}$$

where the functions $g_{ij} = \langle \mathbf{r}_i, \mathbf{r}_j \rangle$ are the entries of the matrix $\mathcal{G}$. Similarly, the length of $\tilde{\gamma} = \tilde{\mathbf{r}} \circ \mathbf{u}$ can be expressed as

$$l_{\tilde{\gamma}} = \int_a^b \sqrt{\sum_{i,j=1}^{n-1} u^{i\prime}(t) u^{j\prime}(t) \tilde{g}_{ij}(\mathbf{u}(t))}\, dt,$$

where $\tilde{g}_{ij} = \langle \tilde{\mathbf{r}}_i, \tilde{\mathbf{r}}_j \rangle$. Comparing these two equations we see at once that if $\mathcal{G} = \tilde{\mathcal{G}}$, then $l_\gamma = l_{\tilde{\gamma}}$ for all smooth curves $\mathbf{u}\colon [a,b] \to \Omega$.
Conversely, assume that $l_\gamma = l_{\tilde{\gamma}}$ holds for any smooth curve $\mathbf{u}\colon [a,b] \to \Omega$. Then, in particular, this holds also for the restriction of the curve $\mathbf{u}$ onto the interval $[a,x]$ for any $x \in (a,b]$, consequently

$$\int_a^x \sqrt{\sum_{i,j=1}^{n-1} u^{i\prime}(t) u^{j\prime}(t) g_{ij}(\mathbf{u}(t))}\, dt = \int_a^x \sqrt{\sum_{i,j=1}^{n-1} u^{i\prime}(t) u^{j\prime}(t) \tilde{g}_{ij}(\mathbf{u}(t))}\, dt \qquad \text{for all } x \in (a,b].$$

Differentiating this equation with respect to x at $x_0 \in (a,b)$ we obtain

$$\sqrt{\sum_{i,j=1}^{n-1} u^{i\prime}(x_0) u^{j\prime}(x_0) g_{ij}(\mathbf{u}(x_0))} = \sqrt{\sum_{i,j=1}^{n-1} u^{i\prime}(x_0) u^{j\prime}(x_0) \tilde{g}_{ij}(\mathbf{u}(x_0))},$$

or equivalently,

$$\sum_{i,j=1}^{n-1} u^{i\prime}(x_0) u^{j\prime}(x_0) g_{ij}(\mathbf{u}(x_0)) = \sum_{i,j=1}^{n-1} u^{i\prime}(x_0) u^{j\prime}(x_0) \tilde{g}_{ij}(\mathbf{u}(x_0)) \quad \text{for all } x_0 \in (a,b). \tag{3.6}$$

Let $\mathbf{u}_0 \in \Omega$ be in arbitrary point of the parameter domain, $\mathbf{v} = (v^1, \dots, v^{n-1}) \in \mathbb{R}^{n-1}$ be an arbitrary vector, and consider the curve $\mathbf{u} : [a,b] \to \mathbb{R}^{n-1}$, $\mathbf{u}(t) = \mathbf{u}_0 + t\mathbf{v}$. If the interval $[a,b]$ is a sufficiently small neighborhood of 0, then the image of $\mathbf{u}$ is in Ω and we can apply (3.6) to $\mathbf{u}$ at $x_0 = 0$. This gives

$$\sum_{i,j=1}^{n-1} v^i v^j g_{ij}(\mathbf{u}_0) = \sum_{i,j=1}^{n-1} v^i v^j \tilde{g}_{ij}(\mathbf{u}_0) \quad \text{for any } \mathbf{u}_0 \in \Omega \text{ and } \mathbf{v} \in \mathbb{R}^{n-1}.$$

For a fixed $\mathbf{u}_0$, both sides of this equation is a quadratic polynomial of the coordinates of $\mathbf{v}$. These two quadratic polynomials are equal, so the coefficients of each monomial are equal. Comparing the coefficients of the monomial $v^i v^j$ we obtain the equality

$$\frac{1}{1+\delta_{ij}}(g_{ij}(\mathbf{u}_0) + g_{ji}(\mathbf{u}_0)) = \frac{1}{1+\delta_{ij}}(\tilde{g}_{ij}(\mathbf{u}_0) + \tilde{g}_{ji}(\mathbf{u}_0)).$$

Since $\mathcal{G}$ and $\tilde{\mathcal{G}}$ are symmetric matrices, this equation gives $\mathcal{G} = \tilde{\mathcal{G}}$. $\square$

According to the proposition, all quantities that can be expressed in terms of the entries of the matrix of the first fundamental form, in particular the Gaussian curvature, are invariant under bendings. The bending invariance of the Gaussian curvature is remarkable because the principal curvatures of a surface do change in general when the surface is bent.

Definition 3.1.31. Let $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ be a hypersurface. Consider the mapping R that assigns four tangential vector fields $\mathbf{X} = \sum_{i=1}^{n-1} X^i \mathbf{r}_i$, $\mathbf{Y} = \sum_{i=1}^{n-1} Y^i \mathbf{r}_i$, $\mathbf{Z} = \sum_{i=1}^{n-1} Z^i \mathbf{r}_i$, $\mathbf{W} = \sum_{i=1}^{n-1} W^i \mathbf{r}_i$ a function according to the formula

$$R(X, Y; Z, W) = \sum_{i=1}^{n-1} \sum_{m=1}^{n-1} \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} R_{imjk} X^i Y^m Z^j W^k.$$

We shall call R the *Riemannian curvature tensor* of the hypersurface, the functions R_{imjk} the *components of the curvature tensor.* ※

Recall that a *tensor of type* (k, l) *over a linear space* V can be identified with a multilinear function

$$T\colon \underbrace{V \times \cdots \times V}_{k} \times \underbrace{V^* \times \cdots \times V^*}_{l} \to \mathbb{R}$$

defined on the Cartesian product of k copies of V and l copies of V. “Multilinear” means that fixing all but one variables, we obtain a linear function of the free variable. If $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is a basis of V, $(\mathbf{e}^1, \ldots, \mathbf{e}^n)$ is its dual basis, then every tensor T is uniquely determined by its values on basis vector combinations, i.e. by the numbers

$$T^{i_1 \ldots i_l}_{j_1 \ldots j_k} = T(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}; \mathbf{e}^{i_1}, \ldots, \mathbf{e}^{i_l}),$$

which are called the *components of the tensor* T *with respect to the basis* $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Now consider a regular parameterized hypersurface M, $\mathbf{r}\colon \Omega \to \mathbb{R}^n$. A *tensor field of type* (k, l) over M is a mapping T that assigns to every point $\mathbf{u} \in \Omega$ a tensor of type (k, l) over the tangent space of M at $\mathbf{r}(\mathbf{u})$. $T(\mathbf{u})$ is uniquely determined by its components $T^{i_1 \ldots i_l}_{j_1 \ldots j_k}(\mathbf{u})$ with respect to the basis $\mathbf{r}_1(\mathbf{u}), \ldots, \mathbf{r}_{n-1}(\mathbf{u})$. The functions $\mathbf{u} \mapsto T^{i_1 \ldots i_l}_{j_1 \ldots j_k}(\mathbf{u})$ are called the *components of the tensor field* T. T is said to be a *smooth tensor field* if its components are smooth.

Examples.

- Functions on M are tensor fields of type $(0, 0)$.
- Tangential vector fields are tensor fields of type $(0, 1)$ (V is isomorphic to V^{**} in a natural way).

- The first and second fundamental forms of a hypersurface are tensor fields of type $(2,0)$.
- The mapping that assigns to every point of a hypersurface the Weingarten map at that point is a tensor of type $(1,1)$. (The linear space of $V \to V$ linear mappings is isomorphic to $T^{(1,1)}V$ in a natural way.)
- Let f be a smooth function on M. Consider the tensor field of type $(1,0)$ defined on a tangent vector X to be the derivative of f in the direction X. This tensor field is the differential of f.
- The Riemannian curvature tensor is a tensor field of type $(4,0)$.

Curvature tensor is one of the most basic objects of study in differential geometry. However, Definition 3.1.31 does not cast any light on its geometric content. To understand its real meaning, we shall introduce the curvature tensor in a more natural way on a more general level in the framework of Riemannian manifolds. For this purpose, we have to get acquainted with some fundamental definitions and constructions. This will be the goal of the next chapter.

Exercise 3.1.32. Determine the components of the Riemannian curvature tensor of the plane with respect to the polar parameterization

$$\mathbf{r}(u,v) = (u\cos v, u\sin v, 0).$$ ¤

Exercise 3.1.33. Compute the Christoffel symbols of the second kind for surfaces of revolution with respect to the parameterization

$$\mathbf{r}(u,v) = (x(u)\cos v, x(u)\sin v, y(u)),$$

where $x(u)$ and $y(u)$ are given functions of one variable. ¤

Exercise 3.1.34. The components of the first fundamental form of a surface with respect to a given parameterization are

$$g_{uu} = g_{vv} = 1, \quad g_{uv} = 0.$$

Show that the Gauss curvature of the surface is constant 0. ¤

Exercise 3.1.35. The components of the first fundamental form of a surface with respect to a given parameterization defined on an open subset of the upper half-plane $\{(u,v) \mid v > 0\}$ are

$$g_{uu} = g_{vv} = \frac{1}{v^2}, \quad g_{uv} = 0.$$

Show that the Gauss curvature of the surface is constant -1. (*Poincaré's half-plane model of the hyperbolic plane.*) ¤

Exercise 3.1.36. Show that the curvature tensor satisfies the following symmetry relations:

$$R(X,Y;Z,W) = -R(Y,X;Z,W) = -R(X,Y;W,Z);$$

$$R(X,Y;Z,W) + R(X,Z;W,Y) + R(X,W;Y,Z) = 0 \text{ (\textit{Bianchi identity})}. \quad ¤$$

3.1.6 Surface Volume

In this section we define the surface volume of a domains on a hypersurface. Let $\mathbf{r}\colon \Omega \to \mathbf{R}^n$ be a regular parameterized hypersurface in $\mathbb{R}^n$, denote by $\mathcal{G}\colon \Omega \to \mathbb{R}^{(n-1)\times(n-1)}$ the matrix of the first fundamental form with respect to the basis $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$.

Definition 3.1.37. $D \subset \Omega$ be a Lebesgue-measurable subset. We define the *surface volume* $\mu_{n-1}(\mathbf{r}(D))$ *of* $\mathbf{r}(D)$ as the integral

$$\mu_{n-1}(\mathbf{r}(D)) = \int_D \sqrt{\det \mathcal{G}}\, d\lambda_{n-1},$$

where the integral is taken with respect to the Lebesgue measure λ_{n-1} on $\mathbb{R}^{n-1}$. In the case $n = 3$, $\mu_2(\mathbf{r}(D))$ is also called the *surface area of* $\mathbf{r}(D)$. ※

The motivation for choosing the function $\sqrt{\det \mathcal{G}}$ for the integrand is that geometrically, $\sqrt{\det \mathcal{G}(\mathbf{u})}$ is the volume of the $(n-1)$-dimensional parallelepiped spanned by the partial derivatives $\mathbf{r}_1(\mathbf{u}), \dots, \mathbf{r}_{n-1}(\mathbf{u})$. Since these partial derivatives are the images of the standard basis vectors of $\mathbb{R}^{n-1}$ under the derivative map of $\mathbf{r}$ at $\mathbf{u}$, and the standard basis vectors of $\mathbb{R}^{n-1}$ span a unit cube of volume 1, $\sqrt{\det \mathcal{G}(\mathbf{u})}$ is the factor by which the derivative map $\mathbf{r}'(\mathbf{u})$ stretches or shrinks $(n-1)$-dimensional volume.
A rigorous statement, which justifies the definition in another way is the following one.

Proposition 3.1.38. *Suppose that the hypersurfaces* $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ *and* $\tilde{\mathbf{r}}\colon \tilde{\Omega} \to \mathbb{R}^n$ *are reparameterizations of one another, that is, assume there is a diffeomorphism* $h\colon \Omega \to \tilde{\Omega}$ *such that* $\mathbf{r} = \tilde{\mathbf{r}} \circ \mathbf{u}$. *Denote by* $\mathcal{G}$ *and* $\tilde{\mathcal{G}}$ *the matrices of the first fundamental forms of* $\mathbf{r}$ *and* $\tilde{\mathbf{r}}$ *respectively. Then for any Lebesgue-measurable subset* $D \subset \Omega$,

$$\int_D \sqrt{\det \mathcal{G}}\, d\lambda_{n-1} = \int_{h(D)} \sqrt{\det \tilde{\mathcal{G}}}\, d\lambda_{n-1}.$$

For this reason, the surface volume of $\mathbf{r}(D) = \tilde{\mathbf{r}}(h(D))$ *does not depend on the way we parameterize the domain.*

Proof. We apply the theorem on the change of variables in multivariable integrals. This gives at once that

$$\int_D |\det h'| \sqrt{\det \tilde{\mathcal{G}}} \circ h \, d\lambda_{n-1} = \int_{h(D)} \sqrt{\det \tilde{\mathcal{G}}} \, d\lambda_{n-1},$$

so we have only to prove $|\det h'| \sqrt{\det \tilde{\mathcal{G}}} \circ h = \sqrt{\det \mathcal{G}}$.
By the chain rule we have

$$\mathbf{r}_i(\mathbf{u}) = \sum_{j=1}^{n-1} \tilde{\mathbf{r}}_j(h(\mathbf{u})) \partial_i h^j(\mathbf{u}),$$

where $(h^1, \dots, h^{n-1})$ are the coordinate functions of h. Taking the wedge product of these equations and applying Proposition 1.2.31 we obtain

$$\mathbf{r}_1(\mathbf{u}) \wedge \cdots \wedge \mathbf{r}_{n-1}(\mathbf{u}) = \det h'(\mathbf{u}) \cdot \tilde{\mathbf{r}}_1(h(\mathbf{u})) \wedge \cdots \wedge \tilde{\mathbf{r}}_{n-1}(h(\mathbf{u})).$$

Computing the length of the $(n-1)$-vectors on the two sides, with the help of equation (1.9), we get the required equality

$$\sqrt{\det \mathcal{G}(\mathbf{u})} = |\det h'(\mathbf{u})| \sqrt{\det \tilde{\mathcal{G}}(h(\mathbf{u}))}. \qquad \square$$

3.2 Surfaces in $\mathbb{R}^3$

A regular parameterized surface $\mathbf{r}\colon \Omega \to \mathbb{R}^3$ (Ω is an open subset of the plane) has two principal curvatures $\kappa_1(u,v)$ and $\kappa_2(u,v)$ at each point $p = \mathbf{r}(u,v)$ of the surface. If $\kappa_1(u,v) \leq \kappa_2(u,v)$ then $\kappa_1(u,v)$ is the minimum of normal curvatures in different directions at p, while $\kappa_2(u,v)$ is the maximum of them. If $\kappa_1(u,v) < \kappa_2(u,v)$ then the principal directions corresponding to $\kappa_1(u,v)$ and $\kappa_2(u,v)$ are uniquely defined, however if $\kappa_1(u,v) = \kappa_2(u,v)$ then the point $\mathbf{r}(u,v)$ is umbilical, the normal curvature is constant in all directions and every direction is principal.

3.2.1 Surfaces of Revolution

The next example shows how to compute the principal curvatures and directions for a surface of revolution. Consider a regular plane curve $\gamma = (x,y)\colon I \to \mathbb{R}^2$, with positive second coordinate function $y > 0$, and rotate the curve about the x-axis. This obtained *surface of revolution* can be parameterized by the mapping

$$\mathbf{r}(u,v) = (x(u), y(u)\cos v, y(u)\sin v).$$

The coordinate lines of this parameterization are the rotations of the generating curve γ, and the circles drawn by points of the generator curve as they rotate about the x-axis. Rotated copies of the generating curve are called the *generators*, *generatrices* or *meridians* of the surface of revolution, while the circles are the so-called *circles of latitude*.
The tangent vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ are obtained by partial differentiation

$$\begin{aligned}\mathbf{r}_u(u,v) &= (x'(u), y'(u)\cos v, y'(u)\sin v),\\ \mathbf{r}_v(u,v) &= (0, -y(u)\sin v, y(u)\cos v).\end{aligned}$$

The matrix of the first fundamental form with respect to the basis $\mathbf{r}_u, \mathbf{r}_v$ is

$$\mathcal{G} = \begin{pmatrix} \langle \mathbf{r}_u, \mathbf{r}_u\rangle & \langle \mathbf{r}_u, \mathbf{r}_v\rangle \\ \langle \mathbf{r}_v, \mathbf{r}_u\rangle & \langle \mathbf{r}_v, \mathbf{r}_v\rangle \end{pmatrix} = \begin{pmatrix} x'^2(u) + y'^2(u) & 0 \\ 0 & y^2(u) \end{pmatrix}.$$

To obtain the matrix of the second fundamental form we need the normal vector field and the second order partial derivatives of $\mathbf{r}$.

$$\begin{aligned}\mathbf{r}_u \times \mathbf{r}_v &= \det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x'(u) & y'(u)\cos v & y'(u)\sin v \\ 0 & -y(u)\sin v & y(u)\cos v \end{pmatrix}\\ &= (y'(u)y(u), -y(u)x'(u)\cos v, -y(u)x'(u)\sin v),\end{aligned}$$

$$\mathbf{N} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|} = \frac{1}{\sqrt{x'^2(u) + y'^2(u)}}(y'(u), -x'(u)\cos v, -x'(u)\sin v).$$

The second order partial derivatives of the parameterization are

$$\begin{aligned}\mathbf{r}_{uu}(u,v) &= (x''(u), y''(u)\cos v, y''(u)\sin v),\\ \mathbf{r}_{uv}(u,v) &= (0, -y'(u)\sin v, y'(u)\cos v),\\ \mathbf{r}_{vv}(u,v) &= (0, -y(u)\cos v, -y(u)\sin v).\end{aligned}$$

The matrix of the second fundamental form is

$$\begin{aligned}\mathcal{B} &= \begin{pmatrix} \langle \mathbf{N}, \mathbf{r}_{uu}\rangle & \langle \mathbf{N}, \mathbf{r}_{uv}\rangle \\ \langle \mathbf{N}, \mathbf{r}_{vu}\rangle & \langle \mathbf{N}, \mathbf{r}_{vv}\rangle \end{pmatrix}\\ &= \frac{1}{\sqrt{x'^2(u) + y'^2(u)}} \begin{pmatrix} x''(u)y'(u) - y''(u)x'(u) & 0 \\ 0 & x'(u)y(u) \end{pmatrix}.\end{aligned}$$

The matrix of the Weingarten map with respect to the basis $\mathbf{r}_u, \mathbf{r}_v$ is

$$\mathcal{L} = \mathcal{B}\mathcal{G}^{-1} = \begin{pmatrix} \dfrac{x''(u)y'(u) - y''(u)x'(u)}{(x'^2(u) + y'^2(u))^{3/2}} & 0 \\ 0 & \dfrac{x'(u)}{y(u)(x'^2(u) + y'^2(u))^{1/2}} \end{pmatrix}.$$

As we see, the matrix of the Weingarten map is diagonal, consequently $\mathbf{r}_u, \mathbf{r}_v$ are eigenvectors and the diagonal elements of $\mathcal{L}$ are eigenvalues of the Weingarten map. Thus the principal curvatures of the surface are

$$\kappa_1(u,v) = \frac{x''(u)y'(u) - y''(u)x'(u)}{(x'^2(u) + y'^2(u))^{3/2}} \qquad \kappa_2(u,v) = \frac{x'(u)}{y(u)(x'^2(u) + y'^2(u))^{1/2}}.$$

We could have obtained this result in a more geometrical way. For any point p on the surface, the plane through p and the x-axis is a symmetry plane of the surface. Thus, reflection of a principal direction of the surface at p is also a principal direction (with the same principal curvature). The principal curvatures at p are either equal and then every direction is principal, or different and then the principal directions are unique. Since a direction is invariant under a reflection in a plane if and only if it is parallel or orthogonal to the plane, we may conclude that $\mathbf{r}_u$ and $\mathbf{r}_v$ are principal directions of the surface. Principal curvatures are the curvatures of the normal sections of the surface in the direction $\mathbf{r}_u, \mathbf{r}_v$.

The normal section of the surface in the direction $\mathbf{r}_u$ is a meridian of the surface. Its curvature can be calculated according to the formula known for plane curves and gives κ_1 up to sign. The difference in sign is due to the fact that the unit normal of the surface and the principal normal of the meridian are opposite to one another.

The plane passing through p perpendicular to the x-axis intersects the surface in a circle of latitude the tangent of which at p is $\mathbf{r}_v$. The curvature of this circle is $\frac{1}{y(u)}$. The normal curvature $\kappa_2 = k(\mathbf{r}_v)$ of the surface in the direction $\mathbf{r}_v$ and the curvature of the circle intersection are related to one another by Meusnier's theorem as follows

$$\frac{1}{y(u)} = \frac{1}{\cos\alpha}\kappa_2,$$

where α is the angle between the normal of the surface and the principal normal of the circle. As it is easy to see, α is the direction angle of the tangent to the meridian at p, that is, by elementary calculus $\tan\alpha = y'(u)/x'(u)$, from which $\cos\alpha = \frac{x'(u)}{\sqrt{x'^2(u) + y'^2(u)}}$.

Therefore, we get

$$\kappa_2(u,v) = \frac{x'(u)}{y(u)\sqrt{x'^2(u) + y'^2(u)}}$$

as before. The equation $\frac{1}{y(u)} = \frac{1}{\cos\alpha}\kappa_2$ has the following consequence.

Corollary 3.2.1. *The second principal radius of curvature* $\dfrac{1}{\kappa_2}$ *of a surface of revolution at a given point* p *is the length of the segment of the normal of the surface between* p *and the* x*-axis intercept.*

As an application, let us show that the surface of revolution generated by the tractrix has constant -1 Gaussian curvature. For this reason the surface is called *pseudosphere*. Its intrinsic geometry is locally the same as that of Bolyai's and Lobachevsky's hyperbolic plane. This local model of hyperbolic geometry was discovered by Eugenio Beltrami (1835-1899).

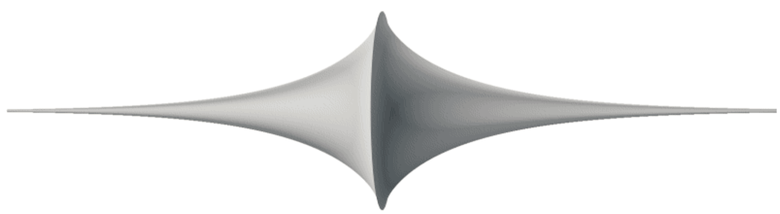

Figure 3.2: The pseudosphere.

The tractrix is defined as the involute of the chain curve $\gamma(t) = (t, \cosh t)$ touching the chain curve at $(0, 1)$. The length of the chain curve arc between $\gamma(0)$ and $\gamma(t)$ is

$$\int_0^t \|\gamma'(\tau)\| d\tau = \int_0^t \sqrt{1 + \sinh^2 \tau} d\tau = \int_0^t \cosh \tau d\tau = \sinh t.$$

This way, the tractrix has the parameterization

$$\begin{aligned}\hat{\gamma}(t) = \gamma(t) - \sinh t \frac{\gamma'(t)}{\|\gamma'(t)\|} &= (t, \cosh t) - \sinh t \frac{(1, \sinh t)}{\cosh t} \\ &= \left(t - \tanh t, \frac{1}{\cosh t}\right).\end{aligned}$$

As we know from the theory of evolutes and involutes, the chain curve is the evolute of the tractrix, the segment $\gamma(t)\hat{\gamma}(t)$ is normal to the tractrix, and its length is the radius of curvature of the tractrix at $\hat{\gamma}(t)$. This implies from one hand that the first principal curvature of the pseudosphere is $\kappa_1 = -\frac{1}{\sinh t}$. On the other hand, we obtain that the equation of the normal line of the tractrix at $\hat{\gamma}(t)$ is

$$\frac{y - \cosh t}{x - t} = \sinh t.$$

The x-intercept is

$$\left(t - \frac{\cosh t}{\sinh t}, 0\right).$$

According to the general results on surfaces of revolution, the second principal radius of curvature of the pseudosphere is the distance between $\hat{\gamma}(t)$ and $(t - \coth t, 0)$, i.e.

$$\kappa_2^{-1} = \|(\coth t - \tanh t, (\cosh t)^{-1})\| = \left(\left(\frac{\cosh^2 t - \sinh^2 t}{\cosh t \sinh t}\right)^2 + \frac{1}{\cosh^2 t}\right)^{1/2}$$
$$= \left(\frac{1}{\sinh^2 t \cosh^2 t} + \frac{1}{\cosh^2 t}\right)^{1/2} = \frac{1}{\sinh t}.$$

This shows that $K = \kappa_1 \kappa_2 \equiv -1$.

3.2.2 Lines of Curvature, Triply Orthogonal Systems

Definition 3.2.2. A regular curve on a surface is said to be a *line of curvature* if the tangent vectors of the curve are principal directions. ※

There are many parameterizations of a hypersurface. In applications we should always try to find a parameterization that makes solving the problem easier. For theoretical purposes, it is good to know the existence of certain parameterizations that have nice properties. In what follows, we study parameterizations, for which coordinate lines are lines of curvature.

Theorem 3.2.3. *The coordinate lines of a regular parameterization* $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ *of a hypersurface are lines of curvature if the matrices of the first and second fundamental forms are diagonal. The converse is also true if the principal curvatures of the hypersurface are different at each point.*

Proof. The matrix $\mathcal{L}$ of the Weingarten map is the quotient $\mathcal{B}\mathcal{G}^{-1}$ of the matrices of the first and second fundamental forms. If these matrices are diagonal, then so is $\mathcal{L}$. Obviously, the matrix of a linear map with respect to a basis is diagonal if and only if the basis consists of eigenvectors of the linear map. In our case, we get that $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$ form an eigenvector basis for the Weingarten map, i.e. these vectors are principal directions. Since $\mathbf{r}_i$ is the tangent vector of the i-th family of coordinate lines, the coordinate lines are lines of curvature.

Now suppose that the coordinate lines are lines of curvature and that the principal curvatures $\kappa_1, \dots, \kappa_{n-1}$ corresponding to the principal directions

$\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$ are different at every point. In this case,

$$\kappa_i\langle \mathbf{r}_i, \mathbf{r}_j\rangle = \langle L\mathbf{r}_i, \mathbf{r}_j\rangle = \langle \mathbf{r}_i, L\mathbf{r}_j\rangle = \kappa_j\langle \mathbf{r}_i, \mathbf{r}_j\rangle,$$
$$(\kappa_i - \kappa_j)\langle \mathbf{r}_i, \mathbf{r}_j\rangle = 0,$$

and since $(\kappa_i - \kappa_j) \neq 0$ for $i \neq j$,

$$g_{ij} = \langle \mathbf{r}_i, \mathbf{r}_j\rangle = 0 \quad \text{if } i \neq j.$$

Hence, the matrix $\mathcal{G}$ of the first fundamental form is diagonal. The matrix $\mathcal{L}$ of the Weingarten map is also diagonal by our assumption, consequently the matrix $\mathcal{B} = \mathcal{L}\mathcal{G}$ of the second fundamental form is diagonal as well. □

Theorem 3.2.4. *Every non-umbilical point of a regular parameterized surface in $\mathbb{R}^3$ has a neighborhood that admits a reparameterization with respect to which coordinate lines are lines of curvature.*

In view of other applications, we formulate the key part of the proof as a separate theorem.

Theorem 3.2.5. *Let X_1 and X_2 be two smooth tangential vector fields on a regular parameterized surface $\mathbf{r}\colon \Omega \to \mathbb{R}^3$. Assume that at a given point $p = \mathbf{r}(u_0, v_0)$, the vectors $X_1(u_0, v_0)$ and $X_2(u_0, v_0)$ are linearly independent. Then there is a diffeomorphism $H\colon W \to V$ between an open neighborhood $W \subset \mathbb{R}^2$ of the origin and an open neighborhood $V \subset \Omega$ of (u_0, v_0) such that the composition $\tilde{\mathbf{r}} = \mathbf{r} \circ H$ is a regular reparameterization of the surface in a neighborhood of p such that the partial derivative vector fields $\tilde{\mathbf{r}}_1$ and $\tilde{\mathbf{r}}_2$ are parallel to X_1 and X_2 respectively.*

Proof. Let us decompose X_1 and X_2 as linear combinations

$$X_1 = X_1^1\mathbf{r}_1 + X_1^2\mathbf{r}_2,$$
$$X_2 = X_2^1\mathbf{r}_1 + X_2^2\mathbf{r}_2$$

of $\mathbf{r}_1$ and $\mathbf{r}_2$. The coefficient functions X_i^j are smooth functions on Ω and they define two smooth vector fields $F_1, F_2\colon \Omega \to \mathbb{R}^2$ by

$$F_i(u, v) = (X_i^1(u, v), X_i^2(u, v)) \quad i = 1, 2.$$

and two ordinary differential equations $\mathbf{u}' = F_i \circ \mathbf{u}$.
Let Φ_t and Ψ_t be the flows generated by F_1 and F_2 respectively (see Definition 1.7.6) and consider the smooth maps

$$H_1(s, t) = \Psi_t(\Phi_s(u_0, v_0)), \qquad H_2(s, t) = \Phi_s(\Psi_t(u_0, v_0)).$$

Since the curve $H_1(s,0) = H_2(s,0) = \Phi_s(u_0, v_0)$ is an integral curve of F_1,

$$\partial_1 H_1(0,0) = \partial_1 H_2(0,0) = F_1(0,0),$$

and similarly

$$\partial_2 H_1(0,0) = \partial_2 H_2(0,0) = F_2(0,0).$$

Thus, the derivative matrix of H_1 and H_2 at the origin is

$$H_1'(0,0) = H_2'(0,0) = \begin{pmatrix} X_1^1(u_0,v_0) & X_2^1(u_0,v_0) \\ X_1^2(u_0,v_0) & X_2^2(u_0,v_0) \end{pmatrix}. \tag{3.7}$$

By our assumptions, $X_1(u_0,v_0)$ and $X_2(u_0,v_0)$ are linearly independent vectors, therefore both H_1 and H_2 has invertible derivative matrix at $(0,0)$. Applying the inverse function theorem to both H_1 and H_2 we conclude that the origin $(0,0)$ has open neighborhoods W_1, W_2, and the point (u_0,v_0) has open neighborhoods $V_1, V_2 \subset \Omega$ such that $H_i|_{W_i} \colon W_i \to V_i$ is a diffeomorphism for $i = 1, 2$.

Denote by $G_i = (G_i^1, G_i^2) \colon V_i \to W_i$ the inverse map of $H_i|_{W_i}$, and consider the map $G = (G_1^1, G_2^2) \colon V_1 \cap V_2 \to \mathbb{R}^2$. The derivative matrices of G_1 and G_2 at (u_0,v_0) are equal to the inverse matrix (3.7), consequently, this inverse matrix is also the derivative matrix of G at (u_0,v_0). Applying again the inverse function theorem to G, we obtain that there are open subsets $V \subset V_1 \cap V_2$ around (u_0,v_0) and $W \subset W_1 \cap W_2$ around the origin, such that $G|V \colon V \to W$ is a diffeomorphism. We can assume for simplicity that W is an open square $W = \{(s,t) \mid |s| < \epsilon, |t| < \epsilon\}$. We claim that the inverse map $H = (G|_V)^{-1} \colon W \to V$ gives a reparameterization with the required property.

For this purpose, examine the coordinate lines of the parameterization $\mathbf{r} \circ H$. Since the first coordinate function of G is the same as the first coordinate function of G_1,

$$G(H_1(s,t)) = (s, \alpha(s,t)) \text{ for all } (s,t) \in G_1(V),$$

where $\alpha \colon W \to \mathbb{R}$ is a smooth function. As $G \circ H$ leaves the first coordinate invariant, its inverse also must have the form

$$G_1(H(s,t)) = (s, \beta(s,t)), \text{ for all } (s,t) \in W,$$

where for all $|s| < \epsilon$, $\beta(s,.) = \alpha(s,.)^{-1}$. Applying H_1 to both sides we get

$$H(s,t) = H_1(s, \beta(s,t)) = \Psi_{\beta(s,t}(\Phi_s(u_0,v_0)).$$

For a fixed s, this is a reparameterization by $\beta(s,.)$ of the integral curve of F_2 starting at $\Phi_s(u_0,v_0))$. Thus,

$$\tilde{\mathbf{r}}_t(s,t) = \beta_t(s,t) \cdot X_2(H(s,t)).$$

This means that $\tilde{\mathbf{r}}_t$ is parallel to the vector field X_2. A similar argument shows that the vector field $\tilde{\mathbf{r}}_s$ is parallel to X_1, that is the coordinate lines of the new parameterization are tangent to X_1 and X_2. □

Remarks.

- The parameterization constructed above has also the property that the coordinate lines through p are integral curves of X_1 and X_2. In particular, if X_1 and X_2 are unit vector fields, then these two parameter lines are parameterized by arc length.
- The theorem does not hold for higher dimensions. (Find why the above proof does not work in higher dimensions.)

Proof of Theorem 3.2.4. Let $\kappa_1 \leq \kappa_2$ be the two principal curvatures arranged in increasing order. Since κ_1 and κ_2 change continuously and they are different at p, they are different also in a small neighborhood of p. Restricting our attention to this neighborhood, we may assume that the surface consists of non-umbilical points. Then at each point q of the surface, there are unique pairs of opposite unit vectors $\pm(X_1)_q$ and $\pm(X_2)_q$ pointing in the principal directions corresponding to κ_1 and κ_2 respectively. If we fix the unit principal directions $(X_1)_p$ and $(X_2)_p$ arbitrarily and then choose the direction of $(X_1)_q$ and $(X_2)_q$ at nearby points so that $(X_i)_q$ enclosed an acute angle with $(X_i)_p$ then we obtain to smooth vector fields X_1 and X_2 on a neighborhood of p. Applying to these two vector fields the above theorem, we obtain the required reparameterization. □

Dupin's Theorem and Curvature Lines on Ellipsoids

Now we give a description of the lines of curvature on ellipsoids. Our approach, which is based on Dupin's theorem, works for any surface of second order. Dupin's theorem claims that if we have three families of surfaces such that the surfaces of any of the families foliate an open domain in $\mathbb{R}^3$, and surfaces from different families intersect one another orthogonally, then surfaces from different families intersect one another in lines of curvature. We can obtain families of surfaces in a natural way considering a curvilinear coordinate system on $\mathbb{R}^3$.

Definition 3.2.6. A *local (curvilinear) coordinate system* on an open subset U of $\mathbb{R}^3$ is a smooth mapping $\phi\colon U \to \mathbb{R}^3$, which maps U onto its image $\Omega = \phi(U)$ bijectively and the inverse $\phi^{-1}\colon \Omega \to \mathbb{R}^3$ of which is also smooth. The inverse $\mathbf{r} = \phi^{-1}$ is a regular parameterization of U. Ω is covered by planes parallel to one of the three coordinate planes in $\mathbb{R}^3$. The images of these families of planes are the *coordinate surfaces* of the local coordinate system ϕ (or that of the parameterization $\mathbf{r}$). There are three families of coordinate surfaces. Each family covers the domain U. Coordinate surfaces from different families intersect one another in *coordinate lines*. We say that ϕ (or $\mathbf{r}$) defines a *triply orthogonal system*, if coordinate surfaces from different families intersect one another orthogonally, or equivalently, if $\langle \mathbf{r}_i, \mathbf{r}_j \rangle = 0$ for $i \neq j$. ※

Theorem 3.2.7 (Dupin's theorem)**.** *If the curvilinear coordinate system $\phi\colon U \to \mathbb{R}^3$ defines a triply orthogonal system on U, then the coordinate lines are lines of curvature on the coordinate surfaces.*

Proof. Let $\mathbf{r}\colon \Omega = \phi(U) \to \mathbb{R}^3$ be the inverse of ϕ. We may consider without loss of generality the coordinate surface parameterized by $(u, v) \mapsto \mathbf{r}(u, v, w_0)$, where w_0 is constant. It is enough to show that the matrices of the first and second fundamental forms of the surface with respect to the given parameterization are diagonal. The matrix of the first fundamental form is diagonal by our assumption $\langle \mathbf{r}_1, \mathbf{r}_2 \rangle = 0$. The nondiagonal element of the matrix of the second fundamental form is $\langle \mathbf{r}_{12}, \mathbf{N} \rangle$, where $\mathbf{N}$ is the unit normal of the surface. Since $\mathbf{r}_3$ is parallel to $\mathbf{N}$, $\langle \mathbf{r}_{12}, \mathbf{N} \rangle = 0$ will follow from $\langle \mathbf{r}_{12}, \mathbf{r}_3 \rangle = 0$. Differentiating the equation $\langle \mathbf{r}_1, \mathbf{r}_3 \rangle = 0$ with respect to the second variable yields

$$\langle \mathbf{r}_{12}, \mathbf{r}_3 \rangle + \langle \mathbf{r}_1, \mathbf{r}_{23} \rangle = 0$$

and similarly, we also have

$$\begin{aligned} \langle \mathbf{r}_{23}, \mathbf{r}_1 \rangle + \langle \mathbf{r}_2, \mathbf{r}_{31} \rangle &= 0, \\ \langle \mathbf{r}_{31}, \mathbf{r}_2 \rangle + \langle \mathbf{r}_3, \mathbf{r}_{12} \rangle &= 0. \end{aligned}$$

Solving this system of linear equations for the unknown quantities $\langle \mathbf{r}_{12}, \mathbf{r}_3 \rangle$, $\langle \mathbf{r}_{23}, \mathbf{r}_1 \rangle$, $\langle \mathbf{r}_{31}, \mathbf{r}_2 \rangle$, we see that they are all zero. □

The canonical equation of an ellipsoid has the form

$$\frac{x^2}{A} + \frac{y^2}{B} + \frac{z^2}{C} = 1.$$

Suppose $A > B > C$. We can embed this surface into a triply orthogonal system of second order surfaces as follows. Consider the surface

$$F_\lambda: \ \frac{x^2}{A+\lambda} + \frac{y^2}{B+\lambda} + \frac{z^2}{C+\lambda} = 1.$$

F_λ is

- an ellipsoid for $\lambda > -C$;
- a one sheeted hyperboloid for $-C > \lambda > -B$;

- a two sheeted hyperboloid for $-B > \lambda > -A$.

In accordance with these cases, we get three families of surfaces. Surfaces obtained by such a perturbation of the equation of a second order surface are called *confocal second order surfaces*.

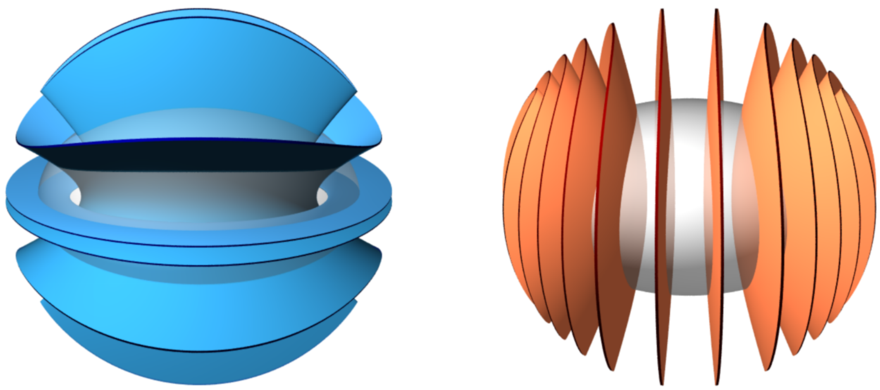

Figure 3.3: One and two sheeted hyperboloids confocal to an ellipsoid.

Proposition 3.2.8. *Let $(x, y, z) \in \mathbb{R}^3$ be a point for which $xyz \neq 0$. Then there exist exactly three λ-s, one from each of the intervals $(-C, +\infty)$, $(-B, -C)$, $(-A, -B)$ such that $(x, y, z) \in F_\lambda$.*

Proof. Condition $(x, y, z) \in F_\lambda$ is equivalent to

$$\begin{aligned} P(\lambda) = {} & (A+\lambda)(B+\lambda)(C+\lambda) - \\ & - \big(x^2(B+\lambda)(C+\lambda) + y^2(A+\lambda)(C+\lambda) + z^2(A+\lambda)(B+\lambda)\big) = 0. \end{aligned}$$

This is an equation of degree three for λ, so the number of solutions is not more than three. To see that all the three solutions are real and located as it is stated, compute P at the nodes.

$$\begin{aligned} P(-A) &= -x^2(B-A)(C-A) < 0, \\ P(-B) &= -y^2(A-B)(C-B) > 0, \\ P(-C) &= -z^2(A-C)(B-C) < 0. \end{aligned}$$

Furthermore, $F(\lambda) = \lambda^3 + \cdots$ implies $\lim_{\lambda\to\infty} F(\lambda) = +\infty$. Thus, by Bolzano's intermediate value theorem, F has at least one root in each of the intervals $(-C, +\infty)$, $(-B, -C)$, $(-A, -B)$. □

Proposition 3.2.9. *If $(x, y, z) \in F_\lambda \cap F_{\lambda'}$, $xyz \neq 0$, $\lambda \neq \lambda'$, then F_λ intersects $F_{\lambda'}$ orthogonally.*

Lemma 3.2.10. *If a nonempty subset M of $\mathbb{R}^3$ is defined by an equation $F(x,y,z)=0$, i.e.,*

$$M = \{(x,y,z) \in \mathbb{R}^3 \mid F(x,y,z) = 0\},$$

so that the gradient vector field $\operatorname{grad} F = (\partial_1 F, \partial_2 F, \partial_3 F)$ *is not zero at points of M, then every point of M has a neighborhood (in M) which is the image of a regular parameterized surface. In this case the tangent plane of M at $p\mathbf{r}(\mathbf{u}_0) \in M$ is orthogonal to* $\operatorname{grad} F(p)$.

Proof. The first part of the lemma is a direct application of the inverse function theorem, we shall prove it later in more generality (see Theorem 4.1.12). Suppose that M admits a regular parameterization $\mathbf{r}$ around $p \in M$. Then $F \circ \mathbf{r} \equiv 0$. Differentiating with respect to the i-th variable ($i = 1,2$) using the chain rule we obtain

$$0 = \partial_i(F \circ \mathbf{r}) = \langle (\operatorname{grad} F) \circ \mathbf{r}, \mathbf{r}_i \rangle,$$

hence $\operatorname{grad} F(p)$ is orthogonal to the tangent vectors $\mathbf{r}_1(\mathbf{u}_0), \mathbf{r}_2(\mathbf{u}_0)$ which span the tangent plane T_pM. $\square$

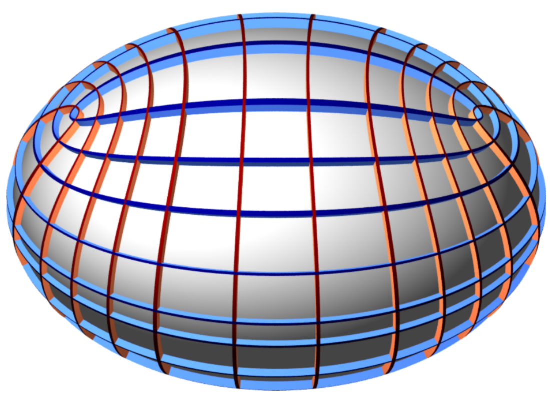

Figure 3.4: Lines of curvature on an ellipsoid.

Proof of Proposition 3.2.9. We need to show

$$\operatorname{grad} F_\lambda(x,y,z) \perp \operatorname{grad} F_{\lambda'}(x,y,z),$$

or equivalently

$$\frac{x^2}{(A+\lambda)(A+\lambda')} + \frac{y^2}{(B+\lambda)(B+\lambda')} + \frac{z^2}{(C+\lambda)(C+\lambda')} = 0.$$

We know that

$$\frac{x^2}{A+\lambda}+\frac{y^2}{B+\lambda}+\frac{z^2}{C+\lambda}=1 \text{ and } \frac{x^2}{A+\lambda'}+\frac{y^2}{B+\lambda'}+\frac{z^2}{C+\lambda'}=1.$$

Subtracting these equalities we obtain $(\lambda-\lambda')$ times the equality to prove. Since $\lambda \neq \lambda'$, we are ready. □

3.2.3 Ruled and Developable Surfaces

Regular surfaces swept out by a moving straight line are ruled surfaces. A bit more generally, we shall call a regular surface *ruled* if every point of the surface has a neighborhood with a regular parameterization of the form

$$\mathbf{r}(u,v)=\gamma(u)+v\delta(u),$$

where γ is a smooth curve, called the *directrix*, δ is a nowhere zero vector field along γ. The straight lines $v \mapsto \gamma(u_0)+v\delta(u_0)$ are the *generators* or *generatrices* of the surface.

Theorem 3.2.11. *The following statements are equivalent for ruled surfaces:*

(i) *the normal vector field* $\mathbf{N}$ *is constant along the generators;*

(ii) $\mathbf{r}_{uv}$ *is tangential for the parameterization* $\mathbf{r}(u)=\gamma(u)+v\delta(u)$*;*

(iii) *the Gaussian curvature* K *is constant* 0.

Proof. (i) $\Longrightarrow$ (iii). If $\mathbf{N}$ is constant along the generators, then $L(\mathbf{r}_v) = -\mathbf{N}_v = \mathbf{0} = 0\mathbf{r}_v$, thus the generators are lines of curvature and the corresponding principal curvature is 0 everywhere. From this follows that the Gaussian curvature is 0.

(iii) $\Longrightarrow$ (i). The normal section of a ruled surface in the direction of a generator is the generator itself. Hence, the normal curvature of the surface in the direction $\mathbf{r}_v$ of the generators is 0. If $\mathbf{r}_v$ were not a principal direction, then 0 would be strictly between the principal curvatures, in which case we would have $K<0$. The contradiction shows that $\mathbf{r}v$ is a principal direction at a given point, and thus $-\mathbf{N}_v = L(\mathbf{r}_v) = 0\mathbf{r}_v = \mathbf{0}$, i.e. $\mathbf{N}$ is constant along the generators.

Remark: We have proved here that $K \leq 0$ *for any ruled surface.*

(ii) $\Longleftrightarrow$ (iii). According to the formula

$$K=\frac{\det \mathcal{B}}{\det \mathcal{G}},$$

$K=0$ if and only if $\det \mathcal{B}=0$. Since $\mathbf{r}_{vv}=0$,

$$\det \mathcal{B} = \det \begin{pmatrix} \langle \mathbf{r}_{uu},\mathbf{N}\rangle & \langle \mathbf{r}_{uv},\mathbf{N}\rangle \\ \langle \mathbf{r}_{vu},\mathbf{N}\rangle & \langle \mathbf{r}_{vv},\mathbf{N}\rangle \end{pmatrix} = \det \begin{pmatrix} \langle \mathbf{r}_{uu},\mathbf{N}\rangle & \langle \mathbf{r}_{uv},\mathbf{N}\rangle \\ \langle \mathbf{r}_{vu},\mathbf{N}\rangle & 0 \end{pmatrix} = -\langle \mathbf{r}_{uv},\mathbf{N}\rangle^2$$

and thus $\det \mathcal{B} = 0$ if and only if $\mathbf{r}_{uv}$ is orthogonal to $\mathbf{N}$ i.e. if $\mathbf{r}_{uv}$ is tangential. □

Definition 3.2.12. A ruled surface that satisfies the equivalent conditions of the previous theorem is called a *developable surface*. ※

Examples.

(1) The following surfaces are ruled but not developable:

- $\mathbf{r}(u, v) = (u, v, uv)$, (hyperbolic paraboloid);
- $\mathbf{r}(u, v) = (a \cos u, b \sin u, 0) + v(-a \sin u, b \cos u, c)$, (one sheeted hyperboloid);
- $\mathbf{r}(u, v) = (0, 0, u) + v(\cos u, \sin u, 0)$ (helicoid).

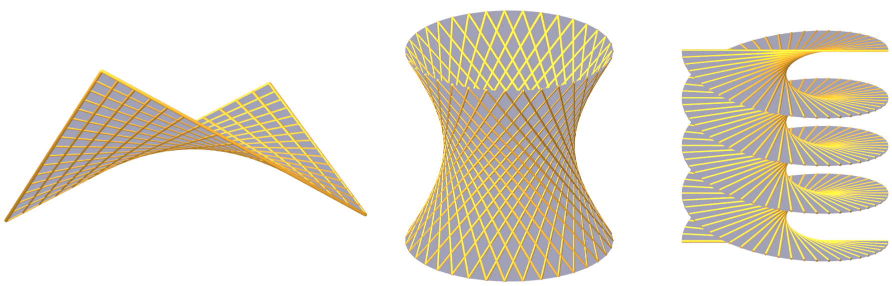

Figure 3.5: A hyperbolic paraboloid, a one-sheeted hyperboloid, and a helicoid.

(2) There are three main types of developable surfaces:

- *Cylinders over a curve.* Let γ be a regular space curve, $\mathbf{v} \neq \mathbf{0}$ a vector nowhere tangent to γ. We define the *cylinder over γ with generators parallel to* $\mathbf{v}$ by the parameterization

$$\mathbf{r}(u, v) = \gamma(u) + v\mathbf{v}.$$

  Since $\mathbf{r}_{uv} = \mathbf{0}$ is tangential, cylinders over a curve are developable.

- *Cones over a curve.* Let γ be a regular curve, $\mathbf{p}$ be the position vector of a point not lying on any tangent to the curve. The *cone over γ with vertex* $\mathbf{p}$ is defined by the parameterization

$$\mathbf{r}(u, v) = v\gamma(u) + (1 - v)\mathbf{p}.$$

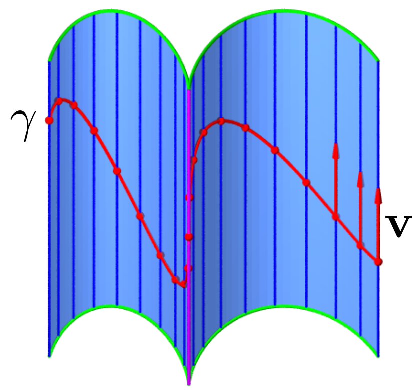

Figure 3.6: Cylinder over a curve with one singular generatrix.

The cone is regular only in the domain $v \neq 0$. The tangent plane of the cone is spanned by the vectors

$$\mathbf{r}_u(u,v) = v\gamma'(u) \quad \text{and} \quad \mathbf{r}_v(u,v) = \gamma(u) - \mathbf{p}.$$

Since

$$\mathbf{r}_{uv}(u,v) = \gamma'(u) = (1/v)\mathbf{r}_u(u,v),$$

cones are developable.

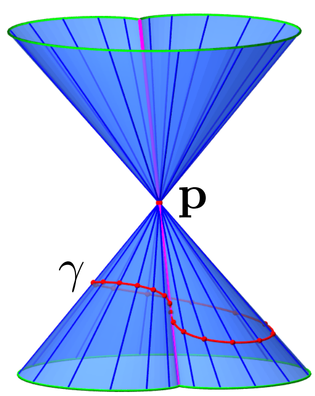

Figure 3.7: Cone over a curve with a singular generatrix.

- *Tangent surfaces.* Let γ be a curve of general type in $\mathbb{R}^3$. We show that the regular part of the surface swept out by the tangent lines of γ is developable. Indeed, the surface can be parameterized by

$$\mathbf{r}(u,v) = \gamma(u) + v\gamma'(u).$$

Partial derivatives of $\mathbf{r}$ are

$$\mathbf{r}_u(u,v) = \gamma'(u) + v\gamma''(u) \quad \text{and} \quad \mathbf{r}_v(u,v) = \gamma'(u).$$

Since γ is of general type, $\gamma'(u)$ and $\gamma''(u)$ are linearly independent, hence singularities of the surface of tangent lines are located along the generating curve γ.

As $\mathbf{r}_{uv}(u,v) = \gamma''(u) = (1/v)\big(\mathbf{r}_u(u,v) - \mathbf{r}_v(u,v)\big)$ for $v \neq 0$, the regular part of the surface of tangent lines is developable. Surfaces of this type are called *tangent surfaces of regular curves.*

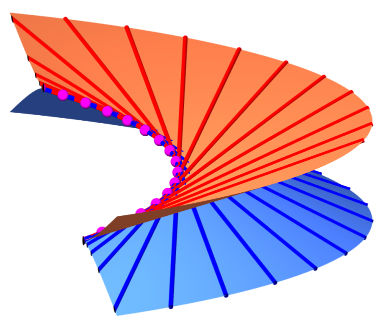

Figure 3.8: The tangent surface of a curve.

According to the definition, a developable surface is a ruled surface with Gaussian curvature equal to zero. The following theorem states that in most cases the condition of being ruled follows from $K \equiv 0$.

Theorem 3.2.13. *If the Gaussian curvature of a surface is* 0 *everywhere and the surface contains no planar point, then it is developable.*

Proof. Gaussian curvature is positive at spherical points so the surface contains no umbilics. Therefore we may consider a parameterization $\mathbf{r}$ around any point $\mathbf{p}$ such that $\mathbf{p} = \mathbf{r}(0,0)$, coordinate lines are lines of curvature and the coordinate lines through $\mathbf{p}$ are unit speed curves (see Theorem 3.2.4). Suppose that $\mathbf{r}_u$ corresponds to the nonzero principal curvature $\kappa_1 = \kappa \neq 0$. Then

$$\mathbf{N}_u = -\kappa \mathbf{r}_u \quad \text{and} \quad \mathbf{N}_v = 0\mathbf{r}_v = \mathbf{0}.$$

The second equation shows that $\mathbf{N}$ is constant along v-coordinate lines. What we have to show is that v-coordinate lines are straight lines. For this purpose, it is enough to prove that $\mathbf{r}$ is linear in v, i.e. $\mathbf{r}_{vv} = \mathbf{0}$. The idea we use is that the only vector which is perpendicular to each vector of a basis is the

zero vector. According to this observation, it suffices to show that $\mathbf{r}_{vv}$ is orthogonal to the vectors $\mathbf{N}$, $\mathbf{r}_u$, $\mathbf{r}_v$.
(i) $\mathbf{r}_{vv} \perp \mathbf{N}$. This orthogonality follows from

$$0 = \partial_v \langle \mathbf{r}_v, \mathbf{N} \rangle = \langle \mathbf{r}_{vv}, \mathbf{N} \rangle + \langle \mathbf{r}_v, \mathbf{N}_v \rangle = \langle \mathbf{r}_{vv}, \mathbf{N} \rangle.$$

(ii) $\mathbf{r}_{vv} \perp \mathbf{r}_u$. Since lines of curvature are orthogonal,

$$0 = \partial_v \langle \mathbf{r}_u, \mathbf{r}_v \rangle = \langle \mathbf{r}_{uv}, \mathbf{r}_v \rangle + \langle \mathbf{r}_u, \mathbf{r}_{vv} \rangle.$$

On the other hand, using the facts that $\mathbf{N}_u = -\kappa \mathbf{r}_u \perp \mathbf{r}_v$ and $\mathbf{N}_{uv} = \partial_u(\mathbf{N}_v) = \mathbf{0}$ we get

$$\langle \mathbf{r}_{uv}, \mathbf{r}_v \rangle = \big\langle -\partial_v\big(\mathbf{N}_u/\kappa\big), \mathbf{r}_v \big\rangle = (\kappa_v/\kappa^2)\langle \mathbf{N}_u, \mathbf{r}_v \rangle - (1/\kappa)\langle \mathbf{N}_{uv}, \mathbf{r}_v \rangle = 0.$$

Combining these two equalities we get $\langle \mathbf{r}_{vv}, \mathbf{r}_u \rangle = 0$.
(iii) $\mathbf{r}_{vv} \perp \mathbf{r}_v$. This will follow from the observation that v-coordinate lines are all parameterized by arc length, i.e. $\|\mathbf{r}_v\| \equiv 1$. We know that $\|\mathbf{r}_v(0, v)\| = 1$ by the construction of $\mathbf{r}$. By our previous equation we also have

$$\partial_u \langle \mathbf{r}_v, \mathbf{r}_v \rangle = 2\langle \mathbf{r}_{uv}, \mathbf{r}_v \rangle = 0,$$

showing that $\langle \mathbf{r}_v, \mathbf{r}_v \rangle$ does not depend on u. Thus,

$$\|\mathbf{r}_v(u, v)\| = \|\mathbf{r}_v(0, v)\| = 1 \text{ for every } u, v.$$

This completes the proof. □

We finish the investigation of developable surfaces with a structure theorem stating that every developable surface is made up of pieces of cylinders, cones and tangent surfaces.

Theorem 3.2.14. *Let $\mathbf{r}\colon [a, b] \times [c, d] \to \mathbb{R}^3$ be a developable surface without planar points and suppose that the parameterization $\mathbf{r}$ of the surface is the one we used in the proof of the previous theorem. Then there exists a nowhere dense closed subset A of $[a, b]$ the complement of which is a union of open intervals $[a, b] \setminus A = I_1 \cup I_2 \cup \cdots$ such that the restriction of $\mathbf{r}$ onto $I_n \times [c, d]$ is a part of a cylinder or cone or a tangent surface.*

Proof. As it was proved above, $\mathbf{r}$ has the form $\mathbf{r}(u, v) = \mathbf{a}(u) + v\mathbf{b}(u)$, where $\mathbf{b}(u)$ is a unit vector field along the curve $\mathbf{a}(u)$. We have

$$\mathbf{r}_u(u, v) = \mathbf{a}'(u) + v\mathbf{b}'(u), \qquad \mathbf{r}_v(u, v) = \mathbf{b}(u), \qquad \mathbf{r}_{uv}(u, v) = \mathbf{b}'(u).$$

By the definition of developable surfaces, $\mathbf{r}_{uv}(u, v) = \mathbf{b}'(u)$ must be tangential to the surface, i.e. it lies in the plane spanned by $\mathbf{r}_u$ and $\mathbf{r}_v$. $\mathbf{r}_u$ is orthogonal to $\mathbf{r}_v$ as they are lines of curvature, furthermore, $\mathbf{r}_{uv}$ is also orthogonal to $\mathbf{r}_v$ as $0 = \langle \mathbf{b}(u), \mathbf{b}(u)\rangle' = 2\langle \mathbf{b}(u), \mathbf{b}'(u)\rangle$. For there is only one direction in a plane which is orthogonal to a given nonzero vector, $\mathbf{r}_u$ and $\mathbf{r}_{uv}$ must be parallel: $\mathbf{b}'(u) \parallel \mathbf{a}'(u)$. Hence, $\mathbf{b}'(u) = c(u)\mathbf{a}'(u)$ for some function $c\colon [a,b] \to \mathbb{R}$. Now let A be the set of those roots of $c \cdot c'$ in $[a,b]$, which do not have a neighborhood containing only roots of $c \cdot c'$. Then A is closed and nowhere dense in $[a,b]$. If $[a,b] \setminus A$ is the union of the disjoint open intervals $I_1, I_2, \dots$, then for the restriction of c onto I_n, we have one of the following possibilities:

(a) the restriction is identically 0;

(b) the restriction is a nonzero constant;

(c) the restriction is strictly monotone and nowhere zero.

In the first case, $\mathbf{b}'(u) = \mathbf{0}$ and thus $\mathbf{b}$ is constant on I_n, thus the restriction of $\mathbf{r}$ onto $I_n \times [c,d]$ is a part of a cylinder.
In the second case, the point $\mathbf{p}(u) = \mathbf{a}(u) - (1/c)\mathbf{b}(u)$ does not depend on u. Indeed, $\mathbf{p}'(u) = \mathbf{a}'(u) - (1/c)\mathbf{b}'(u) = \mathbf{0}$. Furthermore, the point $\mathbf{p}$ lies on every generator of the surface, so this case serves a part of a cone.
Finally, consider the curve $\gamma(u) = \mathbf{a}(u) - (1/c(u))\mathbf{b}(u)$ for the last case. As

$$\gamma'(u) = \mathbf{a}'(u) + (c'(u)/c^2(u))\mathbf{b}(u) - \mathbf{b}'(u)/c(u) = (c'(u)/c^2(u))\mathbf{b}(u) \parallel \mathbf{b}(u),$$

the tangent of γ at $\gamma(u)$ coincides with the generator $v \mapsto \mathbf{r}(u,v)$ of the surface, and γ is of general type too, as

$$\begin{aligned}\gamma'(u) \times \gamma''(u) &= \big(c'(u)/c^2(u)\big)^2 \mathbf{b}(u) \times \mathbf{b}'(u)\\ &= \big(c'(u)/c^2(u)\big)^2 c(u)\mathbf{b}(u) \times \mathbf{a}'(u) \neq \mathbf{0}.\end{aligned}$$

We conclude that in the third case the restriction of $\mathbf{r}$ onto $I_n \times [c,d]$ is a part of the tangent surface of the curve of general type γ. □

Exercise 3.2.15. Prove the inequality

$$H^2 \geq K$$

for the Minkowski curvature H and the Gauss curvature K of a surface in $\mathbb{R}^3$. At which points does equality hold? ¤

Exercise 3.2.16. Find umbilical points on the ellipsoid $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$, $(a > b > c)$ parameterized by

$$\mathbf{r}(u,v) = (a\cos u \cos v, b \cos u \sin v, c \sin u).$$ ¤

Exercise 3.2.17. Show that the Minkowski curvature of the following surfaces is 0 everywhere

(a) $\mathbf{r}(u,v) = (v\cos u, -v\sin u, bu)$ (*helicoid*),

(b) $\mathbf{r}(u,v) = (\cosh v\cos u, \cosh v\sin u, v)$ (*catenoid*: surface of revolution generated by the chain curve). ¤

Exercise 3.2.18. Let M be a surface with a regular parameterization $\mathbf{r}$. Denote by $\mathbf{N}$ the unit normal vector field of M and consider the parallel surface M_d parameterized by

$$\mathbf{r}_d(u,v) = \mathbf{r}(u,v) + d\mathbf{N}(u,v),$$

where $d \neq 0$ is a fixed real number.

(a) Show that the singular points of the parameterization $\mathbf{r}_d$ correspond to those points of M at which $1/d$ is a principal curvature.

(b) Show that $\mathbf{N}(u,v)$ is a unit normal of M_d at $\mathbf{r}_d(u,v)$.

(c) Express the Gaussian and Minkowski curvature of M_d with those of M. ¤

3.2.4 Asymptotic Curves on Negatively Curved Surfaces

Consider a regular parameterized surface in $\mathbf{r}\colon \Omega \to \mathbb{R}^3$, and a point $p = \mathbf{r}(\mathbf{u}_0) \in M = \mathbf{r}(\Omega)$, at which the Gaussian curvature $K_p = K(\mathbf{u}_0)$ negative. Then the product of the two principal curvatures κ_1, κ_2 at p is negative, thus, the principal curvatures have opposite signs. We may assume without loss of generality that $\kappa_1 < 0 < \kappa_2$. Let $\mathbf{e}_1, \mathbf{e}_2$ be unit vectors, pointing in principal directions corresponding to κ_1 and κ_2 respectively. Since κ_1 and κ_2 are different eigenvalues of the Weingarten map, $\mathbf{e}_1$ is orthogonal to $\mathbf{e}_1$.
The principal curvatures at a point are the extremal values of the normal curvature function. Since they have opposite sign in our case, there must be tangent directions at p, in which the normal curvature vanishes.

Definition 3.2.19. If p is a point on a surface, then the tangent directions at p in which the normal curvature vanishes are called the *asymptotic directions at p.* ※

We can find these directions easily with the help of Euler's formula. If the direction angle of a unit vector $\mathbf{v} \in T_pM$ relative to orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2)$ is θ, then $\mathbf{v} = \cos(\theta)\mathbf{e}_1 + \sin(\theta)\mathbf{e}_2$, and

$$k_p(\mathbf{v}) = \cos^2(\theta)\cdot\kappa_1 + \sin^2(\theta)\kappa_2.$$

Solving the equation $k_p(\mathbf{v}) = 0$ for θ, we obtain

$$k_p(\mathbf{v}) = 0 \quad\iff\quad \tan(\theta) = \pm\sqrt{-\frac{\kappa_1}{\kappa_2}}.$$

Thus, there are two linearly independent asymptotic directions, arranged symmetrically in the principal directions. In other words, the principal directions are the bisectors of the angles determined by the principal directions.
We described above how to find asymptotic direction at a point, where the Gaussian curvature is negative. Asymptotic directions exist also at points where the Gaussian curvature is 0, but at such points the picture we obtained for the negative curvature points degenerates in a way. If $\kappa_1 = 0 < \kappa_2$ or $\kappa_1 < 0 = \kappa_2$, then the two asymptotic directions coincide with the principal direction corresponding to the 0 principal curvature. If $\kappa_1 = 0 = \kappa_2$, then p is a planar umbilical point and all tangent directions are asymptotic.
There are no asymptotic directions at a point where the Gaussian curvature is positive.

Definition 3.2.20. A regular curve $\gamma\colon I \to M$ on a surface M is called an *asymptotic curve* if $\gamma'(t)$ is an asymptotic direction for all $t \in I$. ※

Proposition 3.2.21. *A regular curve $\gamma\colon I \to M$ is an asymptotic curve on the surface M if and only if $\gamma''(t)$ is tangent to the surface at $\gamma(t)$ for all $t \in I$.*
If $\gamma'(t)$ and $\gamma''(t)$ are linearly independent, then the second condition is equivalent to the requirement that the osculating plane of γ at t coincides with the tangent plane of the surface at $\gamma(t)$.

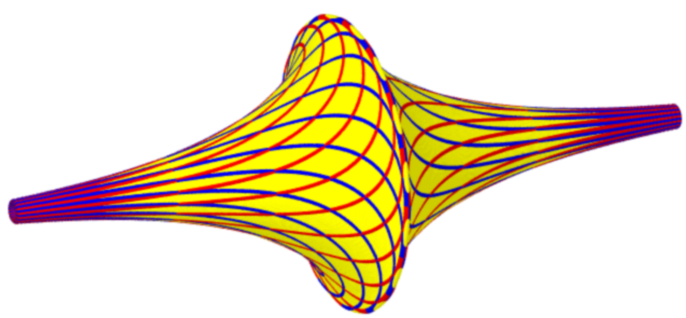

Figure 3.9: Asymptotic curves on the pseudosphere.

Proof. The normal curvature can be expressed with the help of the fundamental forms

$$k_{\gamma(t)}(\gamma'(t)) = \frac{II_{\gamma(t)}(\gamma'(t),\gamma'(t))}{\|\gamma'(t)\|^2} = \frac{-\langle\gamma'(t), L_{\gamma(t)}(\gamma'(t))\rangle}{\|\gamma'(t)\|^2}.$$

Thus, $\gamma'(t)$ points in asymptotic direction if and only if

$$\langle\gamma'(t), L_{\gamma(t)}(\gamma'(t))\rangle = 0.$$

Denote by $\mathbf{N}_p$ the unit normal of the surface at p. Differentiating the equation

$$\langle \gamma'(t), \mathbf{N}_{\gamma'(t)} \rangle \equiv 0$$

with respect to t we obtain

$$\langle \gamma''(t), \mathbf{N}_{\gamma'(t)} \rangle + \langle \gamma'(t), -L_{\gamma(t)}(\gamma'(t)) \rangle = 0,$$

consequently, $\gamma'(t)$ is a principal direction if and only $\gamma''(t)$ is orthogonal to $\mathbf{N}_{\gamma(t)}$. □

Theorem 3.2.22 (Beltrami–Enneper)**.**

(a) *Let $\gamma\colon I \to M$ be an asymptotic line on a surface M in $\mathbb{R}^3$. Then*

$$\langle \gamma' \times \gamma'', \gamma''' \rangle = \epsilon \sqrt{-K_\gamma} \|\gamma' \times \gamma''\|^2, \tag{3.8}$$

where the function $K_\gamma\colon I \to \mathbb{R}$ assigns to $t \in I$ the Gaussian curvature of M at $\gamma(t)$, ϵ is the sign of $\langle \gamma' \times L_\gamma(\gamma'), \mathbf{N}_\gamma \rangle$.

(b) *In particular, if γ' and γ'' are linearly independent at t, therefore the torsion τ of γ is defined at t, then*

$$\tau(t) = \epsilon \sqrt{-K_{\gamma(t)}}.$$

(c) *If two asymptotic curves meet at a point p with nonparallel speed vectors and their torsions are defined at p, then their torsions at p are opposite to one another.*

Proof. (a) Since γ is an asymptotic curve, both $\gamma'(t)$ and $\gamma''(t)$ are in the tangent plane at $\gamma(t)$, consequently, $\gamma' \times \gamma'' = \epsilon_1 \|\gamma' \times \gamma''\| \mathbf{N}_\gamma$ and

$$\langle \gamma' \times \gamma'', \gamma''' \rangle = \epsilon_1 \|\gamma' \times \gamma''\| \langle \mathbf{N}_\gamma, \gamma''' \rangle.$$

Where ϵ_1 is the sign of $\langle \gamma' \times \gamma'', \mathbf{N}_\gamma \rangle$.
Differentiating $\langle \gamma'', \mathbf{N}_\gamma \rangle \equiv 0$, we obtain

$$\langle \gamma''', \mathbf{N}_\gamma \rangle = -\left\langle \gamma'', \frac{d}{dt}\mathbf{N}_\gamma \right\rangle = \langle \gamma'', L_\gamma(\gamma') \rangle.$$

Using the linear algebraic identity

$$L_\gamma(\gamma') \times L_\gamma(\gamma'') = *(L_\gamma(\gamma') \wedge L_\gamma(\gamma'')) = *(\det(L_\gamma)\gamma' \wedge \gamma'') = \det(L_\gamma)\gamma' \times \gamma''$$

and the Lagrange identity we obtain

$$\begin{aligned} \det(L_\gamma)\|\gamma' \times \gamma''\|^2 &= \langle \gamma' \times \gamma'', L_\gamma(\gamma') \times L_\gamma(\gamma'') \rangle \\ &= \det \begin{pmatrix} \langle \gamma', L_\gamma(\gamma') \rangle & \langle \gamma'', L_\gamma(\gamma') \rangle \\ \langle \gamma', L_\gamma(\gamma'') \rangle & \langle \gamma'', L_\gamma(\gamma'') \rangle \end{pmatrix} \\ &= -\langle \gamma'', L_\gamma(\gamma') \rangle^2. \end{aligned}$$

In the last step we used that $\langle \gamma', L_\gamma(\gamma')\rangle = 0$ and $\langle \gamma'', L_\gamma(\gamma')\rangle = \langle \gamma', L_\gamma(\gamma'')\rangle$. The first equation holds since γ' points in asymptotic direction, the second is true because the Weingarten map is self-adjoint. Taking square root results in

$$\sqrt{-\det(L_\gamma)}\|\gamma' \times \gamma''\| = \epsilon_2\langle \gamma'', L_\gamma(\gamma')\rangle,$$

where $\epsilon_2 = \operatorname{sgn}\langle \gamma'', L_\gamma(\gamma')\rangle$. Combining the equations above we get

$$\langle \gamma' \times \gamma'', \gamma'''\rangle = \epsilon\sqrt{-\det(L_\gamma)}\|\gamma' \times \gamma''\|^2,$$

where $\epsilon = \epsilon_1\epsilon_2$.
If $\gamma'(t) \times \gamma''(t) = \mathbf{0}$ at a point, then equation (3.8) holds with any choice of $\epsilon(t)$, as both sides are equal to 0. Assume that $\gamma'(t) \times \gamma''(t) \neq \mathbf{0}$. Then $\gamma''(t)$ can be decomposed into the orthogonal components

$$\gamma''(t) = a\gamma'(t) + b(\mathbf{N}_{\gamma(t)} \times \gamma'(t)), \text{ where } a, b \in \mathbb{R}, \quad b \neq 0,$$

hence $\epsilon_1(t)$ is the sign of

$$\langle \gamma'(t) \times \gamma''(t), \mathbf{N}_{\gamma(t)}\rangle = \langle \gamma'(t) \times b\big(\mathbf{N}_{\gamma(t)} \times \gamma'(t)\big), \mathbf{N}_{\gamma(t)}\rangle = b\|N_{\gamma(t)} \times \gamma'(t)\|^2,$$

while $\epsilon_2(t)$ is the sign of

$$\begin{aligned}\langle \gamma''(t), L_{\gamma(t)}(\gamma'(t))\rangle &= \langle b\big(\mathbf{N}_{\gamma(t)} \times \gamma'(t)\big), L_{\gamma(t)}(\gamma'(t))\rangle \\ &= b\langle \gamma'(t) \times L_{\gamma(t)}(\gamma'(t)), \mathbf{N}_{\gamma(t)}\rangle.\end{aligned}$$

Thus, $\epsilon(t)$ is the sign of $\langle \gamma'(t) \times L_{\gamma(t)}(\gamma'(t)), \mathbf{N}_{\gamma(t)}\rangle$.
Part (b) of the statement follows directly from (a) and the formula for the curvature of curves of general type.
As for part (c), it is clear from (b) that both torsions have absolute value $\sqrt{-K_p}$, where K_p is the Gaussian curvature at p, so we have to show only that the signs are opposite. By the description of the sign ϵ, we need to show that if $\mathbf{v}$ and $\mathbf{w}$ are two nonparallel asymptotic directions at p, then the vectors $\mathbf{v} \times L_p(\mathbf{v})$ and $\mathbf{w} \times L_p(\mathbf{w})$ have opposite orientation.
Let $(\mathbf{e}_1, \mathbf{e}_2)$ be an orthonormal basis of T_pM, consisting of principal directions ordered so that $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{N}_p$. Denote by κ_1, κ_2 the corresponding principal curvatures. If $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2$, and $\mathbf{w} = w_1\mathbf{e}_1 + w_2\mathbf{e}_2$ then $\mathbf{v} \times L_p(\mathbf{v}) = v_1v_2(\kappa_2 - \kappa_1)\mathbf{N}_p$ and $\mathbf{w} \times L_p(\mathbf{w}) = w_1w_2(\kappa_2 - \kappa_1)\mathbf{N}_p$. On the other hand, since the principal directions bisect the angle between the asymptotic directions, $\mathbf{w}$ must be multiple of $v_1\mathbf{e}_1 - v_2\mathbf{e}_2$, therefore, v_1v_2 and w_1w_2 have opposite signs. $\square$

3.2.5 Surfaces of Constant Negative Curvature

Surfaces of constant negative curvature are important because their local geometry coincides with the geometry of the hyperbolic plane. In this section we discuss some important facts on them.

When we scale a surface by a factor $\lambda > 0$, for example by a central homothety, the Gaussian curvature K changes to K/λ^2, so we can focus on surfaces of constant Gaussian curvature $K \equiv -1$ without loss of generality.

Important results can be obtained by considering special parameterizations of these surfaces, especially those for which the coordinate lines are asymptotic lines and those for which the coordinate lines are lines of curvature. Both types of parameterizations exist locally by Theorems 3.2.4 and 3.2.5.

Consider a regular parameterization $\mathbf{r}\colon \Omega \to \mathbb{R}^3$ of a surface with Gaussian curvature $K \equiv -1$ and assume that the coordinate lines are asymptotic lines. We shall denote by u and v the first and second variables of the parameterization respectively. In the standard notations $\mathbf{r}_i$, $\mathbf{r}_{ij}$, g_{ij}, g^{ij} b_{ij} Γ_{ij}^k, Γ_{ijk}, R_{ijk}^l of hypersurface theory, the indices i, j, k run over $1, \dots, n-1$. In the case of surfaces in $\mathbb{R}^3$, these indices are equal to 1 or 2. Below we shall replace in each of these notations all indices by the variable symbols u and v. Index 1 will be replaced by u, index 2 by v. For example, we shall write Γ_{uv}^v instead of Γ_{12}^2.

By the characterization of asymptotic lines, the condition on $\mathbf{r}$ is equivalent to $\mathbf{r}_{uu} \perp \mathbf{N}$ and $\mathbf{r}_{vv} \perp \mathbf{N}$. Therefore, the matrix of the second fundamental form is of the form $\mathcal{B} = \begin{pmatrix} 0 & C \\ C & 0 \end{pmatrix}$. Let us introduce also the notation $A = \|\mathbf{r}_u\|$, $B = \|\mathbf{r}_v\|$ and denote by $\omega(u,v) \in (0,\pi)$ the angle between $\mathbf{r}_u(u,v)$ and $\mathbf{r}_v(u,v)$. Then

$$\mathcal{G} = \begin{pmatrix} A^2 & AB\cos\omega \\ AB\cos\omega & B^2 \end{pmatrix} \tag{3.9}$$

and by the expression of the Gaussian curvature

$$-C^2 = \det(\mathcal{B}) = K\det(\mathcal{G}) = -A^2B^2\sin^2(\omega),$$

so $C = \pm AB\sin(\omega)$. Changing the order of the parameters u and v, $\mathbf{N}$ and $\mathcal{B}$ changes to $-\mathbf{N}$ and $-\mathcal{B}$, so we may assume without loss of generality that

$$\mathcal{B} = \begin{pmatrix} 0 & AB\sin\omega \\ AB\sin\omega & 0 \end{pmatrix}. \tag{3.10}$$

If there is a surface the fundamental forms of which have these special forms, then it is certain, that the coordinate lines are asymptotic lines and the surface has constant -1 Gaussian curvature. However, we know from the fundamental theorem of curve theory, that a surface with given fundamental forms exist (locally) if and only if the prescribed matrices satisfy the Gauss and Codazzi–Mainardi equations. These equations are not automatically satisfied for the above matrices leading to some partial differential equations that must be satisfied by the functions A, B and ω. Let us compute these equations.

First we observe that the inverse of $\mathcal{G}$ is

$$\mathcal{G}^{-1} = \begin{pmatrix} \dfrac{1}{A^2 \sin^2 \omega} & -\dfrac{\cos \omega}{AB \sin^2 \omega} \\ -\dfrac{\cos \omega}{AB \sin^2 \omega} & \dfrac{1}{B^2 \sin^2 \omega} \end{pmatrix}.$$

Lemma 3.2.23. *If a regular parameterized surface of negative curvature has the property that the coordinate lines are asymptotic lines and $b_{uv} > 0$, then the Codazzi–Mainardi equations reduce to the equations*

$$-\partial_u(\log b_{uv}) = \Gamma^v_{uv} - \Gamma^u_{uu}, \tag{3.11}$$

$$\partial_v(\log b_{uv}) = \Gamma^v_{vv} - \Gamma^u_{vu}. \tag{3.12}$$

Proof. The general form of the Codazzi–Mainardi equation is

$$b_{ij,k} - b_{ik,j} = \sum_{l=1}^{2} (\Gamma^l_{ik} b_{lj} - \Gamma^l_{ij} b_{lk}).$$

Since both sides are skew-symmetric in the indices j and k, these equations are true trivially when $j = k$ and the equation for i, j, k is equivalent to the equation for $(i', j', k') = (i, k, j)$. For this reason, when we have only two variables u and v then there are only two independent Codazzi–Mainardi equations for $(i, j, k) = (1, 1, 2)$ and $(i, j, k) = (2, 1, 2)$. The equations for these indices are

$$b_{uu,v} - b_{uv,u} = (\Gamma^u_{uv} b_{uu} - \Gamma^u_{uu} b_{uv}) + (\Gamma^v_{uv} b_{vu} - \Gamma^v_{uu} b_{vv}),$$

$$b_{vu,v} - b_{vv,u} = (\Gamma^u_{vv} b_{uu} - \Gamma^u_{vu} b_{uv}) + (\Gamma^v_{vv} b_{vu} - \Gamma^v_{vu} b_{vv}).$$

Taking into consideration $b_{uu} = b_{vv} = 0$, these equations reduce to

$$-b_{uv,u} = -\Gamma^u_{uu} b_{uv} + \Gamma^v_{uv} b_{vu},$$

$$b_{vu,v} = -\Gamma^u_{vu} b_{uv} + \Gamma^v_{vv} b_{vu}.$$

Dividing these equations by $b_{uv} = b_{vu}$ we obtain the required equalities for the logarithmic derivative of b_{uv} □

Lemma 3.2.24. *The Codazzi–Mainardi equations are satisfied by the matrices* (3.9) *and* (3.10) *if and only if*

$$\partial_v A = \partial_u B = 0.$$

Proof. First we compute the right-hand side of (3.11)

$$\begin{aligned}\Gamma^v_{uv} - \Gamma^u_{uu} &= \Gamma_{uvu}g^{uv} + \Gamma_{uvv}g^{vv} - \Gamma_{uuu}g^{uu} - \Gamma_{uuv}g^{vu} \\ &= \frac{1}{2}g_{uu,v}g^{uv} + \frac{1}{2}g_{vv,u}g^{vv} - \frac{1}{2}g_{uu,u}g^{uu} - g_{uv,u}g^{vu} + \frac{1}{2}g_{uu,v}g^{vu} \\ &= g_{uu,v}g^{uv} + \frac{1}{2}g_{vv,u}g^{vv} - \frac{1}{2}g_{uu,u}g^{uu} - g_{uv,u}g^{vu} \\ &= \frac{\partial_u B - 2\partial_v A\cos\omega}{B\sin^2\omega} - \frac{\partial_u A}{A\sin^2\omega} + \partial_u\big(\log(AB\cos\omega)\big)\cot^2\omega.\end{aligned}$$

Thus equation (3.11) gives

$$\begin{aligned}&-\frac{\partial_u A}{A} - \frac{\partial_u B}{B} - \cot\omega\cdot\partial_u\omega \\ &\qquad = \frac{\partial_u B - 2\partial_v A\cos\omega}{B\sin^2\omega} - \frac{\partial_u A}{A\sin^2\omega} + \partial_u\big(\log(AB\cos\omega)\big)\cot^2\omega,\end{aligned}$$

which is equivalent to

$$0 = \frac{2\partial_u B - 2\partial_v A\cos\omega}{B\sin^2\omega},$$

after some rearrangement and simplifications. This gives $\partial_u B = \partial_v A\cos\omega$. Flipping the role of the two variables, we obtain $\partial_v A = \partial_u B\cos\omega$ similarly. Since ω cannot be a multiple of π, the last two equations imply that $\partial_u B = \partial_v A = 0$. Conversely, reversing the steps of the computation, if $\partial_u B = \partial_v A = 0$, then the Codazzi–Mainardi equations are fulfilled. □

Definition 3.2.25. The net of the coordinate lines of a parametrization $\mathbf{r}\colon\Omega\to\mathbb{R}^3$ of a surface is called a *Chebyshev net* if the opposite sides of any curvilinear quadrangle bounded by segments of the coordinate lines are of equal length. ※

Lemma 3.2.26. *The coordinate lines of a regular parametrization* $\mathbf{r}\colon\Omega\to\mathbb{R}^3$ *form a Chebyshev net if and only if* $\partial_v\|\mathbf{r}_u\| = \partial_u\|\mathbf{r}_v\| \equiv 0$.

Proof. The parameter lines of $\mathbf{r}$ form a Chebyshev net if and only if for any rectangle $[u_0,u_1]\times[v_0,v_1]\subset\Omega$, we have

$$\begin{aligned}\int_{u_0}^{u_1}\|\mathbf{r}_u\|(u,v_0)du &= \int_{u_0}^{u_1}\|\mathbf{r}_u\|(u,v_1)du \\ \text{and}\quad \int_{v_0}^{v_1}\|\mathbf{r}_v\|(u_0,v)dv &= \int_{v_0}^{v_1}\|\mathbf{r}_v\|(u_1,v)dv.\end{aligned}\tag{3.13}$$

The first equation means that the integral

$$I(v) = \int_{u_0}^{u_1}\|\mathbf{r}_u\|(u,v)du$$

is constant on any interval $[v_0, v_1]$ on which it is defined. The latter condition holds if and only if

$$I'(v) = \int_{u_0}^{u_1} \partial_v \|\mathbf{r}_u\|(u,v)du \equiv 0$$

for any v u_0, u_1 such that $[u_0, u_1] \times \{v\} \subset \Omega$. This is clearly equivalent to the vanishing of $\partial_v\|\mathbf{r}_u\|$.
Changing the role of the parameters we can see that the second condition in (3.13) is equivalent to $\partial_u\|\mathbf{r}_v\| \equiv 0$. □

Thus, we obtained as a consequence of Lemma 3.2.24 (the Codazzi–Mainardi equations) that the asymptotic lines on a surface of constant Gaussian curvature -1 form a Chebyshev net. This means that we can reparameterize the surface in a way that all coordinate lines are asymptotic curves *parameterized by arc length*. Choosing such a parameterization, we shall have $A \equiv B \equiv 1$, and the matrices $\mathcal{G}$ and $\mathcal{B}$ take the simpler form

$$\mathcal{G} = \begin{pmatrix} 1 & \cos\omega \\ \cos\omega & 1 \end{pmatrix}, \qquad \mathcal{B} = \begin{pmatrix} 0 & \sin\omega \\ \sin\omega & 0 \end{pmatrix}.$$

Let us compute now what the Gauss equation gives for a surface with these fundamental form matrices. The inverse of $\mathcal{G}$ is

$$\mathcal{G}^{-1} = \begin{pmatrix} \dfrac{1}{\sin^2\omega} & -\dfrac{\cos\omega}{\sin^2\omega} \\ -\dfrac{\cos\omega}{\sin^2\omega} & \dfrac{1}{\sin^2\omega} \end{pmatrix},$$

the Christoffel symbols of the second kind are

$$\begin{aligned}
\Gamma_{uuu} &= \frac{1}{2}g_{uu,u} = 0,\\
\Gamma_{uuv} &= \frac{1}{2}(2g_{uv,u} - g_{uu,v}) = -\sin\omega\cdot\omega_u,\\
\Gamma_{uvu} &= \frac{1}{2}(g_{uv,u} + g_{uu,v} - g_{uv,u}) = 0,\\
\Gamma_{uvv} &= \frac{1}{2}(g_{uv,v} + g_{vv,u} - g_{uv,v}) = 0,\\
\Gamma_{vvu} &= \frac{1}{2}(2g_{uv,v} - g_{vv,u}) = -\sin\omega\cdot\omega_v,\\
\Gamma_{vvv} &= \frac{1}{2}g_{vv,v} = 0,
\end{aligned}$$

and the Christoffel symbols of the first kind are

$$\begin{aligned}
\Gamma^u_{uu} &= \Gamma_{uuu}g^{uu} + \Gamma_{uuv}g^{vu} = \cot\omega \cdot \omega_u,\\
\Gamma^v_{uu} &= \Gamma_{uuu}g^{uv} + \Gamma_{uuv}g^{vv} = -\frac{\omega_u}{\sin\omega},\\
\Gamma^u_{uv} &= \Gamma_{uvu}g^{uu} + \Gamma_{uvv}g^{vu} = 0,\\
\Gamma^v_{uv} &= \Gamma_{uvu}g^{uv} + \Gamma_{uvv}g^{vv} = 0,\\
\Gamma^u_{vv} &= \Gamma_{vvu}g^{uu} + \Gamma_{vvv}g^{vu} = -\frac{\omega_v}{\sin\omega},\\
\Gamma^v_{vv} &= \Gamma_{vvu}g^{uv} + \Gamma_{vvv}g^{vv} = \cot\omega \cdot \omega_v.
\end{aligned}$$

The only nontrivial Gauss equation for this surface is

$$-\sin^2\omega = \det\mathcal{B} = R_{uvvu} = R^u_{uvv}g_{uu} + R^v_{uvv}g_{vu} = R^u_{uvv} + R^v_{uvv}\cos\omega,$$

where the components of the curvature tensor are

$$\begin{aligned}
R^u_{uvv} &= \Gamma^u_{vv,u} - \Gamma^u_{vu,v} + \sum_{s\in\{u,v\}}(\Gamma^s_{vv}\Gamma^u_{su} - \Gamma^s_{vu}\Gamma^u_{sv}) = \Gamma^u_{vv,u} + \Gamma^u_{vv}\Gamma^u_{uu}\\
&= \frac{\cos\omega\cdot\omega_u\cdot\omega_v - \omega_{uv}\sin\omega}{\sin^2\omega} - \frac{\cos\omega\cdot\omega_u\cdot\omega_v}{\sin^2\omega} = -\frac{\omega_{uv}}{\sin\omega},
\end{aligned}$$

and

$$\begin{aligned}
R^v_{uvv} &= \Gamma^v_{vv,u} - \Gamma^v_{vu,v} + \sum_{s=u}^{v}(\Gamma^s_{vv}\Gamma^v_{su} - \Gamma^s_{vu}\Gamma^v_{sv}) = \Gamma^v_{vv,u} + \Gamma^u_{vv}\Gamma^v_{uu}\\
&= -\frac{\omega_u\cdot\omega_v}{\sin^2\omega} + \omega_{uv}\cot\omega + \frac{\omega_u\cdot\omega_v}{\sin^2\omega} = \omega_{uv}\cot\omega.
\end{aligned}$$

Substituting these into the Gauss equation leads to

$$-\sin^2\omega = \omega_{uv}\frac{\cos^2\omega - u}{\sin\omega} = -\omega_{uv}\sin\omega,$$

hence ω satisfies the partial differential equation $\omega_{uv} = \sin\omega$.

Definition 3.2.27. The partial differential equation $\omega_{uv} = \sin\omega$ is called the *sine-Gordon equation.* ※

We can sum up our results as follows.

Theorem 3.2.28. *A surface of constant Gaussian curvature -1 can be parameterized in such a way that the coordinate lines are the asymptotic lines parameterized by arc length. For such a parameterization the angle $\omega\colon \Omega \to (0,\pi)$ between the asymptotic lines gives a solution of the sine-Gordon equation.*

Conversely, if Ω is convex and a function $\omega\colon \Omega \to (0,\pi)$ solves the sine-Gordon equation, then there is a regular parameterized surface $\mathbf{r}\colon \Omega \to \mathbb{R}^3$ with fundamental form matrices

$$\mathcal{G} = \begin{pmatrix} 1 & \cos\omega \\ \cos\omega & 1 \end{pmatrix} \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} 0 & \sin\omega \\ \sin\omega & 0 \end{pmatrix}.$$

This surface has constant Gaussian curvature -1 and the coordinate lines of the parametrization are unit speed asymptotic curves.

Roughly speaking, surfaces with constant Gaussian curvature -1 are in one to one correspondence with solutions of the sine-Gordon equation with values in the interval $(0,\pi)$.
The integrated form of the sine-Gordon equation is the Hazzidakis formula.

Theorem 3.2.29 (Hazzidakis Formula)**.** *Let $\mathbf{r}\colon \Omega \to \mathbb{R}^3$ be a surface of constant Gaussian curvature -1 parameterized as above. Then the surface area A of the curvilinear quadrangle bounded by asymptotic lines obtained as the $\mathbf{r}$ image of the rectangle $[u_0,u_1]\times[v_0,v_1]\subset\Omega$ is*

$$\omega(u_1,v_1)-\omega(u_0,v_1)-\omega(u_1,v_0)+\omega(u_0,v_0).$$

Proof. By the definition of the surface area

$$A=\int_{v_0}^{v_1}\int_{u_0}^{u_1}\sqrt{\det(\mathcal{G}(u,v)}dudv=\int_{v_0}^{v_1}\int_{u_0}^{u_1}\sin(\omega(u,v))dudv.$$

As ω solves the sine-Gordon equation,

$$\begin{aligned}\int_{v_0}^{v_1}\int_{u_0}^{u_1}\sin(\omega(u,v))dudv &= \int_{v_0}^{v_1}\int_{u_0}^{u_1}\omega_{uv}(u,v))dudv \\ &= \int_{v_0}^{v_1}(\omega_v(u_1,v)-\omega_v(u_0,v))dv \\ &= \omega(u_1,v_1)-\omega(u_0,v_1)-\omega(u_1,v_0)+\omega(u_0,v_0). \quad \square\end{aligned}$$

Corollary 3.2.30. *The area A of any curvilinear quadrangle bounded by asymptotic lines is less then 2π. The function ω satisfies the inequality*

$$2\pi > \omega(u_1,v_1)-\omega(u_0,v_1)-\omega(u_1,v_0)+\omega(u_0,v_0) > 0. \tag{3.14}$$

Definition 3.2.31. Let $\mathbf{r}\colon \Omega\to\mathbb{R}^3$ be a regular parameterized surface and define for $\mathbf{u}_0\in\Omega$ and $R>0$ the set $B(\mathbf{u}_0,R)$ as the set of those points $\mathbf{v}\in\Omega$ for which there is a smooth curve $\mathbf{u}\colon[0,1]\to\Omega$ connecting $\mathbf{u}_0=\mathbf{u}(0)$ to $\mathbf{v}=\mathbf{u}(1)$ so that the length of $\mathbf{r}\circ\mathbf{u}$ is less then R. The surface $\mathbf{r}$ is said to be *complete* if for any $\mathbf{u}_0\in\Omega$ and any $R>0$, the closure of $B(\mathbf{u}_0,R)$ is contained in Ω. ※

Intuitively, completeness means that moving along a curve of finite length on the surface we cannot reach (converge to) points that do not belong to the surface.

Theorem 3.2.32 (Hilbert's Theorem)**.** *There is no complete regular surface of constant negative Gaussian curvature in* $\mathbb{R}^3$*.*

Proof. Suppose to the contrary that such a surface exists. Then by completeness, for any $p \in \mathbf{r}(\Omega)$ both maximal asymptotic curves starting at p have a parameterization by arc length *defined on the whole real line*. Let $\gamma_p \colon \mathbb{R} \to \mathbf{r}(\Omega)$ be a unit speed parameterization of one of the asymptotic lines starting at $\gamma(0) = p$, and for each $t \in \mathbb{R}$, denote by $\eta_{\gamma_p(t)} \colon \mathbb{R} \to \mathbf{r}(\Omega)$ the unit speed parameterization of the asymptotic line starting at $\eta_{\gamma_p(t)}(0) = \gamma_p(t)$. Orient the initial speed of $\eta_{\gamma_p(t)}$ so that $\gamma'(t) \times \eta'_{\gamma_p(t)}(0)$ be a positive multiple of the normal vector $\mathbf{N}_{\gamma_p(t)}$ for all t. Then the map $\tilde{\mathbf{r}} \colon \mathbb{R}^2 \to \mathbb{R}^3$, $\tilde{\mathbf{r}}(u, v) = \eta_{\gamma_p(u)}(v)$ is a regular parameterized surface with constant Gaussian curvature -1, for which the coordinate lines are naturally parameterized asymptotic curves. (We do not claim that $\tilde{\mathbf{r}}$ is injective, nor that it has the same image as $\mathbf{r}$.) This map $\tilde{\mathbf{r}}$ would define a solution of the sine-Gordon equation which would take values in the interval $(0, \pi)$. However, as the next lemma states, there are no such solutions. $\square$

Lemma 3.2.33. *The sine-Gordon equation has no solution on the band* $[0, 1] \times \mathbb{R}$ *for which* $0 < \omega < \pi$*.*

Proof. Assume to the contrary that such a solution exists. We discuss three cases separately. In the first case, suppose that $\omega(1, 0) > \omega(0, 0)$. Set $c = (\omega(1, 0) - \omega(0, 0))/3$ and let

$$\hat{u}_1 = \max\{s \in [0, 1] \colon \omega(s, 0) - \omega(0, 0) = c\},$$
$$\hat{u}_2 = \min\{s \in [0, 1] \colon \omega(1, 0) - \omega(s, 0) = c\}.$$

Then it is clear that $0 < \hat{u}_1 < \hat{u}_2 < 1$ and for any $t \in [\hat{u}_1, \hat{u}_2]$, we have

$$c + \omega(0, 0) \leq \omega(t, 0) \leq \omega(1, 0) - c$$

Let $t \in [\hat{u}_1, \hat{u}_2]$ and $v > 0$ be arbitrary. Then applying inequality (3.14), we obtain

$$\omega(t, v) > \omega(t, 0) + \omega(0, v) - \omega(0, 0) > \omega(t, 0) - \omega(0, 0) \geq c,$$

and

$$\omega(t, v) < \omega(1, v) + \omega(t, 0) - \omega(1, 0) \leq \omega(1, v) - c < \pi - c.$$

The last two inequalities imply that $\sin(\omega(t,v)) \geq \sin(c)$ for $(t,v) \in [\hat{u}_1, \hat{u}_2] \times [0, +\infty)$. Integrating on the rectangle $[\hat{u}_1, \hat{u}_2] \times [0, v]$ gives

$$\int_0^v \int_{\hat{u}_1}^{\hat{u}_2} \sin(\omega(u,v))dudv \geq v(\hat{u}_2 - \hat{u}_1)\sin(c).$$

On the other hand, the integral is given by the Hazzidakis formula and

$$\begin{aligned}\int_0^v \int_{\hat{u}_1}^{\hat{u}_2} & \sin(\omega(u,v))dudv \\ &= \omega(\hat{u}_2, v) - \omega(\hat{u}_1, v) \\ &\quad - \omega(\hat{u}_2, 0) + \omega(\hat{u}_1, 0) < 2\pi.\end{aligned}$$

However, this is a contradiction, because the inequality $2\pi > v(\hat{u}_2 - \hat{u}_1)\sin(c)$ cannot be true for any positive v.

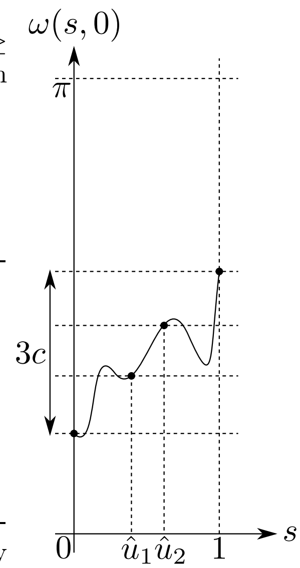

Figure 3.10

If $\omega(0,0) > \omega$, then consider the function $\tilde{\omega}(u,v) = \omega(1-u, -v)$. The function $\tilde{\omega}$ is also defined on $[0,1] \times \mathbb{R}$, it solves the sine-Gordon equation, takes its values in the interval $(0,\pi)$, but $\tilde{\omega}(1,0) > \tilde{\omega}(0,0)$, so the second case can be reduced to the first case.
Finally, assume $\omega(0,0) = \omega(1,0)$. Then introduce the function $\tilde{\omega}(u,v) = \omega(u, v+1)$. The function $\tilde{\omega}$ is again defined on $[0,1] \times \mathbb{R}$, it solves the sine-Gordon equation, takes its values in the interval $(0,\pi)$, furthermore, using inequality (3.14)

$$\tilde{\omega}(1,0) - \tilde{\omega}(0,0) = \omega(1,1) - \tilde{\omega}(0,1) > \omega(1,0) - \omega(0,0) = 0.$$

Applying the arguments of the first case to $\tilde{\omega}$ we obtain a contradiction. $\square$

Bäcklund transform

A Bäcklund transform in contemporary mathematics is a method to find new solutions to certain partial differential equations if one particular solution is already known. Bäcklund transforms have their origins in differential geometry. Namely, in the 1880's L. Bianchi and A.V. Bäcklund introduced a nontrivial geometrical construction of new surfaces of constant negative Gaussian curvature from an initial such surface using a solution of a first order partial differential equation. As surfaces of constant negative curvature can be described by solutions of the sine-Gordon equation, Bäcklund transformation of surfaces can be viewed as a transformation of solutions of the sine-Gordon equation.

Definition 3.2.34. Let $\mathbf{r}\colon \Omega \to \mathbb{R}^3$ be a surface of constant Gaussian curvature -1. We say that another regular parameterized hypersurface $\tilde{\mathbf{r}}\colon \Omega \to \mathbb{R}^3$ is a *Bäcklund transform of* $\mathbf{r}$ *with parameter* σ if

(1) $\|\tilde{\mathbf{r}}(u,v) - \mathbf{r}(u,v)\| \equiv \cos\sigma$,

(2) $\tilde{\mathbf{r}}(u,v) - \mathbf{r}(u,v)$ is tangent to $\mathbf{r}$ at $\mathbf{r}(u,v)$ and $\tilde{\mathbf{r}}$ at $\tilde{\mathbf{r}}(u,v)$,

(3) the angle between $\tilde{\mathbf{N}}$ and $\mathbf{N}$ is $\frac{\pi}{2} \pm \sigma$.

Bianchi transform of a surface is a Bäcklund transform with parameter 0. ※

To describe Bäcklund transforms of a surface, it is convenient to reparameterize it with the lines of curvature as the parameter lines. If $\mathbf{r}$ is the parameterization of the surface for which the parameter lines are naturally parameterized asymptotic curves, then $\hat{\mathbf{r}}(u,v) = \mathbf{r}(\frac{u+v}{2}, \frac{v-u}{2})$ is a reparameterization such that $\hat{\mathbf{r}}_u(u,v) = \frac{1}{2}\mathbf{r}_u(\frac{u+v}{2}, \frac{v-u}{2}) + \frac{1}{2}\mathbf{r}_v(\frac{u+v}{2}, \frac{v-u}{2})$ and $\hat{\mathbf{r}}_v(u,v) = -\frac{1}{2}\mathbf{r}_u(\frac{u+v}{2}, \frac{v-u}{2}) + \frac{1}{2}\mathbf{r}_v(\frac{u+v}{2}, \frac{v-u}{2})$ point in principal directions. Denote by θ the function defined by $\theta(u,v) = \frac{1}{2}\omega(\frac{u+v}{2}, \frac{v-u}{2})$, where ω is the angle between $\mathbf{r}_u$ and $\mathbf{r}_v$. As

$$\begin{aligned}(\theta_{uu} - \theta_{vv})(u,v) &= \frac{1}{8}([\omega_{uu} + 2\omega_{uv} + \omega_{vv}] \\ &\qquad - [\omega_{uu} - 2\omega_{uv} + \omega_{vv}])\left(\frac{u+v}{2}, \frac{v-u}{2}\right) \\ &= 2\omega_{uv}\left(\frac{u+v}{2}, \frac{v-u}{2}\right),\end{aligned}$$

the sine-Gordon equation for ω can be rephrased in terms of θ as follows

$$\theta_{uu} - \theta_{vv} = \sin\theta\cos\theta.$$

The matrices of the fundamental forms of the parameterization $\hat{\mathbf{r}}$ are

$$\hat{\mathcal{G}} = \begin{pmatrix}\cos^2\theta & 0 \\ 0 & \sin^2\theta\end{pmatrix} \qquad \text{and} \qquad \hat{\mathcal{B}} = \begin{pmatrix}\sin\theta\cos\theta & 0 \\ 0 & -\sin\theta\cos\theta\end{pmatrix}.$$

Computing the Christoffel symbols from $\hat{\mathcal{G}}$ we can express the second order partial derivatives of $\hat{\mathbf{r}}$ as linear combinations of the Gauss frame. We omit the details of this mechanical computation, and give just the result

$$\begin{aligned}\hat{\mathbf{r}}_{uu} &= -\tan\theta\cdot\theta_u\cdot\hat{\mathbf{r}}_u + \cot\theta\cdot\theta_v\cdot\hat{\mathbf{r}}_v + \sin\theta\cos\theta\hat{\mathbf{N}} \\ \hat{\mathbf{r}}_{uv} &= -\tan\theta\cdot\theta_v\cdot\hat{\mathbf{r}}_u + \cot\theta\cdot\theta_u\cdot\hat{\mathbf{r}}_v \\ \hat{\mathbf{r}}_{vv} &= -\tan\theta\cdot\theta_u\cdot\hat{\mathbf{r}}_u + \cot\theta\cdot\theta_v\cdot\hat{\mathbf{r}}_v - \sin\theta\cos\theta\hat{\mathbf{N}}\end{aligned} \tag{3.15}$$

From the first two conditions, a Bäcklund transform of $\hat{\mathbf{r}}$ must be of the form

$$\tilde{\mathbf{r}} = \hat{\mathbf{r}} + \cos\sigma\left(\cos\bar{\theta}\cdot\frac{\hat{\mathbf{r}}_u}{\cos\theta} + \sin\bar{\theta}\cdot\frac{\hat{\mathbf{r}}_v}{\sin\theta}\right),$$

where $\bar{\theta}$ is an unknown angle function which encodes the direction of the vector $\tilde{\mathbf{r}}-\hat{\mathbf{r}}$. Let us compute the restrictions $\bar{\theta}$ must obey in order to guarantee that $\tilde{\mathbf{r}}$ is a Bäcklund transformation of $\hat{\mathbf{r}}$.
Applying the formulas in (3.15), one can calculate the expressions

$$\begin{aligned}
\tilde{\mathbf{r}}_u &= \left[1 - \cos\sigma\frac{\sin\bar{\theta}}{\cos\theta}(\bar{\theta}_u + \theta_v)\right]\hat{\mathbf{r}}_u + \left[\cos\sigma\frac{\cos\bar{\theta}}{\sin\theta}(\bar{\theta}_u + \theta_v)\right]\hat{\mathbf{r}}_v \\
&\quad + \left[\cos\sigma\sin\theta\cos\bar{\theta}\right]\hat{\mathbf{N}} \\
\tilde{\mathbf{r}}_v &= \left[-\cos\sigma\frac{\sin\bar{\theta}}{\cos\theta}(\bar{\theta}_v + \theta_u)\right]\hat{\mathbf{r}}_u + \left[1 + \cos\sigma\frac{\cos\bar{\theta}}{\sin\theta}(\theta_u + \bar{\theta}_v)\right]\hat{\mathbf{r}}_v \\
&\quad - \left[\cos\sigma\sin\bar{\theta}\cos\theta\right]\hat{\mathbf{N}}
\end{aligned}$$

for the partial derivatives of $\tilde{\mathbf{r}}$.
Since $\tilde{\mathbf{N}}$ must be orthogonal to $\tilde{\mathbf{r}} - \hat{\mathbf{r}}$ and should enclose angle $\frac{\pi}{2} \pm \sigma$ with $\hat{\mathbf{N}}$, it must be of the form

$$\tilde{\mathbf{N}} = \pm\Big(\pm\cos\sigma\left(-\sin\bar{\theta}\cdot\frac{\hat{\mathbf{r}}_u}{\cos\theta} + \cos\bar{\theta}\cdot\frac{\hat{\mathbf{r}}_v}{\sin\theta}\right) + \sin\sigma\cdot\hat{\mathbf{N}}\Big)$$

The angle function $\bar{\theta}$ of the transformation is restricted by the conditions $\langle\tilde{\mathbf{N}}, \tilde{\mathbf{r}}_u\rangle = \langle\tilde{\mathbf{N}}, \tilde{\mathbf{r}}_v\rangle = 0$. Computing these dot products

$$\begin{aligned}
0 = \langle\tilde{\mathbf{N}}, \tilde{\mathbf{r}}_u\rangle &= \cos\sigma\Big(\pm\big(-\sin\bar{\theta}\cos\theta + \cos\sigma(\bar{\theta}_u + \theta_v)\big) + \sin\sigma\sin\theta\cos\bar{\theta}\Big) \\
0 = \langle\tilde{\mathbf{N}}, \tilde{\mathbf{r}}_v\rangle &= \cos\sigma\Big(\pm\big(\sin\theta\cos\bar{\theta} + \cos\sigma(\theta_u + \bar{\theta}_v)\big) - \sin\sigma\sin\bar{\theta}\cos\theta\Big)
\end{aligned}$$

We can assume $\cos\sigma \neq 0$ otherwise the Bäcklund transform coincides with the original surface and this case is not of interest. Then rearranging the equations for $\bar{\theta}$ we obtain

$$\begin{aligned}
\bar{\theta}_u + \theta_v &= \frac{\sin\bar{\theta}\cos\theta \mp \sin\sigma\cos\bar{\theta}\sin\theta}{\cos\sigma} \\
\theta_u + \bar{\theta}_v &= -\frac{\sin\theta\cos\bar{\theta} \mp \sin\sigma\cos\theta\sin\bar{\theta}}{\cos\sigma}
\end{aligned} \tag{3.16}$$

By Frobenius Theorem (Theorem 1.7.10), this system of partial differential equations has a unique solution for $\bar{\theta}$ in the a neighborhood of any point

$(u_0, v_0) \in \Omega$ with any initial condition $\bar\theta(u_0, v_0) = \bar\theta_0$ if and only if the equations do not contradict the Young theorem $\bar\theta_{uv} = \bar\theta_{vu}$. As

$$\begin{aligned}
(\bar\theta_u)_v &= -\theta_{vv} \\
&+ \frac{\cos\bar\theta\cos\theta\bar\theta_v - \sin\bar\theta\sin\theta\theta_v \mp \sin\sigma(-\sin\bar\theta\sin\theta\bar\theta_v + \cos\bar\theta\cos\theta\theta_v)}{\cos\sigma}, \\
(\bar\theta_v)_u &= -\theta_{uu} \\
&+ \frac{-\cos\theta\cos\bar\theta\theta_u + \sin\bar\theta\sin\theta\bar\theta_u \pm \sin\sigma(-\sin\bar\theta\sin\theta\theta_u + \cos\bar\theta\cos\theta\bar\theta_u)}{\cos\sigma},
\end{aligned}$$

the integrability condition for $\bar\theta$ is

$$\begin{aligned}
0 = \theta_{uu} - \theta_{vv} &+ \frac{\cos\bar\theta\cos\theta \pm \sin\sigma\sin\bar\theta\sin\theta}{\cos\sigma}(\theta_u + \bar\theta_v) \\
&+ \frac{-\sin\bar\theta\sin\theta \mp \sin\sigma\cos\theta\cos\bar\theta}{\cos\sigma}(\bar\theta_u + \theta_v) \\
= \theta_{uu} - \theta_{vv} &+ \frac{\cos\bar\theta\cos\theta \pm \sin\sigma\sin\bar\theta\sin\theta}{\cos\sigma} \cdot \frac{-\sin\theta\cos\bar\theta \pm \sin\sigma\cos\theta\sin\bar\theta}{\cos\sigma} \\
&+ \frac{-\sin\bar\theta\sin\theta \mp \sin\sigma\cos\theta\cos\bar\theta}{\cos\sigma} \cdot \frac{\sin\bar\theta\cos\theta \mp \sin\sigma\cos\bar\theta\sin\theta}{\cos\sigma} \\
= \theta_{uu} - \theta_{vv} &- \sin\theta\cos\theta.
\end{aligned}$$

However, this condition is fulfilled, since this is essentially the sine-Gordon equation written in terms of θ.
This computation proves the existence theorem for Bäcklund transforms. Before we formulate the theorem, observe, that the sign $\pm$ can be incorporated with σ as its sign.

Theorem 3.2.35. *For any choice of the point $(u_0, v_0) \in \Omega$, the parameter $\sigma \in [-\pi/2, \pi/2]$, a vector $\mathbf{v}$ of length $\cos\sigma$ tangent to $\hat{\mathbf{r}}$ at (u_0, v_0), there is a unique Bäcklund transform $\tilde{\mathbf{r}}$ of $\hat{\mathbf{r}}$, defined in a neighborhood of (u_0, v_0), which shifts $\hat{\mathbf{r}}(u_0, v_0)$ to $\tilde{\mathbf{r}}(u_0, v_0) = \hat{\mathbf{r}}(u_0, v_0) + \mathbf{v}$.*

Theorem 3.2.36. *Let $\tilde{\mathbf{r}}$ be a σ parameter Bäcklund transform of $\hat{\mathbf{r}}$. Then*

- *$\tilde{\mathbf{r}}$ is regular at (u, v) if and only if $\tilde{\mathbf{r}}(u, v) - \hat{\mathbf{r}}(u, v)$ is not a principal direction of $\hat{\mathbf{r}}$.*
- *The Bäcklund transform $\tilde{\mathbf{r}}$ has constant Gaussian curvature -1.*
- *the parameter lines of $\tilde{\mathbf{r}}$ are the lines of curvature of the Bäcklund transform.*

- *The angle of the asymptotic curves of* $\tilde{\mathbf{r}}$ *is* $2\bar{\theta}$. *Parameter lines of the reparameterization* $\tilde{\mathbf{r}}(u+v, v-u)$ *are the asymptotic curves of the Bäcklund transform. That is, the natural correspondence between the points of the original surface and its Bäcklund transform moves lines of curvature to lines of curvature, and asymptotic curves to asymptotic curves.*

Proof. A lengthy but elementary computation shows that the matrix of the first fundamental form of $\tilde{\mathbf{r}}$ is

$$\tilde{\mathcal{G}} = \begin{pmatrix} \cos^2\bar{\theta} & 0 \\ 0 & \sin^2\bar{\theta} \end{pmatrix}$$

This shows that singular points of the Bäcklund transform are the points where $\bar{\theta}$ is an integer multiple of $\pi/2$, which means that $\tilde{\mathbf{r}} - \hat{\mathbf{r}}$ is parallel to one of the principal directions $\hat{\mathbf{r}}_u$, $\hat{\mathbf{r}}_v$.
Another use of the knowledge of $\tilde{\mathcal{G}}$ is that we can compute the Gaussian curvature from it by the Theorema Egregium. It gives that the Gaussian curvature of $\tilde{\mathbf{r}}$ is

$$\tilde{K} = \frac{\bar{\theta}_{vv} - \bar{\theta}_{uu}}{sin\bar{\theta}\cos\bar{\theta}}.$$

$\tilde{K} \equiv -1$ is equivalent to the variant of the sine-Gordon equation $\bar{\theta}_{uu} - \bar{\theta}_{vv} = \sin\bar{\theta}\cos\bar{\theta}$. We saw that the sine-Gordon equation for θ was the integrability condition for the partial differential equation (3.16) for the unknown $\bar{\theta}$. The functions θ and $\bar{\theta}$ enter into (3.16) in a symmetric way. Since $\bar{\theta}$ is obtained as a solution of (3.16) for a given θ, (3.16) does have a solution in θ given $\bar{\theta}$. Thus $\bar{\theta}$ satisfies the integrability condition $\bar{\theta}_{uu} - \bar{\theta}_{vv} = \sin\bar{\theta}\cos\bar{\theta}$.
To compute the matrix of the second fundamental form of $\tilde{\mathbf{r}}$ we first compute partial derivative of $\tilde{\mathbf{N}}$ with respect to v

$$\begin{aligned}\pm\tilde{\mathbf{N}}_v = \pm\cos\sigma\left(-\frac{\cos\bar{\theta}}{\cos\theta}(\bar{\theta}_v + \theta_u)\hat{\mathbf{r}}_u - \frac{\sin\bar{\theta}}{\sin\theta}(\bar{\theta}_v + \theta_u)\hat{\mathbf{r}}_v - \cos\bar{\theta}\cos\theta\hat{\mathbf{N}}\right)\\ + \sin\sigma\hat{\mathbf{N}}_v\end{aligned}$$

From this, again some simple but tedious computation shows $\langle\tilde{\mathbf{r}}_u, \tilde{\mathbf{N}}_v\rangle = 0$, which means that for the parametrization $\tilde{\mathbf{r}}$ of the Bäcklund transform the coordinate lines are lines of curvature.
With some more work we can compute also $\langle\tilde{\mathbf{r}}_v, \tilde{\mathbf{N}}_v\rangle = \pm\sin\bar{\theta}\cos\bar{\theta}$. Since the $\tilde{\mathcal{B}}$ is symmetric and its determinant is $\tilde{K}\cdot\det\tilde{\mathcal{G}} = -\cos^2\bar{\theta}\sin^2\bar{\theta}$, $\tilde{\mathcal{B}}$ must be of the form

$$\tilde{\mathcal{B}} = \begin{pmatrix} \pm\sin\bar{\theta}\cos\bar{\theta} & 0 \\ 0 & \mp\sin\bar{\theta}\cos\bar{\theta} \end{pmatrix}.$$

This gives at once that $\tilde{\mathbf{r}}_u \pm \tilde{\mathbf{r}}_v$ point in asymptotic directions, from which the rest of the theorem follows. □

3.2.6 Minimal Surfaces

When a wire frame is dipped into a soap solution, and pulled out, in most cases a soap film is created. When a soap film is stable, its surface area is locally minimal in the sense that small perturbations of the soap film do not decrease the surface area. Our goal below is to find a differential geometric consequence of local minimality. For this purpose, we shall use the notion of variation of a parameterized hypersurface.

Definition 3.2.37. A *smooth variation of a regular parameterized hypersurface* $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ is a smooth map $\mathbf{R}\colon \Omega \times (-\delta, \delta) \to \mathbb{R}^n$, such that for all $t \in (-\delta, \delta)$, the map $\mathbf{r}_t\colon \Omega \to \mathbb{R}^n$ defined by $\mathbf{r}_t(\mathbf{u}) = \mathbf{R}(\mathbf{u}, t)$ is a regular parameterized hypersurface and $\mathbf{r}_0 = \mathbf{r}$.
A variation is *compactly supported* if there is a compact set $K \subset \Omega$, such that $\mathbf{R}(u, t) = \mathbf{r}(u)$ for all $\mathbf{u} \notin K$ and $t \in (-\delta, \delta)$. ※

For each point $\mathbf{r}(\mathbf{u})$ of the hypersurface, the variation $\mathbf{R}$ defines a parameterized curve $\mathbf{R}(\mathbf{u}, .)\colon (-\delta, \delta) \to \mathbb{R}^n$ which describes a motion of that point. The image of the parameterized hypersurface $\mathbf{r}_t$ consists of those points at which the points of the original hypersurface $\mathbf{r}$ arrive after moving on for time t. Assigning to each point of the hypersurface the initial speed vector of its motion, we obtain a vector field along the hypersurface, called the infinitesimal variation associated to the variation $\mathbf{R}$.

Definition 3.2.38. The *infinitesimal variation associated to the variation* $\mathbf{R}$ is the vector field $X = \partial_t \mathbf{R}(., 0)\colon \Omega \to \mathbb{R}^n$ along the hypersurface $\mathbf{r}$, where $\partial_t \mathbf{R} = \partial_n \mathbf{R}$ is the partial derivative of $\mathbf{R}$ with respect to the nth variable, (which we denote by t now). ※

If the variation $\mathbf{R}$ is supported by a compact set K, then X vanishes outside K, so X has compact support.

Theorem 3.2.39 (First Variation of the Surface Volume)**.** *Assume that a variation* $\mathbf{R}\colon \Omega \times (-\delta, \delta) \to \mathbb{R}^n$ *of a hypersurface* $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ *is supported by a compact set* $K \subset \Omega$*. Denote by* X *the infinitesimal variation associated to* $\mathbf{R}$*, by* H *the Minkowski curvature of* $\mathbf{r}$*, and by* $\mathcal{G}$ *the matrix of the first fundamental form of* $\mathbf{r}$ *with respect to the basis* $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$*. Then the derivative of the surface volume* $V_{n-1}(t)$ *of* $\mathbf{r}_t(K)$ *is*

$$V'_{n-1}(0) = -(n-1) \int_K \langle X, \mathbf{N} \rangle H \sqrt{\det \mathcal{G}} \, d\lambda_{n-1}.$$

Proof. By the definition of the surface volume,

$$V_{n-1}(t) = \int_K \sqrt{\det \mathcal{G}_t} \, d\lambda_{n-1},$$

where $\mathcal{G}_t$ is the Gram matrix of the partial derivatives $\mathbf{r}_{t,1}, \dots, \mathbf{r}_{t,n-1}$. Differentiating with respect to t at $t = 0$, we obtain

$$V'_{n-1}(0) = \int_K \frac{d}{dt}\sqrt{\det \mathcal{G}_t}|_{t=0}\, d\lambda_{n-1} = \int_K \frac{d}{dt}\|\mathbf{r}_{t,1} \wedge \dots \wedge \mathbf{r}_{t,n-1}\|\big|_{t=0}\, d\lambda_{n-1}. \tag{3.17}$$

Applying the chain rule, the Leibniz rule for the differentiation of dot and wedge products, and Young's theorem, we get

$$\begin{aligned}\frac{d}{dt}\|\mathbf{r}_{t,1} \wedge \dots \wedge \mathbf{r}_{t,n-1}\|\big|_{t=0} &= \frac{\left\langle \frac{d}{dt}\mathbf{r}_{t,1} \wedge \dots \wedge \mathbf{r}_{t,n-1}\big|_{t=0}, \mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\right\rangle}{\|\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\|} \\ &= \frac{\left\langle \sum_{i=1}^{n-1} \mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{i-1} \wedge \partial_t\partial_i \mathbf{R}(.,0) \wedge \mathbf{r}_{i+1} \wedge \dots \wedge \mathbf{r}_{n-1}, \mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\right\rangle}{\|\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\|} \\ &= \frac{\left\langle \sum_{i=1}^{n-1} \mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{i-1} \wedge X_i \wedge \mathbf{r}_{i+1} \wedge \dots \wedge \mathbf{r}_{n-1}, \mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\right\rangle}{\|\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\|},\end{aligned}$$

and

$$V'_{n-1}(0) = \int_K \frac{\left\langle \sum_{i=1}^{n-1} \mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{i-1} \wedge X_i \wedge \mathbf{r}_{i+1} \wedge \dots \wedge \mathbf{r}_{n-1}, \mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\right\rangle}{\|\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\|}\, d\lambda_{n-1}. \tag{3.18}$$

Decompose X into the sum of a vector field $X^\top$ tangential to the hypersurface $\mathbf{r}$ and a vector field $X^\perp$ orthogonal to it. Obviously, the orthogonal component is $X^\perp = \langle X, \mathbf{N}\rangle\mathbf{N}$. The key observation for the proof of the theorem is that the tangential component $X^\top$ has no contribution to the integral in (3.18), that is,

$$\int_K \frac{\left\langle \sum_{i=1}^{n-1} \mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{i-1} \wedge X_i^\top \wedge \mathbf{r}_{i+1} \wedge \dots \wedge \mathbf{r}_{n-1}, \mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\right\rangle}{\|\mathbf{r}_1 \wedge \dots \wedge \mathbf{r}_{n-1}\|}\, d\lambda_{n-1} = 0.$$

To prove this fact, decompose $X^\top$ as a linear combination of the vector fields $\mathbf{r}_i$. The coefficients in $X^\top = \sum_{i=1}^{n-1} F^i \mathbf{r}_i$ are smooth functions on Ω, thus, they yield a smooth vector field $F = (F^1, \dots, F^{n-1})\colon \Omega \to \mathbb{R}^{n-1}$ on Ω. F vanishes outside K, so a maximal integral curve of the differential equation $\gamma'(t) = F(\gamma(t))$ is constant if it starts at a point $\gamma(0) \notin K$ and it remains in K if it starts at a point of K. This means that the flow of the differential equation consists of diffeomorphisms $\Phi_t\colon \Omega \to \Omega$ for $t \in \mathbb{R}$ such that $\Phi_t(K) = K$.

Consider the variation $\tilde{\mathbf{R}}\colon \Omega \times \mathbb{R} \to \mathbb{R}^n$, $\tilde{\mathbf{R}}(\mathbf{u}, t) = \mathbf{r}(\Phi_t(\mathbf{u}))$ of the hypersurface $\mathbf{r}$. The hypersurfaces $\tilde{\mathbf{r}}_t = \tilde{\mathbf{R}}(., t) = \mathbf{r} \circ \Phi_t$ are simply reparameterizations of the initial hypersurface $\mathbf{r}$. Thus, by Proposition 3.1.38, the surface volume $\tilde{V}_{n-1}(t)$ of $\tilde{\mathbf{r}_t}(K)$ is constant, in particular, its derivative with respect to t is 0.
The infinitesimal variation of $\mathbf{r}$ associated to the variation $\tilde{\mathbf{R}}$ is the vector field

$$\begin{aligned}\partial_t \tilde{\mathbf{R}}(., 0) &= \frac{d}{dt}(\mathbf{r} \circ \Phi_t)|_{t=0} \\ &= (\mathbf{r}' \circ \Phi_0) \cdot \left(\frac{d}{dt}\Phi_t|_{t=0}\right) = (\mathbf{r}') \cdot \begin{pmatrix} F^1 \\ \vdots \\ F^{n-1} \end{pmatrix} = \sum_{i=1}^{n-1} F^i \mathbf{r}_i = X^\top .\end{aligned}$$

Thus, the application of equation (3.18) to the variation $\tilde{\mathbf{R}}$ yields

$$\begin{aligned} 0 &= \tilde{V}'_{n-1}(0) \\ &= \int_K \frac{\left\langle \sum_{i=1}^{n-1} \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{i-1} \wedge X_i^\top \wedge \mathbf{r}_{i+1} \wedge \cdots \wedge \mathbf{r}_{n-1}, \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1} \right\rangle}{\|\mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1}\|}\end{aligned}$$

as we claimed. If we omit the vanishing contribution of $X^\top$ from (3.18), then

$$\begin{aligned} &V'_{n-1}(0) \\ &= \int_K \frac{\left\langle \sum_{i=1}^{n-1} \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{i-1} \wedge X_i^\perp \wedge \mathbf{r}_{i+1} \wedge \cdots \wedge \mathbf{r}_{n-1}, \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1} \right\rangle}{\|\mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1}\|} \, d\lambda_{n-1}\end{aligned}$$

remains. The partial derivatives of $X^\perp = \langle X, \mathbf{N} \rangle \mathbf{N}$ are $X_i^\perp = \partial_i(\langle X, \mathbf{N} \rangle)\mathbf{N} + \langle X, \mathbf{N} \rangle \mathbf{N}_i$. As $\mathbf{N}$ is orthogonal to all partial derivatives of $\mathbf{r}$,

$$\langle \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{i-1} \wedge \mathbf{N} \wedge \mathbf{r}_{i+1} \wedge \cdots \wedge \mathbf{r}_{n-1}, \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1} \rangle = 0$$

by equation (1.9). Write $\mathbf{N}_i = -L(\mathbf{r}_i)$, where L is the Weingarten map of $\mathbf{r}$ as a linear combination

$$\mathbf{N}_i = -L(\mathbf{r}_i) = -\sum_{j=1}^{n-1} l_i^j \mathbf{r}_j$$

of the partial derivatives of $\mathbf{r}$. The coefficients l_i^j in the decomposition are the entries of the matrix of L with respect to the basis $\mathbf{r}_1, \dots, \mathbf{r}_{n-1}$. Then

$$\sum_{i=1}^{n-1} \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{i-1} \wedge \mathbf{N}_i \wedge \mathbf{r}_{i+1} \wedge \cdots \wedge \mathbf{r}_{n-1} = -\left(\sum_{i=1}^{n-1} l_i^i\right) \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1}$$

The sum on the right-hand side is $\operatorname{tr} L = (n-1)H$. Substituting back the results of these partial computations we obtain

$$\begin{aligned} V'_{n-1}(0) &= \int_K \frac{-(n-1)H\langle X, \mathbf{N}\rangle \langle \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1}, \mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1}\rangle}{\|\mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1}\|} \, d\lambda_{n-1} \\ &= -(n-1)\int_K \langle X, \mathbf{N}\rangle H \|\mathbf{r}_1 \wedge \cdots \wedge \mathbf{r}_{n-1}\| \, d\lambda_{n-1} \\ &= -(n-1)\int_K \langle X, \mathbf{N}\rangle H \sqrt{\det \mathcal{G}} \, d\lambda_{n-1}. \qquad \square \end{aligned}$$

The theorem gives a necessary condition for a hypersurface to minimize the surface volume among hypersurfaces with given boundary.

Corollary 3.2.40. *Let $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ be a regular parameterized hypersurface. Assume that for any compact subset $K \subset \Omega$, the surface volume of $\mathbf{r}(K)$ is minimal among the surface volumes of $\tilde{\mathbf{r}}(K)$, where $\tilde{\mathbf{r}}\colon \Omega \to \mathbb{R}^n$ is running over all regular hypersurfaces coinciding with $\mathbf{r}$ outside K. Then $\mathbf{r}$ has constant Minkowski curvature* 0.

Proof. Suppose to the contrary that $H(\mathbf{u}_0) \neq 0$ for a point $\mathbf{u}_0 \in \Omega$. As H is continuous, we can find a closed ball $K = \bar{B}(\mathbf{u}_0, r) \subset \Omega$ around $\mathbf{u}_0$ such that H is positive on K. Let $h\colon \mathbb{R}^{n-1} \to [0,1]$ be a smooth function, which is constant 1 on the ball $B(\mathbf{u}_0, r/2)$ and 0 outside K.
Consider the variation $\mathbf{R}(\mathbf{u}, t) = \mathbf{r}(\mathbf{u}) + t \cdot h(\mathbf{u})\mathbf{N}(\mathbf{u})$ of $\mathbf{r}$. It is supported by the compact set K, and its associated infinitesimal variation is $h\mathbf{N}$. Applying Theorem 3.2.39 to this variation we obtain

$$V'_{n-1}(0) = -(n-1)\int_K hH\sqrt{\det \mathcal{G}} d\lambda_{n-1}.$$

The function hH is 0 outside K, non-negative inside K and strictly positive on a ball around $\mathbf{u}_0$, thus the value of the integral is negative, in particular, $V'_{n-1}(0) < 0$. This means that if $t > 0$ is small enough, then the surface volume of $\mathbf{r}_t(K)$ will be strictly smaller than that of $\mathbf{r}(K)$, which contradicts the minimality condition on $\mathbf{r}$. $\square$

The corollary motivates the following definition.

Definition 3.2.41. A regular parameterized hypersurface is called a *minimal hypersurface* if its Minkowski curvature is constant 0. ※

Exercise 3.2.42. Show that a connected surface of revolution in $\mathbb{R}^3$ is a minimal surface if and only if it is part of a plane orthogonal to the axes of rotation or a that of a catenoid. (A catenoid is any surface similar to the catenoid considered in Exercise 3.2.17. (b).) ¤

A union of some disjoint closed curves in $\mathbb{R}^3$ can bound several minimal surfaces having different surface areas. Thus, among surfaces with a given boundary, not all minimal surfaces minimize the surface area.

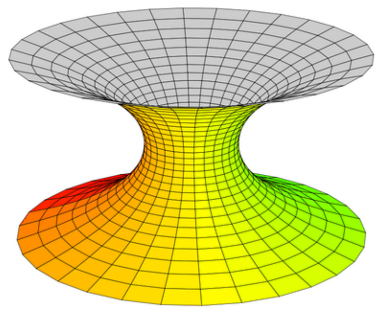

Figure 3.11: The catenoid.

Exercise 3.2.43. Assume that C_1 and C_2 are circles of radius r in parallel planes and that the straight line through their centers O_1 and O_2 is orthogonal to their planes. Show that if $\mathrm{d}(O_1, O_2)/r$ is small enough, then there exists two non-congruent catenoid rings bounded by $C_1 \cup C_2$. ¤

Chapter 4

Manifolds

4.1 Topological and Differentiable Manifolds

4.1.1 Basic Definitions

In the following chain of definitions n is a fixed *positive integer*.

Definition 4.1.1. Let X be an arbitrary set. A *local parameterization* of X is an injective mapping $\varphi\colon \Omega \to X$ from an open subset Ω of $\mathbb{R}^n$ onto a subset of X. The inverse $\varphi^{-1}\colon \varphi(\Omega) \to \Omega$ of such a parameterization is called a *chart* because through φ^{-1} the region $\operatorname{im}\varphi \subset X$ is "charted" on $U \subset \mathbb{R}^n$, just as a region of the earth is charted on a topographic or a political map. φ^{-1} is also called a *local coordinate system* because through φ^{-1} each point $p \in \operatorname{im}\varphi$ corresponds to an n-tuple of real numbers, the *coordinates* of p. ※

Definition 4.1.2. An *atlas* on X is a collection of charts $\mathcal{A} = \{\varphi_i \colon i \in I\}$ such that every point is represented in at least one chart i.e. $\bigcup_{i\in I} \operatorname{dom}\varphi_i = X$. ※

Definition 4.1.3. Two charts $\varphi_1\colon U_1 \to V_1$ and $\varphi_2\colon U_2 \to V_2$ are said to be $\mathcal{C}^r$*-compatible* if the domains $\varphi_1(U_1 \cap U_2)$ and $\varphi_2(U_1 \cap U_2)$ of the *transit mappings* $\varphi_2 \circ \varphi_1^{-1}$ and $\varphi_1 \circ \varphi_2^{-1}$ are open subsets of $\mathbb{R}^n$, and $\varphi_2 \circ \varphi_1^{-1}$ and $\varphi_1 \circ \varphi_2^{-1}$ are r times continuously differentiable. (A mapping is 0 times continuously differentiable if it is continuous). ※

Definition 4.1.4. An atlas is $\mathcal{C}^r$*-compatible* if any two charts in the atlas are $\mathcal{C}^r$-compatible. ※

A $\mathcal{C}^0$-compatible atlas $\mathcal{A}$ on a set X defines a topology on X as follows. Let $U \subset X$ be open if and only if $(U \cap \operatorname{dom}\varphi)$ is open in $\mathbb{R}^n$ with respect to any chart φ from $\mathcal{A}$.

Proposition 4.1.5. *The family of open sets yields a topology on X. The domains of the charts of $\mathcal{A}$ are open subsets of X. Each chart $\varphi \in \mathcal{A}$ is a homeomorphism between its domain and image with respect to these topologies.*

Proof. The first part of the statement follows directly from the following set theoretical identities.

(i) $\varphi(\emptyset \cap \operatorname{dom}\varphi) = \emptyset$,

(ii) $\varphi(X \cap \operatorname{dom}\varphi) = \operatorname{im}\varphi$,

(iii) $\varphi(U \cap V \cap \operatorname{dom}\varphi) = \varphi(U \cap \operatorname{dom}\varphi) \cap \varphi(V \cap \operatorname{dom}\varphi)$,

(iv) $\varphi((\bigcup_{i\in I} U_i) \cap \operatorname{dom}\varphi) = \bigcup_{i\in I} \varphi(U_i \cap \operatorname{dom}\varphi)$.

The rest of the statement is also a straightforward consequence of the definitions. □

Corollary 4.1.6. *The topology of a topological space X is induced by a $\mathcal{C}^0$-compatible atlas if and only if each point of X has a neighborhood homeomorphic to $\mathbb{R}^n$.*

Definition 4.1.7. An *n-dimensional topological manifold* is a pair $(X, \mathcal{A})$ consisting of a point set X and a $\mathcal{C}^0$-compatible atlas $\mathcal{A}$ on it, such that the topology induced by the atlas on X is Hausdorff and second countable. The

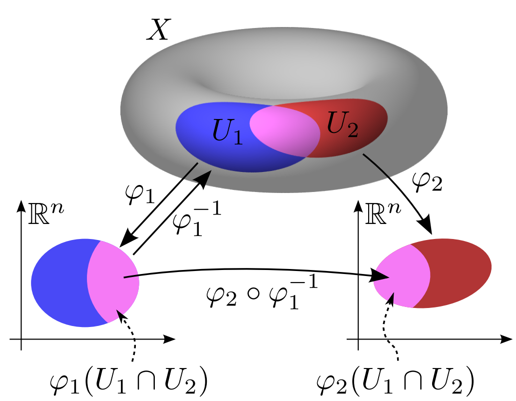

Figure 4.1: The transit map between two charts.

topological manifolds $(X, \mathcal{A}_1)$ and $(X, \mathcal{A}_2)$ are considered to be the same if $\mathcal{A}_1$ and $\mathcal{A}_2$ induce the same topology on X, i.e. if the combined atlas $\mathcal{A}_1 \cup \mathcal{A}_2$ is $\mathcal{C}^0$-compatible. ※

Among all atlases defining the same topological manifold structure on a set X, there is a unique maximal atlas.

Definition 4.1.8. A $\mathcal{C}^r$*-manifold* is a set X equipped with a $\mathcal{C}^r$-compatible atlas $\mathcal{A}$ turning X into a topological manifold. Two atlases define the same $\mathcal{C}^r$-manifold structure on X if their union is also $\mathcal{C}^r$-compatible. ※

It is clear that the class of atlases defining the same $\mathcal{C}^r$-manifold structure on X contains a unique *maximal atlas*.

Definition 4.1.9. A mapping $f : X \to Y$ from a $\mathcal{C}^r$-manifold $(X, \mathcal{A})$ into the $\mathcal{C}^r$-manifold $(Y, \mathcal{B})$ is said to be r *times continuously differentiable* or *of class* $\mathcal{C}^r$ if for any two charts $\varphi \in \mathcal{A}$ and $\psi \in \mathcal{B}$, the mapping $\psi \circ f \circ \varphi^{-1}$ is r times continuously differentiable. The map f is a $\mathcal{C}^r$*-diffeomorphism* if it is a bijection and both f and f^{-1} are of class $\mathcal{C}^r$. ※

Definition 4.1.10. Two $\mathcal{C}^r$-manifolds are $\mathcal{C}^r$*-diffeomorphic* if there is a $\mathcal{C}^r$-diffeomorphism between them. ※

As before, the adjectives *smooth* and *differentiable* will be used in the sense *infinitely many times differentiable* or $\mathcal{C}^\infty$. Unless otherwise stated, manifolds and mappings between them are assumed to be smooth.

Examples.

$\mathbb{R}^n$ equipped with the atlas consisting of only one chart, the identity mapping of $\mathbb{R}^n$, is an n-dimensional differentiable manifold.

Open subsets $U \subset X$ of an n-dimensional $\mathcal{C}^r$-manifold $(X, \mathcal{A})$ become n-dimensional $\mathcal{C}^r$-manifolds with the atlas $\{\varphi\big|_{\mathrm{dom}(\varphi)\cap U} : \varphi \in \mathcal{A}\}$.

If $(X_1, \mathcal{A}_1)$ and $(X_2, \mathcal{A}_2)$ are two $\mathcal{C}^r$-manifolds of dimension n and m respectively, then the product space $X_1 \times X_2$ has a natural $(n+m)$-dimensional $\mathcal{C}^r$-manifold structure given by the atlas

$$\{(\varphi_1 \times \varphi_2) : \mathrm{dom}(\varphi_1) \times \mathrm{dom}(\varphi_2) \to \mathbb{R}^{n+m} : \varphi_1 \in \mathcal{A}_1, \varphi_2 \in \mathcal{A}_2\}.$$

We have introduced the topology on the Grassmann manifolds $Gr(n, k)$ in Section 1.4. The topologies of these spaces come from a $k(n-k)$-dimensional differentiable manifold structure. We construct a chart $\varphi_{\mathcal{B}}$ on $Gr(n, k)$ to every ordered basis $\mathcal{B} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ of $\mathbb{R}^n$. Let us denote by V the subspace spanned by the first k vectors of $\mathcal{B}$ and by W the subspace spanned by the last $(n-k)$ vectors. It is clear that $\mathbb{R}^n = V \oplus W$. Denote by $\pi\colon \mathbb{R}^n \to V$ the projection of $\mathbb{R}^n$ onto V along W. The chart $\varphi_{\mathcal{B}}$ will be defined on the set

$$\mathrm{dom}(\varphi_{\mathcal{B}}) = \{L \in Gr(n,k) : L \cap W = \{\mathbf{0}\}\}.$$

$\varphi_{\mathcal{B}}$ assigns to $L \in \mathrm{dom}(\varphi_{\mathcal{B}})$ a $k \times (n-k)$ matrix in the following way. The restriction of π onto L is a linear isomorphism between L and V. The preimages of the vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ yield a basis of L which has the form

$$(\pi\big|_L)^{-1}(\mathbf{x}_i) = \mathbf{x}_i + \sum_{j=k+1}^{n} \alpha_{ij}\mathbf{x}_j.$$

It is clear that setting $\varphi_{\mathcal{B}}(L)$ equal to the matrix of coefficients (α_{ij}), $i = 1, \dots, k$; $j = k+1, \dots, n$, we obtain a bijection between $\mathrm{dom}(\varphi_{\mathcal{B}})$ and the set of all $k \times (n-k)$ matrices. The family of all charts of the form $\varphi_{\mathcal{B}}$ is a $\mathcal{C}^\infty$-compatible atlas on $Gr(n,k)$.
$Gr(n,k)$ is a compact manifold, it can be covered by a finite number of charts. Indeed we get a finite atlas on $Gr(n,k)$ if we let $\mathcal{B}$ run through different permutations of the standard basis of $\mathbb{R}^n$. The Grassmann manifold $Gr(n+1,1)$ is the *n-dimensional projective space*. The geometrical way to introduce projective spaces is the following. We take an n-dimensional Euclidean space and add to it a collection of extra points, called ideal points or points at infinity, in such a way, that we attach one point at infinity to each straight line and two straight lines get the same point at infinity if and only if they are parallel. If we put the n-dimensional space into the $(n+1)$-dimensional one and fix a point O outside it, then every straight line through O intersects the projective closure of the n-dimensional Euclidean space in a unique ordinary or ideal point and this is the natural correspondence between the two ways of introducing projective spaces.

Definition 4.1.11. A *Lie group* is a differentiable manifold G with a group operation '$\cdot$' such that the mapping

$$G \times G \to G, \quad (x,y) \mapsto x \cdot y^{-1}$$

is differentiable. ※

Example. $\mathrm{GL}(n,\mathbb{R})$ and $\mathrm{GL}(n,\mathbb{C})$ are open subsets in the linear spaces of all $n \times n$ real or complex matrices, hence they have a differentiable manifold structure. They also have a group structure, which is smooth since the entries of the quotient of two matrices are rational functions of the entries of the original matrices and rational functions are smooth. This way, general linear groups are Lie groups.

4.1.2 Configuration Spaces

Many examples of manifolds can be obtained as the configuration space of a mechanical system.
As the motion of a particle in $\mathbb{R}^3$ corresponds to a parameterized space curve, a motion of a system of n points can be described by n parameterized curves $\mathbf{x}_i : [a,b] \to \mathbb{R}^3$, $i = 1, 2, \ldots, n$.
Putting these mappings together, we obtain a curve

$$(\mathbf{x}_1, \ldots, \mathbf{x}_n) : [a,b] \to \mathbb{R}^3 \times \mathbb{R}^3 \times \cdots \times \mathbb{R}^3 \quad (n \text{ times})$$

in the direct product of n copies of $\mathbb{R}^3$, the projections of which on the i^{th} factor of the product is just the curve $\mathbf{x}_i$.
For there is a one-to-one correspondence between points of the product $\mathbb{R}^3 \times \mathbb{R}^3 \times \cdots \times \mathbb{R}^3$ (n times) and the possible configurations of n points in the space, we can call the direct product of n copies of $\mathbb{R}^3$ the *configuration space* of the system of n points.
In general, the *configuration space of a mechanical system* is the set of all of its possible positions, equipped with some natural additional structures such as topology or the structure of a differentiable manifold.
The advantage of introducing the configuration space is that the motion of the system can be interpreted as one single curve in the configuration space instead of a set of space curves.
Non-trivial examples can be obtained by putting some constraints on a system of n points. For example, some pairs of points can be connected by a rigid rod, some points can be fixed or forced to move along a line or a surface. Further constraints can be obtained by specifying the type of joint at the points where two or more rods meet.
The configuration space of a system of n points with constraints is a subspace of $\mathbb{R}^{3n}$ and it is quite natural to furnish it with the subspace topology inherited from $\mathbb{R}^{3n}$.

Remark. In physics, the dimension of the configuration space of a mechanical system (provided that it is a manifold) is called the *number of degrees of freedom.*

Examples.

- The configuration space of the planar pendulum is the circle S^1.
- The configuration space of the spherical pendulum is the two-dimensional sphere S^2.
- The configuration space of a planar double pendulum is the direct product of two circles, i.e. the torus $T^2 = S^1 \times S^1$.

- The configuration space of a spherical double pendulum is the direct product of two spheres $S^2 \times S^2$.
- A rigid segment in the plane has for its configuration space the direct product $\mathbb{R}^2 \times S^1$, which is homeomorphic to the open solid torus.

In the above examples, the configuration space carries a manifold topology. Theorem 4.1.12 below explains why the configuration space of a system of n points with constraints so often happens to be a manifold.

4.1.3 Submanifolds of $\mathbb{R}^n$

Theorem 4.1.12. *Let $F\colon \mathbb{R}^n \to \mathbb{R}^k$ be a smooth mapping, the image of which contains $\mathbf{0} \in \mathbb{R}^k$. Consider the preimage $X = \{\mathbf{x} \in \mathbb{R}^n : F(\mathbf{x}) = 0\}$ of the point $\mathbf{0}$.*
Let us suppose that the gradient vectors

$$\operatorname{grad} f_i(\mathbf{x}) = (\partial_1 f_i(\mathbf{x}), \partial_2 f_i(\mathbf{x}), \dots, \partial_n f_i(\mathbf{x}))$$

of the coordinate functions of $F = (f_1, f_2, \dots, f_n)$ are linearly independent at each point $\mathbf{x}$ of X.
Then X is an $(n-k)$-dimensional topological manifold, furthermore, there is a well-defined differentiable manifold structure on X.

Remark. The condition on the independence of the gradient vectors of the coordinate functions is essential. By a theorem of Whitney, for any closed set $C \subset \mathbb{R}^n$ there exists a smooth function f on $\mathbb{R}^n$ such that $C = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = 0\}$.

Proof. The linear space $\mathbb{R}^{n-k}$ can be embedded into $\mathbb{R}^n$ by the mapping

$$\iota\colon (x_1, x_2, \dots, x_{n-k}) \mapsto (\underbrace{0, \dots, 0}_{k \text{ zeros}}, x_1, x_2, \dots, x_{n-k}).$$

Consider the set $\tilde{\mathcal{A}}$ of those diffeomorphisms $\Phi\colon V \to U$ between open subsets of $\mathbb{R}^n$ through which the set $X \cap V$ is mapped onto the intersection $\iota(\mathbb{R}^{n-k}) \cap U$.
Put

$$\mathcal{A} = \{\varphi : \varphi \text{ has the form } \varphi = \iota^{-1} \circ \Phi\big|_{X \cap V} : X \cap V \to \mathbb{R}^{n-k}, \text{ where } \Phi \in \tilde{\mathcal{A}}, \operatorname{dom}\Phi = V\}.$$

Obviously, elements of $\mathcal{A}$ define a homeomorphism between open subsets of X and that of $\mathbb{R}^{n-k}$. It is also clear from the construction that the mappings

$\varphi \circ \psi^{-1}$, defined on $\psi(\operatorname{dom}\varphi \cap \operatorname{dom}\psi)$, are smooth for any $\varphi, \psi \in \mathcal{A}$. In such a way, to prove that $\mathcal{A}$ is a $\mathcal{C}^\infty$-compatible atlas on X one has only to check that each point of X is represented in at least one chart belonging to $\mathcal{A}$.
For this purpose, consider an arbitrary point $\mathbf{x}$ of X and the gradient vectors of the coordinate functions of F at $\mathbf{x}$. Since they are linearly independent, we can obtain a basis by joining further $n-k$ vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n-k}$ to them. The gradient vector of the linear function $g_i(\mathbf{x}) = \langle \mathbf{a}_i, \mathbf{x}\rangle$ is $\mathbf{a}_i$ at any point, consequently, the inverse function theorem (Theorem 1.5.12) can be applied to the mapping

$$\Phi = (f_1, \ldots, f_k, g_1, ..., g_{n-k})\colon \mathbb{R}^n \to \mathbb{R}^n$$

at the point $\mathbf{x}$. According to the theorem, Φ is a diffeomorphism between a neighborhood V of $\mathbf{x}$ and an open neighborhood of $\Phi(\mathbf{x})$. Denote by $\tilde{\Phi}$ the restriction of Φ onto V. The mapping $\tilde{\Phi}$ belongs to $\tilde{\mathcal{A}}$ and the chart $\iota^{-1} \circ (\tilde{\Phi}|_{V \cap M})$ is defined in a neighborhood of $\mathbf{x}$, so the proof is finished. □

Exercise 4.1.13. Check that the topology of $(X, \mathcal{A})$ coincides with the subspace topology inherited from $\mathbb{R}^n$, consequently, it is Hausdorff and second countable. ¤

Definition 4.1.14. We say that $X \subset \mathbb{R}^n$ is an *embedded* $(n-k)$*-dimensional submanifold in* $\mathbb{R}^n$ if in a neighborhood U of every point $\mathbf{x} \in X$ there are functions $f_1, f_2, \ldots, f_k\colon U \to \mathbb{R}$ such that the intersection $U \cap X$ is given by the equations $f_1 = f_2 = \cdots = f_k = 0$ and the vectors $\operatorname{grad} f_1(\mathbf{x}), \ldots, \operatorname{grad} f_k(\mathbf{x})$ are linearly independent. ※

The study of higher dimensional manifolds was launched at the beginning of the 20-th century by H. Poincaré (1854 - 1912). At that time topology was in its cradle and the abstract definition of a topological space had not been created. Poincaré worked with embedded submanifolds in $\mathbb{R}^n$. This was not a real loss of generality since by *Whitney's theorem* every differentiable manifold of dimension n can be embedded into $\mathbb{R}^{2n}$ as a submanifold.

Examples. As an application, consider the set S^{n-1} of points in $\mathbb{R}^n$ the distance of which from the origin is equal to one. They are characterized by the equality

$$f(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2 - 1 = 0.$$

The gradient vector of f at the point $\mathbf{x} \in \mathbb{R}^n$ is just $2\mathbf{x}$. It is zero only at the origin, which does not belong to S^{n-1}, so S^{n-1} is a topological manifold with a natural differentiable structure on it. S^{n-1} is called *the* $(n-1)$*-dimensional sphere with the standard differentiable structure.*

Most important Lie groups are obtained as closed subgroups of $\mathrm{GL}(n,\mathbb{R})$, defined by some equalities on the matrix entries. For example, $\mathrm{GL}(n,\mathbb{C})$ is isomorphic to the subgroup of $\mathrm{GL}(2n,\mathbb{R})$ consisting of matrices of the form $\begin{pmatrix} A & B \\ -B & A \end{pmatrix}$, where A, B are $n\times n$ matrices. The following example illustrates how we can prove that a closed subgroup of $\mathrm{GL}(n,\mathbb{R})$ is a Lie group. (Remark that according to Cartan's theorem, every closed subgroup of a Lie group is a Lie group, therefore the computation below is an elementary verification of a special case of Cartan's theorem.)

Consider the orthogonal group

$$O(n) = \{A \in \mathrm{GL}(n,\mathbb{R}) : AA^T = I\}.$$

We claim that it is an $\frac{n(n-1)}{2}$-dimensional Lie group. If $A = (a_{ij})$, then the equation $AA^T = I$ is equivalent to the system of equations

$$F_{ij}(A) = \sum_{k=1}^{n} a_{ik}a_{jk} - \delta_{ij} = 0, \quad (i,j = 1,\dots,n).$$

Since $F_{ij} = F_{ji}$, these equations are not independent. We get however an independent system of equations if we restrict ourselves to the equations with $1 \le i \le j \le n$. The number of these equations is $\frac{n(n+1)}{2}$, so if we show the independence of the gradients, then we obtain the required expression for the dimension of $O(n)$.

The gradient of F_{ij} at $A \in O(n)$ is an $n \times n$ matrix with entries

$$\frac{\partial F_{ij}}{\partial a_{rs}} = \sum_{k=1}^{n} \left(\frac{\partial a_{ik}}{\partial a_{rs}} a_{jk} + a_{ik}\frac{\partial a_{jk}}{\partial a_{rs}} \right) = a_{js}\delta_{ir} + a_{is}\delta_{jr}.$$

We show that these vectors, are orthogonal with respect to the usual scalar product on $\mathbb{R}^{n^2} \cong \mathrm{Mat}(n,\mathbb{R})$. Indeed, taking the scalar product of the gradient vectors of F_{ij} and F_{kl} at A we obtain

$$\begin{aligned}
&\sum_{r,s=1}^{n} (a_{js}\delta_{ir} + a_{is}\delta_{jr})(a_{ls}\delta_{kr} + a_{ks}\delta_{lr}) \\
&\qquad = \sum_{r,s=1}^{n} (a_{js}a_{ls}\delta_{ir}\delta_{kr} + a_{js}a_{ks}\delta_{ir}\delta_{lr} + a_{is}a_{ls}\delta_{jr}\delta_{kr} + a_{is}a_{ks}\delta_{jr}\delta_{lr}) \\
&\qquad = \sum_{r=1}^{n} (\delta_{jl}\delta_{ir}\delta_{kr} + \delta_{jk}\delta_{ir}\delta_{lr} + \delta_{il}\delta_{jr}\delta_{kr} + \delta_{ik}\delta_{jr}\delta_{lr}) \\
&\qquad = \delta_{jl}\delta_{ik} + \delta_{jk}\delta_{il} + \delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl} = 2\delta_{ik}\delta_{jl}(1 + \delta_{ij}\delta_{kl}).
\end{aligned}$$

(In the last step, we used that as $i \leq j$ and $k \leq l$, $\delta_{il}\delta_{jk} \neq 0$ only if $i = j = k = l$.)
We know that the determinant of an orthogonal matrix is ± 1. Thus $O(n)$ has two components. The connected component on which determinant is 1, is the *special orthogonal group* $SO(n)$, and it has also dimension $\frac{n(n-1)}{2}$. $SO(n)$ is the configuration space of a rigid n-dimensional body with one fixed point. The configuration space of an n-dimensional body without a fix point is the space $SO(n) \times \mathbb{R}^n$.
For $n = 2$, $SO(2)$ is the 1-dimensional group of rotations of the plane. For $n = 3$, $SO(3)$ is a 3-dimensional manifold. With the help of quaternions we can show that this group is homeomorphic to the 3-dimensional projective space.
Let $\mathbb{H}$ denote the 4-dimensional space of quaternions $x + yi + zj + wk$, and let us identify $\mathbb{R}^3$ with the space of pure imaginary quaternions. For $0 \neq q \in \mathbb{H}$ let us denote by ρ_q the transformation $\rho_q(h) = q^{-1}hq$. Since $|q^{-1}hq| = |h|$, ρ_q is an orthogonal transformation. If h is a real number, then $\rho_q(h) = h$, thus ρ_q maps $\mathbb{R} \subset \mathbb{H}$ into itself. Consequently, it maps $\mathbb{R}^3$, the orthogonal complement of $\mathbb{R}$ also into itself. The assignment $q \mapsto \rho_q\big|_{\mathbb{R}^3}$ is a group homomorphism from the multiplicative group $\mathbb{H}\setminus\{\mathbf{0}\}$ to $SO(3)$. The kernel of this homomorphism is the center of $\mathbb{H}\setminus\{\mathbf{0}\}$, i.e. $\mathbb{R}\setminus\{\mathbf{0}\}$. This homomorphism is also surjective as it follows from the following two exercises.

Exercise 4.1.15. Show that every element of $SO(3)$ is a rotation about a straight line.

¤

Exercise 4.1.16. Show that if $a \in \mathbb{R}^3$ is a pure imaginary quaternion of length 1, $\alpha \in \mathbb{R}$ and $q = \cos\alpha + a\sin\alpha$, then $\rho_q\big|_{\mathbb{R}^3}$ is a rotation about a with angle 2α. ¤

We conclude that $SO(3)$ is isomorphic to the factor group $\mathbb{H} \setminus \{\mathbf{0}\}/\mathbb{R} \setminus \{\mathbf{0}\}$, but cosets of $\mathbb{R}\setminus\{\mathbf{0}\}$ in $\mathbb{H}\setminus\{\mathbf{0}\}$ are in one to one correspondence with straight lines through $\mathbf{0}$ in $\mathbb{H}$, i.e. there is a natural bijection between $SO(3)$ and the projective space $Gr(4,1)$.

4.1.4 Remarks on the Classification of Manifolds

The classification problem of n-dimensional manifolds can be formulated in different levels. For each r we may consider the category of $\mathcal{C}^r$-manifolds and r-times differentiable mappings. The following two theorems show that the classification problem of $\mathcal{C}^r$-manifolds is the same problem for all $1 \leq r \leq \infty$.

Theorem 4.1.17. *For $r \geq 1$, every maximal $\mathcal{C}^r$-compatible atlas contains a $\mathcal{C}^\infty$-compatible atlas.*

Theorem 4.1.18. *If two $\mathcal{C}^\infty$-manifolds are $\mathcal{C}^1$-diffeomorphic, then they are $\mathcal{C}^\infty$-diffeomorphic.*

None of the above theorems can be extended to $\mathcal{C}^0$-manifolds. There exist topological manifolds which have no $\mathcal{C}^1$-compatible atlas, and there exist homeomorphic but not diffeomorphic differentiable manifolds.

In 1956 J.W. Milnor constructed a differentiable manifold which is homeomorphic to the standard 7-dimensional sphere but not diffeomorphic to it. Such manifolds were given the name *exotic spheres.* Later on an even more surprising result was published by Milnor and Kervaire. There are exactly 28 mutually non-diffeomorphic differentiable structures on a topological 7-sphere. Since then many examples of topological manifolds having many different differentiable structures have been obtained. One of the most interesting constructions is due to Donaldson, who invented exotic differentiable structures on $\mathbb{R}^4$.

We do not meet exotic differentiable structures in dimension two. The classification of compact surfaces is the same up to homeomorphism and diffeomorphism. The classification theorem of compact 2-dimensional manifolds gives a list of non-diffeomorphic compact surfaces and asserts that every compact surface is diffeomorphic to one of the surfaces in the list. The list of compact surfaces consists of two lists in fact. The first list contains the orientable compact surfaces, the second contains the non-orientable ones.

Orientable compact surfaces. The simplest orientable closed surface is the sphere $S^2 = \{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| = 1\}$. The next example is the torus $T^2 = S^1 \times S^1$. Cutting a small disc out of the torus, we get a surface with boundary, called a handle. A typical orientable compact surface is a sphere with g handles. We can obtain this surface if we cut g holes on the surface of the sphere and glue a handle to each of them.

Non-orientable compact surfaces. The first member of this list is the real projective plane. To understand the topological structure of the projective plane, let us cut the projective plane into two parts by a hyperbola. The interior of the hyperbola has two components in the Euclidean plane, but these components are glued together along a segment of the line at infinity, so the interior is a topological disc. The exterior of the hyperbola is a long infinite band in the Euclidean plane, the "ends" of which are glued together along another segment of the line at infinity. One can see that the ends of the band are glued together by a half twist so what we get is a Möbius band. We conclude that the projective plane is the union of a disc and a Möbius band glued together along their boundaries. A typical non-orientable compact surface is a sphere with g Möbius bands. We obtain this surface cutting g discs out of the sphere and gluing to the boundary of each hole a Möbius band. Non-orientable compact surfaces can not be embedded into the 3-dimensional Euclidean space, so although one can easily construct a 3-dimensional model

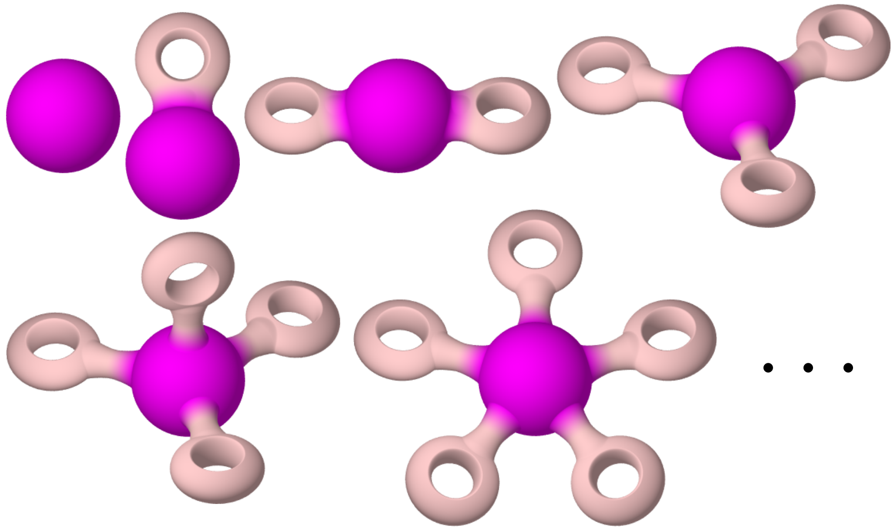

Figure 4.2: Orientable closed surfaces.

of a Möbius band and that of a sphere with g holes, it is impossible to glue the Möbius bands to the sphere in practice. If however we could try this in a 4-dimensional space there would be no difficulty.

Exercise 4.1.19. The configuration space of the pentagon (closed chain of five rods in the plane) with one edge fixed is a compact surface (sometimes with singularities). What kind of surfaces can we obtain? ¤

Exercise 4.1.20. Give an example of a set X with a $\mathcal{C}^0$-compatible atlas $\mathcal{A}$ on it such that the topology induced on X by $\mathcal{A}$ is

(a) second countable but not Hausdorff;

(b) Hausdorff but not second countable. ¤

Exercise 4.1.21. Show that the special unitary group

$$SU(n) = \{A \in \mathrm{GL}(n, \mathbb{C}) : AA^* = I, \det A = 1\}$$

is a Lie group and determine its dimension. Prove that $SU(2)$ is diffeomorphic to the 3-dimensional sphere S^3. ¤

Exercise 4.1.22. Which surface shall we get from the classification list if we glue to the sphere $k \geq 1$ Möbius bands and l handles? ¤

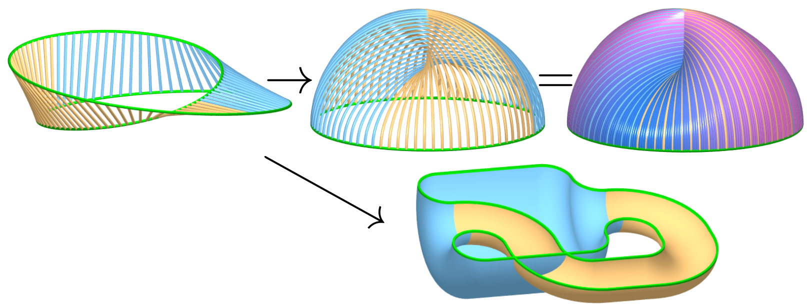

Figure 4.3: Self-intersecting images of the Möbius band with planar boundary: the cross cap and half of the Klein bottle.

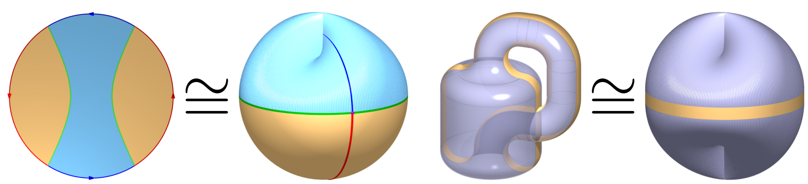

Figure 4.4: A sphere with one and two Möbius bands: the projective plane and the Klein bottle.

Exercise 4.1.23. Let P be a complex polynomial of degree k having k different roots. Consider the subset of $\mathbb{C}^2$ defined by

$$M = \{(z, w) \in \mathbb{C}^2 : z^l = P(w)\}.$$

Show that for some natural numbers g and N, M is diffeomorphic to a sphere with g handles from which N points are omitted. Express g and N in terms of k and l. ¤

4.2 The Tangent Bundle

The aim of this section is to give and reconcile different commonly used definitions of a tangent vector to a manifold. Before passing over to the abstract situation, we shall deal with submanifolds of $\mathbb{R}^n$.

Definition 4.2.1. Smooth mappings $\gamma\colon [a, b] \to M$ of an interval into a differentiable manifold $(M, \mathcal{A})$ are called *smooth parameterized curves* in the manifold. ※

Definition A. Let M be a differentiable manifold embedded in $\mathbb{R}^n$, $\mathbf{x}_0 \in M$. A vector $\mathbf{v}$ is called a *tangent vector to* M *at* $\mathbf{x}_0$ if there is a smooth curve $\gamma\colon [-\varepsilon, \varepsilon] \to M$ passing through $\mathbf{x}_0 = \gamma(0)$ with speed vector $\mathbf{v} = \gamma'(0)$.
The *(linear) tangent space* $T_{\mathbf{x}_0}M$ *of* M *at* $\mathbf{x}_0 \in M$ is the set of all tangent vectors to M at $\mathbf{x}_0$.
The *affine tangent space of* M *at* $\mathbf{x}_0$ is the translate of $T_{\mathbf{x}_0}M$ with the vector $\mathbf{x}_0$. ※

Theorem 4.2.2. *Let us suppose that a k-dimensional manifold M embedded in $\mathbb{R}^n$ is given in a neighborhood U of $\mathbf{x}_0 \in M$ by a system of equalities $f_1 = \cdots = f_{n-k} = 0$, where $f_1, \dots, f_{n-k}$ are smooth functions on U such that the vectors* $\operatorname{grad} f_1(\mathbf{x}_0), \dots, \operatorname{grad} f_{n-k}(\mathbf{x}_0)$ *are linearly independent. Then the tangent space of M at $\mathbf{x}_0$ consists of the vectors orthogonal to* $\operatorname{grad} f_1(\mathbf{x}_0), \dots, \operatorname{grad} f_{n-k}(\mathbf{x}_0)$.

Corollary 4.2.3. *The tangent space of a k-dimensional submanifold M of $\mathbb{R}^n$ is a k-dimensional linear subspace of $\mathbb{R}^n$. The affine tangent space of M at a point is an affine subspace of $\mathbb{R}^n$.*

Proof. If $\mathbf{x}\colon [-\varepsilon, \varepsilon] \to M$ is a smooth curve on M passing through $\mathbf{x}_0 = \gamma(0)$, and having coordinate functions $\gamma = (x_1, \dots, x_n)$, then we have

$$f_i(x_1(t), \dots, x_n(t)) = 0 \qquad i = 1, \dots, n-k$$

for each $t \in [-\varepsilon, \varepsilon]$. Differentiating with respect to t we get

$$\partial_1 f_i(\gamma(0))x_1'(0) + \cdots + \partial_n f_i(\gamma(0))x_n'(0) = 0,$$

which means that the vectors $\operatorname{grad} f_i(\mathbf{x}_0)$ and $\gamma'(0)$ are orthogonal.
Now we prove that if a vector $\mathbf{v}$ is orthogonal to the vectors $\operatorname{grad} f_i(\mathbf{x}_0)$, $1 \le i \le n-k$, then $\mathbf{v}$ is a tangent vector. Let us take a smooth local parameterization $F\colon \mathbb{R}^k \to M \subset \mathbb{R}^n$ of M around the point $\mathbf{x}_0$. The curve $\gamma\colon t \mapsto F(F^{-1}(\mathbf{x}_0) + t\mathbf{y})$, where $\mathbf{y} \in \mathbb{R}^k$ is fixed, is a curve on M passing through $\mathbf{x}_0 = \gamma(0)$. The speed vector of this curve at $t = 0$ is

$$\partial_1 F(F^{-1}(\mathbf{x}_0))y_1 + \cdots + \partial_k F(F^{-1}(\mathbf{x}_0))y_k,$$

where $y_1, \dots, y_k$ are the coordinates of $\mathbf{y}$. By the construction of local parameterizations of embedded manifolds, presented in the proof of Theorem 4.1.12, F is a restriction onto $\mathbb{R}^k$ of a diffeomorphism between open subsets of $\mathbb{R}^n$, consequently, the vectors $\partial_i F(F^{-1}(\mathbf{x}_0))$ are linearly independent. We conclude, that the tangent space is contained in the k-dimensional linear subspace orthogonal to $\operatorname{grad} f_1(\mathbf{x}_0), \dots, \operatorname{grad} f_{n-k}(\mathbf{x}_0)$ and contains the k-dimensional linear subspace spanned by the vectors $\partial_i F(F^{-1}(\mathbf{x}_0))$ $1 \le i \le k$, which means that both linear subspaces coincide with the tangent space. $\square$

The definition of tangent vectors can also be given in intrinsic terms, independent of the embedding of M into $\mathbb{R}^n$. Let us define an *equivalence relation* on the set

$$\Gamma(M,p) = \{\gamma\colon [-\varepsilon,\varepsilon] \to M \mid \gamma(0) = p,\ \gamma \text{ is smooth}\},$$

consisting of curves passing through $p \in M$, by calling two curves $\gamma_1, \gamma_2 \in \Gamma(M,p)$ equivalent if $(\phi\circ\gamma_1)'(0) = (\phi\circ\gamma_2)'(0)$ for some chart ϕ around p. Then this condition is true for any chart from the atlas of M. (Prove this!)

Definition B. A *tangent vector* to a manifold M at the point $p \in M$ is an equivalence class of curves in $\Gamma(M,p)$. The set of equivalence classes is called the *tangent space of M at p* and is denoted by T_pM. ※

In this setting, the *speed vector* $\gamma'(s) \in T_{\gamma(s)}M$ of a smooth curve is the equivalence class of the curve $t \mapsto \gamma(s+t)$ in $\Gamma(M,s)$.
Given a chart ϕ around p, we can establish a one-to-one correspondence between the equivalence classes and points of $\mathbb{R}^m$, ($m = \dim M$), assigning to the equivalence class of a curve $\gamma \in \Gamma(M,p)$ the vector $(\phi\circ\gamma)'(0) \in \mathbb{R}^m$. With the help of this identification, we can introduce a vector space structure on the tangent space, not depending on the choice of the chart.
For embedded manifolds, Definition B agrees with Definition A. The advantage of Definition B is that it is applicable also for abstract manifolds, not embedded anywhere.

Definition 4.2.4. If $\phi = (x^1,\dots,x^m)$ is a chart on the manifold M around the point p, $\gamma \in \Gamma(M,p)$, then the coordinates $(x^1\circ\gamma)'(0),\dots,(x^m\circ\gamma)'(0)$ of $(\phi\circ\gamma)'(0)$ are called the *components of the speed vector $\gamma'(0)$ with respect to the chart ϕ.* ※

The main difficulty of defining tangent vectors to a manifold is due to the fact that an abstract manifold might not be naturally embedded into a fixed finite dimensional linear space. Nevertheless, there is a universal embedding of each differentiable manifold into an infinite dimensional linear space.
Let us denote by $\mathcal{F}(M)$ the linear vector space of smooth functions on M, and by $\mathcal{F}^*(M)$ the dual space of $\mathcal{F}(M)$ that is the space of linear functions on $\mathcal{F}(M)$, and consider the embedding ι of M into $\mathcal{F}^*(M)$ defined by the formula

$$[\iota(p)](f) = f(p), \qquad \text{where } p \in M,\ f \in \mathcal{F}(M).$$

Having embedded the manifold M into $\mathcal{F}^*(M)$, we can define tangent vectors to M as elements of the linear space $\mathcal{F}^*(M)$.

Definition 4.2.5. Let M be a differentiable manifold, $p \in M$. We say that a linear function $D \in \mathcal{F}^*(M)$ defined on $\mathcal{F}(M)$ is a *derivation of $\mathcal{F}(M)$ at p* if the equality

$$D(fg) = D(f)g(p) + f(p)D(g)$$

holds for every $f, g \in \mathcal{F}(M)$. ※

Derivations at a point $p \in M$ form a linear subspace $\operatorname{Der}_p M$ of $\mathcal{F}^*(M)$.
Each curve $\gamma \in \Gamma(M, p)$ defines a derivation at the point p by the formula $D_{\gamma'(0)}(f) = (f \circ \gamma)'(0)$, where $f \in \mathcal{F}(M)$. $D_{\gamma'(0)}$ is the speed vector of the curve $\iota \circ \gamma$ in $\mathcal{F}^*(M)$. Since two curves define the same derivation if and only if they are equivalent, there is a one-to-one correspondence between the equivalence classes of curves and the derivations obtained as $D_{\gamma'(0)}$ for some γ.

Definition C. A *tangent vector* to a manifold M at the point $p \in M$ is a derivation of the form $D_{\gamma'(0)}$, where $\gamma \in \Gamma(M, p)$.
The *tangent space* T_pM *of* M *at the point* p is the set of derivations $D_{\gamma'(0)}$ along curves in M passing through $p = \gamma(0)$. ※

We are going to show that although Definition C defines T_pM just as a subset of $\operatorname{Der}_p M$, we have $T_pM = \operatorname{Der}_p M$ in fact. To prove this we need some preliminary steps.

Lemma 4.2.6. *If $f \in \mathcal{F}(M)$ is a constant function and D is a derivation at a point $p \in M$, then $D(f) = 0$.*

Proof. Because of linearity, it is enough to show that $D(\underline{1}) = 0$, where $\underline{1}$ is the constant 1 function on M. This comes from

$$D(\underline{1}) = D(\underline{1} \cdot \underline{1}) = D(\underline{1})\underline{1}(p) + \underline{1}(p)D(\underline{1}) = 2D(\underline{1}). \qquad \square$$

Lemma 4.2.7. *If two functions $f, g \in M$ coincide on a neighborhood U of $p \in M$ and D is a derivation at p then $D(f) = D(g)$.*

Proof. Using the bump function construction of Proposition 1.5.13, we can define a smooth function h on M which is zero outside U and such that $h(p) = 1$. In this case $h(f - g)$ is the constant 0 function on M. Thus we have

$$0 = D(\underline{0}) = D(h(f-g)) = D(h)(f(p) - g(p)) + h(p)D(f-g) = D(f) - D(g).$$
$\square$

Remarks.

(i) The bump function construction shows that the mapping $\iota \colon M \to \mathcal{F}^*(M)$ is an inclusion. Indeed, if $p \neq q$ are distinct points of M, then there is a smooth function h on M such that

$$[\iota(p)](h) = h(p) = 1 \neq [\iota(q)](h) = h(q) = 0.$$

(ii) We can extend a derivation $D \in Der_p$ to functions f defined only in a neighborhood U of p by taking a smooth function h on M such that h is zero outside U and constant 1 in a neighborhood of p and putting $D(f) := D(\tilde{f})$, where

$$\tilde{f}(x) = \begin{cases} f(x)h(x) & \text{if } x \in U, \\ 0 & \text{if } x \notin U. \end{cases}$$

By Lemma 4.2.7, this extension of D is correctly defined.

Lemma 4.2.8. *Let $f \colon B \to \mathbb{R}$ be a smooth function defined on an open ball $B \subset \mathbb{R}^n$ around the origin. Then there exist smooth functions g_i, $1 \leq i \leq n$ on B such that*

$$f(\mathbf{x}) = f(\mathbf{0}) + \sum_{i=1}^{n} x_i g_i(\mathbf{x}) \text{ for } \mathbf{x} = (x_1, \dots, x_n) \in B$$

and

$$g_i(\mathbf{0}) = \partial_i f(\mathbf{0}).$$

Proof. Since

$$f(\mathbf{x}) - f(\mathbf{0}) = \int_0^1 \frac{d\,f(t\mathbf{x})}{d\,t} dt = \int_0^1 \sum_{i=1}^{n} x_i \partial_i f(t\mathbf{x}) dt$$
$$= \sum_{i=1}^{n} x_i \int_0^1 \partial_i f(t\mathbf{x}) dt,$$

we may take $g_i(\mathbf{x}) = \int_0^1 \partial_i f(t\mathbf{x}) dt$. □

Now we are ready to prove the main theorem.

Theorem 4.2.9. *The tangent space to a differentiable manifold M at the point $p \in M$ coincides with the space of all derivations of $\mathcal{F}(M)$ at p, which is a linear space having the same dimension as M has.*

Proof. Let us take a differentiable manifold $(M, \mathcal{A})$ and a chart $\phi = (x^1, \dots, x^n) \in \mathcal{A}$ defined in a neighborhood of $p \in M$. Define the derivations $\partial_i^{\phi}(p)$ as follows

$$[\partial_i^{\phi}(p)](f) := \partial_i (f \circ \phi^{-1})(\phi(p)).$$

We prove that the derivations $\partial_i^{\phi}(p)$ form a basis in the space of derivations at p. They are linearly independent since if we have

$$\sum_{i=1}^{n} \alpha^i \partial_i^{\phi}(p) = 0,$$

then applying this derivation to the j-th coordinate function x^j we get

$$\sum_{i=1}^{n} \alpha^i \partial_i (x^j \circ \phi^{-1})(\phi(p)) = \alpha^j = 0.$$

On the other hand, if D is an arbitrary derivation at p, then we have $D = \sum_{i=1}^{n} D(x^i)\partial_i^{\phi}(p)$. Indeed, let $f \in \mathcal{F}(M)$ be an arbitrary smooth function on M and apply Lemma 4.2.8 to the function $\mathbf{x} \mapsto f \circ \phi^{-1}(\mathbf{x} + \phi(p))$ around $\mathbf{0}$. We obtain the existence of functions g_i defined around $\phi(p)$ such that

$$f = f(p) + \sum_{i=1}^{n} (x^i - x^i(p)) g_i \circ \phi \quad \text{and} \quad g_i(\phi(p)) = \partial_i (f \circ \phi^{-1})(\phi(p)).$$

In this case, however, we have

$$\begin{aligned} D(f) &= D(f(p)) + \sum_{i=1}^{n} D((x^i - x^i(p))) g_i(\phi(p)) + (x^i(p) - x^i(p)) D(g_i \circ \phi) \\ &= \sum_{i=1}^{n} D(x^i) \partial_i (f \circ \phi^{-1})(\phi(p)) = \sum_{i=1}^{n} D(x^i)[\partial_i^{\phi}(p)](f). \end{aligned}$$

To finish the proof, we only have to show that every derivation at the point p can be obtained as a speed vector of a curve passing through p. Define the curve $\gamma \colon [-\varepsilon, \varepsilon] \to M$ by the formula $\gamma(t) := \phi^{-1}(\phi(p) + (t\alpha^1, \dots, t\alpha^n))$. Then obviously the speed vector $\gamma'(0)$ is just $\sum_{i=1}^{n} \alpha^i \partial_i^{\phi}(p)$. □

Definition 4.2.10. The basis $(\partial_1^{\phi}(p), \dots, \partial_n^{\phi}(p))$ of T_pM defined in the proof of Theorem 4.2.9 is called the *basis induced by the chart* ϕ. ※

4.2.1 The Tangent Bundle

The disjoint union of the tangent spaces of M at the various points, $\bigcup_{p \in M} T_pM$, has a natural differentiable manifold structure, the dimension of which is twice the dimension of M. This manifold is called the *tangent bundle of* M and is denoted by TM. A point of this manifold is a vector D, tangent to M at some point p. Local coordinates on TM are constructed as follows. Let $\phi = (x^1, \dots, x^n)$ be a chart on M the domain U of which contains p, and $D(x^1), \dots, D(x^n)$ be the components of D in the basis $\partial_i^{\phi}(p)$. Then the mapping

$$D \mapsto (x^1(p), \dots, x^n(p), D(x^1), \dots, D(x^n))$$

give a local coordinate system on $\bigcup_{p \in U} T_pM \subset TM$. The set of all local coordinate systems constructed this way forms a $\mathcal{C}^\infty$-compatible atlas on TM, that turns TM into a differentiable manifold.

Exercise 4.2.11. Check the last statement. ¤

The mapping $\pi\colon TM \to M$ which takes a tangent vector $D \in T_pM$ to the point $p \in M$ at which the vector is tangent to M is called the *natural projection of the tangent bundle*. The inverse image of a point $p \in M$ under the natural projection is the tangent space T_pM. This space is called the *fiber of the tangent bundle over the point p*.

4.2.2 The Derivative of a Smooth Map

Definition 4.2.12. Let $f\colon M \to N$ be a smooth mapping between the differentiable manifolds $(M, \mathcal{A})$, $(N, \mathcal{B})$, and let $p \in M$. The *derivative* of f at the point p is the linear map $T_pf\colon T_pM \to T_{f(p)}N$ given in the following way. Let $D \in T_pM$ and consider a curve $\gamma\colon [-\varepsilon, \varepsilon] \to M$ with $\gamma(0) = p$ and speed vector $\gamma'(0) = D$. Then $T_pf(D)$ is the speed vector $(f \circ \gamma)'(0)$. ※

The tangent vector $T_pf(D)$ as a derivation at $f(p)$ assigns to a smooth function $h \in \mathcal{F}(N)$ the number

$$(h \circ (f \circ \gamma))'(0) = ((h \circ f) \circ \gamma)'(0) = D(h \circ f).$$

Proposition 4.2.13. *The derivative T_pf is correctly defined (does not depend on the choice of γ) and is linear.*

Proof. We derive a formula for T_pf using local coordinates which will show both parts of the proposition.
Let $\phi = (x^1, \dots, x^m)$ and $\psi = (y^1, \dots, y^n)$ be local coordinates in a neighborhood of $p \in M$ and $f(p) \in N$ respectively.
If the components of D in the basis $\partial_i^{\phi}(p)$ corresponding to the chart ϕ are $\{\alpha^i : 1 \le i \le m\}$ then we have $(x^i \circ \gamma)'(0) = \alpha^i$. Observe, that α^i depends only on D and the fixed chart ϕ, but not on the curve γ. The components $\{\beta^j : 1 \le j \le n\}$ of $T_pf(D)$ in the basis $\partial_j^{\psi}(f(p))$ induced by the chart ψ can be computed by the formula $\beta^j = (y^j \circ f \circ \gamma)'(0)$. Denote by $\tilde{f}^j$ the j-th coordinate function of the mapping $\tilde{f} = \psi \circ f \circ \phi^{-1}$, i.e.

$$\tilde{f}^j = y^j \circ f \circ \phi^{-1}.$$

Then we have

$$\begin{aligned}\beta^j &= (y^j \circ f \circ \gamma)'(0) = [(y^j \circ f \circ \phi^{-1}) \circ \phi \circ \gamma]'(0) = [\tilde{f}^j \circ (\phi \circ \gamma)]'(0)\\ &= \sum_{i=1}^{m} \partial_i \tilde{f}^j(\phi(p))(x^i \circ \gamma)'(0) = \sum_{i=1}^{m} \partial_i \tilde{f}^j(\phi(p))\alpha^i,\end{aligned}$$

which shows that $T_pf(D)$ depends only on D but not on γ and that T_pf is a linear mapping the matrix of which in the bases $(\partial_1^\phi(p), \dots, \partial_m^\phi(p))$ and $(\partial_1^\psi(f(p)), \dots, \partial_n^\psi(f(p)))$ is

$$\left(\partial_i \tilde{f}^j(\phi(p))\right)_{\substack{1\le i\le m\\ 1\le j\le n}}. \qquad \square$$

We remark that although the matrix representation of T_pf we have just computed depends on the choice of the charts ϕ and ψ, the linear map T_pf itself does not depend these choices. This is clear from the definition of T_pF, which does not use local coordinates.

Proposition 4.2.14 (Chain Rule)**.** *If $g\colon M \to N$ and $f\colon N \to P$ are smooth maps, $p \in M$, then $T_p(f \circ g) = T_{g(p)}f \circ T_pg$.*

Proof. If a tangent vector of M at p is represented by the curve γ, then

$$T_p(f \circ g)([\gamma]) = (f \circ g \circ \gamma)'(0) = T_{g(p)}f([g \circ \gamma]) = T_{g(p)}f(T_pg([\gamma])). \qquad \square$$

4.3 The Lie Algebra of Vector Fields

Definition 4.3.1. A *smooth vector field* X over a differentiable manifold M is a smooth mapping of M into its tangent bundle, such that $X(p) \in T_pM$ for each $p \in M$. ※

Obviously, smooth vector fields over M form a real vector space with respect to the operations

$$(X + Y)(p) := X(p) + Y(p), \qquad (\lambda X)(p) := \lambda X(p),$$

where X, Y are vector fields, $\lambda \in \mathbb{R}$, $p \in M$. We can multiply vector fields by smooth functions as well, by the rule

$$(fX)(p) := f(p)X(p).$$

We denote by $\mathfrak{X}(M)$ the vector space of smooth vector fields.
Associated to a local coordinate system $\phi = (x^1, \dots, x^n)$ on M, there is a basis of T_pM at each $p \in \operatorname{dom}\phi$, formed by the tangent vectors $\{\partial_i^\phi(p) : 1 \le i \le n\}$. The mapping $\partial_i^\phi\colon p \mapsto \partial_i^\phi(p)$ gives a local smooth vector field in the domain of the chart for each i. Thus, every smooth vector field X can be written in the form

$$X = \sum_{i=1}^n X^i \partial_i^\phi,$$

where the X^i-s are smooth functions on the domain of ϕ. The functions X^i are called the *components* of the vector field X with respect to the chart ϕ.

Given a smooth vector field X on a manifold M, we may pose the following problem. Find those smooth curves $\gamma\colon (a,b) \to M$ in M for which the speed of γ at $t \in (a,b)$ is $X(\gamma(t))$. Such curves are called the *integral curves* of the vector field. Obviously, a restriction of an integral curve onto a subinterval is also an integral curve, therefore, it is enough to look for the *maximal integral curves* which can not be extended to an integral curve defined on a larger interval.

The solution of this problem reduces to an ordinary differential equation of first order. Indeed, γ is an integral curve if and only if for each chart $\phi = (x^1, \dots, x^n)$, the "vector-valued" function

$$\mathbf{f} = \phi \circ \gamma\colon (a,b) \to \mathbb{R}^n$$

satisfies the differential equation

$$\mathbf{f}'(t) = (X^1 \circ \phi^{-1} \circ \mathbf{f}(t), \dots, X^n \circ \phi^{-1} \circ \mathbf{f}(t)).$$

Actually, finding integral curves of a vector field is the same problem as solving an ordinary differential equation, only the language of formulation is different. Translating the basic results of the theory of ordinary differential equations into the language of geometry we get the following theorems, we mention without proof.

Theorem 4.3.2.

(i) (Existence and uniqueness of solutions). *Let X be a smooth vector field on a differentiable manifold M. Then for each point $p \in M$, there exists a unique maximal integral curve $\gamma_p\colon (a,b) \to M$ of the vector field X such that $0 \in (a,b)$ and $\gamma_p(0) = p$ (a and b are allowed to be $-\infty$ and ∞ respectively).*

(ii) (Unboundedness of solutions "in time or space"). *If a (or b) is finite then no compact subset of M contains the image $\gamma_p((a,0))$ (or $\gamma_p((0,b))$).*

(iii) (Differentiable dependence on the initial point). *Let us define the set $U_t \subset M$ for $t \in \mathbb{R}$ as follows*

$$U_t = \{p \in M : t \in \operatorname{dom} \gamma_p\}.$$

Then U_t is an open subset of M and the mapping $H_t\colon U_t \to M$ defined by $H_t(p) = \gamma_p(t)$ is a diffeomorphism between U_t and U_{-t}. If, furthermore, the expression $H_{t_1}(H_{t_2}(p))$ is defined, then so is $H_{t_1+t_2}(p)$

and $H_{t_1}(H_{t_2}(p)) = H_{t_1+t_2}(p)$. The family $\{H_t : t \in \mathbb{R}\}$ is called the one-parameter family of diffeomorphisms *or the* flow generated by the vector field X.

A useful consequence of the fundamental theorems is the following statement on the straightening of vector fields.

Proposition 4.3.3. *Let X be an arbitrary vector field on a manifold M and $p \in M$ be a point such that $X(p) \neq \mathbf{0}$. Then there exists a chart $\phi = (x^1, \dots, x^n)\colon U \to \mathbb{R}^n$ around p for which $X|_U = \partial_1^\phi$.*

This means that the derivative of the mapping ϕ turns the vector field X into a constant vector field on $\mathbb{R}^n$.

Proof. Let H_t be the flow generated by the vector field X. Choose a chart $\psi\colon V \to \mathbb{R}^n$ around p such that $\psi(p) = \mathbf{0}$. Then the image of ψ contains a ball B_r of radius r about the origin. Decompose X as a linear combination $X = X^1\partial_1^\psi + \dots + X^n\partial_n^\psi$. Since $X(p) \neq 0$, one of the coefficients $X^i(p)$ should also be different from 0. We may assume that $X^1(p) \neq 0$ as this can be reached by permuting the coordinates.
Let $B_r^{n-1} \subset B_r$ be the $(n-1)$-dimensional ball consisting of those point of B_r the first coordinate of which is 0, and let $j\colon B_r^{n-1} \to M$ be the inverse ψ restricted onto B_r^{n-1}. The set W_ϵ of those points $q \in M$ for which the maximal integral curve starting at q is defined on a fixed open interval $(-\epsilon, \epsilon)$ is open in M and these open sets for all positive $\epsilon > 0$ cover M. Since $j(B_r^{n-1})$ is compact, it is covered by a finite number of such sets W_{ϵ_i}. Thus, taking for $\epsilon > 0$ the smallest of the ϵ_i's, we obtain a positive number such that all maximal integral curves of X starting at a point of $j(B_r^{n-1})$ is defined on the interval $(-\epsilon, \epsilon)$.
Consider the smooth map $F\colon (-\epsilon, \epsilon) \times B_r^{n-1} \to M$ which is defined by the formula

$$F(x_1, \dots, x_n) = H_{x_1}(j(0, x_2, \dots, x_n)).$$

The partial derivatives of F at the origin are $\partial_1 F(\mathbf{0}) = X(p)$, and $\partial_i F(\mathbf{0}) = \partial_i^\psi(p)$ if $i \geq 2$. These vectors are linearly independent, therefore, by the inverse function theorem, there is an open neighborhood U of p and $W \subset (-\epsilon, \epsilon) \times B_r^{n-1}$ of the origin in $\mathbb{R}^n$ such that $F|W$ is a diffeomorphism between W and U. In other words, $F|W$ is a local parameterization of M and its inverse ϕ is a local coordinate system on M. We claim that ϕ has the required property. Indeed, if $q \in U$ is an arbitrary point, and $\phi(q) = (q_1, \dots, q_n)$, then the parameterized coordinate line $(-\epsilon, \epsilon) \to M$, $t \mapsto F(t, q_2, \dots, q_n)$ is an integral curve of X, therefore its speed vector at $t = q_1$ is $\partial_1^\phi(q) = X(q)$. □

Since tangent vectors to a manifold at a point are identified with derivations at the point, vector fields can be regarded as differential operators called *derivations of smooth functions* assigning to a smooth function another smooth function by the formula

$$[X(f)](p) = [X(p)](f), \text{ where } X \in \mathfrak{X}(M), f \in \mathcal{F}(M), p \in M.$$

In this sense, a vector field X is a linear mapping $X\colon \mathcal{F}(M) \to \mathcal{F}(M)$, satisfying the *Leibniz rule*

$$X(fg) = X(f)g + fX(g).$$

Definition 4.3.4. Let A and B be two linear endomorphisms of a vector space V. Then the linear mapping $[A, B] = A \circ B - B \circ A$ is called the *commutator* of them. ※

Proposition 4.3.5. *The commutator of linear mappings satisfies the following identities ($A, B, C \in End(V)$, $\lambda \in \mathbb{R}$)*

(i) $[A + B, C] = [A, C] + [B, C]$, $\quad [C, A + B] = [C, A] + [C, B]$
$[\lambda A, B] = [A, \lambda B] = \lambda[A, B]$ *(bilinearity),*

(ii) $[A, B] = -[B, A]$ *(skew-commutativity),*

(iii) $[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$ *(Jacobi identity).*

Proof. We prove only (iii), the rest is left to the reader.

$$\begin{aligned}
&[A,[B, C]] + [B, [C, A]] + [C, [A, B]] \\
&\quad= [A, (B \circ C - C \circ B)] + [B, (C \circ A - A \circ C)] + [C, (A \circ B - B \circ A)] \\
&\quad= A \circ (B \circ C - C \circ B) - (B \circ C - C \circ B) \circ A + B \circ (C \circ A - A \circ C) - \\
&\qquad - (C \circ A - A \circ C) \circ B + C \circ (A \circ B - B \circ A) - (A \circ B - B \circ A) \circ C \\
&\quad= A \circ B \circ C - A \circ C \circ B - B \circ C \circ A + C \circ B \circ A + \\
&\qquad + B \circ C \circ A - B \circ A \circ C - C \circ A \circ B + A \circ C \circ B + \\
&\qquad + C \circ A \circ B - C \circ B \circ A - A \circ B \circ C + B \circ A \circ C \\
&\quad= 0.
\end{aligned}$$
□

Definition 4.3.6. Let us suppose that a linear space L is endowed with a bilinear operation $[\,,]\colon L \times L \to L$ satisfying conditions (i), (ii), and (iii) of the above proposition. Then the pair $(L, [\,,])$ is called a *Lie algebra*. ※

Proposition 4.3.7. *Let X and Y be two smooth vector fields on a manifold M. Considering them as linear endomorphisms of the vector space of smooth functions $\mathcal{F}(M)$, the commutator $[X, Y]$ of them is also a vector field.*

Proof. The commutator $[X, Y]$ is a linear endomorphism of $\mathcal{F}(M)$ so we have only to check that it satisfies the Leibniz rule. For $f, g \in \mathcal{F}(M)$ we have

$$\begin{aligned}[X, Y](fg) &= (X \circ Y - Y \circ X)(fg) = X(Y(fg)) - Y(X(fg)) \\ &= X(Y(f)g + fY(g)) + Y(X(f)g + fX(g)) \\ &= X \circ Y(f)g + Y(f)X(g) + X(f)Y(g) + fX \circ Y(g) - Y \circ X(f)g - \\ &\quad - X(f)Y(g) - Y(f)X(g) - fY \circ X(g) \\ &= [X, Y](f)g + f[X, Y](g).\end{aligned}$$ □

As a corollary, we obtain that he commutator of vector fields, which is generally called the *Lie bracket* of them, is a binary operation on $\mathfrak{X}(M)$, giving the space of vector fields a Lie algebra structure.

Proposition 4.3.8. *Let us choose a local coordinate system $\phi = (x^1, \dots, x^n)$ on M and denote by $\partial_1^\phi, \dots, \partial_n^\phi$ the associated coordinate vector fields. Then we have*

(i) $[\partial_i^\phi, \partial_j^\phi] = 0$*;*

(ii) $[fX, gY] = fg[X, Y] + fX(g)Y - gY(f)X$ *for each* $X, Y \in \mathfrak{X}(M)$, $f, g \in \mathcal{F}(M)$*;*

(iii) *if* $X = \sum_{i=1}^n X^i \partial_i^\phi$, $Y = \sum_{i=1}^n Y^i \partial_i^\phi$ *are arbitrary vector fields, then*

$$[X, Y] = \sum_{i=1}^n (X(Y^i) - Y(X^i))\partial_i^\phi = \sum_{i=1}^n \Big(\sum_{j=1}^n X^j \partial_j^\phi Y^i - Y^j \partial_j^\phi X^i \Big) \partial_i^\phi.$$

Proof. (i) The first part of the proposition is equivalent to Young's theorem since for any smooth function f on the domain of ϕ, we have

$$\begin{aligned}\partial_i^\phi(\partial_j^\phi(f)) &= \partial_i^\phi(\partial_j(f \circ \phi^{-1}) \circ \phi) = \partial_i(\partial_j(f \circ \phi^{-1})) \circ \phi \\ &= \partial_j(\partial_i(f \circ \phi^{-1})) \circ \phi = \partial_j^\phi(\partial_i^\phi(f)).\end{aligned}$$

(ii) Let h be an arbitrary smooth function on M, and apply the operator $[fX, gY]$ to it.

$$\begin{aligned}[fX, gY](h) &= fX(gY(h)) - gY(fX(h)) \\ &= fX(g)Y(h) + fgX(Y(h)) - gY(f)X(h) - gfY(X(h)) \\ &= (fg[X, Y] + fX(g)Y - gY(f)X)(h).\end{aligned}$$

(iii) Using (i) and (ii) we get

$$[X,Y]=\Big[\sum_{i=1}^n X^i\partial_i^\phi,\sum_{j=1}^n Y^j\partial_j^\phi\Big]=\sum_{i=1}^n\sum_{j=1}^n[X^i\partial_i^\phi,Y^j\partial_j^\phi]=$$

$$=\sum_{i=1}^n\sum_{j=1}^n X^i\partial_i^\phi(Y^j)\partial_j^\phi-Y^j\partial_j^\phi(X^i)\partial_i^\phi=\sum_{i=1}^n(X(Y^i)-Y(X^i))\partial_i^\phi.\ \square$$

Suppose that we are given two vector fields X and Y on an open subset of $\mathbb{R}^n$. The corresponding flows H_s and G_t do not commute in general: $H_s\circ G_t\neq G_t\circ H_s$.
To measure the lack of commutation of the flows H_s and G_t, we consider the difference

$$\Phi(s,t;p)=G_t\circ H_s(p)-H_s\circ G_t(p)$$

for a fixed point $p\in\mathbb{R}^n$. Φ is a differentiable function of s and t and it is $\mathbf{0}$ if t or s is zero. This means, that in the Taylor expansion of Φ around $(0,0;p)$

$$\Phi(s,t;p)=\Phi(0,0;p)+\left(s\frac{\partial\Phi}{\partial s}(0,0;p)+t\frac{\partial\Phi}{\partial t}(0,0;p)\right)+$$
$$+\left(\frac{s^2}{2}\frac{\partial^2\Phi}{\partial s^2}(0,0;p)+st\frac{\partial^2\Phi}{\partial s\partial t}(0,0;p)+\frac{t^2}{2}\frac{\partial^2\Phi}{\partial t^2}(0,0;p)\right)+o(s^2+t^2)$$

the only non-zero partial derivative is $\dfrac{\partial^2\Phi}{\partial s\partial t}(0,0;p)$.

Proposition 4.3.9. *Through the natural identification of the tangent space of $\mathbb{R}^n$ at p with the vectors of $\mathbb{R}^n$, the vector $\dfrac{\partial^2\Phi}{\partial s\partial t}(0,0;p)$ corresponds to the tangent vector $[X,Y](p)$.*

Proof. Put $X=\sum_{i=1}^n X^i\partial_i$, $Y=\sum_{i=1}^n Y^i\partial_i$. Let us compute first the vector $\frac{\partial^2}{\partial s\partial t}H_s\circ G_t(p)$ at $s=t=0$. Since H_s is the flow of X, for any point q, the curve $\gamma_q(s)=H_s(q)$ is an integral curve of X. Thus, we have

$$\left(\frac{\partial}{\partial s}H_s\circ G_t(p)\right)(0,t)=\left(\frac{\partial}{\partial s}\gamma_{G_t(p)}(s)\right)(0,t)=X(G_t(p)).$$

Differentiating with respect to t,

$$\frac{\partial^2}{\partial s\partial t}H_s\circ G_t(p)\big|_{s=t=0}=\frac{d}{dt}X(G_t(p))\big|_{t=0}=\sum_{i=1}^n Y(X^i)\partial_i(p)$$

A similar computation shows that

$$\frac{\partial^2}{\partial s\partial t}G_t\circ H_s(p)\big|_{s=t=0}=\sum_{i=1}^n X(Y^i)\partial_i(p)$$

Subtracting these equalities we get

$$\frac{\partial^2\Phi}{\partial s\partial t}(0,0,;p) = \sum_{i=1}^{n}(X(Y^i) - Y(X^i))\partial_i(p) = [X,Y](p). \qquad \square$$

Now returning to the Taylor expansion of Φ, we see that

$$\Phi(s,t;p) = st\frac{\partial^2\Phi}{\partial s\partial t}(0,0,;p) + o(s^2+t^2) = st[X,Y](p) + o(s^2+t^2)$$

In particular, we obtain the following expression for $[X,Y](p)$.

$$[X,Y](p) = \lim_{t\to 0}\frac{G_t\circ H_t(p) - H_t\circ G_t(p)}{t^2}$$

Definition 4.3.10. We say that two vector fields are *commuting* if their Lie bracket is the zero vector field. ※

Theorem 4.3.11. *Let $\{H_t : t\in\mathbb{R}\}$ and $\{G_t : t\in\mathbb{R}\}$ be the one-parameter families of diffeomorphisms generated by the vector fields X and Y respectively and suppose that the vector fields X and Y are commuting. Then the diffeomorphisms H_s and G_t are commuting as well in the following sense.*
For each point p of the manifold there exists a positive ε (depending on p) such that for any pair of real numbers s,t satisfying the inequality $|s|+|t|\le\varepsilon$ the expressions $H_s(G_t(p))$ and $G_t(H_s(p))$ are defined and coincide: $H_s(G_t(p)) = G_t(H_s(p))$.

Proof. If both X and Y vanish at p then $H_s(p) = G_t(p) = p$ for any s and t and thus the assertion holds trivially. We may thus suppose that one of the vectors $X(p)$, $Y(p)$, say $X(p)$ is not zero. By the theorem on the straightening of vector fields we may suppose that the manifold is an open subset of $\mathbb{R}^n$, with coordinates $(x^1,\dots,x^n)$, and the vector field X coincides with the basis vector field ∂_1. Let $Y = \sum_{i=1}^n Y^i\partial_i$ be the decomposition of Y into a linear combination of the basis vector fields ∂_i. By the formula for the Lie bracket of vector fields we have

$$0 = [X,Y] = \sum_{i=1}^{n}(\partial_1 Y^i)\partial_i \quad\iff\quad \partial_1 Y^i = 0 \ \text{ for each } i.$$

Consequently, the functions Y^i do not depend on x^1, thus the vector field Y is invariant under translations parallel to the vector $\mathbf{e}_1 = (1,0,\dots,0)$. This implies that if γ is an integral curve of the vector field Y, then so is $\gamma + s\mathbf{e}_1$ for any s (the domain of $\gamma + s\mathbf{e}_1$ is an open subset of the domain of γ). On the other hand, the diffeomorphism H_s is just a translation by the vector $s\mathbf{e}_1$. Hence, for small s and t, we have

$$\begin{aligned} G_t(H_s(p)) &= G_t(p + s\mathbf{e}_1) = (\gamma_p + s\mathbf{e}_1)(t) \\ &= \gamma_p(t) + s\mathbf{e}_1 = G_t(p) + s\mathbf{e}_1 = H_s(G_t(p)). \end{aligned}$$ □

If $f\colon M \to N$ is a smooth map, then the derivative of f can be used to "send forward" tangent vectors of M to tangent vectors of N. However this, in general, cannot be used to "send forward" vector fields. Taking a smooth vector field on M, and sending forward its vectors by the derivative map of f we obtain a bunch of tangent vectors of N which may not be tangent vectors of a smooth vector field of N. One obstruction of this can be that if f is not injective, then we can get different tangent vectors at some points of N coming from different preimages. If we get at most one vector at each point of N then it is still possible that this partially defined vector field does not extend to a global smooth vector field. If there is a smooth extension but f is not surjective, then the smooth extension is definitely not unique. These problems are eluded by the notion of f-related vector fields which makes the assumption that the image vectors extend to a smooth vector field a part of the definition.

Definition 4.3.12. Let $f\colon M \to N$ be a smooth map between two smooth manifolds. We say that the vector fields $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(N)$ are *f-related* if $T_pf(X_p) = Y_{f(p)}$ for all $p \in M$. ※

Thinking of vector fields as derivations of smooth functions, X and Y are f-related if and only if for any smooth function $h \in \mathcal{C}^\infty$ on N, the equality $Y(h)(f(p)) = Y_{f(p)}h = X_p(h \circ f) = X(h \circ f)(p)$ holds for all $p \in M$ that is, if $Y(h) \circ f = X(h \circ f)$. This equivalent version of the definition is convenient to prove the following proposition.

Proposition 4.3.13. *If $f\colon M \to N$ is a smooth map, $X_1, X_2 \in \mathfrak{X}(M)$ and $Y_1, Y_2 \in \mathfrak{X}(N)$ are smooth vector fields such that X_1 is f-related to Y_1 and X_2 is f-related to Y_2, then $[X_1, X_2]$ is f-related to $[Y_1, Y_2]$.*

Proof. Let $h \in \mathcal{C}^\infty$ be an arbitrary smooth function on N, then

$$\begin{aligned} [Y_1, Y_2](h) \circ f &= Y_1(Y_2(h)) \circ f - Y_2(Y_1(h)) \circ f \\ &= X_1(Y_2(h) \circ f) - X_2(Y_1(h) \circ f) \\ &= X_1(X_2(h \circ f)) - X_2(X_1(h \circ f)) = [X_1, X_2](h \circ f). \end{aligned}$$

as we wanted to show. □

Corollary 4.3.14. *When $f\colon M \to N$ is a diffeomorphism, then for any smooth vector field X on M, there is a unique smooth vector field $f_*(X)$*

on N which is f-related to X. The push forward f_* of vector fields is a Lie-algebra isomorphism between $\mathfrak{X}(M)$ and $\mathfrak{X}(N)$.

4.3.1 The Lie Algebra of Lie Groups

Let $F\colon M \to N$ be a diffeomorphism between two manifolds. Then F_* defines a bijection between $\mathfrak{X}(M)$ and $\mathfrak{X}(N)$, (which we shall denote also by TF) and this bijection is a Lie algebra isomorphism.
Let G be a Lie group, $g \in G$. Let L_g denote the left translation by g, i.e., $L_g\colon G \to G$, $L_g(h) = gh$. L_g is a diffeomorphism, its inverse is $L_{g^{-1}}$. A vector field $X \in \mathfrak{X}(G)$ is called *left invariant* if $TL_g(X) = X$ for all $g \in G$. Since $TL_g(X) = X$ and $TL_g(Y) = Y$ imply $TL_g[X,Y] = [TL_g(X), TL_g(Y)] = [X,Y]$, left invariant vector fields form a Lie subalgebra of $\mathfrak{X}(G)$.

Definition 4.3.15. The Lie algebra of left invariant vector fields of a Lie group is called the *Lie algebra of the Lie group.* ※

If $X \in \mathfrak{X}(G)$ is left invariant, then $X(g) = T_eL_g(X(e))$, thus, a left invariant vector field is uniquely determined by the vector $X(e) \in T_eG$ (e is the unit element of the group G). Since every vector in T_eG extends to a left invariant vector field this way, the assignment $X \mapsto X(e)$ yields a linear isomorphism between the vector space of left invariant vector fields on G and T_eG. As a consequence, we obtain that the Lie algebra of a Lie group is finite dimensional and its dimension is the same as that of the Lie group.
As an example, let us determine the Lie algebra of $\mathrm{GL}(n,\mathbb{R})$. $\mathrm{GL}(n,\mathbb{R})$ is an open subset in $\mathrm{Mat}(n,\mathbb{R})$, so its manifold structure is given by one chart, the embedding. Tangent spaces at different points can be identified with the linear space $\mathrm{Mat}(n,\mathbb{R})$. For $A \in \mathrm{GL}(n,\mathbb{R})$, the left translation $M \mapsto AM$ extends to a linear transformation of the whole linear space $\mathrm{Mat}(n,\mathbb{R})$. The derivative of a linear transformation of a linear space is the linear transformation itself, if we identify the tangent spaces at different points with the linear space, so a left invariant vector field $X\colon \mathrm{GL}(n,\mathbb{R}) \to \mathrm{Mat}(n,\mathbb{R})$ has the form $X(A) = AM$, where $M \in \mathrm{Mat}(n,\mathbb{R})$ is a fixed matrix.
The integral curves of a left invariant vector field on $\mathrm{GL}(n,\mathbb{R})$ can be described with the help of the exponential function for matrices. If M is an arbitrary square matrix, then we define e^M as the sum of the infinite series

$$\sum_{k=0}^{\infty} \frac{M^k}{k!}.$$

If we define the curve $\gamma_A\colon \mathbb{R} \to \mathrm{GL}(n,\mathbb{R})$ by $\gamma_A(t) = Ae^{Mt}$, then we obtain an integral curve of the vector field $X(A) = AM$. Indeed,

$$\gamma_A'(t) = Ae^{Mt}M = X(Ae^{Mt}) = X(\gamma_A(t)).$$

The flow generated by the left invariant vector field X consists of the diffeomorphisms

$$H_t(A) = Ae^{Mt},$$

that is, H_t is a right translation by e^{Mt}.
Now let us take two left invariant vector fields $X(A) = AM$ and $Y(A) = AN$ and consider the flows H_t and G_t generated by them.
Computing $G_t \circ H_t(A) - H_t \circ G_t(A)$ up to $o(t^2)$, we get

$$\begin{aligned} G_t \circ H_t(A) - H_t \circ G_t(A) &= A(e^{Mt}e^{Nt} - e^{Nt}e^{Mt}) \\ &= A\Big(I + Mt + \frac{1}{2}(Mt)^2\Big)\Big(I + Nt + \frac{1}{2}(Nt)^2\Big) - \\ &\quad - A\Big(I + Nt + \frac{1}{2}(Nt)^2\Big)\Big(I + Mt + \frac{1}{2}(Mt)^2\Big) + o(t^2) = \\ &= A(MN - NM)t^2 + o(t^2). \end{aligned}$$

We obtain, that *the Lie algebra of* $\mathrm{GL}(n,\mathbb{R})$ *is isomorphic to the Lie algebra of all matrices with Lie bracket* $[M,N] = MN - NM$.

Exercise 4.3.16. Let ∂_1 and ∂_2 be the two coordinate vector fields on $\mathbb{R}^2$ determined by the identity mapping. Describe the vector fields

$$\begin{aligned} X(x^1,x^2) &= x^1\partial_1 + x^2\partial_2 \\ Y(x^1,x^2) &= x^2\partial_1 - x^1\partial_2, \end{aligned}$$

compute their Lie bracket, and determine the flows generated by them. ¤

Exercise 4.3.17. Show that the Lie algebra of $SO(n)$ is isomorphic to the Lie algebra of skew-symmetric $n\times n$ matrices with Lie bracket $[X,Y] = XY - YX$. ¤

Exercise 4.3.18. Show that $\mathbb{R}^3$ endowed with the cross-product $\times$ is a 3-dimensional Lie algebra isomorphic to the Lie algebra of $SO(3)$. ¤

Exercise 4.3.19. For $\mathbf{v} \in \mathbb{R}^3$, let $X_{\mathbf{v}}$ denote the vector field on $\mathbb{R}^3$, defined by

$$X_{\mathbf{v}}(\mathbf{x}) = \mathbf{v} \times \mathbf{x}.$$

Describe the flow generated by $X_{\mathbf{v}}$, and prove that $[X_{\mathbf{v}}, X_{\mathbf{w}}] = -X_{\mathbf{v}\times\mathbf{w}}$. ¤

Exercise 4.3.20. Show that the Lie algebra of left invariant vector fields on a Lie group is isomorphic to the Lie algebra of right invariant vector fields. ¤

4.3.2 Subspace Distributions and the Frobenius Theorem

Frobenius theorem for subspace distributions is a geometric formulation of theorem giving necessary and sufficient condition for the solvability of certain systems of partial differential equations.

Definition 4.3.21. Let M be an n-dimensional manifold, $0 \leq k \leq n$ a fixed integer. A *k-dimensional smooth subspace distribution on M* is a map $\mathcal{D}$, which assigns to each point p of M a k-dimensional subspace $\mathcal{D}_p$ of the tangent space T_pM in such a way, that each point $p \in M$ has an open neighborhood U on which one can define k smooth vector fields $X_1, \dots, X_k$ with the property that for any $q \in U$, the subspace $\mathcal{D}_q$ is spanned by the vectors $(X_1)_q, \dots, (X_k)_q$. ※

The word *distribution* is used in many different senses in mathematics. It can mean a generalized function, a probability distribution, a subspace distribution. In what follows, it will always mean a subspace distribution.

Definition 4.3.22. We say that *a vector field X is tangent to the distribution* $\mathcal{D}$ if $X_p \in \mathcal{D}_p$ for all $p \in M$.
A distribution $\mathcal{D}$ is called *involutive* or *completely integrable*, if for any two vector fields that are tangent to $\mathcal{D}$, the Lie bracket of the vector fields is also tangent to $\mathcal{D}$. ※

Definition 4.3.23. An injectively immersed submanifold of M is a pair (N, ι), where N is a manifold and $\iota \colon N \to M$ is an injective immersion, that is an injective map the derivative map $T_q\iota \colon T_qN \to T_{\iota(q)}M$ of which is injective for each $q \in N$. We say that (N, ι) contains, or goes through a point p of M if $p \in \iota(N)$. ※

Definition 4.3.24. An *integral manifold* of a distribution $\mathcal{D}$ is an injectively immersed submanifold (N, ι) for which $T\iota(T_qN) = \mathcal{D}_{\iota(q)}$ for all $q \in N$. ※

Proposition 4.3.25. *If the distribution $\mathcal{D}$ on M has the property that each point of M is contained in an integral manifold of $\mathcal{D}$, then $\mathcal{D}$ is involutive.*

Proof. Let X and Y be two vector fields tangent to $\mathcal{D}$, $p \in M$ be an arbitrary point. To show that $[X, Y]_p \in \mathcal{D}_p$, consider an integral manifold (N, ι) of $\mathcal{D}$ passing through p. From the definition of an integral manifold, there exists two uniquely defined smooth vector fields $\tilde{X}$ and $\tilde{Y}$ on N which are ι-related with X and Y respectively. By Proposition 4.3.13, then $[\tilde{X}, \tilde{Y}]$ is also ι-related to $[X, Y]$, so if $q = \iota^{-1}(p)$, then

$$[X, Y]_p = T\iota([\tilde{X}, \tilde{Y}]_q) \in T\iota(T_qN) = \mathcal{D}_p,$$

which was to be proved. □

The theorem of Frobenius claims that the converse of this proposition is also true.

Theorem 4.3.26 (Frobenius' Theorem)**.** *Let $\mathcal{D}$ be an involutive smooth k-dimensional subspace distribution on the manifold M. Then every point of M is contained in an integral manifold of $\mathcal{D}$. Moreover, there is a local coordinate system $\phi = (x^1, \dots, x^n)$ in an open neighborhood U of any point $p \in M$ such that for any $q \in U$, the submanifold $N_q = \{\bar{q} \in U \mid x^i(\bar{q}) = x^i(q),\ \text{for } i = k+1, \dots, n\}$ of M together with the inclusion map $\iota\colon N_q \to M$ is an integral manifold of $\mathcal{D}$ passing through q. Integral manifolds also have the local uniqueness property that for any connected integral submanifold (N, ι) of $\mathcal{D}$ with image in U the functions $x^i \circ \iota$ are constant for $k+1 \le i \le n$.*

Proof. Observe first that the properties imposed on the chart $(x^1, \dots, x^n)$ are equivalent to the condition that the subspace $\mathcal{D}_q$ is spanned for all $q \in U$ by the vectors $\partial_i^\phi|_q$ for $i = 1, \dots, k$. This means geometrically that any involutive distribution can be "straightened" locally, that is locally diffeomorphic to a subspace distribution on $\mathbb{R}^n$ assigning to each point a translate of a given subspace.

We prove the theorem by induction on the dimension of the distribution. If $k = 0$, then the statement is obvious. Assume that $k \ge 1$ and that the theorem is proved for $(k-1)$-dimensional distributions. In order to construct a suitable chart around p, choose an open set V around p, over which the distribution can be spanned by the pointwise linearly independent smooth vector fields $X_1, \dots, X_k$.

By Proposition 4.3.3, there is a a local coordinate system $\psi = (y^1, \dots, y^n)$ defined on an open neighborhood $W \subset V$ of p for which $X_1|_W = \partial_1^\psi$. Denote by $\Sigma \subset W$ the $(n-1)$-dimensional submanifold defined by the equation $y^1 = 0$.

Define the vector fields $Y_1, \dots, Y_k$ on W by the equations $Y_1 = X_1$, $Y_i = X_i - X_i(y^1)X_1$, $(i = 2, \dots, k)$. It is clear that $Y_1, \dots, Y_k$ span the same subspace at each point of W as $X_1, \dots, X_k$. For $i \ge 2$, we also have

$$Y_i(y^1) = X_i(y^1) - X_i(y^1)X_1(y^1) = 0,$$

which means that the vector field Y_i is tangent to the level sets of the first coordinate function y^1, in particular, it is tangent to Σ.

Let Z_i be the restriction of the vector field Y_i onto Σ. We claim that the distribution $\mathcal{D}'$ spanned by the vector fields $Z_2, \dots, Z_k$ on Σ is involutive. Indeed, as $\mathcal{D}$ is involutive, the Lie bracket $[Y_r, Y_s]$ of the vector fields Y_r and Y_s is tangent to $\mathcal{D}$ hence can be written as

$$[Y_r, Y_s] = \sum_{i=1}^k c_{rs}^i Y_i \tag{4.1}$$

where the coefficients c^i_{rs} are smooth functions on W. Evaluating both sides of this equation on the function y^1 for $r, s \geq 2$ we obtain $0 = c^1_{rs}$. Thus restricting the equation (4.1) onto Σ gives the decomposition $[Z_r, Z_s] = \sum_{i=2}^k c^i_{rs} Z_i$ which implies easily that $\mathcal{D}'$ is involutive.

Applying the induction hypothesis for the distribution $\mathcal{D}'$ we obtain a local coordinate system $\eta = (z^2, \dots, z^n)$ on Σ defined around p, such that $\mathcal{D}'$ is spanned by the vector fields ∂^η_i, $i = 2, \dots, k$. (For convenience, we shifted the indices by 1. This means that ∂^η_i denotes derivation with respect to the $(i-1)$st coordinate z^i of η.) Suppose that the transit map between the coordinates $(y^2, \dots, y^n)|_\Sigma$ and $(z^2, \dots, z^n)$ is given by the smooth functions f^i in such a way that

$$z^i = f^i \circ (y^2, \dots, y^n)|_\Sigma, \quad (i = 2, \dots, n). \tag{4.2}$$

Introduce a new local coordinate system $\phi = (x^1, \dots, x^n)$ on M defined in some open neighborhood of p with the help of the equations

$$x^1 = y^1, \tag{4.3}$$

$$x^i = f^i \circ (y^2, \dots, y^n), \quad (i = 2, \dots, n). \tag{4.4}$$

Shrink the open domain on which ϕ is defined to an open neighborhood U of p in such a way that the image of U under the chart ϕ be equal to an open cube K the edges of which are parallel to the coordinate axes. We can assume without loss of generality that K is centered at the origin, which is the image $\phi(p)$ of p as well.

We show that the local coordinate system ϕ satisfies the requirements of the theorem. For this purpose, it is enough to check that $Y_i(x^r) = 0$, if $i \leq k$ and $k+1 \leq r \leq n$. Computing the derivative matrix of the transit map (4.3-4.4) we can see that $\partial^\phi_1 = \partial^\psi_1 = Y_1$ on U. This implies at once that $Y_1(x^r) = \partial^\phi_1(x^r) = 0$ for $k+1 \leq r \leq n$. Suppose $2 \leq i \leq k$. Using the decomposition (4.1) for $[Y_1, Y_i]$ and applying this derivation to x^r with $(k+1 \leq r \leq n)$ we get

$$\partial^\phi_1(Y_i(x^r)) = Y_1(Y_i(x^r)) = [Y_1, Y_i](x^r) = \sum_{s=1}^k c^s_{1i} Y_s(x^r). \tag{4.5}$$

Let the projection of the cube K onto the first coordinate axis be the interval $(-a, a)$ and denote by K' its orthogonal projection onto the hyperplane $x^1 = 0$. Fix an arbitrary point $\mathbf{c} = (0, c_2, \dots, c_n) \in K'$ and denote by $q_t \in U$ the point the coordinates of which are $x^1(q_t) = t$, $x^2(q_t) = c_2, \dots, x^n(q_t) = c_n$. Consider the functions $h_{ir} \colon (-a, a) \to \mathbb{R}$, $h_{ir}(t) = (Y_i(x^r))(q_t)$, $(2 \leq i \leq k < r \leq n)$. According to the equations (4.5), the functions h_{ir} $(i = 2, \dots, k)$

satisfy the linear ordinary differential equation

$$h'_{ir} = \sum_{s=2}^{k} \hat{c}_{1i}^{s} h_{sr}, \quad (i = 2, \dots, k), \tag{4.6}$$

where $\hat{c}_{1i}^s(t) = c_{1i}^s(q_t)$. We know also that at the point $q_0 \in \Sigma$, the vector $Y_i(q_0) \in \mathcal{D}'_{q_0}$ is a linear combination of the vectors $\partial_s^\phi|_{q_0} = \partial_s^\eta|_{q_0}$ $s = 2, \dots, k$. For this reason, $(Y_i(x^r))(q_0) = 0$, that is $h_{ir}(0) = 0$ for all $2 \le i \le k < r \le n$. The functions $h_{ir} \equiv 0$ also solve the differential equation (4.6) with initial condition $h_{ir}(0) = 0$, thus, by the uniqueness of solutions with a given initial condition, $h_{ir}(t) = (Y_i(x^r))(q_t) \equiv 0$ for all $2 \le i \le k < r \le n$. As choosing the constants $c_2, \dots, c_n$ and $t \in (-a, a)$ properly q_t can be any point of U, $Y_i(x^r) \equiv 0$ for all $2 \le i \le k < r \le n$, and this is what we wanted to prove. □

Definition 4.3.27. A *maximal integral manifold* of a distribution is a connected integral manifold (N, ι) the image $\iota(N)$ of which is not a proper subset of the image of any other connected integral manifold. ※

Theorem 4.3.28. *Let $\mathcal{D}$ be a k-dimensional involutive distribution on the manifold M. Then there is a unique maximal integral manifold through any point p in M. Uniqueness should be understood as follows. If (N_1, ι_1) and (N_2, ι_2) are two maximal integral manifolds through a given point p, then there is a diffeomorphism $h\colon N_1 \to N_2$ such that $\iota_1 = \iota_2 \circ h$.*

Proof. Existence. Let $N \subset M$ be the subset of those points $q \in M$ that can be connected to p by a piecewise smooth continuous curve the speed vectors of which are in the distribution $\mathcal{D}$. It is clear that N contains the image of any connected integral manifold passing through p, therefore it is sufficient to create a smooth manifold structure on N, which makes N together with the inclusion map $\iota\colon N \to M$ an integral manifold of $\mathcal{D}$.

By the Frobenius theorem, M can be covered by a countable family of local coordinate systems

$$\phi_i = (x_i^1, \dots, x_i^n)\colon U_i \to \mathbb{R}^n$$

in such a way that $\phi_i(U_i)$ is a cube the edges of which are parallel to the coordinate axes, furthermore, the connected integral manifolds of $\mathcal{D}$ lying in U_i are exactly the connected open subsets of a submanifold of U that can be defined by a system of equations $x_i^r = c_r$, $(k+1 \le r \le n)$, with some constants c_r.

Call the submanifolds of U_i that can be defined by a system of equations of the form $x_i^r = c_r$, $(k+1 \le r \le n)$ the *slices* of U_i.

Choose for all $q \in N$ an index i_q such that $q \in U_{i_q}$, and let S_{i_q} denote the slice of U_{i_q} which contains q. Observe that open subsets of the subspace topologies of the slices $\{S_{i_q} \mid q \in N\}$ give a base of a locally Euclidean topology on N. Furthermore, the charts

$$(x^1_{i_q}, \ldots, x^k_{i_q})|_{S_{i_q}} : S_{i_q} \to \mathbb{R}^k$$

form a $\mathcal{C}^\infty$-atlas on N. It is more or less a routine work to check that this is a right smooth manifold structure on N. Probably the only point which requires some idea is the proof of the fact that N is second countable. This is a corollary of the fact that N contains at most a countable number of slices. The latter is true, because due to the connectedness of N one can choose for any two slices S, S' contained in N a sequence $S = S_1, S_2, \ldots, S_m = S'$ of slices in N such that any two consecutive slices in the sequence have a point in common. Thus, it is sufficient to prove that each slice intersects at most countable number of other slices. Since each slice is contained in one of the countable number of sets U_i, it suffices to show that for any pair of indices $i, j \in \mathbb{N}$ a slice S of U_i can intersect at most countable number of slices of U_j. If S cuts the slices $\{S_\alpha | \alpha \in I\}$ of U_j, then the intersections $S \cap S_\alpha$ are disjoint open subsets in S, however, as S is second countable it can contain at most a countable number of pairwise disjoint open subsets, hence I is countable. The proof of *uniqueness* is simple and left to the reader. □

4.4 Tensor Bundles and Tensor Fields

We gave definition of tensor fields on hypersurfaces earlier. In this section, we are going to extend the definition to manifolds. The general definition will introduce tensor fields as sections of the tensor bundles over the manifold. Tensor bundles are some vector bundles produced from the tangent bundle.

Definition 4.4.1. A smooth real *k-dimensional vector bundle* (E, π) *over a manifold* M consists of

- a smooth manifold E called the *total space* of the bundle,
- a smooth map $\pi : E \to M$, called the *projection* of the bundle,
- a given k-dimensional linear space structure on the preimage $F_p = \pi^{-1}(p)$ of each point $p \in M$.

The linear space F_p is called the *fiber over* p, M itself is the *base space* of the bundle. In order to form a vector bundle, these data should satisfy the following local triviality condition: for any point $p \in M$ there is an open neighborhood U of p and a diffeomorphism $h : \pi^{-1}(U) \to U \times \mathbb{R}^k$ such that for any point $q \in U$, the restriction of h onto the fiber F_q is a linear isomorphism between F_q and $\{q\} \times \mathbb{R}^k$, where the linear space structure of the latter space is essentially the standard linear structure of $\mathbb{R}^k$. ※

Examples.

- A trivial vector bundle over M is a product space $E = M \times V$ with the projection map $\pi = \pi_M$ onto the first component, where V is a k-dimensional linear space over $\mathbb{R}$.
- The tangent bundle $\pi\colon TM \to M$ of an n-dimensional manifold M is an n-dimensional vector bundle. It is usually not a trivial bundle.

There are several methods to construct new vector bundles from given ones. To construct tensor bundles we can apply the following general construction for the tangent bundle of a manifold.
Consider the category $\mathcal{V}_k$ of k-dimensional linear spaces in which the morphisms between two k-dimensional linear spaces are the linear isomorphisms between them and let Φ be a covariant functor from $\mathcal{V}_k$ to $\mathcal{V}_l$ which is smooth in the sense that for any two k-dimensional linear spaces V W, the map $\Phi\colon \operatorname{Mor}(V,W) \to \operatorname{Mor}(\Phi(V),\Phi(W))$ is smooth. Smoothness means that if we fix some bases in V ,W, $\Phi(V)$ and $\Phi(W)$ to identify linear maps with their matrices, then the coefficients of the matrix of $\Phi(L)$ are smooth functions of the matrix coefficients of $L \in \operatorname{Mor}(V,W)$.
If $\pi\colon E \to M$ is a k-dimensional smooth vector bundle over M then applying Φ to each fiber of E, we can construct a new smooth vector bundle over M with l-dimensional fibers. The total space of this bundle is the disjoint union

$$E^{\Phi} = \bigcup_{p\in M}^{*} \Phi(F_p),$$

where F_p is the fiber over p in E. The projection π^{Φ} of E^{Φ} assigns to each point of $\Phi(F_p)$ the base point p. Thus, the fiber over p in the new bundle is $\Phi(F_p)$ which is a linear space of dimension l. It remains to define the smooth structure on the total space E^{Φ} and to show that the new bundle is locally trivial.
Fix a linear isomorphism between the linear spaces $I\colon \Phi(\mathbb{R}^k) \to \mathbb{R}^l$.
Choose for any point p in M an open neighborhood which is both the domain of a chart $\phi\colon U \to \mathbb{R}^n$ and a neighborhood over which the bundle E is trivialized by a diffeomorphism $h\colon \pi^{-1}(U) \to U \times \mathbb{R}^k$ as in the definition of local triviality. Then we can define a chart

$$\widetilde{(\phi,h)} = (\phi\circ\pi^{\Phi}, I\circ(\Phi h))\colon (\pi^{\Phi})^{-1}(U) \to \mathbb{R}^n \times \mathbb{R}^l$$

from ϕ and h on the set $(\pi^{\Phi})^{-1}(U) = \bigcup_{p\in M}^{*} \Phi(F_p)$ in the following way. The first component $\phi\circ\pi^{\Phi}$ of the chart $\widetilde{(\phi,h)}$ assigns to each element e in $(\pi^{\Phi})^{-1}(U)$ the coordinates of the base point $\pi^{\Phi}(e)$ with respect to the chart ϕ. These coordinates are the first n coordinates of e. The symbolic notation $I \circ (\Phi h))$ in the second component of the chart means the following. Since h is a local trivialization of the original bundle, it defines for each $q \in U$ a linear isomorphism $h_q \colon F_q \to \mathbb{R}^k$. The functor Φ assigns to this isomorphism an isomorphism $\Phi(h_q) \colon \Phi(F_q) \to \Phi(\mathbb{R}^k)$. Composing $\Phi(h_q)$ with the isomorphism I, we obtain a linear isomorphisms $\Phi(\mathbb{R}^k) \to \mathbb{R}^l$. The last l coordinates of $\widetilde{(\phi, h)}(e)$ for $e \in \Phi(F_q)$ are the coordinates of the vector $I(\Phi(h_q)(e))$.

The set of all charts constructed this way form an atlas on E^Φ. The transit map between any two charts is defined on an open subset of $\mathbb{R}^{n+l}$ and it is smooth. Indeed, if ψ and $\hat{h}$ are another chart and local trivialization over an open subset V, then the transit $\widetilde{(\psi, \hat{h})} \circ \widetilde{(\phi, h)}^{-1}$ map is defined on the open set $\phi(U \cap V) \times \mathbb{R}^l$, and maps $(\mathbf{x}, \mathbf{v}) \in \phi(U \cap V) \times \mathbb{R}^l$ to

$$\widetilde{(\psi, \hat{h})} \circ \widetilde{(\phi, h)}^{-1}(\mathbf{x}, \mathbf{v}) = \big(\psi(\phi^{-1}(\mathbf{x})), I \circ \Phi(\hat{h}_{\phi^{-1}(\mathbf{x})}) \circ (\Phi(h_{\phi^{-1}(\mathbf{x})}))^{-1} \circ I^{-1}(\mathbf{v})\big).$$

This transit map is obviously smooth, since it is built up from compositions of smooth functions. Thus, this atlas defines a smooth manifold structure on E^Φ. The vector bundle we have constructed is locally trivial, as it is shown by the diffeomorphisms

$$(\pi^\Phi, I \circ (\Phi(h))) \colon (\pi^\Phi)^{-1}(U) \to U \times \mathbb{R}^l.$$

Let $k, l \geq 0$ be two natural numbers. The construction of the tensor space $T^{(k,l)}V$ from a linear space V can be thought of as a functor from $\mathcal{V}_n$ to $\mathcal{V}_{n^{k+l}}$. If $L \colon V \to W$ is a linear isomorphism, then the associated isomorphism between $T^{(k,l)}V$ and $T^{(k,l)}W$ is

$$T^{(k,l)}(L) = \underbrace{(L^*)^{-1} \otimes \dots \otimes (L^*)^{-1}}_{k \text{ times}} \otimes \underbrace{L \otimes \dots \otimes L}_{l \text{ times}} \colon T^{(k,l)}V \to T^{(k,l)}W.$$

Applying the functor $T^{(k,l)}$ to the tangent bundle of a manifold, we obtain the the *tensor bundle* $T^{(k,l)}M \to M$ or the *bundle of tensors of type* (k, l) *over* M.

There is another important vector bundle over M, *the bundle of k-forms.* This is obtained by applying the functor $\Phi \colon V \mapsto \Lambda^k(V^*)$ to the tangent bundle. The isomorphism corresponding to a linear isomorphism $L \colon V \to W$ under this functor is the isomorphism $\Phi(L)$ uniquely characterized by the identity

$$\Phi(L)(l_1 \wedge \dots \wedge l_k) = (L^*)^{-1}(l_1) \wedge \dots \wedge (L^*)^{-1}(l_k), \text{ for all } l_1, \dots, l_k \in V^*.$$

The resulting vector bundle of k-forms will be denoted by $\Lambda^k T^* M$

Definition 4.4.2. A *smooth section of a vector bundle* (E, π) over M is a smooth map $s\colon M \to E$ such that $\pi \circ s = \mathrm{id}_M$. In other words, a section assign to each point of the manifold an element of the fiber F_p over that point. ※

Definition 4.4.3. Smooth sections of the tensor bundle $T^{(k,l)}M$ are called *smooth tensor fields of type* (k, l) *over* M. ※

Definition 4.4.4. A *differential* k*-form over a manifold* M is a smooth section of the bundle $\Lambda^k T^*M$. The set of all differential k-forms over M are denoted by $\Omega^k(M)$. ※

Smooth sections of a vector bundle, in particular, tensor fields and differential forms can be added and multiplied by a smooth function pointwise. With these operations they are modules over the ring of smooth functions.

Using the derivative of a smooth map between two manifolds, we can push forward or pull back certain types of tensors. Let $f\colon M \to N$ be a smooth map. Then for any point $p \in M$, the derivative map $T_p f$ is a linear map from $T_p M$ to $T_{f(p)}N$.

If $S\colon M \to T^{(0,l)}M$ is a smooth tensor field of type $(0, l)$, then it assigns to each point $p \in M$ an element $S(p)$ of the tensor product $T_pM \otimes \cdots \otimes T_pM$. Applying the kth tensor power $T_pf \otimes \cdots \otimes T_pf$ of the derivative map T_pf we can push $S(p)$ forward to $T_pf \otimes \cdots \otimes T_pf(S(p)) \in T^{(0,l)}(T_{f(p)N})$. Thus, the tensor assigned to a point by a tensor fields of type $(0, l)$ on M can be pushed forward to N, but, in general, these pushed forward tensors may not extend to a smooth tensor field on N or if they extend, the extension may not be unique. Thus, tensor fields of type $(0, l)$ may not be pushed forward to tensor fields of the same type on N. The situation is similar to the special case of vector fields, which was discussed when we introduced f-related vector fields.

Assume now that $S\colon N \to T^{(k,0)}N$ is a tensor field of type $(k, 0)$ on the image manifold N. Then for any $q \in N$, $S(q)$ is an element of $(T_qN)^* \otimes \cdots \otimes (T_qN)^*$. For $p \in M$, applying the linear transformation $(T_pf)^* \otimes \cdots \otimes (T_pf)^*$to $S(f(p))$ we obtain a tensor of type $(k, 0)$ over the linear space T_pM. In contrast to type $(k, 0)$ tensors, pulling back a tensor field of type $(0, l)$ on N pointwise, as described, we obtain a smooth tensor field on M. The pull back of the tensor field S by the smooth function f will be denoted by f^*S.

The pull-back of a tensor field of type $(k, 0)$ can also be explained as follows. The tensor space $T^{(k,0)}(T_qM)$ is naturally isomorphic to the linear space of k-linear functions on T_qN. Therefore, we can think of $S(q)$ as a k-linear function on T_qN. Similarly, for $p \in M$ $f^*S(p)$ is a k-linear function on T_pM. By the definition of $f * S$, the value of $f^*S(p)$ on the tangent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in T_pM$ can be computed by the formula

$$f^*S(p)(\mathbf{v}_1, \ldots, \mathbf{v}_k) = S(f(p))(T_pf(\mathbf{v}_1), \ldots, T_pf(\mathbf{v}_k)).$$

Differential forms can be interpreted as alternating tensors of type $(k,0)$, therefore, as a special case of the pull back of type $(k,0)$ tensors, differential forms on N can be pulled back to a differential form on M. Pull back for differential forms will also be denoted by $f^*\colon \Omega^k N \to \Omega^k M$.
In general, tensor fields of mixed type (k,l), where $k,l \geq 1$, can neither be pushed forward, nor be pulled back by a smooth map. However, there is a special case, when we can both push forward and pull back tensor fields of any type. This is the case, when f is a diffeomorphism. The idea is that a diffeomorphism identifies the two manifolds, so any tensor field on one of them corresponds to a tensor field on the other.
More formally, if $f\colon M \to N$ is a diffeomorphism S is a tensor field of type (k,l) on M, then S can be pushed forward to a type (k,l) tensor field f_*S on N by the formula

$$(f_*S)(f(p)) = \underbrace{((T_pf)^*)^{-1} \otimes \cdots \otimes ((T_pf)^*)^{-1}}_{k \text{ times}} \otimes \underbrace{T_pf \otimes \cdots \otimes T_pf}_{l \text{ times}}(S(p)).$$

To pull back a type (k,l) tensor field S on N, we essentially push it forward by f^{-1}, that is

$$(f^*S)(p) = \underbrace{(T_pf)^* \otimes \cdots \otimes (T_pf)^*}_{k \text{ times}} \otimes \underbrace{(T_{f(p)}f^{-1})^* \otimes \cdots \otimes (T_{f(p)}f^{-1})^*}_{l \text{ times}}(S(f(p))).$$

It is useful to know how one can work with tensor fields in local coordinates. If $\phi = (x^1,\dots,x^n)\colon U \to \mathbb{R}^n$ is a local coordinate system on the manifold, then the vector fields $\partial_1^\phi,\dots,\partial_n^\phi$ induced by ϕ give a basis in the tangent space T_pM for all $p \in M$. Taking the dual basis of the basis at each point we obtain n tensor fields of type $(1,0)$, which are usually denoted by $dx^1,\dots,dx^n$. The tensor fields $dx^{i_1} \otimes \cdots \otimes dx^{i_k} \otimes \partial^\phi_{j_1} \otimes \cdots \otimes \partial^\phi_{j_l}$, $(1 \leq i_1,\dots,i_k, j_1 \dots . j_l \leq n)$ yield a basis of the tensor space $T^{(k,l)}(T_pM)$ at each point $p \in U$, therefore, any smooth tensor field S of type (k,l), can be written uniquely as a linear combination

$$S = \sum_{i_1,\dots,i_k,j_1\dots.j_l=1}^{n} S^{j_1\dots j_l}_{i_1\dots i_k} dx^{i_1} \otimes \cdots \otimes dx^{i_k} \otimes \partial^\phi_{j_1} \otimes \cdots \otimes \partial^\phi_{j_l}$$

of these tensor fields, where the coefficients $S^{j_1\dots j_l}_{i_1\dots i_k}$ are smooth functions on U.
Similarly, any differential k-form ω on U can be decomposed uniquely as a linear combination

$$\omega = \sum_{1\leq i_1<\cdots<i_k\leq n} \omega_{i_1\dots i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k},$$

where the coefficients $\omega_{i_1\dots i_k}$ are smooth functions on U.

4.5 The Lie Derivative

Let M be a smooth manifold and X be a smooth vector field on M. Let $\Phi_t : U_t \to U_{-t}$, $(t \in \mathbb{R})$ be the flow generated by the vector field.

Definition 4.5.1. The *Lie derivative* $\mathcal{L}_X S$ of a tensor field S of type (k, l) on M is the type (k, l) tensor field

$$\mathcal{L}_X S = \lim_{t \to 0} \frac{\Phi_t^* S - S}{t}. \quad ※$$

The pulled back tensor field $\Phi_t^* S$ is defined only on U_t. Nevertheless, $\mathcal{L}_X S$ is properly defined on the whole manifold M, since for each $p \in M$, there is a positive t_0, such that $p \in U_{t_0} \cap U_{-t_0}$, and then $\Phi_t^* S$ is defined at p for $|t| \leq t_0$, so the limit can be taken at p.
The definition shows exactly the key idea behind the Lie derivative. It is the initial speed at which the tensor field changes when we start moving it by the flow of the vector field X. However, the limit on the right-hand side is not the most practical way to compute the Lie derivative, since to compute the flow, we have to solve an ordinary differential equation. Our goal now is to work out formulae for the computation of the Lie derivative without having to compute the flow of the vector field explicitly. We start with the simplest types of tensor fields.
Smooth functions on M are the tensor fields of type $(0, 0)$, or differential 0-forms. The Lie derivative for them is the well known differentiation of smooth functions with respect to vector fields.

Proposition 4.5.2. *For a smooth function f on M, $\mathcal{L}_X f = Xf$.*

Proof. In general, if $F : M \to N$ is a smooth map, and f is a smooth function on N, then the pull back of f with F is the composition $F^*(f) = f \circ F$. Applying this for $M = N$, and $F = \Phi_t$, we get

$$\mathcal{L}_X f(p) = \lim_{t \to 0} \frac{f(\Phi_t(p)) - f(p)}{t} = \frac{d}{dt} f(\Phi_t(p))|_{t=0} = X_p(f).$$

In the last step, we used the fact that $\gamma_p(t) = \Phi_t(p)$ is an integral curve of X, therefore its speed vector at 0 is $\gamma'(0) = X_p$. □

Vector fields are tensor fields of type $(0, 1)$. Lie derivation of vector fields is also a familiar operation.

Proposition 4.5.3. *If X and Y are smooth vector fields on M, then $\mathcal{L}_X Y = [X, Y]$.*

Proof. Let Φ_t and Ψ_t denote the flows generated by X and Y respectively. Choose a smooth function h on M and a point $p \in M$. Then

$$(\mathcal{L}_X Y(h))(p) = \lim_{t \to 0} \frac{((\Phi_t^* Y)(h))(p) - (Y(h))(p)}{t} = \frac{d}{dt} T_{\Phi_t(p)} \Phi_{-t}(Y_{\Phi_t(p)})(h)|_{t=0}$$

Since $\eta_p(s) = \Psi_s(\Phi_t(p))$ is an integral curve of Y, $\eta'(0) = Y_{\Phi_t(p)}$, and $T_{\Phi_t(p)}\Phi_{-t}(Y_{\Phi_t(p)}) = \frac{d}{ds}\Phi_{-t} \circ \Psi_s \circ \Phi_t(p)|_{s=0}$, thus

$$(\mathcal{L}_X Y(h))(p) = \frac{\partial^2}{\partial t \partial s} h(\Phi_{-t} \circ \Psi_s \circ \Phi_t(p))|_{t=s=0}.$$

Consider the function F defined on a neighborhood of the origin in $\mathbb{R}^3$ by the formula

$$F(u, v, w) = h(\Phi_u \circ \Psi_v \circ \Phi_w(p)).$$

Using the chain rule

$$\begin{aligned}
&\frac{\partial^2}{\partial t \partial s} h(\Phi_{-t} \circ \Psi_s \circ \Phi_t(p))|_{t=s=0} \\
&\quad = \frac{\partial^2}{\partial t \partial s} F(-t, s, t)|_{t=s=0} \\
&\quad = -\frac{\partial^2}{\partial u \partial v} F(u, v, 0)|_{u=v=0} + \frac{\partial^2}{\partial w \partial v} F(0, v, w)|_{w=v=0} \\
&\quad = \frac{\partial^2}{\partial t \partial s} [h(\Psi_s \circ \Phi_t(p)) - h(\Phi_t \circ \Psi_s(p))]|_{t=s=0} \\
&\quad = X_p(Y(h)) - Y_p(X(h)) = [X, Y]_p(h).
\end{aligned}$$

□

The Lie derivative of a tensor field of type $(1, 0)$, which is the same as a differential 1-form, it can be obtained from a Leibniz-type differentiation rule. If $\alpha \in \Omega^1(M)$ is a differential 1-form, then it assigns to each $p \in M$ a linear function α_p on T_pM. Therefore, if Y is a vector field on M, we can evaluate α on X pointwise and get a smooth function $\alpha(X)$, the value of which at p is $\alpha_p(X_p)$. The smooth function $\alpha(X)$ depends on α and X in a bilinear way over $\mathbb{R}$. For this reason, the Lie derivative of $\alpha(X)$ can be also computed by the following Leibniz-type formula.

Proposition 4.5.4. *For any two vector fields X, Y and any differential 1-form α on M, we have*

$$\mathcal{L}_X(\alpha(Y)) = (\mathcal{L}_X \alpha)(Y) + \alpha(\mathcal{L}_X Y).$$

In particular, rearranging, and using the previous two propositions

$$(\mathcal{L}_X \alpha)(Y) = X(\alpha(Y)) - \alpha([X, Y]).$$

Proof. We use the "standard trick" to prove all Leibniz-type rules as follows

$$\begin{aligned}\mathcal{L}_X(\alpha(Y)) &= \lim_{t\to 0}\frac{\Phi_t^*(\alpha(Y))-\alpha(Y)}{t} = \\ &\lim_{t\to 0}\frac{\Phi_t^*\alpha(\Phi_t^*Y)-\alpha(\Phi_t^*Y)+\alpha(\Phi_t^*Y)-\alpha(Y)}{t} \\ &= \lim_{t\to 0}\frac{\Phi_t^*\alpha-\alpha}{t}(\lim_{t\to 0}\Phi_t^*Y)+\alpha\left(\lim_{t\to 0}\frac{\Phi_t^*Y-Y}{t}\right) \\ &= (\mathcal{L}_X\alpha)(Y)+\alpha(\mathcal{L}_XY).\end{aligned}$$ □

Tensor fields can always be multiplied by the tensor product, differential forms can be multiplied by the wedge product pointwise. The Leibniz rule also holds for these operations.

Proposition 4.5.5. *If S_1 and S_2 are two arbitrary tensor fields, not necessarily of the same type, ω, η are arbitrary differential forms, then for any smooth vector field X, we have*

$$\mathcal{L}_X(S_1\otimes S_2) = (\mathcal{L}_XS_1)\otimes S_2 + S_1\otimes(\mathcal{L}_XS_2),$$

and

$$\mathcal{L}_X(\omega\wedge\eta) = (\mathcal{L}_X\omega)\wedge\eta+\omega\wedge(\mathcal{L}_X\eta).$$

Both equations can be proved by the standard trick. We leave the details to the reader.

Proposition 4.5.6. *If in addition to the assumptions of the previous proposition S_1 and S_2 have the same type, and the differential forms ω and η are of the same degree, then for any two real numbers $\lambda,\mu\in\mathbb{R}$ we have*

$$\mathcal{L}_X(\lambda S_1+\mu S_2) = \lambda\mathcal{L}_XS_1+\mu\mathcal{L}_XS_2,$$

and

$$\mathcal{L}_X(\lambda\omega+\mu\eta) = \lambda\mathcal{L}_X\omega+\mu\mathcal{L}_X\eta.$$

Locally, any tensor field can be written as the sum of a finite number of tensor products of some vector fields and differential one-forms. Similarly, any differential form can be written locally as a sum of wedge products of some differential 1-forms. Thus, the above rules are enough to compute the Lie derivative of any tensor field and differential form.
The following proposition tells us how the Lie derivation operator $\mathcal{L}_X$ depends on the vector field X.

Proposition 4.5.7. *The operator $\mathcal{L}_X$ depends $\mathbb{R}$-linearly on X, and satisfies the commutator relation*

$$\mathcal{L}_X\circ\mathcal{L}_Y-\mathcal{L}_Y\circ\mathcal{L}_X = \mathcal{L}_{[X,Y]}.$$

Proof. The proof follows the steps we went through to work out formulae to compute $\mathcal{L}_X$ on more and more complicated tensor fields.
If the Lie derivations in the statement are applied to smooth functions, then linearity in X is obvious, the commutator relation is just a reformulation of the definition of the Lie bracket.
When Z is a vector field,

$$\mathcal{L}_{\lambda X+\mu Y}Z = [\lambda X + \mu Y, Z] = \lambda[X,Z] + \mu[Y,Z] = (\lambda\mathcal{L}_X + \mu\mathcal{L}_Y)Z,$$

and

$$(\mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X)Z = [X[Y,Z]] - [Y,[X,Z]] = [[X,Y],Z] = \mathcal{L}_{[X,Y]}Z.$$

by the Jacobi identity.
If α is a 1-form, and Z is an arbitrary vector field, then

$$\begin{aligned}(\mathcal{L}_{\lambda X+\mu Y}\alpha)(Z) &= (\lambda X + \mu Y)(\alpha(Z)) - \alpha([\lambda X + \mu Y, Z]) \\ &= \lambda X(\alpha(Z)) + \mu Y(\alpha(Z)) - \lambda\alpha([X,Z]) - \mu\alpha([Y,Z]) \\ &= ((\lambda\mathcal{L}_X + \mu\mathcal{L}_Y)\alpha)(Z),\end{aligned}$$

furthermore,

$$\begin{aligned}&(\mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X)(\alpha)(Z) \\ &\quad = X((\mathcal{L}_Y\alpha)(Z)) - (\mathcal{L}_Y\alpha)([X,Z]) - Y((\mathcal{L}_X\alpha)(Z)) + (\mathcal{L}_X\alpha)([Y,Z]) \\ &\quad = X(Y(\alpha(Z))) - X(\alpha([Y,Z])) - Y(\alpha([X,Z])) + \alpha([Y,[X,Z]]) - \\ &\quad\quad - Y(X(\alpha(Z))) + Y(\alpha([X,Z])) + X(\alpha([Y,Z])) - \alpha([X,[Y,Z]]) \\ &\quad = [X,Y](\alpha(Z)) - \alpha([[X,Y],Z]) = (\mathcal{L}_{[X,Y]}\alpha)(Z).\end{aligned}$$

Next we show that if the statement is true when the Lie derivatives are evaluated on the tensor fields S_1, S_2, or on the differential forms ω and η, then it also holds on $S_1 \otimes S_2$ or on $\omega \wedge \eta$ respectively. We check only for the tensor product. The wedge product can be treated in the same way.

$$\begin{aligned}(\mathcal{L}_{\lambda X+\mu Y})(S_1 \otimes S_2) &= (\mathcal{L}_{\lambda X+\mu Y})(S_1) \otimes S_2 + S_1 \otimes (\mathcal{L}_{\lambda X+\mu Y})(S_2) \\ &= (\lambda\mathcal{L}_X + \mu\mathcal{L}_Y)(S_1) \otimes S_2 + S_1 \otimes (\lambda\mathcal{L}_X + \mu\mathcal{L}_Y)(S_2) \\ &= \lambda(\mathcal{L}_X S_1) \otimes S_2 + \mu(\mathcal{L}_Y S_1) \otimes S_2 + S_1 \otimes \lambda(\mathcal{L}_X S_2) \\ &\quad + S_1 \otimes \mu(\mathcal{L}_Y S_2) \\ &= (\lambda\mathcal{L}_X + \mu\mathcal{L}_Y)(S_1 \otimes S_2).\end{aligned}$$

As for the commutator,

$$\begin{aligned}
&(\mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X)(S_1 \otimes S_2) \\
&\quad = \mathcal{L}_X(\mathcal{L}_Y(S_1) \otimes S_2 + S_1 \otimes \mathcal{L}_Y S_2) - \mathcal{L}_Y(\mathcal{L}_X(S_1) \otimes S_2 + S_1 \otimes \mathcal{L}_X S_2) \\
&\quad = \mathcal{L}_X(\mathcal{L}_Y(S_1)) \otimes S_2 + \mathcal{L}_Y S_1 \otimes \mathcal{L}_X S_2 + \mathcal{L}_X S_1 \otimes \mathcal{L}_Y S_2 \\
&\qquad + S_1 \otimes \mathcal{L}_X(\mathcal{L}_Y(S_2)) - \mathcal{L}_Y(\mathcal{L}_X(S_1)) \otimes S_2 - \mathcal{L}_X S_1 \otimes \mathcal{L}_Y S_2 \\
&\qquad - \mathcal{L}_Y S_1 \otimes \mathcal{L}_X S_2 - S_1 \otimes \mathcal{L}_Y(\mathcal{L}_X(S_2)) \\
&\quad = (\mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X)(S_1)) \otimes S_2 + S_1 \otimes (\mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X)(S_2)) \\
&\quad = (\mathcal{L}_{[X,Y]} S_1)) \otimes S_2 + S_1 \otimes (\mathcal{L}_{[X,Y]} S_2) = \mathcal{L}_{[X,Y]}(S_1 \otimes S_2).
\end{aligned}$$

It is easy to see that if the identities we want to show hold on two tensor fields or differential forms of the same type, then the identity holds also on their sum. Since locally, e.g. in the domain of any chart, tensor fields can be written as sums of tensor products of vector fields and 1-forms, and similarly, differential forms can be written as sums of wedge-products of 1-forms, the proposition is proved. □

Definition 4.5.8. We say that *(the flow of) a vector field X leaves a tensor field S invariant* if $\Phi_t^*(S) = S|_{U_t}$ for all $t \in \mathbb{R}$, where $\Phi_t \colon U_t \to U_{-t}$ $(t \in \mathbb{R})$ is the flow generated by X. ※

Proposition 4.5.9. *The flow of a vector field X leaves the tensor field S invariant if and only if $\mathcal{L}_X S = 0$.*

Proof. If the flow leaves S invariant, then

$$\mathcal{L}_X S = \lim_{t\to 0} \frac{\Phi_t^* S - S}{t} = \lim_{t \to 0} \frac{S - S}{t} = 0.$$

To prove the converse, assume that $\mathcal{L}_X S = 0$. Denote by S_t the one-parameter family of tensor fields $S_t = \Phi_t^* S$. $S_0 = S$ and for each $p \in M$, $S_t(p)$ is defined on an open interval around 0, so to prove the statement, it suffices to check that $\frac{d}{dt} S_t \equiv 0$. However, this is true since

$$\lim_{t \to t_0} \frac{S_t - S_{t_0}}{t - t_0} = \lim_{h \to 0} \Phi_{t_0}^* \left(\frac{\Phi_h^* S - S}{h} \right) = \Phi_{t_0}^*(\mathcal{L}_X S) = 0. \qquad \square$$

Corollary 4.5.10. *Vector fields leaving a tensor field invariant form a Lie subalgebra of the Lie algebra of all vector fields.*

Proof. Indeed, if X and Y leave S invariant, then $\mathcal{L}_{\lambda X + \mu Y} S = \lambda \mathcal{L}_X S + \mu \mathcal{L}_Y S = 0$ and

$$\mathcal{L}_{[X,Y]} S = \mathcal{L}_X(\mathcal{L}_Y S) - \mathcal{L}_Y(\mathcal{L}_X S) = 0. \qquad \square$$

4.6 Differential Forms

A differential k-form ω on a manifold M assigns to each point $p \in M$ an element of $\Lambda^k(T_pM)^*$. As elements of $\Lambda^k V^*$ for a linear space V can be identified with the linear space $A^k(V)$ of alternating k-linear functions, we can evaluate $\omega(p)$ on any k tangent vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in T_pM$. In particular, if we are given k smooth vector fields $X_1, \dots, X_k$ on M, then we can evaluate ω on $X_1, \dots, X_k$ pointwise and get a smooth function on M. This function will be denoted by $\omega(X_1, \dots, X_k)$.

The pointwise wedge product of a differential k-form $\omega \in \Omega^k(M)$ and a differential l-form $\eta \in \Omega^l(M)$ is a differential $(k+l)$-form. Wedge product of differential forms is not a commutative operation. Instead of that it satisfies the *supercommutativity* relation

$$\omega \wedge \eta = (-1)^{k \cdot l} \eta \wedge \omega.$$

4.6.1 Interior Product by a Vector Field

Definition 4.6.1. The *interior product of a smooth vector field X and a differential k-form* ω on a differentiable manifold M is a differential $(k-1)$-form $\iota_X \omega$. It is defined by the property that

$$\iota_X \omega(X_1, \dots, X_{k-1}) = \omega(X, X_1, \dots, X_{k-1}).$$

When $k = 0$, we set $\iota_X \omega = 0$. ※

Interior product with the vector field X can be thought of as a linear map from $\Omega^*(M) = \bigoplus_{k=0}^{\dim M} \Omega^k(M)$ into itself. This is a *degree -1* map, as it decreases the degree of any form by 1.

Definition 4.6.2. In general, a linear map $L\colon \Omega^*(M) \to \Omega^*(M)$ is a *degree d* map, if $L(\Omega^k(M)) \subset \Omega^{k+d}(M)$ for all k. ※

It is interesting that although interior product has a linear algebraic character, it satisfies a Leibniz-type rule.

Definition 4.6.3. A linear map $L\colon \Omega^*(M) \to \Omega^*(M)$ is called a *superderivation* if for any $\omega \in \Omega^k(M)$ and $\eta \in \Omega^l(M)$, we have

$$L(\omega \wedge \eta) = L(\omega) \wedge \eta + (-1)^k \omega \wedge L(\eta).$$ ※

Proposition 4.6.4. *Interior product with the vector field X is a superderivation.*

Proof. Both sides are bilinear functions of ω and η, therefore, it is enough to check the graded version of the Leibniz rule for forms that generate $\Omega^k(M)$ and $\Omega^l(M)$ additively. It is also enough to show the graded Leibniz identity on domains of charts, since any point is contained in the domain of a chart. In the domain of a chart, any k-form can be written as the sum of wedge products of k-tuples of 1-forms, so it is enough to show the graded Leibniz rule for the special case when both ω and η are wedge products of 1-forms. Let $l_1, \dots, l_s$ be 1-forms. Then

$$\iota_X(l_1 \wedge \dots \wedge l_s)(X_1, \dots, X_{s-1}) = \det \begin{pmatrix} l_1(X) & l_1(X_1) & \dots & l_1(X_{s-1}) \\ l_2(X) & l_2(X_1) & \dots & l_2(X_{s-1}) \\ \vdots & & \ddots & \vdots \\ l_s(X) & l_s(X_1) & \dots & l_s(X_{s-1}) \end{pmatrix}.$$

Computing the determinant by Laplace expansion with respect to the first column we obtain

$$\begin{aligned} &\iota_X(l_1 \wedge \dots \wedge l_s)(X_1, \dots, X_{s-1}) \\ &\qquad = \sum_{i=1}^{s} (-1)^{i+1} l_i(X) \cdot (l_1 \wedge \dots \wedge \widehat{l_i} \wedge \dots \wedge l_s)(X_1, \dots, X_{s-1}), \end{aligned}$$

where the hat over l_i means that it is omitted. This means that

$$\iota_X(l_1 \wedge \dots \wedge l_s) = \sum_{i=1}^{s} (-1)^{i+1} l_i(X) \cdot (l_1 \wedge \dots \wedge \widehat{l_i} \wedge \dots \wedge l_s).$$

Consider the special case of the statement, when $\omega = l_1 \wedge \dots \wedge l_k$ and $\eta = l_{k+1} \wedge \dots \wedge l_{k+l}$. Then

$$\begin{aligned} \iota_X(\omega \wedge \eta) = \iota_X(l_1 \wedge \dots \wedge l_{k+l}) &= \sum_{i=1}^{k+l} (-1)^{i+1} l_i(X) \cdot (l_1 \wedge \dots \wedge \widehat{l_i} \wedge \dots \wedge l_{k+l}) \\ &= \sum_{i=1}^{k} (-1)^{i+1} l_i(X) \cdot (l_1 \wedge \dots \wedge \widehat{l_i} \wedge \dots \wedge l_{k+l}) + \\ &\quad + \sum_{i=1}^{l} (-1)^{k+i+1} l_{k+i}(X) \cdot (l_1 \wedge \dots \wedge \widehat{l_{k+i}} \wedge \dots \wedge l_{k+l}) \\ &= \iota_X(\omega) \wedge \eta + (-1)^k \omega \wedge \iota_X(\eta). \end{aligned}$$

By the introductory arguments, this special case implies the general one. □

Exercise 4.6.5. Show that for any two vector fields X an Y, we have $\iota_X \circ \iota_Y = -\iota_Y \circ \iota_X$. ¤

Exercise 4.6.6. Show that the Lie derivation of differential forms and the interior product with a vector field satisfies the commutation relation $\mathcal{L}_X \circ \iota_Y - \iota_Y \circ \mathcal{L}_X = \iota_{[X,Y]}$. ¤

4.6.2 Exterior Differentiation

We can associate to each smooth function a differential 1-form, its differential.

Definition 4.6.7. The *differential df of a smooth function* $f \in \Omega^0(M)$ is the differential 1-form $df \in \Omega^1(M)$, the evaluation of which on a vector field X is given by $df(X) = X(f)$. ※

Remark. If $\phi = (x^1, \dots, x^n) \colon U \to \mathbb{R}^n$ is a chart, then the 1-forms determined by the pointwise dual basis of the basis vector fields $\partial_1^\phi, \dots, \partial_n^\phi$ were denoted by $dx^1, \dots, dx^n$. The notation is motivated by the fact that these 1-forms can also be obtained as the differentials of the coordinate functions. Indeed,

$$dx^i(\partial_j^\phi) = \partial_j^\phi(x^i) = \delta_j^i.$$

Our goal is to extend the differential d from functions to all differential forms.

Theorem 4.6.8. *There is a unique superderivation $d \colon \Omega^*(M) \to \Omega^*(M)$ of degree 1, which has the following properties.*

- *For $f \in \Omega^0(M)$, df is the differential of the function f.*
- *$d \circ d = 0$.*

The superderivation d defined uniquely by the theorem is called the *exterior differentiation of differential forms.*

Proof. To prove *uniqueness*, first observe that if two k forms ω_1, ω_1 coincide on an open neighborhood V of a point p, then $d\omega_1$ and $d\omega_2$ must coincide at p. Indeed, choose a smooth bump function $h \in \Omega^0(M)$ such that $\operatorname{supp} h \subset V$ and $h \equiv 1$ on a neighborhood of p. Then $h(\omega_1 - \omega_2) = 0$, therefore,

$$0 = d(h(\omega_1 - \omega_2)) = dh \wedge (\omega_1 - \omega_2) + h(d\omega_1 - d\omega_2).$$

Evaluating at p gives $(d\omega_1)(p) = (d\omega_1)(p)$.
Let p be an arbitrary point in M and choose a chart $\phi = (x^1, \dots, x^n) \colon U \to \mathbb{R}^n$ around p. Fix also a smooth function $g \in \Omega^0(M)$ such that $\operatorname{supp} g \subset U$ and $g|V \equiv 1$ on an open neighborhood $V \subset U$ of p. Introduce the notation that for a smooth function f defined on U, denote by $\tilde{f}$ the smooth function on M defined by

$$\tilde{f}(q) = \begin{cases} f(q)g(q), & \text{if } q \in U, \\ 0, & \text{otherwise.} \end{cases}$$

Let ω be an arbitrary differential k-form. We can write $\omega|U$ uniquely as

$$\omega|U = \sum_{1\le i_1<\dots<i_k\le n} \omega_{i_1\dots i_k} dx^{i_1}\wedge\dots\wedge dx^{i_k}.$$

Consider the differential form $\tilde{\omega}\in\Omega^k(M)$ defined by

$$\tilde{\omega} = \sum_{1\le i_1<\dots<i_k\le n} \tilde{\omega}_{i_1\dots i_k} \widetilde{dx^{i_1}}\wedge\dots\wedge \widetilde{dx^{i_k}}.$$

The forms ω and $\tilde{\omega}$ coincide on V and the computation of d on $\tilde{\omega}$ is uniquely dictated by the properties of d, thus,

$$(d\omega)(p) = (d\tilde{\omega})(p) = \left(\sum_{1\le i_1<\dots<i_k\le n} d\tilde{\omega}_{i_1\dots i_k}\wedge \widetilde{dx^{i_1}}\wedge\dots\wedge \widetilde{dx^{i_k}}\right)(p).$$

Since by this method we can compute the value of $d\omega$ at any point p, d is uniquely defined by its properties.
For the existence, we first construct for each chart $\phi=(x^1,\dots,x^n)\colon U\to\mathbb{R}^n$ an exterior differentiation $d^\phi\colon\Omega^*(U)\to\Omega^*(U)$ on U. For $\omega\in\Omega^*(U)$ write ω as

$$\omega = \sum_{1\le i_1<\dots<i_k\le n} \omega_{i_1\dots i_k} dx^{i_1}\wedge\dots\wedge dx^{i_k},$$

and set

$$d^\phi\omega = \sum_{1\le i_1<\dots<i_k\le n} d\omega_{i_1\dots i_k}\wedge dx^{i_1}\wedge\dots\wedge dx^{i_k}.$$

Let us check that d^ϕ satisfies all the requirements. It is obviously linear and increases the degree of forms by 1, and coincides with d on smooth functions. It is also a superderivation. Indeed, if $\eta\in\Omega^l(M)$ has the decomposition

$$\eta = \sum_{1\le j_1<\dots<j_l\le n} \eta_{j_1\dots j_l} dx^{j_1}\wedge\dots\wedge dx^{j_l},$$

then

$$\begin{aligned}
&d^\phi(\omega\wedge\eta) = \\
&d^\phi\Big(\sum_{1\le i_1<\dots<i_k\le n}\ \sum_{1\le j_1<\dots<j_l\le n} \omega_{i_1\dots i_k}\cdot\eta_{j_1\dots j_l} dx^{i_1}\wedge\dots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\dots\wedge dx^{j_l}\Big)\\
&= \sum_{1\le i_1<\dots<i_k\le n}\ \sum_{1\le j_1<\dots<j_l\le n} d(\omega_{i_1\dots i_k}\cdot\eta_{j_1\dots j_l})\wedge dx^{i_1}\wedge\dots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\dots\wedge dx^{j_l}
\end{aligned}$$

For smooth function f and g, the Leibniz-type formula $d(f\cdot g)=df\cdot g+f\cdot dg$ holds, since for any smooth vector field X,

$$d(f\cdot g)(X)=X(f\cdot g)=X(f)\cdot g+f\cdot X(g)=(df\cdot g+f\cdot dg)(X).$$

Applying this special case to the general one, we obtain

$$\begin{aligned}&d^\phi(\omega\wedge\eta)=\\&\sum_{1\le i_1<\dots<i_k\le n}\ \sum_{1\le j_1<\dots<j_l\le n}\eta_{j_1\dots j_l}\cdot d\omega_{i_1\dots i_k}\wedge dx^{i_1}\wedge\dots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\dots\wedge dx^{j_l}\\&+\sum_{1\le i_1<\dots<i_k\le n}\ \sum_{1\le j_1<\dots<j_l\le n}\omega_{i_1\dots i_k}\cdot d\eta_{j_1\dots j_l}\wedge dx^{i_1}\wedge\dots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\dots\wedge dx^{j_l}\\&=d\omega\wedge\eta+(-1)^k\omega\wedge d\eta.\end{aligned}$$

It remains to show that $d^\phi\circ d^\phi=0$. As d^ϕ is a superderivation,

$$\begin{aligned}&d^\phi(d^\phi(\omega))=\\&\sum_{1\le i_1<\dots<i_k\le n}\left(d^\phi(d\omega_{i_1\dots i_k})\wedge dx^{i_1}\wedge\dots\wedge dx^{i_k}-d\omega_{i_1\dots i_k}\wedge d^\phi(dx^{i_1}\wedge\dots\wedge dx^{i_k})\right)\\&=\sum_{1\le i_1<\dots<i_k\le n}d^\phi(d\omega_{i_1\dots i_k})\wedge dx^{i_1}\wedge\dots\wedge dx^{i_k},\end{aligned}$$

it is enough to prove that $d^\phi(d(f))=0$ for any smooth function f. Since $dx^1,\dots,dx^n$ is the pointwise dual basis of $\partial_1^\phi,\dots,\partial_n^\phi$, df can be decomposed as a linear combination of the 1-forms dx^i as

$$df=\sum_{i=1}^n(\partial_i^\phi f)dx^i.$$

From this formula,

$$\begin{aligned}d^\phi(df)&=\sum_{i=1}^n d(\partial_i^\phi f)\wedge dx^i=\sum_{i=1}^n\sum_{j=1}^n(\partial_j^\phi\partial_i^\phi f)dx^j\wedge dx^i\\&=\sum_{1\le i<j\le n}(\partial_i^\phi\partial_j^\phi f-\partial_j^\phi\partial_i^\phi f)dx^i\wedge dx^j=0.\end{aligned}$$

If we consider two charts, $\phi\colon U\to\mathbb{R}^n$ and $\psi\colon V\to\mathbb{R}^n$, then the restrictions of both ϕ and ψ onto $U\cap V$ give an exterior differentiation on $U\cap V$. But the uniqueness part of the theorem implies that these exterior derivations must be the same. This means that if ω is a globally defined k-form on M, then the forms $d^\phi(\omega|U)$ taken for all charts $\phi\colon U\to\mathbb{R}^n$ glue together to a global $(k+1)$-form $d\omega$ on M. The map $\omega\mapsto d\omega$ is an exterior derivation, since all the properties can be checked locally, using a local coordinate system. □

In the rest of this section we prove some useful formulae involving the exterior differentiation.

Proposition 4.6.9. *If $f\colon M \to N$ is a smooth map, $\omega \in \Omega^k(N)$ is a differential k-form on N, then $d(f^*(\omega)) = f^*(d\omega)$.*

Proof. The statement is true for smooth functions $h \in \Omega^0(M)$, because for any smooth vector field X on M and for any $p \in M$, we have

$$\begin{aligned} d(f^*(h))(X_p) &= X_p(f^*(h)) = X_p(h \circ f) = (T_p f(X_p))(h) \\ &= dh(T_p f(X_p)) = (f^*(dh))(X_p). \end{aligned}$$

It is also clear that a wedge product can be pulled back componentwise, i.e., $f^*(\omega \wedge \eta) = f^*(\omega) \wedge f^*(\eta)$.
If $p \in M$ is an arbitrary point, then we can introduce a local coordinate system $\phi = (x^1, \dots, x^n)\colon U \to \mathbb{R}^n$ around $f(p)$, where $n = \dim N(!)$. By continuity of f, we can also find an open neighborhood V of p, such that $f(V) \subset U$. Then for any $\omega \in \Omega^k(N)$, we can write

$$\omega|_U = \sum_{1 \le i_1 < \dots < i_k \le n} \omega_{i_1 \dots i_k} dx^{i_1} \wedge \dots \wedge dx^{i_k},$$

hence

$$\begin{aligned} (f^*\omega)|_V &= \sum_{1 \le i_1 < \dots < i_k \le n} f^*(\omega_{i_1 \dots i_k}) f^*(dx^{i_1}) \wedge \dots \wedge f^*(dx^{i_k}) \\ &= \sum_{1 \le i_1 < \dots < i_k \le n} (\omega_{i_1 \dots i_k} \circ f) d(x^{i_1} \circ f) \wedge \dots \wedge d(x^{i_k} \circ f). \end{aligned}$$

and

$$(f^*(d\omega))|_V = \sum_{1 \le i_1 < \dots < i_k \le n} d(\omega_{i_1 \dots i_k} \circ f) \wedge d(x^{i_1} \circ f) \wedge \dots \wedge d(x^{i_k} \circ f) = d(f^*(\omega)).$$

□

Proposition 4.6.10. *Lie derivation and exterior derivation of differential forms commute, that is*

$$\mathcal{L}_X \circ d = d \circ \mathcal{L}_X.$$

Proof. Using the usual notation, for any differential form ω, we have

$$\mathcal{L}_x(d\omega) = \frac{d}{dt}(\Phi_t^*(d\omega))|_{t=0} = \frac{d}{dt} d(\Phi_t^*(\omega))|_{t=0}.$$

In the last expression, we can change the order of differentiation with respect to the parameter t and the exterior differentiation by Young's theorem. Indeed, if we write the one-parameter family of forms $\omega_t = \Phi_t^*\omega$ using local coordinates as

$$\omega_t(p) = \sum_{1\leq i_1<\cdots<i_k\leq n} \omega_{i_1\ldots i_k}(p,t)dx^{i_1}\wedge\cdots\wedge dx^{i_k},$$

then

$$\begin{aligned}\frac{d}{dt}d(\omega_t)(p) &= \sum_{1\leq i_1<\cdots<i_k\leq n}\sum_{j=1}^{n}\partial_t\partial_j^{\phi}\omega_{i_1\ldots i_k}(p,t)dx^j\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\\ &= \sum_{1\leq i_1<\cdots<i_k\leq n}\sum_{j=1}^{n}\partial_j^{\phi}\partial_t\omega_{i_1\ldots i_k}(p,t)dx^j\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\\ &= d\left(\frac{d}{dt}\omega_t\right).\end{aligned}$$

Consequently, $\mathcal{L}_x(d\omega) = \frac{d}{dt}d(\Phi_t^*(\omega))|_{t=0} = d\left(\frac{d}{dt}\Phi_t^*(\omega)|_{t=0}\right) = d(\mathcal{L}_x\omega)$. □

Proposition 4.6.11 (Cartan's Formula)**.** *Exterior differentiation and interior product with a vector field are related to Lie derivation of differential forms by the formula*

$$\mathcal{L}_X = \iota_X\circ d + d\circ\iota_X.$$

Proof. The formula is true for functions $f\in\Omega^0(M)$, since

$$\mathcal{L}_X(f) = X(f) = df(X) = \iota_X(df) = \iota_X(df) + d(\iota_X f).$$

In the last step we used that ι_X is 0 on smooth functions.
If the formula holds for a form ω, then it holds for $d\omega$. Indeed,

$$\mathcal{L}_X(d\omega) = d(\mathcal{L}_X\omega) = d\circ\iota_X\circ d(\omega) + d\circ d\circ\iota_X(\omega) = d\circ\iota_X\circ d(\omega),$$

on the other hand,

$$(\iota_X\circ d + d\circ\iota_X)(d\omega) = \iota_X\circ d\circ d(\omega) + d\circ\iota_X\circ d(\omega) = d\circ\iota_X\circ d(\omega).$$

If the identity holds for the forms $\omega\in\Omega^k(M)$ and $\eta\in\Omega^l(M)$, then it holds for their wedge product as well. Using the Leibniz rule for $\mathcal{L}_X$, we get

$$\begin{aligned}\mathcal{L}_X(\omega\wedge\eta) &= \mathcal{L}_X(\omega)\wedge\eta + \omega\wedge\mathcal{L}_X(\eta)\\ &= \iota_X(d\omega)\wedge\eta + d(\iota_X\omega)\wedge\eta + \omega\wedge\iota_X(d\eta) + \omega\wedge d(\iota_X\eta).\end{aligned}$$

Computing the other side

$$\begin{aligned}
&(\iota_X \circ d + d \circ \iota_X)(\omega \wedge \eta) \\
&= \iota_X(d\omega \wedge \eta + (-1)^k \omega \wedge d\eta) + d(\iota_X \omega \wedge \eta + (-1)^k \omega \wedge \iota_X \eta) \\
&= \iota_X(d\omega) \wedge \eta + (-1)^{k+1} d\omega \wedge \iota_X \eta + (-1)^k \iota_X \omega \wedge d\eta + (-1)^{2k} \omega \wedge \iota_X(d\eta) + \\
&\quad + d(\iota_X \omega) \wedge \eta + (-1)^{k-1} \iota_X \omega \wedge d\eta + (-1)^k d\omega \wedge \iota_X \eta + (-1)^{2k} \omega \wedge d(\iota_X \eta) \\
&= \iota_X(d\omega) \wedge \eta + d(\iota_X \omega) \wedge \eta + \omega \wedge \iota_X(d\eta) + \omega \wedge d(\iota_X \eta).
\end{aligned}$$

Since locally, in the domain of a chart, any differential form can be written as the sum of wedge products of functions and differentials of functions, the proposition is true. □

The formula below can also be used as a coordinate free definition of the exterior differential of a differential form.

Proposition 4.6.12. *If $\omega \in \Omega^k(M)$, and $X_0, \dots, X_k$ are smooth vector fields on M, then*

$$\begin{aligned}
d\omega(X_0, \dots, X_k) = &\sum_{i=0}^{k} (-1)^i X_i(\omega(X_0, \dots, \hat{X}_i, \dots, X_k)) + \\
&+ \sum_{0 \le i < j \le k} (-1)^{i+j} \omega([X_i, X_j], X_0, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_k).
\end{aligned}$$

Proof. We prove the formula by induction on k. For a smooth function $f \in \Omega^0(M)$, the second sum is empty, and $df(X) = X(f)$, so the formula holds. Assume that the formula is true for $(k-1)$-forms. Using Cartan's formula

$$\begin{aligned}
d\omega(X_0, \dots, X_k) &= (\iota_{X_0} d\omega)(X_1 \dots, X_k) \\
&= (\mathcal{L}_{X_0} \omega)(X_1 \dots, X_k) - d(\iota_{X_0} \omega)(X_1 \dots, X_k).
\end{aligned}$$

$(\mathcal{L}_{X_0} \omega)(X_1 \dots, X_k)$ can be expressed by a Leibniz-type rule as follows

$$\begin{aligned}
(\mathcal{L}_{X_0} \omega)(X_1 \dots, X_k) &= \mathcal{L}_{X_0}(\omega(X_1 \dots, X_k)) - \sum_{i=1}^{k} \omega(X_1, \dots, \mathcal{L}_{X_0} X_i, \dots, X_k) \\
&= X_0(\omega(X_1 \dots, X_k)) + \sum_{i=1}^{k} (-1)^k \omega([X_0, X_i], X_1, \dots, \hat{X}_i, \dots, X_k).
\end{aligned}$$

The induction hypothesis applied to the $(k-1)$-form $\iota_{X_0}\omega$ provides

$$\begin{aligned}-d(\iota_{X_0}\omega)(X_1\dots,X_k) &= \sum_{i=1}^{k}(-1)^i X_i(\iota_{X_0}\omega(X_1,\dots,\hat{X}_i,\dots,X_k))+\\ &\quad+\sum_{1\le i<j\le k}(-1)^{i+j+1}\iota_{X_0}\omega([X_i,X_j],X_1,\dots,\hat{X}_i,\dots,\hat{X}_j,\dots,X_k)\\ &=\sum_{i=1}^{k}(-1)^i X_i(\omega(X_0,\dots,\hat{X}_i,\dots,X_k))+\\ &\quad+\sum_{1\le i<j\le k}(-1)^{i+j}\omega([X_i,X_j],X_0,\dots,\hat{X}_i,\dots,\hat{X}_j,\dots,X_k).\end{aligned}$$

Adding the last two equations we obtain the statement for ω. □

4.6.3 De Rham Cohomology

Definition 4.6.13. A differential form ω is said to be *closed* if $d\omega = 0$. It is *exact* if there is a differential form η such that $\omega = d\eta$. ※

Exact differential forms are closed because of $d \circ d = 0$, but the converse is not always true. De Rham cohomology spaces measure in some sense how much the converse statement fails to be true. Let $k \geq 0$ be an integer, and consider the spaces of closed and exact k-forms on M

$$Z^k(M) = \Omega^k(M) \cap (\ker d) \qquad \text{and} \qquad B^k(M) = \Omega^k(M) \cap (\operatorname{im} d).$$

Definition 4.6.14. The *k-th de Rham cohomology space* is the factor space

$$H^k_{dR}(M) = Z^k(M)/B^k(M).$$ ※

The kth de Rham cohomology space is a linear space. Its elements are equivalence classes of closed differential k-forms, where two forms represent the same equivalence class, also called *cohomology class* if and only if their difference is exact.

It is useful to consider the direct sum $H^*_{dR}(M) = \bigoplus_{k=0}^{\infty} H^k_{dR}(M)$ of all de Rham cohomology spaces, because this linear space has also a multiplicative structure induced by the wedge product of forms. Let $[\omega]$ and $[\eta]$ be two cohomology classes represented by the closed forms ω and η. Then $\omega\wedge\eta$ is a closed form and the cohomology class $[\omega\wedge\eta]$ depends only on the cohomology classes $[\omega]$ and $[\eta]$. Indeed, if we choose other representatives $\omega+d\alpha$ and $\eta+d\beta$ from these cohomology classes, then their wedge product

$$(\omega+d\alpha)\wedge(\eta+d\beta) = \omega\wedge\eta + d\big(\alpha\wedge(\eta+d\beta)+(-1)^{\deg\omega}\omega\wedge\beta\big)$$

is in the same cohomology class as $\omega\wedge\eta$.

Definition 4.6.15. The linear space $H^*_{dR}(M)$ together with its multiplicative structure defined by
$$[\omega][\eta] = [\omega \wedge \eta]$$
is called the *de Rham cohomology algebra of M*. ※

Multiplication of cohomology classes is associative, distributive with respect to addition. It has a unit element, represented by the constant 1 function in $H^0_{dR}(M)$. Instead of commutativity, it is supercommutative, that is, for $[\omega] \in H^k_{dR}(M)$ and $[\eta] \in H^l_{dR}(M)$, we have
$$[\omega][\eta] = (-1)^{kl}[\eta][\omega].$$

Every smooth map $f\colon M \to N$ between two manifolds induce an algebra homomorphism $f^*\colon H^*_{dR}(N) \to H^*_{dR}(M)$, which maps $H^k_{dR}(N)$ to $H^k_{dR}(M)$. First observe, that if $\omega \in Z^k(N)$ is a closed k-form on N, then its pull back $f^*\omega$ is also closed, since $d(f^*\omega) = f^*(d\omega) = 0$. The cohomology class of $f^*\omega$ depends only on the cohomology class of ω since $f^*(\omega + d\alpha) = f^*\omega + d(f^*\alpha)$. Thus, $f^*\colon H^*_{dR}(N) \to H^*_{dR}(M)$ is properly defined by $f^*[\omega] = [f^*\omega]$ and it is an algebra homomorphism because f^* on differential forms is linear and satisfies $f^*(\omega \wedge \eta) = f^*(\omega) \wedge f^*(\eta)$.
We can also say that H^k_{dR} is a contravariant functor from the category of smooth manifolds to the category of linear spaces, while H^*_{dR} is a contravariant functor from the category of smooth manifolds to the category of algebras. Diffeomorphic manifolds have isomorphic de Rham cohomology spaces and de Rham algebra. In fact, de Rham cohomology spaces are homotopy invariant as well. We shall prove here the seemingly weaker smooth homotopy invariance.

Definition 4.6.16. Two smooth maps $f_0\colon M \to N$ and $f_1\colon M \to N$ between the manifolds M and N are said to be *piecewise smoothly homotopic* if there is a continuous map $H\colon M \times [0,1] \to N$ such that $f_0(p) = H(p,0)$ and $f_1(p) = H(p,1)$ for all $p \in M$, and the interval has a subdivision $0 = t_0 < t_1 < \cdots < t_N = 1$, for which the restriction of H onto $M \times [t_{i-1}, t_i]$ is smooth for all $1 \leq i \leq N$. ※

Theorem 4.6.17. *If two smooth maps $f_0\colon M \to N$ and $f_1\colon M \to N$ are piecewise smoothly homotopic, then they induce the same homomorphism $f_0^* = f_1^*\colon H^*_{dR}(N) \to H^*_{dR}(M)$.*

Proof. We may suppose that the homotopy map $H\colon M \times [0,1] \to N$ connecting f_0 and f_1 is smooth, as any piecewise smooth homotopy is a concatenation of smooth ones. For $t \in [0,1]$, define the embedding $i_t\colon M \to M \times [0,1]$ by $i_t(p) = (p,t)$. Then $f_0 = H \circ i_0$ and $f_1 = H \circ i_1$. Set in general $f_t = H \circ i_t$.

We have to show that if $\omega \in \Omega^k(N)$ is a closed form, then $f_1^*\omega - f_0^*\omega$ is exact. For this purpose, set $\eta = H^*\omega$. Let X be the vector field on $M \times [0,1]$ the flow Φ_τ of which acts on $M \times [0,1]$ by the formula $\Phi_\tau(p,t) = (p, t+\tau)$. Then we have

$$\begin{aligned} f_1^*\omega - f_0^*\omega = i_1^*\eta - i_0^*\eta &= \int_0^1 \frac{d}{dt} i_t^*\eta dt \\ &= \int_0^1 i_t^*(\mathcal{L}_X\eta)dt = \int_0^1 i_t^*(\iota_X(d\eta) + d(\iota_X\eta))dt \\ &= \int_0^1 i_t^*(d(\iota_X\eta))dt = \int_0^1 d(i_t^*(\iota_X\eta))dt = d\left(\int_0^1 i_t^*(\iota_X\eta)dt\right) \end{aligned}$$

as we wanted to show. □

Definition 4.6.18. Two smooth manifolds are *(piecewise smoothly) homotopy equivalent* if there are smooth maps $f\colon M \to N$ and $g\colon N \to M$ such that $f \circ g$ is piecewise smoothly homotopic to id_N and $g \circ f$ is piecewise smoothly homotopic to id_M. ※

Corollary 4.6.19. *Piecewise smoothly homotopy equivalent smooth manifolds have isomorphic de Rham cohomology spaces and algebras.*

Proof. Let f and g be the maps from the definition of homotopy equivalence. Then the algebra homomorphisms $f^*\colon H^*_{dR}(N) \to H^*_{dR}(M)$ and $g^*\colon H^*_{dR}(M) \to H^*_{dR}(N)$ have the property that $f^* \circ g^* = (g \circ f)^* = (\mathrm{id}_M)^* = \mathrm{id}_{H^*_{dR}(M)}$ and similarly, $g^* \circ f^* = \mathrm{id}_{H^*_{dR}(N)}$, which means that f^* and g^* are the inverses of one another, in particular, they are isomorphisms. □

Recall that a subset U of $\mathbb{R}^n$ is called *star-shaped* if U has a point $\mathbf{p}$ such that the segment $[\mathbf{p}, \mathbf{q}] = \{t\mathbf{p} + (1-t)\mathbf{q} \mid t \in [0,1]\}$ is contained in U for all $\mathbf{q} \in U$.

Corollary 4.6.20 (Poincaré Lemma)**.** *If $U \subset \mathbb{R}^n$ is a star-shaped open subset of $\mathbb{R}^n$, then every closed $k(\geq 1)$-form on U is exact.*

Proof. In terms of the de Rham cohomology spaces, Poincaré Lemma claims that $H^k_{dR}(U) = 0$ for $k \geq 1$. Choose a point $\mathbf{p} \in U$ from which all other points of U "can be seen", and let V be the 0-dimensional manifold consisting the single point $\mathbf{p}$. We claim the U and V are smoothly homotopy equivalent. Indeed, let $f\colon V \to U$ be the embedding of V into U, $g\colon U \to V$ be the constant map sending each point to $\mathbf{p}$. Then $g \circ f = \mathrm{id}_V$. The composition $f \circ g$ is not the identity of U but it is smoothly homotopic to it as it is shown by the smooth homotopy map $H\colon U \times [0,1] \to U$, $H(\mathbf{q}, t) = t\mathbf{p} + (1-t)\mathbf{q}$.

The smooth homotopy equivalence implies that $H^k_{dR}(U) \cong H^k_{dR}(V)$ for all k. However, since V is 0-dimensional, there are no non-zero differential k-forms on it for $k \geq 1$, therefore $H^k_{dR}(V) = 0$ for $k \geq 1$. □

Exercise 4.6.21. Show that a function f is a closed differential 0-form if and only if it is locally constant. From this fact, prove that the dimension of the cohomology space $H^0_{dR}(M)$ is the number of connected components of M. In particular, for a star-shaped open domain U in $\mathbb{R}^n$, $H^0_{dR}(U) \cong \mathbb{R}$. ¤

Remark. Applying some theorems about the approximation of continuous functions by smooth ones, smoothness conditions in the above theorems can be weakened. The following two approximation theorems are essential. Any continuous map is homotopic to a smooth one, where homotopic means that the homotopy map H connecting the two maps is assumed to be continuous only. Secondly, if two smooth maps are homotopic, then there is also a smooth homotopy between them.
This means that we can introduce the homomorphism $f^*\colon H^*_{dR}(N) \to H^*_{dR}(M)$ for any continuous map $f\colon M \to N$ taking a smooth map $\tilde{f}\colon M \to N$ homotopic to f and setting $f^* = \tilde{f}^*$.
Homotopy equivalence of topological spaces is defined as the notion of piecewise smooth homotopy equivalence replacing the smoothness assumptions by continuity. Using the above approximation theorems, we can also show that if two manifolds are homotopy equivalent then they have isomorphic de Rham cohomology spaces and algebras.
The latter theorem also enables us to speak about the de Rham cohomology spaces and algebras of topological spaces homotopy equivalent to a manifold.

4.7 Integration of Differential Forms

To integrate a function, we need a measure. Since manifolds do not come with a measure, we cannot integrate functions on a manifold, unless there is an additional structure which produces a measure. On the other hand, differential k-forms can be integrated on parameterized k-dimensional submanifolds (more precisely on k-dimensional chains) without any extra structure. We can also integrate differential n-forms on domains of an n-dimensional manifolds, but that integration requires an orientation of the manifold.

4.7.1 Integration on Chains

Definition 4.7.1. The *standard k-dimensional simplex* Δ^k is the simplex in $\mathbb{R}^k$ defined by the inequalities

$$\Delta^k = \left\{(x_1, \dots, x_k) \in \mathbb{R}^n \,\middle|\, x_i \geq 0 \text{ for } i = 1, \dots, k \text{ and } \sum_{i=1}^k x_i \leq 1\right\}.$$ ※

The standard simplex is the convex hull of the affinely independent points $v_0^k = \mathbf{0}$, $v_1^k = \mathbf{e}_1, \dots, v_k^k = \mathbf{e}_k$, where $\mathbf{0}$ is the origin, $\mathbf{e}_1, \dots, \mathbf{e}_k$ is the standard basis of $\mathbb{R}^k$. These points are the vertices of Δ^k.

Definition 4.7.2. A *smooth singular k-dimensional simplex in a manifold M* is a smooth map $\sigma\colon \Delta^k \to M$ from the standard k-dimensional simplex to M. ※

The word *singular* does not mean that σ must be singular. It means that σ is allowed to be singular.

Definition 4.7.3. A *smooth singular k-dimensional chain in a manifold M* is a *formal* linear combination $\lambda_1\sigma_1 + \cdots + \lambda_s\sigma_s$ of smooth singular k-dimensional simplices $\sigma_i\colon \Delta^k \to M$ with real coefficients $\lambda_i \in \mathbb{R}$. Smooth singular k-dimensional chains in M form a linear space which we denote by $C_k(M)$.

※

The standard k-dimensional simplex Δ^k has $(k+1)$ facets. Let the i-th be the convex hull of the vertices $v_0^k, \dots, \widehat{v_i^k}, \dots, v_k^k$. The standard $(k-1)$-dimensional simplex can be mapped onto the ith facet of Δ^k by a unique affine transformation $l_i^k\colon \Delta^{k-1} \to \Delta^k$ which maps the vertices $v_0^{k-1}, \dots, v_{k-1}^{k-1}$ of Δ^{k-1} onto the vertices $v_0^k, \dots, \widehat{v_i^k}, \dots, v_k^k$ of the ith facet of Δ^k in the given order. We can give l_i^k explicitly by the formulae

$$\begin{aligned} l_0^k(x_1, \dots, x_{k-1}) &= (1 - (x_1 + \cdots + x_{k-1}), x_1, \dots, x_{k-1}) \\ l_i^k(x_1, \dots, x_{k-1}) &= (x_1, \dots, x_{i-1}, 0, x_i, \dots, x_{k-1}) \qquad \text{for } 1 \leq i \leq k-1. \end{aligned}$$

If $0 \leq i < j \leq k$, then both $l_j^k \circ l_i^{k-1}$ and $l_i^k \circ l_{j-1}^{k-1}$ are affine maps from Δ^{k-2} onto a $(k-2)$-dimensional face of Δ^k, and both maps the vertices $v_0^{k-2}, \dots, v_{k-2}^{k-2}$ to the vertices $v_0^k, \dots, \widehat{v_i^k}, \dots, \widehat{v_j^k}, \dots, v_k^k$ in an increasing order, consequently,

$$l_j^k \circ l_i^{k-1} = l_i^k \circ l_{j-1}^{k-1}.$$

Definition 4.7.4. The *boundary of a smooth singular k-dimensional simplex* $\sigma\colon \Delta^k \to M$ is the $(k-1)$-dimensional smooth singular chain

$$\partial\sigma = \sum_{i=0}^k (-1)^i \sigma \circ l_i^k.$$

The *boundary of a smooth singular k-dimensional chain* $\sum_{j=1}^{s} \lambda_s \sigma_s$ is the $(k-1)$-dimensional chain

$$\partial\left(\sum_{j=1}^{s} \lambda_s \sigma_s\right) = \sum_{j=1}^{s} \lambda_s \partial\sigma_s.$$ ※

The boundary map is a linear map $\partial\colon C_k(M) \to C_{k-1}(M)$ for all k.

Proposition 4.7.5. *The boundary of the boundary of a chain is* 0*, that is* $\partial \circ \partial = 0$.

Proof. Since the linear space of smooth singular chains is generated by smooth singular simplices and ∂ is linear, it is enough to check the statement for singular simplices. Let $\sigma\colon \Delta^k \to M$ be a smooth singular simplex. Then

$$\partial(\partial\sigma) = \sum_{i=0}^{k} (-1)^j \partial(\sigma \circ l_j^k) = \sum_{j=0}^{k}\sum_{i=0}^{k-1} (-1)^{i+j} \sigma \circ l_j^k \circ l_i^{k-1}.$$

For $i < j$, the compositions $\sigma \circ l_j^k \circ l_i^{k-1} = \sigma \circ l_i^k \circ l_{j-1}^{k-1}$ enter the last sum with opposite signs, they cancel each other. After these cancellations nothing remains. □

The previous proposition gives rise to the definition of smooth singular homology groups of a manifold.

Definition 4.7.6. A smooth singular chain is called a *smooth singular cycle* if its boundary is 0. A smooth singular chain is called a *boundary*, if it is the boundary of a smooth singular chain.
Denote by $Z_k(M)$ and $B_k(M)$ the linear spaces of k dimensional smooth singular cycles and boundaries respectively. Then the *kth smooth singular homology space with real coefficients* is the factor space $H_k(M) = Z_k(M)/B_k(M)$. ※

Now let us define integration of differential forms on singular chains.

Definition 4.7.7. Let $\omega \in \Omega^k(M)$ be a differential k-form on the manifold M, $\sigma\colon \Delta^k \to M$ be a smooth singular k-dimensional simplex in M. Let $(x^1, \dots, x^k)\colon \mathbb{R}^k \to \mathbb{R}^k$ be the identity chart of $\mathbb{R}^k$ and let us write the differential k-form $\sigma^*\omega$ as

$$\sigma^*\omega = f \cdot dx^1 \wedge \cdots \wedge dx^k,$$

where $f\colon \Delta^k \to \mathbb{R}$ is a smooth function. Then we define the *integral of ω over the smooth singular simplex σ* as

$$\int_\sigma \omega = \int_{\Delta^k} f(\mathbf{x})d\mathbf{x},$$

where the second integral is the integral of f over the standard simplex Δ^k with respect to the Lebesgue measure.
The *integral of* ω *over a k-dimensional smooth singular chain* $c = \sum_{i=1}^{s} \lambda_i \sigma_i$ is defined by

$$\int_c \omega = \sum_{i=1}^{s} \lambda_i \int_{\sigma_i} \omega. \qquad ※$$

Stokes' Theorem is the most fundamental integral formula, a far-reaching generalization of the Newton–Leibniz formula. Though it is much more general then the latter one, we shall see that its proof, in fact, reduces to a finite number of Newton–Leibniz formulas.

Theorem 4.7.8 (Stokes' Theorem for Chains)**.** *Let $c \in C_k(M)$ be a k-dimensional smooth singular chain in the manifold M, $\omega \in \Omega^{k-1}(M)$ be a differential $(k-1)$-form. Then*

$$\int_{\partial c} \omega = \int_c d\omega.$$

Proof. Since both integrals depend depend linearly on c, it is enough to prove the theorem for k-dimensional smooth singular simplices. Let $\sigma \colon \Delta^k \to M$ be one. Denote by $\eta = \sigma^*\omega$ the pull-back of ω onto Δ^k. Since η is a $(k-1)$-form, it can be written as

$$\eta = \sum_{i=1}^{k} f_i \cdot dx^1 \wedge \dots \wedge \widehat{dx^i} \wedge \dots \wedge dx^k.$$

The pull-back of $d\omega$ by σ is

$$\begin{aligned}
\sigma^*(d\omega) = d(\sigma^*\omega) = d\eta &= \sum_{i=1}^{k} df_i \wedge dx^1 \wedge \dots \wedge \widehat{dx^i} \wedge \dots \wedge dx^k \\
&= \sum_{i=1}^{k} \left(\sum_{j=1}^{k} \partial_j f_i dx^j \right) \wedge dx^1 \wedge \dots \wedge \widehat{dx^i} \wedge \dots \wedge dx^k \\
&= \left(\sum_{i=1}^{k} (-1)^{i-1} \partial_i f_i \right) dx^1 \wedge \dots \wedge dx^k.
\end{aligned}$$

Thus, by the definition of the integral of a form on a smooth singular simplex,

$$\int_\sigma d\omega = \sum_{i=1}^{k} (-1)^{i-1} \int_{\Delta^k} \partial_i f_i(\mathbf{x}) d\mathbf{x}.$$

To compute the ith integral on the right let us integrate with respect to the ith variable first, and then with respect to the others. In the first step compute the integral using the Newton-Leibniz rule. This gives us

$$\int\limits_{\Delta^k} \partial_i f_i(\mathbf{x}) d\mathbf{x}$$

$$= \int\limits_{(x_1,\dots,\widehat{x_i},\dots,x_k)\in\Delta^{k-1}} \Big(\int\limits_0^{1-(x_1+\cdots+\widehat{x_i}+\cdots+x_k)} \partial_i f_i(x_1,\dots,x_k) dx_i \Big) dx_1 \dots \widehat{dx_i} \dots dx_k$$

$$= \int\limits_{(x_1,\dots,\widehat{x_i},\dots,x_k)\in\Delta^{k-1}} f_i(x_1,\dots,1-(x_1+\cdots+\widehat{x_i}+\cdots+x_k),\dots,x_k) dx_1 \dots \widehat{dx_i} \dots dx_k -$$

$$- \int\limits_{(x_1,\dots,\widehat{x_i},\dots,x_k)\in\Delta^{k-1}} f_i(x_1,\dots,x_{i-1},0,x_{i+1},\dots,x_k) dx_1 \dots \widehat{dx_i} \dots dx_k.$$

In the first integral of the last expression, make the substitution

$$x_1 = 1-(y_1+\cdots+y_{k-1}) \text{ if } i \neq 1,$$
$$x_j = y_{j-1} \text{ for } 2 \leq j \leq k \text{ and } j \neq i,$$

in the second integral make the substitution

$$(x_1,\dots,\widehat{x_i},\dots,x_k) = (y_1,\dots,y_{k-1}).$$

The determinants of the derivative matrices are equal to ± 1 for both substitutions, therefore

$$\int_{\Delta^k} \partial_i f_i(\mathbf{x}) d\mathbf{x} = \int_{\Delta^{k-1}} f_i(l_0^k(\mathbf{y})) d\mathbf{y} - \int_{\Delta^{k-1}} f_i(l_i^k(\mathbf{y})) d\mathbf{y},$$

and

$$\int_\sigma d\omega = \int_{\Delta^{k-1}} \left(\sum_{i=1}^k (-1)^{i-1} f_i(l_0^k(\mathbf{y})) \right) d\mathbf{y} + \sum_{i=1}^k (-1)^i \int_{\Delta^{k-1}} f_i(l_i^k(\mathbf{y})) d\mathbf{y}.$$

Now let us compute the integrals of ω on the facets of σ. To avoid confusion with coordinates on $\mathbb{R}^k$ and $\mathbb{R}^{k-1}$, let $(y^1,\dots,y^{k-1})$ be the identity chart on $\mathbb{R}^{k-1}$. Assume first that $1 \leq i \leq k$. Since $x^i \circ l_i^k \equiv 0$ and $(l_i^k)^*(dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k) = dy^1 \wedge \cdots \wedge dy^{k-1}$,

$$(\sigma \circ l_i^k)^* \omega = (l_i^k)^* \eta = (f_i \circ l_i^k) dy^1 \wedge \cdots \wedge dy^{k-1}$$

and

$$\int_{\sigma\circ l_i^k}\omega = \int_{\Delta^{k-1}} f_i(l_i^k(\mathbf{y}))d\mathbf{y}.$$

The pull-back of the form $dx^1\wedge\cdots\wedge\widehat{dx^i}\wedge\cdots\wedge dx^k$ by l_0^k is

$$\begin{aligned}
&(l_0^k)^*(dx^1\wedge\cdots\wedge\widehat{dx^i}\wedge\cdots\wedge dx^k)\\
&\qquad= d(1-y^1-\cdots-y^{k-1})\wedge dy^1\wedge\cdots\wedge\widehat{dy^{i-1}}\wedge\cdots\wedge dy^{k-1}\\
&\qquad= (-1)^{i-1}dy^1\wedge\cdots\wedge dy^{k-1},
\end{aligned}$$

consequently,

$$(\sigma\circ l_0^k)^*\omega = (l_0^k)^*\eta = \sum_{i=1}^k(-1)^i(f_i\circ l_0^k)dy^1\wedge\cdots\wedge dy^{k-1},$$

and

$$\int_{\sigma\circ l_0^k}\omega = \int_{\Delta^{k-1}}\left(\sum_{i=1}^k(-1)^{i-1}f_i(l_0^k(\mathbf{y}))\right)d\mathbf{y}.$$

Adding the integrals of ω on the facets of σ with alternating signs we get

$$\begin{aligned}
\int_{\partial\sigma}\omega &= \sum_{i=0}^k(-1)^i\int_{\sigma\circ l_i^k}\omega\\
&= \int_{\Delta^{k-1}}\left(\sum_{i=1}^k(-1)^{i-1}f_i(l_0^k(\mathbf{y}))\right)d\mathbf{y} + \sum_{i=1}^k(-1)^i\int_{\Delta^{k-1}}f_i(l_i^k(\mathbf{y}))d\mathbf{y}\\
&= \int_\sigma d\omega,
\end{aligned}$$

as we wanted to show. □

Corollary 4.7.9. *If $c\in Z_k(M)$ is a cycle, $\omega\in\Omega^k(M)$ is a closed differential form, then the integral $\int_c\omega$ depends only on the homology class $[c]\in H_k(M)$ of c and the de Rham cohomology $[\omega]\in H^k_{dR}(M)$ of ω.*

Proof. Choose other elements $c' = c+\partial a$ and $\omega' = \omega + d\alpha$ from the same homology and cohomology class respectively. Then

$$\int_{c'}\omega' = \int_c\omega + \int_{\partial c}(\omega+d\alpha) + \int_c d\alpha = \int_c\omega + \int_c(d\omega + d(d\alpha)) + \int_{\partial c}\alpha = \int_c\omega.\ \square$$

Definition 4.7.10. The *integral of the de Rham cohomology class* $[\omega]\in H^k_{dR}(M)$ *over the homology class* $[c]\in H_k(M)$ is defined by the equation

$$\int_{[c]}[\omega] = \int_c\omega.$$

※

4.7.2 Integration on Regular Domains

If M is an n-dimensional oriented manifold, then we can integrate a differential n-form on regular domains of M. Below we give details of this integration.

Definition 4.7.11. An *orientation of a* 0*-dimensional manifold* is an assignment of a sign $+$ or $-$ to each of its points.
An *orientation of a manifold* M *of dimension* $n \geq 1$ is a choice of an orientation of each tangent space T_pM of M in a continuous way. Continuity means that for each point $p \in M$, there is a chart $\phi \colon U \to M$ around p, such that the vectors $(\partial_1^\phi|_q, \dots, \partial_n^\phi|_q)$ form a positively oriented basis of the tangent space T_qM for all $q \in U$. A chart having this property is called a *positive chart with respect to the orientation.* A manifold is *orientable* if it has an orientation. ※

The Möbius band is probably the simplest example of an non-orientable manifold. An orientation of a manifold of dimension ≥ 1 can also be given by an atlas consisting of positive charts.

Definition 4.7.12. A subset D of a manifold M is called a *regular domain* if D is closed and for each boundary point $p \in \partial D$ of D, there is a chart $\phi \colon U \to \mathbb{R}^n$ around p such that

$$\phi(U \cap D) = \phi(U) \cap \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_n \geq 0\}. \qquad ※$$

Denote the half-space $\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_n \geq 0\}$ by $\mathbb{R}^n_+$. It is clear that for the chart in the definition

$$\phi(U \cap \partial D) = \phi(U) \cap \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_n = 0\},$$

Thus, if the coordinate functions of ϕ are $(x^1, \dots, x^n)$, then $(x^1, \dots, x^{n-1}) \colon U \cap \partial D \to \mathbb{R}^{n-1}$ is a chart on ∂D. In particular, the boundary of a regular domain in M is an $(n-1)$-dimensional smooth manifold.
A tangent vector $\mathbf{v} \in T_pM$ at a boundary $p \in \partial D$ is said to *point inward* D or, respectively, *outward* D if the last last coordinate of $\mathbf{v}$ in the basis $\partial_1^\phi|_p, \dots, \partial_n^\phi|_p$ positive or, respectively, negative, where ϕ is the chart form the definition of regular domains. The notion does not depend on the choice of the chart as we have the following equivalent characterization. The tangent vector points inward D (or outward D) if and only if $\mathbf{v}$ is not tangent to the boundary ∂D and for any curve $\gamma \in \Gamma_p$ representing the tangent vector $\mathbf{v}$, there is a positive $\delta > 0$ such that $\gamma((-\delta, 0)) \subset \operatorname{ext} D$ and $\gamma((0, \delta)) \subset \operatorname{int} D$ (or conversely, $\gamma((-\delta, 0)) \subset \operatorname{int} D$ and $\gamma((0, \delta)) \subset \operatorname{ext} D$).
An orientation of M induces an orientation on the boundary ∂D of any regular domain as follows. Let $p \in \partial D$, and choose a tangent vector $\mathbf{v} \in T_pM$

pointing outward D. If ∂D is 0-dimensional, $\dim M = 1$, assign to p the sign $+$ if and only if $\mathbf{v}$ is a positively oriented basis of T_pM. If $\dim M > 1$, then let the ordered basis $(\mathbf{v}_1, \dots, \mathbf{v}_{k-1})$ of $T_p(\partial M)$ be positively oriented if and only if $(\mathbf{v}, \mathbf{v}_1, \dots, \mathbf{v}_{k-1})$ is a positively oriented basis of T_pM.
Consider now a regular domain D in an oriented manifold M. To prepare integration of differential forms on D, we cover D by relatively small open sets.
Call a smooth singular n-dimensional simplex $\sigma\colon \Delta^n \to M$ a *positive regular n-simplex*, if σ extends to an open neighborhood $V \supset \Delta^n$ of Δ^n to a diffeomorphism $\sigma\colon V \to U$ the inverse of which is a positive chart of M.
If $p \in \operatorname{int} D$, then choose a positive chart $\phi\colon U \to \mathbb{R}^n$ around p, the domain of which is contained in $\operatorname{int} D$. Composing ϕ with a translation, we may assume that $\phi(p) \in \operatorname{int} \Delta^n$. Then composing ϕ with a central homothety with center at p and with sufficiently large ratio, we may also assume that $\phi(U) \subset \Delta^n$. Then denote by $\sigma_p\colon \Delta^n \to \operatorname{int} D \subset M$ the positive regular n-simplex obtained from the inverse of ϕ. Choose also for p an open neighborhood $p \in U_p \subset U$ such that $\phi(U_p) \subset \operatorname{int} \Delta^n$.
If $p \in \partial D$, then choose a chart $\phi = (x^1, \dots, x^n)\colon U \to \mathbb{R}^n$ around p as in Definition 4.7.12. Restricting ϕ to the connected component of U, we may assume that the vector fields $(\partial_1^\phi, \dots, \partial_n^\phi)$ define bases with constant positive or negative orientation on U. If $\dim M \geq 2$, we can make ϕ a positive chart by replacing x^1 with $-x^1$. If $\dim M = 1$, then let ϵ_p denote the sign corresponding to the orientation of the chart ϕ. Observe, that Δ^n intersects the boundary hyperplane of the half-space $\mathbb{R}^n_+$ in its nth facet $\Delta^{n-1} \times \{0\}$. Composing ϕ with a translation we may assume that $\phi(p)$ is a point of the relative interior of the nth facet. Then composing ϕ with a homothety with sufficiently large ratio and center at $\phi(p)$ we may assume that $\phi(U) \supset \Delta^n$. For $n \geq 2$ let σ_p be the positive regular n-simplex in M obtained from the inverse of ϕ. For $n = 1$, let σ_p be the chain equal to ϵ_p times the regular 1-simplex obtained from the inverse of ϕ. Choose also an open neighborhood $U_p \subset U$ of p such that $\phi(U_p \cap D) \subset \Delta^n$ and $\phi(U_p \cap D)$ intersects only the nth face of Δ^n.

Definition 4.7.13. The *support* $\operatorname{supp} \omega$ *of a differential k-form* $\omega \in \Omega^k(M)$ is the *closure* of the set $\{p \in M \mid \omega(p) \neq 0\}$. ※

Assume that ω is a differential n-form on the oriented manifold M the support of which intersects the regular domain D in a compact set $K = \operatorname{supp} \omega \cap D$. K is covered by the open subsets U_p $(p \in D)$ constructed above, therefore it is also covered by a finite number of them. Let $U_i = U_{p_i}$ $(i = 1, \dots, s)$ be a finite subcovering of K and set $U_0 = M \setminus K$. Denote by σ_i the regular n-simplex σ_{p_i}. Recall that σ_i is carrying a sign ϵ_i in the 1-dimensional case, otherwise it is positive. For the sake of uniform handling of different dimensions, we

can set $\epsilon_i = +1$ in dimensions not less than 2. The open sets $U_0, U_1, \dots, U_s$ cover M, therefore, there is a smooth partition of the unity $h_0, h_1, \dots, h_s$ subordinated to this covering. This means that the h_i's are smooth function on M, $h_i\colon M \to [0,1]$, $\operatorname{supp} h_i \subset U_i$ and $\sum_{i=0}^s h_i \equiv 1$.
We are going to define $\int_D \omega$ as

$$\int_D \omega = \sum_{i=1}^s \int_{\epsilon_i \sigma_i} h_i \omega. \tag{4.7}$$

However, to justify this definition we have to show that the expression on the right-hand side does not depend on the many arbitrary choices of the process. Assume that we repeat the same process but we choose the open sets U_p and the regular simplices σ_p in another way choose another finite subcovering for K and another partition of unity. Denote all the objects we choose the second time by the same letters as in the first time, but decorated with a tilde.
The key lemma is the following.

Lemma 4.7.14. *Assume that $\omega \in \Omega^n(M)$ and σ_i, $\tilde\sigma_j$ are regular n simplices from the above processes. Then*

$$\int_{\epsilon_i \sigma_i} h_i \tilde h_j \omega = \int_{\tilde\epsilon_j \tilde\sigma_j} h_i \tilde h_j \omega.$$

Proof. The regular n-dimensional simplices σ_i and $\tilde\sigma_j$ were obtained as the inverses of some charts $\phi_i\colon U \to \mathbb{R}^n$ and $\tilde\phi_i\colon \tilde U \to \mathbb{R}^n$. Let $H = \phi\circ\tilde\phi^{-1}\colon \tilde\phi(U\cap \tilde U) \to \phi(U \cap \tilde U)$ be the transit map between the two charts.
Since $\operatorname{supp} h_i \subset U_i \subset U$ and $\operatorname{supp} \tilde h_j \subset \tilde U_j \subset \tilde U$, $\operatorname{supp}(h_i \tilde h_j \omega) \subset U_i \cap \tilde U_j \subset U \cap \tilde U$. Assume that

$$\sigma_i^*(h_i \tilde h_j \omega) = f \cdot dx^1 \wedge \dots \wedge dx^n \qquad \tilde\sigma_j^*(h_i \tilde h_j \omega) = \tilde f \cdot dx^1 \wedge \dots \wedge dx^n.$$

Then

$$\int_{\sigma_i} h_i \tilde h_j \omega = \int_{\Delta^n \cap \phi(U_i \cap \tilde U_j)} f(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^n_+ \cap \phi(U_i \cap \tilde U_j)} f(\mathbf{x}) d\mathbf{x}$$

and similarly,

$$\int_{\tilde\sigma_j} h_i \tilde h_j \omega = \int_{\mathbb{R}^n_+ \cap \tilde\phi(U_i \cap \tilde U_j)} \tilde f(\mathbf{x}) d\mathbf{x}$$

The transit map H defines a diffeomorphism $\mathbb{R}^n_+ \cap \tilde\phi(U_i \cap \tilde U_j) \to \mathbb{R}^n_+ \cap \phi(U_i \cap \tilde U_j)$, thus, we can compute the integral of f making substitution $\mathbf{x} = H(\mathbf{u})$. This gives

$$\int_{\sigma_i} h_i \tilde h_j \omega = \int_{\mathbb{R}^n_+ \cap \phi(U_i \cap \tilde U_j)} f(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^n_+ \cap \tilde\phi(U_i \cap \tilde U_j)} f(H(\mathbf{u})) |\det H'(\mathbf{u})| d\mathbf{u}. \tag{4.8}$$

On the other hand, since $\tilde{\sigma}_j = \sigma_i \circ H$.

$$\tilde{f} \cdot dx^1 \wedge \cdots \wedge dx^n = \tilde{\sigma}_j^*(h_i \tilde{h}_j \omega) = H^*(\sigma_i^*(h_i \tilde{h}_j \omega)) = H^*(f \cdot dx^1 \wedge \cdots \wedge dx^n).$$

Denote by $(H^1, \ldots, H^n)$ the coordinate functions $H^i = x^i \circ H$ of H. Then

$$H^*(f \cdot dx^1 \wedge \cdots \wedge dx^n) = (f \circ H) \cdot dH^1 \wedge \cdots \wedge dH^n.$$

As $dH^i = \partial_1 H^i dx^1 + \cdots + \partial_n H^i dx^n$,

$$H^*(f \cdot dx^1 \wedge \cdots \wedge dx^n) = (f \circ H) \cdot (\det H') \cdot dx^1 \wedge \cdots \wedge dx^n,$$

which yields $\tilde{f} = (f \circ H) \cdot (\det H')$, and

$$\tilde{\sigma}_j^*(h_i \tilde{h}_j \omega) = \tilde{f} \cdot dx^1 \wedge \cdots \wedge dx^n = \int_{\mathbb{R}^n_+ \cap \tilde{\phi}(U_i \cap \tilde{U}_j)} f(H(\mathbf{x}))(\det H'(\mathbf{x})) d\mathbf{x}. \quad (4.9)$$

Comparing equations (4.8) and (4.9) we see that the equation we want to show holds up to sign. To have the equality without any sign corrections, it remains to show that $\epsilon_i \tilde{\epsilon}_j = \operatorname{sgn} \det H'$. If $\dim M \geq 2$, then $\epsilon_i = \tilde{\epsilon}_j = 1$, furthermore, ϕ and $\tilde{\phi}$ are positive charts, so the derivative of the transit map between them preserves orientation, that is, it has positive determinant.
For $\dim M = 1$, the derivative of the transit map has positive determinant if and only if either both ϕ and $\tilde{\phi}$ are positive or none of them is positive. This completes the proof. □

Proposition 4.7.15. *For any two different choices of the simplices and the partition of unity*

$$\sum_{i=1}^{s} \int_{\epsilon_i \sigma_i} h_i \omega = \sum_{j=1}^{\tilde{s}} \int_{\tilde{\epsilon}_j \tilde{\sigma}_j} \tilde{h}_j \omega.$$

Proof. The supports of $h_0 \omega$ and $\tilde{h}_0 \omega$ are in $(M \setminus K) \cap \operatorname{supp} \omega \subset \operatorname{ext} D$, thus, the integrals of the forms $h_0 \tilde{h}_j \omega$ and $\tilde{h}_0 h_i \omega$ on any of the regular n-simplices σ_k and $\tilde{\sigma}_l$ are equal to 0. Consequently,

$$\begin{aligned}
\sum_{i=1}^{s} \int_{\epsilon_i \sigma_i} h_i \omega &= \sum_{i=1}^{s} \int_{\epsilon_i \sigma_i} h_i \left(\sum_{j=0}^{\tilde{s}} \tilde{h}_j \right) \omega \\
&= \sum_{i=1}^{s} \sum_{j=1}^{\tilde{s}} \int_{\epsilon_i \sigma_i} h_i \tilde{h}_j \omega = \sum_{i=1}^{s} \sum_{j=1}^{\tilde{s}} \int_{\tilde{\epsilon}_j \tilde{\sigma}_j} h_i \tilde{h}_j \omega \\
&= \sum_{j=1}^{\tilde{s}} \int_{\tilde{\epsilon}_j \tilde{\sigma}_j} \left(\sum_{i=0}^{s} h_i \right) \tilde{h}_j \omega = \sum_{j=1}^{\tilde{s}} \int_{\tilde{\epsilon}_j \tilde{\sigma}_j} \tilde{h}_j \omega.
\end{aligned}$$

□

Definition 4.7.16. Using the notations introduced above, we define the *integral of ω over the regular domain D* by (4.7). ※

Corollary 4.7.17. *Denote by $\underline{\Delta}^n$ the union of the interior of Δ^n and the relative interior of its nth facet. This set is what remains if we remove from Δ^n all its facets except for the nth one.*
If $\operatorname{supp}\omega \cap D$ is covered by $\sigma_i(\underline{\Delta}^n)$, then $\int_D \omega = \int_{\epsilon_i\sigma_i}\omega$.

Proof. If $\sigma_i(\underline{\Delta}^n)$ covers $K = \operatorname{supp}\omega \cap D$, then U_i can be chosen so that U_i covers K alone, so there will be no need for other simplices to compute the integral. Then h_i will be constant 1 on K, so $(h_i\omega)|_D = \omega|D$. □

Theorem 4.7.18 (Stokes' Theorem for Regular Domains)**.** *If D is a regular domain in an oriented n-dimensional manifold M, $\omega \in \Omega^{n-1}(M)$ is a differential $(n-1)$-form such that $\operatorname{supp}\omega \cap D$ is compact, and ∂D is endowed with the induced orientation, then*

$$\int_D d\omega = \int_{\partial D}\omega.$$

Proof. Since $\sum_{i=0}^s h_i \equiv 1$, $\sum_{i=0}^s dh_i \equiv 0$, and

$$d\omega = \sum_{i=0}^s h_i d\omega = \sum_{i=0}^s d(h_i\omega) - \left(\sum_{i=0}^s dh_i\right)\wedge\omega = \sum_{i=0}^s d(h_i\omega).$$

The support of $d(h_0\omega)$ is disjoint from D, so integrating over D we obtain

$$\int_D d\omega = \sum_{i=1}^s \int_D d(h_i\omega) = \sum_{i=1}^s \int_{\epsilon_i\sigma_i} d(h_i\omega).$$

Applying the Stokes' theorem for chains, and the fact that $\operatorname{supp}(h_i\omega)$ intersects only the image of the nth facet of σ_i,

$$\begin{aligned}\int_D d\omega = \sum_{i=1}^s \int_D d(h_i\omega) = \sum_{i=1}^s \int_{\epsilon_i\sigma_i} d(h_i\omega) &= \sum_{i=1}^s \int_{\epsilon_i\partial\sigma_i} h_i\omega \\ &= \sum_{i=1}^s \epsilon_i(-1)^n \int_{\sigma_i\circ l_n^n} h_i\omega.\end{aligned}$$

The system of regular $(n-1)$-simplices $\sigma_i\circ l_n^n$ the relative open subsets $U_i\cap\partial D$ of ∂D and the partition of unity $h_i|_{\partial D}$ form a system to compute the integral $\int_{\partial D}\omega$, however the orientation of the simplices $\sigma_i\circ l_n^n$ is not positive. The orientation of it is given exactly by the sign $\epsilon_i(-1)^n$. Indeed, the derivative of σ_i at the origin takes the standard basis of $\mathbb{R}^n$ to the tangent vectors

$(\mathbf{v}_1, \dots, \mathbf{v}_n)$ at $\sigma_i(\mathbf{0})$. The sign of the orientation of this basis relative to the orientation of M is ϵ_i. The first $(n-1)$ vectors $(\mathbf{v}_1, \dots, \mathbf{v}_{n-1})$ are the images of the standard basis vectors of $\mathbb{R}^{n-1}$ under the derivative of $\sigma_i \circ l_n^n$ at the origin, $\mathbf{v}_n$ is pointing inward D. Thus, the sign of the orientation of $(\mathbf{v}_1, \dots, \mathbf{v}_{n-1})$ relative to the orientation of the boundary is the sign of the orientation of $(-\mathbf{v}_n, \mathbf{v}_1, \dots, \mathbf{v}_{n-1})$ which is exactly $\epsilon_i(-1)^n$. In conclusion we obtain

$$\sum_{i=1}^{s} \epsilon_i (-1)^n \int_{\sigma_i \circ l_n^n} h_i \omega = \int_{\partial D} \omega,$$

which completes the proof. □

4.7.3 Integration on Riemannian Manifolds

Definition 4.7.19. A *Riemannian manifold* is a smooth manifold M endowed with a smooth tensor field g of type $(2,0)$ for which $g(p)$ is a positive definite symmetric bilinear function on T_pM for each $p \in M$. The tensor field g is called the *(Riemannian) metric on M.* ※

Example. A hypersurface of $\mathbb{R}^n$ together with the first fundamental form is a Riemannian manifold.

The bilinear function $g(p)$ is often denoted by $\langle\, , \rangle_p$, and if X and Y are two vector fields on M the function $g(X,Y)$ is also denoted by $\langle X, Y \rangle$.
If M is an n-dimensional oriented Riemannian manifold, then there is a unique differential n-form ω on it, called its *volume form.* The volume form is defined in the following way.
Let $p \in M$ be an arbitrary point and choose a positively oriented orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of T_pM. Let $\mathbf{e}^1, \dots, \mathbf{e}^n$ be the dual basis of the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ and set $\omega(p) = \mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^n$. The form $\omega(p)$ does not depend on the choice of the orthonormal basis. Indeed, if $\mathbf{f}_1, \dots, \mathbf{f}_n$ is another positively oriented orthonormal basis of T_pM with dual basis $\mathbf{f}^1, \dots, \mathbf{f}^n$. Then $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^n$ and $\mathbf{f}^1 \wedge \cdots \wedge \mathbf{f}^n$ both have unit length with respect to the induced Euclidean structure of the 1-dimensional linear space $\Lambda^n(T_pM)^*$, (see Proposition 1.2.67), and they must have the same orientation, hence they ought to be equal.
Denote by $\mathcal{C}_c(M)$ the linear space of continuous functions with compact support on M. Consider the functional $I \colon \mathcal{C}_c(M) \to \mathbb{R}$ defined by

$$I(f) = \int_M f\omega.$$

The right-hand side makes sense, because the integral of compactly supported forms can be defined in the same manner as we did also in the case when the differential form is continuous but not necessarily smooth. The functional I is a positive linear functional on $\mathcal{C}_c(M)$, which means that it is linear and assigns positive numbers to nonnegative functions not identically equal to 0. By the Riesz representation theorem from measure theory, (see Theorem 7.2.8 in [5]), any such positive linear functional corresponds to a unique regular Borel measure μ on M, for which $I(f) = \int_M f d\mu$. The measure obtained this way is the *volume measure of the oriented Riemannian manifold* M.

Though the volume form cannot be defined for non-orientable Riemannian manifolds, there is a volume measure even in that case.

For this purpose, we construct an orientable double covering $\rho\colon \tilde{M} \to M$ of M. $\tilde{M}$ as a set consists of all pairs (p, o), where p is a point of M, o is an orientation of T_pM. The covering map $\rho\colon \tilde{M} \to M$ assigns to the pair (p, o) the point p. Since a linear space has exactly two orientations, each point p in M has exactly two preimages. We define a manifold structure on $\tilde{M}$ with the help of an atlas. Choose a chart $\phi\colon U \to \mathbb{R}^n$ from the atlas $\mathcal{A}$ of M. We can split $\rho^{-1}(U)$ into two disjoint pieces U_+ and U_- as follows. Let U_+ contain those pairs (p, o), for which $p \in U$ and the basis $(\partial_1^\phi|_p, \dots, \partial_n^\phi|_p)$ is positively oriented with respect to o, and let $U_- = \rho^{-1}(U) \setminus U_+$ be the remaining subset. The charts $\phi \circ \rho|_{U_\pm}\colon U_\pm \to \mathbb{R}^n$ $(\phi \in \mathcal{A})$ give an atlas on $\tilde{M}$ turning it into a differentiable manifold. $\tilde{M}$ has a unique orientation for which all the charts $\phi \circ \rho|_{U_+}$ are positive. The restriction of ρ onto $U_\pm$ is a diffeomorphism between $U_\pm$ and U, in particular, ρ is a covering map. $\tilde{M}$ can be endowed by the Riemannian metric $\rho^* g$, obtained as the pull-back of the Riemannian metric of M by the covering map ρ.

As $\tilde{M}$ can be oriented, we can produce a volume form $\tilde{\omega}$ and a volume measure $\tilde{\mu}$ on it as above. Then a subset A of M will be μ measurable with respect to the volume measure of M if and only if $\rho^{-1}(A)$ is $\tilde{\mu}$-measurable, and if this is fulfilled, then we set $\mu(A) = \tilde{\mu}(\rho^{-1}(A))/2$.

4.8 Differentiation of Vector Fields

Although there is a natural way to differentiate a smooth function defined on a manifold with respect to a tangent vector, there is no analogous natural way to differentiate vector fields. In fact, there are infinitely many ways to differentiate vector fields with respect to a tangent vector, and to choose one of them, (the most appropriate one in some sense), the differentiable manifold structure alone is not enough. A fixed rule for the differentiation of vector fields is itself an additional structure on the manifold, called an *affine connection.* Later we shall see that on Riemannian manifolds, i.e., on manifolds the tangent spaces of which are equipped with a dot product we can introduce a special differentiation of vector fields in a natural way. A precise

formulation of this statement is the *Fundamental Theorem of Riemannian Geometry*.
As far as only vector fields on an open domain of $\mathbb{R}^n$ are considered, the rule

$$\partial_Y X = (X \circ \gamma)'(0) \tag{4.10}$$

seems to be quite natural for the derivation of a vector field X over an open subset $U \subset \mathbb{R}^n$ with respect to the tangent vector $Y \in T_p\mathbb{R}^n$, where $\gamma\colon [-\varepsilon, \varepsilon] \to U$ is any smooth curve that satisfies $\gamma(0) = p$ and $\gamma'(0) = Y$. In this formula, using the natural isomorphisms $T_{\mathbf{p}}\mathbb{R}^n \cong \mathbb{R}^n$, X is thought of as a map $X\colon \mathbb{R}^n \to \mathbb{R}^n$. This interpretation of X enables us to compute the derivative of the vector valued function $X \circ \gamma$ at 0.
We see that

$$\partial_Y X = \sum_{i=1}^{n} Y(X^i)\partial_i(p), \tag{4.11}$$

where ∂_i denotes the i-th coordinate vector field on $\mathbb{R}^n$, X^i are the components of the vector field X with respect to ∂_i. In particular, the value of $\partial_Y X$ does not depend on the choice of γ.
It is easy to check that this differentiation of vector fields has the properties.

(1) $$\partial_{Y_1+Y_2} X = \partial_{Y_1} X + \partial_{Y_2} X,$$

(2) $$\partial_{cY} X = c\partial_Y X,$$

(3) $$\partial_Y (X_1 + X_2) = \partial_Y X_1 + \partial_Y X_2,$$

(4) $$\partial_Y (fX) = Y(f)X + f(p)\partial_Y X,$$

(5) $$\partial_{X_1} X_2 - \partial_{X_2} X_1 = [X_1, X_2],$$

(6) $$Y\langle X_1, X_2\rangle = \langle \partial_Y X_1, X_2\rangle + \langle X_1, \partial_Y X_2\rangle,$$

where $X_1, X_2 \in \mathfrak{X}(\mathbb{R}^n)$, $Y \in T_p\mathbb{R}^n$, $f \in \mathcal{F}(\mathbb{R}^n)$, $c \in \mathbb{R}$.
Now consider the general case. Let M be a smooth manifold. As we mentioned, there is no natural rule for derivation of vector fields on M, so we introduce such rules axiomatically, as operations satisfying some of the above properties.

Definition 4.8.1. An *affine connection* ∇ *at* $p \in M$ is a mapping which assigns to each tangent vector $Y \in T_pM$ and each vector field $X \in \mathfrak{X}(M)$ a new tangent vector $\nabla_Y X \in T_pM$ called the *covariant derivative* of X with respect to Y and satisfies the following identities

(i) $$\nabla_{Y_1+Y_2}X = \nabla_{Y_1}X + \nabla_{Y_2}X,$$

(ii) $$\nabla_{cY}X = c\nabla_Y X,$$

(iii) $$\nabla_Y(X_1 + X_2) = \nabla_Y X_1 + \nabla_Y X_2,$$

(iv) $$\nabla_Y(fX) = Y(f)X + f(p)\nabla_Y X,$$

where $X_1, X_2 \in \mathfrak{X}(M)$, $Y, Y_1, Y_2 \in T_pM$, $f \in \mathcal{F}(M)$, $c \in \mathbb{R}$. ※

Definition 4.8.2. A *global affine connection* (or briefly a *connection*) ∇ on M is a mapping which assigns to two smooth vector fields Y and X a new one $\nabla_Y X$ called the *covariant derivative of the vector field X with respect to the vector field Y*, satisfying the identities (i), (ii), (iii), and

(iv') $$\nabla_Y(fX) = Y(f)X + f\nabla_Y X,$$

for all $X_1, X_2, Y, Y_1, Y_2 \in \mathfrak{X}(M)$, and $f, c \in \mathcal{F}(M)$. ※

Lemma 4.8.3. *For a global affine connection ∇, $X, Y \in \mathfrak{X}(M)$, $p \in M$, the tangent vector $(\nabla_Y X)(p)$ depends only on the behavior of X and Y in an open neighborhood of p.*

Proof. Let us suppose that the vector fields X_1 and X_2 coincide on an open neighborhood U of p. Choose a smooth function $h \in \mathcal{F}(M)$ which is zero outside U and constant 1 on a neighborhood of p. Then we have $h(X_1 - X_2) = 0$, consequently

$$0 = \nabla_Y(h(X_1 - X_2)) = h\nabla_Y(X_1 - X_2) + Y(h)(X_1 - X_2).$$

Computing the right-hand side at p we get

$$0 = (\nabla_Y X_1)(p) - (\nabla_Y X_2)(p).$$

Similarly, if the vector fields Y_1 and Y_2 coincide on an open neighborhood U of p and h is chosen as above, then we have

$$0 = \nabla_{h(Y_1-Y_2)}X = h\nabla_{Y_1}X - h\nabla_{Y_2}X,$$

which yields $0 = \nabla_{Y_1}X(p) - \nabla_{Y_2}X(p)$. □

The lemma implies that an affine connection can be restricted onto any open subset and can be recovered from its restrictions onto the elements of an open cover.

Fix a local coordinate system $\phi = (x^1, \dots, x^n)$ on an open subset U of M and let $\partial_1 = \partial_1^\phi, \dots, \partial_n = \partial_n^\phi$ be the basis vector fields on U induced by ϕ. (As we work with a fixed chart ϕ, omission of ϕ from the notation should not lead to confusion.) Given an affine connection ∇ on U, we can express the vector field $\nabla_{\partial_i}\partial_j$ as a linear combination of the basis vector fields

$$\nabla_{\partial_i}\partial_j = \sum_{i=1}^{n} \Gamma_{ij}^k \partial_k.$$

The components Γ_{ij}^k are smooth functions on U, called the *Christoffel symbols* of the covariant derivation with respect to the given chart.

Proposition 4.8.4. *The restriction of a global affine connection onto an open coordinate neighborhood U is uniquely determined by the Christoffel symbols. Any n^3 smooth functions Γ_{ij}^k on U are the Christoffel symbols of an appropriate affine connection on U.*

Proof. Let $X = \sum_{i=1}^n X^i\partial_i$, $Y = \sum_{j=1}^n Y^j\partial_j$ be two smooth vector fields on U. Then by the properties of affine connections, $\nabla_Y X$ can be computed as follows

$$\begin{aligned}
\nabla_Y X &= \nabla_{\left(\sum_{j=1}^n Y^j\partial_j\right)}\left(\sum_{i=1}^n X^i\partial_i\right) = \sum_{j=1}^n Y^j \nabla_{\partial_j}\left(\sum_{i=1}^n X^i\partial_i\right)\\
&= \sum_{i=1}^n\sum_{j=1}^n Y^j\left(\partial_j(X^i)\partial_i + X^i\nabla_{\partial_j}\partial_i\right)\\
&= \sum_{i=1}^n\sum_{j=1}^n Y^j\left(\partial_j(X^i)\partial_i + X^i\sum_{k=1}^n \Gamma_{ji}^k\partial_k\right)\\
&= \sum_{k=1}^n\left(Y(X^k) + \sum_{i=1}^n\sum_{j=1}^n X^iY^j\Gamma_{ji}^k\right)\partial_k.
\end{aligned}$$

This formula shows that the knowledge of the Christoffel symbols enables us to compute the covariant derivative of any vector field with respect to any other one. On the other hand, if Γ_{ij}^k are an arbitrary smooth functions on U for $1 \leq i, j, k \leq n$, then defining the covariant derivative of a vector field by the above formula, we obtain an affine connection on U. □

Note that the connection ∂ defined by (4.11) on $\mathbb{R}^n$ has vanishing Christoffel-symbols with respect to the identity chart of $\mathbb{R}^n$.
Observe also, that in fact, the tangent vector $(\nabla_Y X)(p)$ depends only on the vector $Y(p)$, so a global affine connection on a manifold defines an affine connection at each of its points. Furthermore, we do not need to know the vector field X everywhere on U to compute $(\nabla_Y X)(p)$. It is enough to know X at the points of a curve $\gamma\colon [-\varepsilon, \varepsilon] \to M$ for which $\gamma(0) = p$, $\gamma'(0) = Y(p)$.

Definition 4.8.5. Let $\gamma\colon [a,b] \to M$ be a smooth curve in M. A smooth *vector field* X *along the curve* γ is a smooth mapping $X\colon [a,b] \to TM$ which assigns to each $t \in [a,b]$ a tangent vector $X(t) \in T_{\gamma(t)}M$. ※

Now suppose that M is provided with a global affine connection ∇. Then any vector field X along γ determines a new vector field $\nabla_t X$ along γ called the *covariant derivative of* X *with respect to the curve parameter* t as follows. For $\tau \in [a,b]$, choose a local coordinate system ϕ around $\gamma(\tau)$. Let Γ_{ij}^k be the Christoffel symbols of ∇ with respect to ϕ decompose X and γ' as

$$X(t) = \sum_{i=1}^n X^i(t)\partial_i^\phi(\gamma(t)) \quad \text{ and } \quad \gamma'(t) = \sum_{i=1}^n Y^i(t)\partial_i^\phi(\gamma(t)).$$

Then $\nabla_t X(\tau)$ is given by

$$\nabla_t X(\tau) := \sum_{k=1}^n \left(X^{k\prime}(\tau) + \sum_{i=1}^n \sum_{j=1}^n X^i(\tau)Y^j(\tau)\Gamma_{ji}^k(\gamma(\tau)) \right) \partial_k^\phi(\gamma(\tau)).$$

Remark. Several different notations are used for the covariant derivative of a vector field X along a curve γ: $\nabla_t X$, $\frac{\nabla X}{dt}$, X', $\nabla_{\gamma'}X$. The most concise notation X' does not show the dependence of this operation on ∇. Nevertheless, it is convenient to use when we work with a fixed connection ∇ on M. Notations $\nabla_t X$ and $\frac{\nabla X}{dt}$ can be used only when the parameter of the curve is denoted by a fixed symbol, in our case by t. Of course, if another symbol is used for the parameter, then t should be replaced by that. The notation $\nabla_{\gamma'}X$ is motivated by the fact that if $\tilde{X}$ is a vector field defined in an open neighborhood of $\gamma(\tau)$ which extends X in the sense that $\tilde{X}(\gamma(t))) = X(t)$ for all t close enough to τ, then $\nabla_t X(\tau) = \nabla_{\gamma'(\tau)}\tilde{X}$. In the rest of this book we shall mainly use the notations $\nabla_t X$ and $\nabla_{\gamma'}X$.

Definition 4.8.6. A vector field X along a curve γ is said to be a *parallel vector field* if the covariant derivative $\nabla_{\gamma'}X$ is identically zero. ※

Proposition 4.8.7. *Given a curve* γ *and a tangent vector* X_0 *at the point* $\gamma(0)$*, there is a unique parallel vector field* X *along* γ *which extends* X_0*.*

Proof. The proposition follows from results on ordinary differential equations. Using local coordinates, condition $\nabla_{\gamma'}X = 0$ yields a system of ordinary differential equations for the components of X

$$X^{k\prime} + \sum_{i=1}^n \sum_{j=1}^n X^i Y^j \Gamma_{ji}^k \circ \gamma = 0.$$

Since these equations are linear, the existence and uniqueness theorem for linear differential equations guaranties that the solutions of this system of differential equations are uniquely determined by the initial values $X^k(0)$ and can be defined for all values of t. □

If X is a parallel vector field along the curve γ, then the vector $X(\gamma(t))$ is said to be obtained from X_0 by *parallel transport along* γ.

Definition 4.8.8. A connection is called *symmetric* or *torsion free* if it satisfies the identity

$$\nabla_X Y - \nabla_Y X = [X, Y].$$

※

Applying this identity to the case $X = \partial_i^\phi$, $Y = \partial_j^\phi$, from $[\partial_i^\phi, \partial_j^\phi] = 0$ one obtains the relation

$$\Gamma_{ij}^k = \Gamma_{ji}^k$$

for the Christoffel symbols of a torsion free connection. Conversely, if $\Gamma_{ij}^k = \Gamma_{ji}^k$ then using the expression of covariant derivative with the help of Christoffel symbols we get

$$\begin{aligned}
&\nabla_X Y - \nabla_Y X \\
&\quad = \sum_{k=1}^n \left(X(Y^k) + \sum_{i=1}^n \sum_{j=1}^n Y^i X^j \Gamma_{ji}^k \right) \partial_k^\phi \\
&\quad\quad - \sum_{k=1}^n \left(Y(X^k) + \sum_{i=1}^n \sum_{j=1}^n X^i Y^j \Gamma_{ji}^k \right) \partial_k^\phi \\
&\quad = \sum_{i=1}^n \big(X(Y^k) - Y(X^k) \big) \partial_k^\phi = [X, Y].
\end{aligned}$$

There is a useful characterization of symmetry. Consider a parameterized surface in M that is a smooth mapping $s\colon R \to M$ from a rectangular domain R of the plane $\mathbb{R}^2$ into M.

A *vector field* X *along* s is a mapping which assigns to each $(x, y) \in R$ a tangent vector $X(x, y) \in T_{s(x,y)}M$.

As examples, the two standard vector fields ∂_x and ∂_y on the plane give rise to vector fields $Ts(\partial_x)$ and $Ts(\partial_y)$ along s. These will be denoted briefly by $\partial_x s$ and $\partial_y s$. Here $Ts\colon T\mathbb{R}^2 \to TM$ denotes the derivative map of s.

For any smooth vector field X along s the *covariant partial derivatives* $\nabla_x X$ and $\nabla_y X$ are new vector fields along s constructed as follows. For each fixed y_0, the map $x \mapsto X(x, y_0)$ is a vector field along the curve $x \mapsto s(x, y_0)$. The vector $\nabla_x X(x_0, y_0)$ is defined to be its covariant derivative with respect to x at x_0. This defines $\nabla_x X$ along the entire parameterized surface s. $\nabla_y X$ is defined similarly.

Proposition 4.8.9. *A connection is symmetric if and only if*

$$\nabla_x(\partial_y s) = \nabla_y(\partial_x s)$$

for any parameterized surface s in M.

Proof. Let us choose a local coordinate system $\phi = (x^1, \dots, x^n)$ on M around a point $s(x_0, y_0)$. The mapping s is given locally by the functions $s^i = x^i \circ s$. The vector field $\partial_y s$ has the form

$$\partial_y s = \sum_{i=1}^n \partial_y s^i \cdot (\partial_i^\phi \circ s).$$

The partial covariant derivative of this vector field with respect to x is equal to

$$\begin{aligned}\nabla_x(\partial_y s) &= \nabla_x \left(\sum_{i=1}^n \partial_y s^i \cdot (\partial_i \circ s) \right) \\ &= \sum_{i=1}^n \partial_x(\partial_y(s^i)) \cdot (\partial_i^\phi \circ s) + \partial_y s^i \cdot \nabla_x(\partial_i^\phi \circ s) \\ &= \sum_{i=1}^n \partial_x(\partial_y(s^i)) \cdot (\partial_i^\phi \circ s) + \sum_{k=1}^n \left(\sum_{i=1}^n \sum_{j=1}^n \partial_y s^i \cdot \partial_x s^j \cdot (\Gamma_{ji}^k \circ s) \right) \partial_k^\phi \circ s.\end{aligned}$$

This formula shows that interchanging the role of x and y we obtain the same vector field for any s if and only if $\Gamma_{ij}^k = \Gamma_{ji}^k$. □

Roughly speaking, the torsion free condition halves the degree of freedom in the choice of Christoffel symbols, a symmetric connection is uniquely determined by $\frac{n^2(n+1)}{2}$ arbitrarily chosen functions, nevertheless, the space of symmetric affine connections on a manifold is still infinite dimensional. We can reduce the degree of freedom further on Riemannian manifolds.

Definition 4.8.10. A connection ∇ on a Riemannian manifold M is said to be *compatible with the Riemannian metric* if parallel transport along curves preserves inner products. In other words, for any curve γ and any pair X, Y of parallel vector fields along γ, the inner product $\langle X, Y \rangle$ is constant. ※

Proposition 4.8.11. *Suppose that the connection is compatible with the metric. Let V, W be any two vector fields along γ. Then*

$$\langle V, W \rangle' = \langle \nabla_t V, W \rangle + \langle V, \nabla_t W \rangle.$$

Proof. Choose parallel vector fields $X_1, \dots, X_n$ along γ which are orthonormal at one point of γ and hence at every point of γ. Then the given fields V and W can be expressed as $\sum_{i=1}^n v_i X_i$ and $\sum_{i=1}^n w_i X_i$ respectively (where $v_i = \langle V, X_i \rangle$ is a real valued function). It follows that $\langle V, W \rangle = \sum_{i=1}^n v_i w_i$ and that

$$\nabla_t V = \sum_{i=1}^n v_i' X_i, \qquad \nabla_t W = \sum_{i=1}^n w_i' X_i.$$

Therefore

$$\langle \nabla_t V, W \rangle + \langle V, \nabla_t W \rangle = \sum_{i=1}^n (v_i' w_i + v_i w_i') = \langle V, W \rangle'. \qquad \square$$

Corollary 4.8.12. *An affine connection on a Riemannian manifold is compatible with the metric if and only if for any vector fields X_1, X_2 on M and any tangent vector $Y \in T_pM$ we have*

$$Y(\langle X_1, X_2 \rangle) = \langle \nabla_Y X_1, X_2 \rangle + \langle X_1, \nabla_Y X_2 \rangle.$$

Theorem 4.8.13 (Fundamental Theorem of Riemannian Geometry)**.** *A Riemannian manifold possesses one and only one symmetric connection which is compatible with its metric.*

Proof. Applying the compatibility condition to the basis vector fields $\partial_i, \partial_j, \partial_k$ induced by a fixed chart on the manifold and setting $\langle \partial_j, \partial_k \rangle = g_{jk}$ one obtains the identity

$$\partial_i g_{jk} = \langle \nabla_{\partial_i} \partial_j, \partial_k \rangle + \langle \partial_j, \nabla_{\partial_i} \partial_k \rangle.$$

Permuting i, j and k this gives three linear equations relating the three quantities

$$\langle \nabla_{\partial_i} \partial_j, \partial_k \rangle, \qquad \langle \nabla_{\partial_j} \partial_k, \partial_i \rangle, \qquad \langle \nabla_{\partial_k} \partial_i, \partial_j \rangle.$$

(There are only three such quantities since $\nabla_{\partial_i} \partial_j = \nabla_{\partial_j} \partial_i$.) These equations can be solved uniquely; yielding the *first Christoffel identity*

$$\langle \nabla_{\partial_i} \partial_j, \partial_k \rangle = \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}).$$

The left-hand side of this identity is equal to $\sum_{l=1}^n \Gamma_{ij}^l g_{lk}$. Multiplying by the inverse (g^{kl}) of the matrix (g_{lk}) this yields the *second Christoffel identity*

$$\Gamma_{ij}^k = \sum_{l=1}^n \frac{1}{2}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}) g^{lk}.$$

Thus the connection is uniquely determined by the metric.

Conversely, defining Γ_{ij}^k by this formula, one can verify that the resulting connection is symmetric and compatible with the metric. This completes the proof. □

Definition 4.8.14. The unique symmetric affine connection which is compatible with the metric on a Riemannian manifold is called the *Levi-Cività connection.* ※

Since the formula for the Christoffel symbols of the Levi-Cività connection involves elements of the inverse of the matrix of the metric, it is not a convenient formula to express $\nabla_X Y$ in a coordinate free way. However, there is an explicit coordinate free expression of $\langle \nabla_X Y, Z\rangle$ for an arbitrary vector field Z. To obtain this formula, start with the equation

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_Z X + [X,Z]\rangle.$$

Rearranging and permuting the role of X, Y and Z cyclically, we obtain the following system of equations

$$\begin{aligned} X\langle Y, Z\rangle - \langle Y, [X,Z]\rangle &= \langle \nabla_X Y, Z\rangle + \langle \nabla_Z X, Y\rangle,\\ Y\langle Z, X\rangle - \langle Z, [Y,X]\rangle &= \langle \nabla_Y Z, X\rangle + \langle \nabla_X Y, Z\rangle,\\ Z\langle X, Y\rangle - \langle X, [Z,Y]\rangle &= \langle \nabla_Z X, Y\rangle + \langle \nabla_Y Z, X\rangle. \end{aligned}$$

Solving this system for the unknown quantities $\langle \nabla_X Y, Z\rangle$, $\langle \nabla_Y Z, X\rangle$ and $\langle \nabla_Z X, Y\rangle$ we obtain *Koszul's formula*

$$\begin{aligned} \langle \nabla_X Y, Z\rangle = \frac{1}{2}\big(&X\langle Y, Z\rangle - \langle Y, [X,Z]\rangle + Y\langle Z, X\rangle \\ &- \langle Z, [Y,X]\rangle - Z\langle X, Y\rangle + \langle X, [Z,Y]\rangle\big). \end{aligned} \tag{4.12}$$

The connection defined by formula (4.11) on open subsets of $\mathbb{R}^n$ is just the Levi-Cività connection of $\mathbb{R}^n$.

Consider now a parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$. It is a Riemannian manifold with the first fundamental form as metric. The basis vector fields $\mathbf{r}_i$ through suitable identifications are the same as the basis vector fields ∂_i^ϕ corresponding to the chart $\phi = \mathbf{r}^{-1}$. Comparing the formulae

$$\partial_{\mathbf{r}_i}\mathbf{r}_j = \mathbf{r}_{ij} = \sum_{k=1}^{n-1} \Gamma_{ij}^k \mathbf{r}_k + b_{ij}\mathbf{N} \tag{4.13}$$

and

$$\Gamma_{ij}^k = \sum_{l=1}^{n-1} \frac{1}{2}(g_{jl,i} + g_{il,j} - g_{ij,l})g^{lk}.$$

proved in Section 3.1.5 with the formulae derived for the Levi-Cività connection we can conclude that the Christoffel symbols of a hypersurface introduced previously are the Christoffel symbols of the Levi-Cività connection of the hypersurface. Therefore, denoting by ∇ the Levi-Cività connection of the hypersurface, we can rewrite equation (4.13) as

$$\partial_{\mathbf{r}_i}\mathbf{r}_j = \nabla_{\mathbf{r}_i}\mathbf{r}_j + I\!I(\mathbf{r}_i, \mathbf{r}_j)\mathbf{N}.$$

This formula extends directly to tangential vector fields X, Y, as follows

$$\partial_Y X = \nabla_Y X + I\!I(X,Y)\mathbf{N} = \nabla_Y X + \langle \partial_Y X, \mathbf{N}\rangle \mathbf{N}. \tag{4.14}$$

Exercise 4.8.15. Show that if ∇ and $\tilde{\nabla}$ are two affine connections on a manifold M, then their difference $S(X,Y) = \nabla_X Y - \tilde{\nabla}_X Y$ is an $\mathcal{F}(M)$-bilinear mapping. (In other words, S is a tensor field of type $(2,1)$). Conversely, the sum of a connection and an $\mathcal{F}(M)$-bilinear mapping $S\colon \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ is a connection. (According to these statements, affine global connections form an affine space over the linear space of tensor fields of type $(2,1)$.) ¤

Exercise 4.8.16. Show that the *torsion* $T\colon \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ of a connection ∇ defined by

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$$

is an $\mathcal{F}(M)$-bilinear mapping (i.e., T is a tensor field). ¤

Exercise 4.8.17. Show that if T is the torsion of an affine connection ∇ then $\nabla - T/2$ is a symmetric connection. ¤

Exercise 4.8.18. Check that the connection defined by the Christoffel symbols

$$\Gamma_{ij}^k = \frac{1}{2}\sum_{l=1}^{n-1}(g_{li,j} + g_{lj,i} - g_{ij,l})g^{lk}$$

is symmetric and compatible with the metric. ¤

4.9 Curvature

If ∇ is an affine connection on a manifold M, then we may consider the operator

$$R(X,Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}\colon \mathfrak{X}(M) \to \mathfrak{X}(M),$$

where $[\nabla_X, \nabla_Y] = \nabla_X \circ \nabla_Y - \nabla_Y \circ \nabla_X$ is the usual commutator of operators. The operator $R(X,Y)$, depending on X and Y, is called the *curvature operator* of the connection. The assignment

$$\begin{gathered}\mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M) \\ (X,Y,Z) \mapsto R(X,Y)(Z)\end{gathered}$$

is called the *curvature tensor* of the connection. To reduce the number of brackets, we shall denote $R(X,Y)(Z)$ simply by $R(X,Y;Z)$. Thus, the letter R is used in two different meanings, later it will denote also a third mapping, but the number of arguments of R makes always clear which meaning is considered.

Proposition 4.9.1. *The curvature tensor is linear over the ring of smooth functions in each of its arguments, and it is skew symmetric in the first two arguments.*

Proof. Skew symmetry in the first two arguments is clear, since

$$R(X,Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]} = -[\nabla_Y, \nabla_X] + \nabla_{[Y,X]} = -R(Y,X).$$

According to this, it suffices to check linearity of the curvature tensor in the first and third arguments.
Linearity in the first argument is proved by the following identities.

$$\begin{aligned}R(X_1+X_2,Y) &= [\nabla_{X_1+X_2}, \nabla_Y] - \nabla_{[X_1+X_2,Y]} \\ &= [\nabla_{X_1} + \nabla_{X_2}, \nabla_Y] - \nabla_{[X_1,Y]+[X_2,Y]} \\ &= [\nabla_{X_1}, Y] + [\nabla_{X_2}, \nabla_Y] - \nabla_{[X_1,Y]} - \nabla_{[X_2,Y]} \\ &= R(X_1,Y) + R(X_2,Y).\end{aligned}$$

and

$$\begin{aligned}R(fX,Y;Z) &= ([\nabla_{fX}, \nabla_Y] - \nabla_{[fX,Y]})(Z) \\ &= f\nabla_X\nabla_Y Z - \nabla_Y(f\nabla_X Z) - \nabla_{f[X,Y]-Y(f)X}(Z) \\ &= f\nabla_X\nabla_Y Z - f\nabla_Y\nabla_X Z - Y(f)\nabla_X Z - f\nabla_{[X,Y]}Z + Y(f)\nabla_X(Z) \\ &= f(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z) = fR(X,Y;Z).\end{aligned}$$

Additivity in the third argument is clear, since $R(X,Y)$ is built up of the additive operators ∇_X, ∇_Y and their compositions. To have linearity, we need

$$\begin{aligned}R(X,Y;fZ) &= \nabla_X\nabla_Y(fZ) - \nabla_Y\nabla_X(fZ) - \nabla_{[X,Y]}(fZ)\\ &= \nabla_X(Y(f)Z + f\nabla_Y Z) - \nabla_Y(X(f)Z + f\nabla_X Z)\\ &\quad - [X,Y](f)Z - f\nabla_{[X,Y]}Z\\ &= XY(f)Z + Y(f)\nabla_X Z + X(f)\nabla_Y Z + f\nabla_X\nabla_Y Z -\\ &\quad - YX(f)Z - X(f)\nabla_Y Z - Y(f)\nabla_X Z - f\nabla_Y\nabla_X Z -\\ &\quad - XY(f)Z + YX(f)Z - f\nabla_{[X,Y]}Z\\ &= f(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z) = fR(X,Y;Z). \qquad \square\end{aligned}$$

Proposition 4.9.1 is interesting, because the curvature tensor is built up from covariant derivations, which are not linear operators over the ring of smooth functions.

We have already introduced tensor fields over a hypersurface. We can introduce tensor fields over a manifold in the same manner. A tensor field T of type (k,l) is an assignment to every point p of a manifold M a tensor $T(p)$ of type (k,l) over the tangent space T_pM. If $\partial_1,\dots,\partial_n$ are the basis vector fields defined by a chart over the domain of the chart, and we denote by $dx^1(p),\dots,dx^n(p)$ the dual basis of $\partial_1(p),\dots,\partial_n(p)$, then a tensor field is uniquely determined over the domain of the chart by the components

$$T^{i_1\dots i_l}_{j_1\dots j_k}(p) = T(p)(dx^{i_1},\dots,dx^{i_l};\partial_{j_1},\dots,\partial_{j_k}).$$

We say that the tensor field is *smooth*, if for any chart from the atlas of M, the functions $T^{i_1\dots i_l}_{j_1\dots j_k}$ are smooth. We shall consider only smooth tensor fields. Tensor fields of type $(0,1)$ are the vector fields, tensor fields of type $(1,0)$ are the *differential* 1-*forms*. Thus, a differential 1-form assigns to every point of the manifold a linear function on the tangent space at that point. Differential 1-forms form a module over the ring of smooth functions, which we denote by $\Omega^1(M)$.

Every tensor field of type (k,l) defines an $\mathcal{F}(M)$-multilinear mapping

$$\underbrace{\Omega^1(M)\times\dots\times\Omega^1(M)}_{l\text{ times}}\times\underbrace{\mathfrak{X}(M)\times\dots\times\mathfrak{X}(M)}_{k\text{ times}}\to\mathcal{F}(M)$$

and conversely, every such $\mathcal{F}(M)$-multilinear mapping comes from a tensor field. (Check this!) Therefore, tensor fields can be identified with $\mathcal{F}(M)$-multilinear mappings $\Omega^1(M)\times\dots\times\Omega^1(M)\times\mathfrak{X}(M)\times\dots\times\mathfrak{X}(M)\to\mathcal{F}(M)$.
Tensor fields of type $(k,1)$, that is $\mathcal{F}(M)$-multilinear mappings

$$\Omega^1(M)\times\mathfrak{X}(M)\times\dots\times\mathfrak{X}(M)\to\mathcal{F}(M)$$

can be identified in a natural way with $\mathcal{F}(M)$-multilinear mappings

$$\mathfrak{X}(M)\times\dots\times\mathfrak{X}(M)\to\mathfrak{X}(M).$$

By this identification, $R\colon \mathfrak{X}(M)\times\cdots\times\mathfrak{X}(M)\to\mathfrak{X}(M)$ corresponds to $\tilde{R}\colon \Omega^1(M)\times\mathfrak{X}(M)\times\cdots\times\mathfrak{X}(M)\to\mathcal{F}(M)$, defined by $\tilde{R}(\omega;X_1,\dots,X_k)=\omega(R(X_1,\dots,X_k))$.
Using these identifications, the curvature tensor is a tensor field of type $(3,1)$ by Proposition 4.9.1. It is a remarkable consequence, that although the vectors $\nabla_X Z(p)$ and $\nabla_Y Z(p)$ are not determined by the vectors $X(p),Y(p),Z(p)$, to compute the value of $R(X,Y;Z)$ at p it suffices to know $X(p),Y(p),Z(p)$. Beside skew-symmetry in the first two arguments, the curvature tensor has many other symmetry properties.

Theorem 4.9.2 (First Bianchi Identity)**.** *If R is the curvature tensor of a torsion free connection, then*

$$R(X,Y;Z)+R(Y,Z;X)+R(Z,X;Y)=0$$

for any three vector fields X,Y,Z.

Proof. Let us introduce the following notation. If $F(X,Y,Z)$ is a function of the variables X,Y,Z, then denote by $\sum_{\circlearrowleft}F(X,Y,Z)$ or $\sum_{\circlearrowleft XYZ}F(X,Y,Z)$ the sum of the values of F at all cyclic permutations of the variables (X,Y,Z)

$$\sum_{\circlearrowleft}F(X,Y,Z)=F(X,Y,Z)+F(Y,Z,X)+F(Z,X,Y).$$

We shall use several times that behind the cyclic summation $\sum_{\circlearrowleft}$ we may cyclically rotate X,Y,Z in any expression

$$\sum_{\circlearrowleft}F(X,Y,Z)=\sum_{\circlearrowleft}F(Y,Z,X)=\sum_{\circlearrowleft}F(Z,X,Y).$$

The theorem claims vanishing of

$$\begin{aligned}\sum_{\circlearrowleft}&R(X,Y;Z)\\ &=\sum_{\circlearrowleft}(\nabla_X\nabla_Y Z-\nabla_Y\nabla_X Z-\nabla_{[X,Y]}Z)\\ &=\sum_{\circlearrowleft}(\nabla_X\nabla_Y Z-\nabla_X\nabla_Z Y-\nabla_{[X,Y]}Z)\\ &=\sum_{\circlearrowleft}(\nabla_X[Y,Z]-\nabla_{[X,Y]}Z)\\ &=\sum_{\circlearrowleft}(\nabla_Z[X,Y]-\nabla_{[X,Y]}Z)=\sum_{\circlearrowleft}[Z,[X,Y]],\end{aligned}$$

but the last expression is 0 according to the Jacobi identity on the Lie bracket of vector fields. (At the third and fifth equality we used the torsion free property of ∇.) $\square$

The presence of an affine connection on a manifold allows us to differentiate not only vector fields, but also tensor fields of any type.

Definition 4.9.3. Let (M, ∇) be a manifold with an affine connection. If $\omega \in \Omega^1(M)$ is a differential 1-form, X is a vector field, then we define the *covariant derivative* $\nabla_X\omega$ *of* ω *with respect to* X to be the 1-form

$$(\nabla_X\omega)(Y) = X(\omega(Y)) - \omega(\nabla_X Y), \quad Y \in \mathfrak{X}(M).$$

In general, the *covariant derivative* $\nabla_X T$ *of a tensor field*

$$T\colon \Omega^1(M) \times \cdots \times \Omega^1(M) \times \mathfrak{X}(M) \times \cdots \times \mathfrak{X}(M) \to \mathcal{F}(M)$$

of type (k,l) with respect to a vector field X is a tensor field of the same type, defined by

$$\begin{aligned}(\nabla_X T)(\omega_1, \dots, \omega_l; X_1, \dots, X_k) = X(T(\omega_1, \dots, \omega_l; X_1, \dots, X_k)) - \\ - \sum_{i=1}^{k} T(\omega_1, \dots, \nabla_X\omega_i, \dots, \omega_l; X_1, \dots, X_k) - \\ - \sum_{j=1}^{l} T(\omega_1, \dots, \omega_l; X_1, \dots, \nabla_X X_j, \dots, X_k).\end{aligned}$$

※

For the case of the curvature tensor, this definition gives

$$\begin{aligned}(\nabla_X R)(Y, Z; W) = \nabla_X(R(Y, Z; W)) - R(\nabla_X Y, Z; W) \\ - R(Y, \nabla_X Z; W) - R(Y, Z; \nabla_X W).\end{aligned}$$

Theorem 4.9.4 (Second Bianchi Identity)**.** *The curvature tensor of a torsion free connection satisfies*

$$\sum_{\circlearrowleft XYZ} (\nabla_X R)(Y, Z; W) = 0.$$

Proof. $(\nabla_X R)(Y, Z; W)$ is the value of the operator

$$\nabla_X \circ R(Y, Z) - R(\nabla_X Y, Z) - R(Y, \nabla_X Z) - R(Y, Z) \circ \nabla_X : \mathfrak{X}(M) \to \mathfrak{X}(M)$$

on the vector field W, hence we have to prove vanishing of the operator

$$\sum_{\circlearrowleft XYZ} \big(\nabla_X \circ R(Y, Z) - R(\nabla_X Y, Z) - R(Y, \nabla_X Z) - R(Y, Z) \circ \nabla_X\big). \quad (4.15)$$

First, we have

$$
\begin{aligned}
&\sum_{\circlearrowleft XYZ} \big(\nabla_X \circ R(Y,Z) - R(Y,Z)\circ \nabla_X\big)\\
&= \sum_{\circlearrowleft XYZ} \big((\nabla_X\nabla_Y\nabla_Z - \nabla_X\nabla_Z\nabla_Y - \nabla_X\nabla_{[Y,Z]})\\
&\qquad\qquad - (\nabla_Y\nabla_Z\nabla_X - \nabla_Z\nabla_Y\nabla_X - \nabla_{[Y,Z]}\nabla_X)\big)\\
&= \sum_{\circlearrowleft XYZ} \big((\nabla_X\nabla_Y\nabla_Z - \nabla_X\nabla_Z\nabla_Y - \nabla_X\nabla_{[Y,Z]})\\
&\qquad\qquad - (\nabla_X\nabla_Y\nabla_Z - \nabla_X\nabla_Z\nabla_Y - \nabla_{[Y,Z]}\nabla_X)\big)\\
&= \sum_{\circlearrowleft XYZ} \big(\nabla_{[Y,Z]}\nabla_X - \nabla_X\nabla_{[Y,Z]}\big).
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&\sum_{\circlearrowleft XYZ} \big(-R(\nabla_X Y, Z) - R(Y, \nabla_X Z)\big)\\
&\quad= \sum_{\circlearrowleft XYZ} \big(R(\nabla_X Z, Y) - R(\nabla_X Y, Z)\big)\\
&\quad= \sum_{\circlearrowleft XYZ} \big(R(\nabla_Y X, Z) - R(\nabla_X Y, Z)\big)\\
&\quad= \sum_{\circlearrowleft XYZ} R(\nabla_Y X - \nabla_X Y, Z) = \sum_{\circlearrowleft XYZ} R([Y,X], Z).
\end{aligned}
$$

Combining these results, operator (4.15) equals

$$
\begin{aligned}
&\sum_{\circlearrowleft XYZ} \big(\nabla_{[Y,Z]}\nabla_X - \nabla_X\nabla_{[Y,Z]} + R([Y,X],Z)\big)\\
&= \sum_{\circlearrowleft XYZ} \big(\nabla_{[Y,Z]}\nabla_X - \nabla_X\nabla_{[Y,Z]} + R[Z,Y],X)\big)\\
&= \sum_{\circlearrowleft XYZ} \big(\nabla_{[Y,Z]}\nabla_X - \nabla_X\nabla_{[Y,Z]} + \nabla_{[Z,Y]}\nabla_X - \nabla_X\nabla_{[Z,Y]} - \nabla_{[[Z,Y],X]}\big)\\
&= \sum_{\circlearrowleft XYZ} \nabla_{[[Y,Z],X]} = \nabla_{\sum_{\circlearrowleft}[[X,Y],Z]} = 0.
\end{aligned}
$$

□

Curvature of Riemannian Manifolds

In the remaining part of this unit, we shall deal with Riemannian manifolds, therefore from now on assume that $(M, \langle\,,\rangle)$ is a Riemannian manifold with Levi-Cività connection ∇ and R is the curvature tensor of ∇.

Introduce the tensor $\tilde{R}$ of type $(4,0)$, related to R by the equation

$$\tilde{R}(X,Y;Z,W) = \langle R(X,Y;Z), W\rangle,$$

which is called the *Riemann–Christoffel curvature tensor* of the Riemannian manifold. To simplify notation, we shall denote $\tilde{R}$ also by R. This will not lead to confusion, since the Riemann–Christoffel tensor and the ordinary curvature tensor have different number of arguments.
Levi-Cività connections are connections of special type, so it is not surprising, that the curvature tensor of a Riemannian manifold has stronger symmetries than that of an arbitrary connection. Of course, the general results can be applied to Riemannian manifolds as well yielding

$$R(X,Y;Z,W) = -R(Y,X;Z,W) \quad \text{and} \quad \sum_{\circlearrowleft XYZ} R(X,Y;Z,W) = 0.$$

In addition to these symmetries, we have the following ones.

Theorem 4.9.5. *The Riemann–Christoffel curvature tensor is skew-symmetric in the last two arguments*

$$R(X,Y;Z,W) = -R(X,Y;W,Z).$$

Proof. By the compatibility of the connection with the metric, we have

$$\begin{aligned} X(Y(\langle Z,W\rangle)) &= X(\langle \nabla_Y Z, W\rangle + \langle Z, \nabla_Y W\rangle) \\ &= \langle \nabla_X\nabla_Y Z, W\rangle + \langle \nabla_Y Z, \nabla_X W\rangle \\ &\quad + \langle \nabla_X Z, \nabla_Y W\rangle + \langle Z, \nabla_X\nabla_Y W\rangle, \end{aligned}$$

and similarly,

$$\begin{aligned} Y(X(\langle Z,W\rangle)) &= \langle \nabla_Y\nabla_X Z, W\rangle \\ &\quad + \langle \nabla_X Z, \nabla_Y W\rangle + \langle \nabla_Y Z, \nabla_X W\rangle + \langle Z, \nabla_Y\nabla_X W\rangle. \end{aligned}$$

We also have

$$[X,Y](\langle Z,W\rangle) = \langle \nabla_{[X,Y]} Z, W\rangle + \langle Z, \nabla_{[X,Y]} W\rangle.$$

Subtracting from the first equality the second and the third one and applying $[X,Y] = X\circ Y - Y\circ X$, we obtain

$$\begin{aligned} 0 &= \langle \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]} Z, W\rangle \\ &\quad + \langle Z, \nabla_X\nabla_Y W - \nabla_Y\nabla_X W - \nabla_{[X,Y]} W\rangle \\ &= R(X,Y;Z,W) + R(X,Y;W,Z). \end{aligned}$$

□

For tensors having the symmetries of a Riemannian curvature tensor, we introduce the following

Definition 4.9.6. Let V be a finite dimensional linear space over $\mathbb{R}$. An *algebraic curvature tensor* or *Bianchi tensor* over V is a 4-linear map $S\colon V \times V \times V \times V \to \mathbb{R}$ satisfying the symmetry relations

$$S(X,Y,Z,W) = -S(Y,X,W,Z) = -S(X,Y,W,Z)$$

and the Bianchi identity

$$S(X,Y,Z,W) + S(Y,Z,X,W) + S(Z,X,Y,W) = 0.$$

An *algebraic curvature tensor field* or *Bianchi tensor field* over a manifold M is a tensor field of type $(4,0)$, which assigns to each point p of M a Bianchi tensor over T_pM. ※

We know from linear algebra (see equation (1.8)) that a symmetric bilinear form is uniquely determined by its quadratic form. More generally, when a tensor has some symmetries, it can be reconstructed from its restriction to a suitable linear subspace of its domain. For algebraic curvature tensors, we have the following

Proposition 4.9.7. *Let S_1 and S_2 be algebraic curvature tensors (or tensor fields). If $S_1(X,Y;Y,X) = S_2(X,Y;Y,X)$ for every X and Y, then $S_1 = S_2$.*

Proof. The difference $S = S_1 - S_2$ is also an algebraic tensor (field), and $S(X,Y;Y,X) = 0$ for all X, Y. We have to show $S = 0$.
We have for any X, Y, Z

$$\begin{aligned}
0 &= S(X, Y+Z; Y+Z, X) \\
&= S(X,Y;Y,X) + S(X,Y;Z,X) + S(X,Z;Y,X) + S(X,Z;Z,X) \\
&= S(X,Y;Z,X) + S(X,Z;Y,X) + [S(X,Y;Z,X) + \underbrace{S(Y,Z;X,X)}_{=0} \\
&\quad + S(Z,X;Y,X)] = 2S(X,Y;Z,X).
\end{aligned}$$

Now taking four arbitrary vectors (vector fields) X, Y, Z, W and using $S(X,Y;Z,X) \equiv 0$, we obtain

$$\begin{aligned}
0 &= S(X+W, Y; Z, X+W) \\
&= S(X,Y;Z,X) + S(X,Y;Z,W) + S(W,Y;Z,X) + S(W,Y;Z,W) \\
&= S(X,Y;Z,W) + S(W,Y;Z,X),
\end{aligned}$$

i.e., S is skew symmetric in the first and fourth variables. Thus,

$$S(X,Y;Z,W) = S(Y,X;W,Z) = -S(Z,X;W,Y) = S(Z,X;Y,W),$$

in other words, S is invariant under cyclic permutations of the first three variables. But the sum of the three equal quantities $S(X,Y;Z,W)$, $S(Y,Z;X,W)$ and $S(Z,X;Y,W)$ is 0 because of the Bianchi symmetry, thus $S(X,Y;Z,W)$ is 0. □

Exercise 4.9.8. Let S be an algebraic curvature tensor, and let $Q_S(X,Y) := S(X,Y;Y,X)$. Prove that $Q_S(X,Y) = Q_S(Y,X)$ and

$$\begin{aligned}6S(X,Y;Z,W) =& Q_S(X+W,Y+Z) - Q_S(Y+W,X+Z) + Q_S(Y+W,X) - \\ &- Q_S(X+W,Y) + Q_S(Y+W,Z) - Q_S(X+W,Z) + \\ &+ Q_S(X+Z,Y) - Q_S(Y+Z,X) + Q_S(X+Z,W) - \\ &- Q_S(Y+Z,W) + Q_S(X,Z) - \\ &- Q_S(Y,Z) + Q_S(Y,W) - Q_S(X,W).\end{aligned}$$ ¤

Theorem 4.9.9. *Assume that S is an algebraic curvature tensor. Then*

$$S(X,Y;Z,W) = S(Z,W;X,Y).$$

Proof. Label the vertices of an octahedron as shown in the figure. It follows

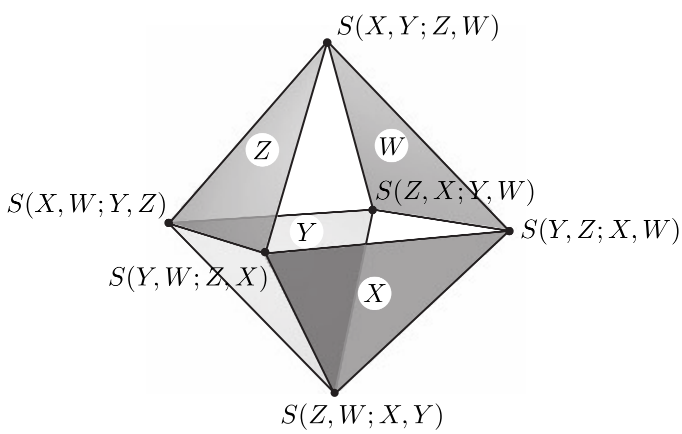

Figure 4.5

easily from our assumptions that the sum of the labels of the vertices of any of the shaded faces of the octahedron is 0 by a Bianchi identity, in which the label of the face stands in the last place and the sum goes over cyclic permutations of the other three variables. If we add these identities for the upper two shaded triangles "Z" and "W" and subtract the identities corresponding to the two lower triangles "X" and "Y", we obtain

$$2S(X,Y;Z,W) - 2S(Z,W;X,Y) = 0,$$

as we wanted to prove. □

Corollary 4.9.10. *For the Riemann-Christoffel tensor, the identity* $R(X,Y;Z,W) = R(Z,W;X,Y)$ *holds.*

Definition 4.9.11. Let M be a Riemannian manifold, p a point on M, X and Y two non-parallel tangent vectors at p. The number

$$K(X,Y) = \frac{R(X,Y;Y,X)}{|X|^2|Y|^2 - \langle X,Y\rangle^2}$$

is called the *sectional curvature* of M at p, in the direction of the plane spanned by the vectors X and Y in T_pM. ※

Proposition 4.9.12. *If M is a Riemannian manifold with sectional curvature K, X and Y are linearly independent tangent vectors at $p \in M$, α and β are non-zero scalars, then*

(i) $K(X,Y) = K(X+Y,Y)$;

(ii) $K(X,Y) = K(\alpha X, \beta Y)$;

(iii) $K(X,Y) = K(Y,X)$.

Furthermore, K depends only on the plane spanned by its arguments.

Proof. (i) follows from

$$\begin{aligned} R(X+Y,Y;Y,X+Y) &= R(X,Y;Y,X) + R(X,Y;Y,Y) + R(Y,Y;Y,X) \\ &\quad + R(Y,Y;Y,Y) = R(X,Y;Y,X) \end{aligned}$$

and

$$\begin{aligned} &|X+Y|^2|Y|^2 - \langle X+Y,Y\rangle^2 \\ &\quad = (|X|^2 + |Y|^2 + 2\langle X,Y\rangle)|Y|^2 - (\langle X,Y\rangle^2 + 2\langle X,Y\rangle|Y|^2 + |Y|^4) \\ &\quad = |X|^2|Y|^2 - \langle X,Y\rangle^2. \end{aligned}$$

(ii) follows from

$$R(\alpha X, \beta Y; \beta Y, \alpha X) = \alpha^2\beta^2 R(X,Y;Y,X)$$

and

$$|\alpha X|^2|\beta Y|^2 - \langle \alpha X, \beta Y\rangle^2 = \alpha^2\beta^2(|X|^2|Y|^2 - \langle X,Y\rangle^2).$$

(iii) comes from the equalities $R(X, Y; Y, X) = R(Y, X; X, Y)$ and $|X|^2|Y|^2 - \langle X, Y\rangle^2 = |Y|^2|X|^2 - \langle Y, X\rangle^2$.
Finally, the last statement follows by (i), (ii), and (iii) as if $\mathbf{x}_1, \mathbf{y}_1$ and $\mathbf{x}_2, \mathbf{y}_2$ are two bases of a 2-dimensional linear space, then we can transform one of them into the other by a finite number of elementary basis transformations of the form

$$(\mathbf{x}, \mathbf{y}) \rightsquigarrow (\mathbf{x}+\mathbf{y}, \mathbf{y}); \quad (\mathbf{x}, \mathbf{y}) \rightsquigarrow (\alpha\mathbf{x}, \beta\mathbf{y}), \text{ where } \alpha\beta \neq 0; \quad (\mathbf{x}, \mathbf{y}) \rightsquigarrow (\mathbf{y}, \mathbf{x}).$$

□

Definition 4.9.13. Riemannian manifolds, the sectional curvature function of which is constant, called *spaces of constant curvature* or simply *space forms*. A space form is *elliptic* or *spherical* if $K > 0$, it is *parabolic* or *Euclidean* if $K = 0$ and is *hyperbolic* if $K < 0$. ※

Typical examples are the n-dimensional sphere, Euclidean space and hyperbolic space. Further examples can be obtained by factorization with fixed point free actions of discrete groups.
The following remarkable theorem resembles Theorem 3.1.25, but its proof is not so simple.

Theorem 4.9.14 (Schur). *If M is a connected Riemannian manifold,* $\dim M \geq 3$ *and the sectional curvature $K(X_p, Y_p)$ $(X_p, Y_p \in T_pM)$ depends only on p (and does not depend on the plane spanned by X_p and Y_p), then K is constant, that is, as a matter of fact, it does not depend on p either.*

Proof. By the assumption,

$$R(X, Y; Y, X) = f(|X|^2|Y|^2 - \langle X, Y\rangle^2)$$

for some function f. Our goal is to show that f is constant. Consider the tensor field of type $(4, 0)$ defined by

$$S(X, Y; Z, W) = f(\langle X, W\rangle\langle Y, Z\rangle - \langle X, Z\rangle\langle Y, W\rangle).$$

It is clear from the definition that S is skew-symmetric in the first and last two arguments. It has also the Bianchi symmetry — indeed,

$$\begin{aligned}\sum_{\circlearrowleft XYZ} S(X, Y; Z, W) &= \sum_{\circlearrowleft XYZ} f(\langle X, W\rangle\langle Y, Z\rangle - \langle X, Z\rangle\langle Y, W\rangle) \\ &= \sum_{\circlearrowleft XYZ} f(\langle Y, W\rangle\langle Z, X\rangle - \langle X, Z\rangle\langle Y, W\rangle) = 0,\end{aligned}$$

thus S is an algebraic curvature tensor field. We also have $R(X, Y; Y, X) = S(X, Y; Y, X)$, therefore $R = S$. Set

$$\tilde{S}(X, Y; Z) = f(\langle Y, Z\rangle X - \langle X, Z\rangle Y).$$

Then for any vector field W, we have

$$\langle R(X,Y;Z),W\rangle = R(X,Y;Z,W) = S(X,Y;Z,W) = \langle \tilde{S}(X,Y;Z),W\rangle,$$

that is,

$$R(X,Y;Z) = \tilde{S}(X,Y;Z) \text{ for all } X,Y,Z.$$

Differentiating with respect to a vector field U we get

$$\begin{aligned}(\nabla_U R)(X,Y;Z) &= (\nabla_U \tilde{S})(X,Y;Z)\\ &= \nabla_U(\tilde{S}(X,Y;Z)) - \tilde{S}(\nabla_U X,Y;Z)\\ &\quad - \tilde{S}(X,\nabla_U Y;Z) - \tilde{S}(X,Y;\nabla_U Z).\end{aligned}$$

Since

$$\begin{aligned}\nabla_U(\tilde{S}(X,Y;Z)) &= U(f)(\langle Y,Z\rangle X - \langle X,Z\rangle Y) + f\nabla_U(\langle Y,Z\rangle X - \langle X,Z\rangle Y)\\ &= U(f)(\langle Y,Z\rangle X - \langle X,Z\rangle Y) + f(U\langle Y,Z\rangle X + \langle Y,Z\rangle \nabla_U X\\ &\quad - U\langle X,Z\rangle Y - \langle X,Z\rangle \nabla_U Y)\\ &= U(f)(\langle Y,Z\rangle X - \langle X,Z\rangle Y) + f(\langle \nabla_U Y,Z\rangle X + \langle Y,\nabla_U Z\rangle X +\\ &\quad + \langle Y,Z\rangle \nabla_U X - \langle \nabla_U X,Z\rangle Y - \langle X,\nabla_U Z\rangle Y - \langle X,Z\rangle \nabla_U Y) =\\ &= U(f)(\langle Y,Z\rangle X - \langle X,Z\rangle Y) + \tilde{S}(\nabla_U X,Y;Z) + \tilde{S}(X,\nabla_U Y;Z) +\\ &\quad + \tilde{S}(X,Y;\nabla_U Z),\end{aligned}$$

we obtain

$$(\nabla_U R)(X,Y;Z) = (\nabla_U \tilde{S})(X,Y;Z) = U(f)(\langle Y,Z\rangle X - \langle X,Z\rangle Y).$$

Using the second Bianchi identity, this gives us

$$\sum_{\circlearrowleft UXY} U(f)(\langle Y,Z\rangle X - \langle X,Z\rangle Y) = \sum_{\circlearrowleft UXY} (\nabla_U R)(X,Y;Z) = 0.$$

If $X \in T_pM$ is an arbitrary tangent vector, then we can find non-zero vectors Y, $Z = U \in T_pM$ such that X, Y and U are orthogonal ($\dim M \geq 3$!). Then

$$0 = \sum_{\circlearrowleft UXY} U(f)(\langle Y,Z\rangle X - \langle X,Z\rangle Y) = X(f)\langle U,U\rangle Y - Y(f)\langle U,U\rangle X.$$

Since X and Y are linearly independent and $\langle U,U\rangle$ is positive, $X(f) = Y(f) = 0$ follows, yielding that the derivative of f with respect to an arbitrary tangent vector X is 0. This means that f is locally constant, and since M is connected, f is constant. □

The curvature tensor is a complicated object containing a lot of information about the geometry of the manifold. There are some obvious ways to derive simpler tensor fields from the curvature tensor. Of course, simplicity is paid by losing information.

Definition 4.9.15. Let (M, ∇) be a manifold with an affine connection, R be the curvature tensor of ∇. The *Ricci tensor* Ric of the connection is a tensor field of type $(2,0)$ assigning to the vector fields X and Y the function $\mathrm{Ric}(X,Y)$ the value of which at $p \in M$ is the trace of the linear mapping

$$\begin{aligned} T_pM &\to T_pM \\ Z_p &\mapsto R(Z_p, X(p); Y(p)), \text{ where } Z_p \in T_pM. \end{aligned}$$

※

Proposition 4.9.16. *The Ricci tensor of a Riemannian manifold is a symmetric tensor*

$$\mathrm{Ric}(X,Y) = \mathrm{Ric}(Y,X).$$

Proof. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be an orthonormal basis in T_pM, where p is an arbitrary point in the Riemannian manifold M. We can compute the trace of a linear mapping $A\colon T_pM \to T_pM$ by the formula

$$\mathrm{Tr}\, A = \sum_{i=1}^{n} \langle A(\mathbf{e}_i), \mathbf{e}_i \rangle.$$

In particular,

$$\begin{aligned} \mathrm{Ric}(X,Y)(p) &= \sum_{i=1}^{n} \langle R(\mathbf{e}_i, X(p); Y(p)), \mathbf{e}_i \rangle = \sum_{i=1}^{n} R(\mathbf{e}_i, X(p); Y(p), \mathbf{e}_i) \\ &= \sum_{i=1}^{n} R(Y(p), \mathbf{e}_i; \mathbf{e}_i, X(p)) = \sum_{i=1}^{n} R(\mathbf{e}_i, Y(p); X(p), \mathbf{e}_i) \\ &= \mathrm{Ric}(Y,X)(p). \end{aligned}$$

□

Since the Ricci tensor of a Riemannian manifold is symmetric, it is uniquely determined by its quadratic form $X \to \mathrm{Ric}(X,X)$ (see equation (1.8)).

Definition 4.9.17. Let $X_p \in T_pM$ be a non-zero tangent vector of a Riemannian manifold M. The *Ricci curvature* of M at p in the direction X_p is the number

$$r(X_p) = \frac{\mathrm{Ric}(X_p, X_p)}{|X_p|^2}.$$

※

Fixing an orthonormal basis $\frac{X_p}{|X_p|} = \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ we can express the Ricci curvature as follows

$$r(X_p) = \frac{\mathrm{Ric}(X_p, X_p)}{|X_p|^2} = \sum_{i=1}^{n} \frac{R(\mathbf{e}_i, X_p; X_p, \mathbf{e}_i)}{|X_p|^2} = \sum_{i=2}^{n} K(X_p, \mathbf{e}_i).$$

The meaning of this formula is that the Ricci curvature in the direction X_p is the sum of the sectional curvatures in the directions of the planes spanned by the vectors X_p and $\mathbf{e}_i$, where $\mathbf{e}_i$ runs over an orthonormal basis of the orthogonal complement of X_p in T_pM. It is a nice geometrical corollary that this sum is independent of the choice of the orthogonal basis.
With the help of a scalar product, one can associate to every bilinear function a linear transformation. For the case of the Riemannian metric and the Ricci tensor, we can find a unique $\mathcal{F}(M)$-linear transformation $\overline{\mathrm{Ric}}\colon \mathfrak{X}(M) \to \mathfrak{X}(M)$ defined by

$$\mathrm{Ric}(X,Y) = \langle X, \overline{\mathrm{Ric}}(Y)\rangle \text{ for every } X, Y \in \mathfrak{X}(M).$$

Definition 4.9.18. The *scalar curvature* $s(p)$ of a Riemannian manifold M at a point p is the trace of the linear mapping $\overline{\mathrm{Ric}}\colon T_pM \to T_pM$. ※

Let us find an expression for the scalar curvature in terms of the Ricci curvature and the sectional curvature. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be an orthonormal basis in T_pM. Then

$$s(p) = \mathrm{Tr}\,\overline{\mathrm{Ric}} = \sum_{i=1}^{n} \langle \overline{\mathrm{Ric}}(\mathbf{e}_i), \mathbf{e}_i\rangle = \sum_{i=1}^{n} \mathrm{Ric}(\mathbf{e}_i, \mathbf{e}_i) = \sum_{i=1}^{n} r(\mathbf{e}_i),$$

i.e. $s(p)$ is the sum of Ricci curvatures in the directions of an orthogonal basis. Furthermore,

$$s(p) = \sum_{i=1}^{n} r(\mathbf{e}_i) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} K(\mathbf{e}_i, \mathbf{e}_j) = 2 \sum_{1\le i<j\le n} K(\mathbf{e}_i, \mathbf{e}_j),$$

that is, the scalar curvature is twice the sum of sectional curvatures taken in the directions of all coordinate planes of an orthonormal coordinate system in T_pM.
To finish this section with, let us study the curvature tensor of a hypersurface M in $\mathbb{R}^n$. According to (4.14), the Levi-Cività connection ∇ of a hypersurface can be expressed as $\nabla = P \circ \partial$, where ∂ is the derivation rule of vector fields along the hypersurface as defined in Definition 3.1.7, P is the orthogonal projection of a tangent vector of $\mathbb{R}^n$ at a hypersurface point onto the tangent space of the hypersurface at that point. Comparing Definition 3.1.7 to formula (4.10), we see that the derivation ∂ of vector fields along a hypersurface is induced by the Levi-Cività connection of $\mathbb{R}^n$, which we also denoted by ∂. As the curvature of $\mathbb{R}^n$ is 0,

$$\partial_X \circ \partial_Y - \partial_Y \circ \partial_X = \partial_{[X,Y]}$$

holds for any tangential vector fields $X, Y \in \mathfrak{X}(M)$.

We have

$$\begin{aligned}\nabla_X\nabla_Y Z &= P(\partial_X\nabla_Y Z) = P(\partial_X(\partial_Y Z - \langle\partial_Y Z,\mathbf{N}\rangle\mathbf{N}))\\ &= P(\partial_X\partial_Y Z) - P(X(\langle\partial_Y Z,\mathbf{N}\rangle)\mathbf{N}) - P(\langle\partial_Y Z,\mathbf{N}\rangle\partial_X\mathbf{N})\\ &= P(\partial_X\partial_Y Z) - \langle\partial_Y Z,\mathbf{N}\rangle\partial_X\mathbf{N},\end{aligned}$$

where $X, Y, Z \in \mathfrak{X}(M)$. Similarly,

$$\nabla_Y\nabla_X Z = P(\partial_Y\partial_X Z) - \langle\partial_X Z,\mathbf{N}\rangle\partial_Y\mathbf{N}.$$

Combining these equalities with

$$\nabla_{[X,Y]}Z = P(\partial_{[X,Y]}Z)$$

we get the following expression for the curvature tensor R of M

$$\begin{aligned}R(X,Y;Z) &= (\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z) - \nabla_{[X,Y]}Z =\\ &= P((\partial_X\partial_Y Z - \partial_Y\partial_X Z) - \partial_{[X,Y]}Z)\\ &\quad - \langle\partial_Y Z,\mathbf{N}\rangle\partial_X\mathbf{N} + \langle\partial_X Z,\mathbf{N}\rangle\partial_Y\mathbf{N}\\ &= \langle\partial_X Z,\mathbf{N}\rangle\partial_Y\mathbf{N} - \langle\partial_Y Z,\mathbf{N}\rangle\partial_X\mathbf{N}.\end{aligned}$$

Since $\langle Z,\mathbf{N}\rangle$ is constant zero,

$$0 = X(\langle Z,\mathbf{N}\rangle) = \langle\partial_X Z,\mathbf{N}\rangle + \langle Z,\partial_X\mathbf{N}\rangle$$

and

$$0 = Y(\langle Z,\mathbf{N}\rangle) = \langle\partial_Y Z,\mathbf{N}\rangle + \langle Z,\partial_Y\mathbf{N}\rangle.$$

Putting these equalities together we deduce that

$$\begin{aligned}R(X,Y;Z) &= \langle Z,\partial_Y\mathbf{N}\rangle\partial_X\mathbf{N} - \langle Z,\partial_X\mathbf{N}\rangle\partial_Y\mathbf{N}\\ &= \langle Z,L(Y)\rangle L(X) - \langle Z,L(X)\rangle L(Y).\end{aligned}$$

Comparing the formula

$$R(X,Y;Z) = \langle Z,L(Y)\rangle L(X) - \langle Z,L(X)\rangle L(Y)$$

relating the curvature tensor to the Weingarten map on a hypersurface with Gauss' equations we see that the curvature tensor R coincides with the curvature tensor defined there. This way, the last equation can also be considered as a coordinate free display of Gauss' equations.

Exercise 4.9.19. Prove that if X_1 and X_2 are two nonparallel principal directions at a given point p of a hypersurface M, κ_1, κ_2 are the corresponding principal curvatures, then

$$K(X_1, X_2) = \kappa_1\kappa_2.$$

What is the minimum and maximum of $K(X,Y)$, when X and Y run over T_pM? ¤

Exercise 4.9.20. Express the Ricci curvature of a hypersurface in $\mathbb{R}^{n+1}$ in a principal direction in terms of the principal curvatures. ¤

Exercise 4.9.21. Express the scalar curvature of a hypersurface in terms of the principal curvatures. ¤

4.10 Decomposition of Algebraic Curvature Tensors

Proposition 4.10.1. *Algebraic curvature tensors over an n-dimensional linear space V form a linear subspace $\mathcal{R}$ of the tensor space $T^{(4,0)}V$, the dimension of which is $(n^4 - n^2)/12$.*

Proof. Skew symmetry in the first two and last two variables implies that every algebraic curvature R tensor can be obtained from a uniquely defined bilinear function $\hat{R}$ on $\Lambda^2 V$ by $R(X, Y, Z, W) = \hat{R}(X\wedge Y, Z\wedge W)$. By Theorem 4.9.9, $\hat{R}$ must also be a symmetric. bilinear function.
Let $\tilde{\mathcal{R}} \subset T^{(4,0)}V$ be the linear space of 4-linear functions S which have the symmetry properties

$$S(X, Y, Z, W) = -S(Y, X, Z, W) = -S(X, Y, W, Z) = S(Z, W, X, Y).$$

It is clear that $\tilde{\mathcal{R}}$ is isomorphic to the space of symmetric bilinear functions on $\Lambda^2 V$. Since $\dim \Lambda^2 V = \frac{n(n-1)}{2}$,

$$\dim \tilde{\mathcal{R}} = \frac{\frac{n(n-1)}{2} \cdot \left(\frac{n(n-1)}{2} + 1\right)}{2} = \frac{n(n-1)(n^2 - n + 2)}{8}.$$

By Theorem 4.9.9 $\mathcal{R} \subset \tilde{\mathcal{R}}$, and $\mathcal{R}$ can be given as the kernel of the map $B\colon \tilde{\mathcal{R}} \to \tilde{\mathcal{R}}$ defined by

$$B(S)(X, Y, Z, W) = \sum_{\circlearrowleft XYZ} S(X, Y, Z, W).$$

The image of B contains only alternating tensors. Indeed, $B(S)(X, Y, Z, W)$ is clearly invariant under cyclic permutations of (X, Y, Z), furthermore,

$$\begin{aligned} B(S)(X, Y, Z, W) &= S(X, Y, Z, W) + S(Y, Z, X, W) + S(Z, X, Y, W) \\ &= -S(X, Y, W, Z) - S(W, X, Y, Z) - S(Y, W, X, Z) \\ &= -B(X, Y, W, Z), \end{aligned}$$

and the cyclic permutation of (X, Y, Z) and the transposition of (Z, W) generates the full permutation group of (X, Y, Z, W).

On the other hand, if S is alternating, then $B(S) = 3S$, so the image of B contains all alternating tensors, that is, $\operatorname{im} B = A^4(V)$. Applying the dimension formula (1.2), we obtain

$$\begin{aligned}\dim \mathcal{R} &= \dim \tilde{\mathcal{R}} - \dim \operatorname{im} B \\ &= \frac{n^4 - 2n^3 + 3n^2 - 2n}{8} - \frac{n(n-1)(n-2)(n-3)}{24} = \frac{n^4 - n^2}{12}.\end{aligned}$$

□

Assume that V is a Euclidean linear space and denote by g the inner product of V. Then the orthogonal group $O(V)$ acts on $\mathcal{R}$ by

$$(\Phi(R))(X, Y; Z, W) = R(\Phi^{-1}(X), \Phi^{-1}(Y); \Phi^{-1}(Z), \Phi^{-1}(W)).$$

We want to study the question how to decompose $\mathcal{R}$ into irreducible $O(V)$-invariant subspaces. We can find $O(V)$-invariant subspaces in $\mathcal{R}$ with the help of the Kulkarni–Nomizu product.

Definition 4.10.2. Let V be a finite dimensional linear space, $h, k \in S^2(V)$ be two symmetric bilinear functions on V. Then the *Kulkarni–Nomizu product* of h and k is the 4-linear map $h \oslash k \in T^{(4,0)}(V)$ given by

$$\begin{aligned} h \oslash k(X, Y; Z, W) &= \begin{vmatrix} h(X,Z) & h(X,W) \\ k(Y,Z) & k(Y,W) \end{vmatrix} + \begin{vmatrix} k(X,Z) & k(X,W) \\ h(Y,Z) & h(Y,W) \end{vmatrix} \\ &= h(X,Z)k(Y,W) - h(X,W)k(Y,Z) + k(X,Z)h(Y,W) - k(X,W)h(Y,Z). \end{aligned}$$

※

The Kulkarni–Nomizu product is an $\mathbb{R}$-bilinear commutative operation. Its relevance to algebraic curvature tensors is that the Kulkarni–Nomizu product of two symmetric bilinear functions is always an algebraic curvature tensor.

Proposition 4.10.3. *$S^2(V) \oslash g = \{h \oslash g \mid h \in S^2(V)\}$ is an $O(V)$ invariant linear subspace of $\mathcal{R}$.*

Proof. Since $h \oslash g$ is linear in h, $S^2(V) \oslash g$ is a linear subspace. The orthogonal group $O(V)$ has a natural action also on $S^2(V)$, defined by $\Phi(h)(X, Y) = h(\Phi^{-1}X, \Phi^{-1}Y)$. If $\Phi \in O(V)$ is an orthogonal transformation, then $\Phi(h \oslash k) = \Phi(h) \oslash \Phi(k)$, and since $\Phi(g) = g$, $\Phi(h \oslash g) = \Phi(h) \oslash g$. As a consequence $\Phi(S^2(V) \otimes g) = S^2(V) \otimes g$. □

Now we construct an $O(V)$-invariant subspace in $\mathcal{R}$ complementary to $S^2(V) \oslash g$.

Definition 4.10.4. The *trace of a bilinear function* $h \in S^2(V)$ defined on a *Euclidean linear space* V with inner product g is $\operatorname{tr}(h) = \langle h, g\rangle$. ※

Proposition 4.10.5. *If* $\mathbf{e}_1, \dots, \mathbf{e}_n$ *is an orthonormal basis of* V, $h \in S^2(V)$, *then* $\operatorname{tr}(h) = \sum_{i=1}^n h(\mathbf{e}_i, \mathbf{e}_i)$. *In particular, the sum on the right-hand side does not depend on the choice of the orthonormal basis.*

Proof. Using the natural embedding $S^2(V) \to T^{(2,0)}V$, if $\mathbf{e}^1, \dots \mathbf{e}^n$ is the dual basis of the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, then

$$\langle h, g\rangle = \left\langle \sum_{i,j=1}^n h(\mathbf{e}_i, \mathbf{e}_j)\mathbf{e}^i \otimes \mathbf{e}^j, \sum_{k=1}^n \mathbf{e}^k \otimes \mathbf{e}^k \right\rangle = \sum_{i=1}^n h(\mathbf{e}_i, \mathbf{e}_i). \qquad \square$$

Definition 4.10.6. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be an orthonormal basis of V and take $R \in \mathcal{R}$. The *Ricci tensor of* R is the bilinear function $\operatorname{Ric}(R) \in T^{(2,0)}(V)$ given by

$$\operatorname{Ric}(R)(X, Y) = \sum_{i=1}^n R(X, \mathbf{e}_i, \mathbf{e}_i, Y). \qquad ※$$

Proposition 4.10.7.

- *The Ricci tensor of* R *does not depend on the choice of the orthonormal basis.*
- $\operatorname{Ric}(R)$ *is symmetric and the linear map* $\operatorname{Ric}\colon \mathcal{R} \to S^2(V)$ *is* $O(V)$ *equivariant, that is,* $\operatorname{Ric}(\Phi(R)) = \Phi(\operatorname{Ric} R))$ *for any* $R \in \mathcal{R}$.
- $\operatorname{Ric}(h \oslash g) = (2-n)h - \operatorname{tr}(h)g$, *in particular,* $\operatorname{Ric}(g \oslash g) = 2(1-n)g$ *and* $\operatorname{tr}(\operatorname{Ric}(h \oslash g)) = 2(1-n)\operatorname{tr}(h)$.

Proof. $\operatorname{Ric}(R)(X, Y)$ can also be obtained as the trace of the bilinear function $h(Z, W) = R(X, Z, W, Y)$, thus it does not depend on the choice of orthonormal basis.
Symmetry of $\operatorname{Ric}(R)$ follows from

$$R(X, \mathbf{e}_i, \mathbf{e}_i, Y) = R(\mathbf{e}_i, Y, X, \mathbf{e}_i) = (-1)^2 R(Y, \mathbf{e}_i, \mathbf{e}_i, X) = R(Y, \mathbf{e}_i, \mathbf{e}_i, X).$$

The $O(V)$ equivariance follows from the first part of the proposition and the fact that an orthogonal transformation maps an orthonormal basis to an

$$\begin{aligned}\operatorname{Ric}(\Phi(R))(X, Y) &= \sum_{i=1}^n R(\Phi^{-1}(X), \Phi^{-1}(\mathbf{e}_i), \Phi^{-1}(\mathbf{e}_i), \Phi^{-1}(Y)) \\ &= \sum_{i=1}^n (R(\Phi^{-1}(X), \mathbf{e}_i, \mathbf{e}_i, \Phi^{-1}(Y)) = \Phi(\operatorname{Ric}(R))(X, Y).\end{aligned}$$

Finally, using the decompositions $X = \sum_i^n g(X, \mathbf{e}_i)\mathbf{e}_i$ and $Y = \sum_i^n g(Y, \mathbf{e}_i)\mathbf{e}_i$, we obtain

$$\begin{aligned}
&\operatorname{Ric}(h \oslash g)(X, Y) \\
&\qquad = \sum_i^n (h \oslash g)(X, \mathbf{e}_i, \mathbf{e}_i, Y) \\
&\qquad = \sum_i^n h(X, \mathbf{e}_i) g(\mathbf{e}_i, Y) - h(X, Y) g(\mathbf{e}_i, \mathbf{e}_i) \\
&\qquad\quad + g(X, \mathbf{e}_i) h(\mathbf{e}_i, Y) - g(X, Y) h(\mathbf{e}_i, \mathbf{e}_i) \\
&\qquad = (2 - n) h(X, Y) - \operatorname{tr}(h) g(X, Y).
\end{aligned}$$

□

Definition 4.10.8. Algebraic curvature tensors with vanishing Ricci curvature are called *algebraic Weyl tensors*. They form an $O(V)$-invariant linear subspace $\mathcal{W}$ of $\mathcal{R}$. ※

Proposition 4.10.9. *The space of algebraic curvature tensors decomposes as*

$$\mathcal{R} = (S^2(V) \oslash g) \oplus \mathcal{W}.$$

Proof. If $\dim V = 1$, then there is nothing to prove as $\mathcal{R}$ is 0-dimensional. For $\dim V = 2$, $\dim \mathcal{R} = 1$, thus $\mathcal{R} = \mathbb{R} g \oslash g = S^2(V) \oslash g$. Since $\operatorname{Ric}(g \oslash g) = -2g \neq 0$, $\mathcal{W}$ is 0-dimensional.

Assume now that $\dim V = n > 2$. Then for any $h \in S^2(V)$, we have

$$h = \frac{\operatorname{Ric}(h \oslash g) + \operatorname{tr}(h) g}{2 - n} = \frac{\operatorname{Ric}(h \oslash g)}{2 - n} + \frac{\operatorname{tr} \operatorname{Ric}(h \oslash g) g}{2(1 - n)(2 - n)},$$

therefore the linear map $P \colon \mathcal{R} \to S^2(V) \oslash g$ defined by

$$P(R) = \left(\frac{\operatorname{Ric}(R)}{2 - n} + \frac{\operatorname{tr} \operatorname{Ric}(R) g}{2(1 - n)(2 - n)} \right) \oslash g \tag{4.16}$$

is a projection onto $S^2(V) \oslash g$, that is, its restriction onto $S^2(V) \oslash g$ is the identity. This implies that $\mathcal{R} = (S^2(V) \oslash g) \oplus \ker P$. It is also clear from the definition of $\mathcal{W}$ that $\ker P \subset \mathcal{W}$. Conversely, if R is in the kernel of the projection, then

$$2(n - 1) \operatorname{Ric}(R) = \operatorname{tr}(\operatorname{Ric} R) g.$$

Taking the trace of both sides $2(n-1) \operatorname{tr}(\operatorname{Ric}(R)) = n \operatorname{tr}(\operatorname{Ric}(R))$ is obtained, from which $\operatorname{tr}(\operatorname{Ric} R) = 0$ and $\operatorname{Ric}(R) = 0$, thus, $R \in \mathcal{W}$. As $\mathcal{R} = (S^2(V) \oslash g) \oplus \ker P$, equation $\ker P = \mathcal{W}$ completes the proof. □

Corollary 4.10.10. *The components of an algebraic curvature tensor* $R \in \mathcal{R}$*, corresponding to the decomposition* $\mathcal{R} = (S^2(V) \oslash g) \oplus \mathcal{W}$ *are* $P(R) \in S^2(V) \oslash g$ *defined in* (4.16)*, and*

$$W(R) = R - P(R) = R - \left(\frac{\mathrm{Ric}(R)}{2-n} + \frac{\mathrm{tr}\,\mathrm{Ric}(R) g}{2(1-n)(2-n)}\right) \oslash g \in \mathcal{W}. \quad (4.17)$$

Definition 4.10.11. We call $P(R)$ the *Ricci component* of the algebraic curvature tensor R as it depends only on the Ricci tensor $\mathrm{Ric}(R)$. The algebraic Weyl tensor $W(R) \in \mathcal{W}$ given by equation (4.17) is called the *Weyl tensor* or *Weyl component of R.* ※

The action of $O(V)$ on $S^2(V)$ is not irreducible. Since $\Phi(g) = g$ for any $\Phi \in O(V)$, the one dimensional subspace spanned by g is an $O(V)$ invariant subspace. The trace is an $O(V)$ invariant linear function on $S^2(V)$, so its kernel $S_0^2(V) = \{h \in S^2(V) \mid \mathrm{tr}(h) = 0\}$, the linear space of 0-trace symmetric bilinear functions, is an $O(V)$ invariant linear subspace of codimension 1.

Proposition 4.10.12. *If* V *is an* n*-dimensional Euclidean linear space with inner product* g*, then the linear space* $S^2(V)$ *splits into the direct sum of the* $O(V)$ *invariant subspaces* $\mathbb{R}g$ *and* $S_0^2(V)$*. The components of* $h \in S^2(V)$ *corresponding to this direct sum decomposition are* $(\mathrm{tr}(h)/n)g \in \mathbb{R}$ *and* $h - (\mathrm{tr}(h)/n)g \in S_0^2(V)$.

We leave the proof as an exercise to the reader.

Definition 4.10.13. The *scalar curvature of an algebraic curvature tensor* R is the trace $s(R) = \mathrm{tr}(\mathrm{Ric}(R))$ of the Ricci tensor of R. ※

Applying Proposition 4.10.12 to decompose the Ricci component of a curvature tensor, we obtain the following theorem summarizing the above calculations.

Theorem 4.10.14. *Let* V *be a Euclidean linear space with inner product* g.

- *If* $\dim V = 2$*, then* $\mathcal{R} = \mathbb{R}g \oslash g$*, therefore, every algebraic curvature tensor* R *can be written as* $R = -\frac{s(R)}{4} g \oslash g$ *and* $S_0^2(V) \oslash g$ *and* $\mathcal{W}$ *are* 0*-dimensional.*

- *If* $\dim V \geq 3$*, then the space of algebraic curvature tensors decomposes into the direct sum of the* $O(V)$ *invariant subspaces* $\mathbb{R}g \oslash g$*,* $S_0^2(V) \oslash g$ *and* $\mathcal{W}$*. The corresponding decomposition of an algebraic curvature tensor* R *is*

$$R = \frac{s(R)}{2(1-n)n} g \oslash g + \frac{1}{2-n}\left(\mathrm{Ric}(R) - \frac{s(R)}{n} g\right) \oslash g + W(R).$$

- *In the special case when* $\dim V = 3$*,* $\dim \mathcal{R} = 6 = \dim S^2(V)$*, consequently,* $\dim \mathcal{W} = 0$*, so the Weyl component of any algebraic curvature tensor is* 0*.*

Definition 4.10.15. The component $\frac{s(R)}{2(1-n)n} g \oslash g \in \mathbb{R} g \oslash g$, depending only on the scalar curvature of R is called the *scalar component of* R, while the second component $\left(\mathrm{Ric}(R) - \frac{s(R)}{n} g\right) \oslash g \in S_0^2(V) \oslash g$ is the *trace free Ricci component of* R as it comes from the trace free part $\mathrm{Ric}(R) - \frac{s(R)}{n} g$ of the Ricci tensor of R. ※

We mention without proof that the representations of the group $O(V)$ on the spaces $\mathbb{R} g \oslash g$, $S_0^2(V) \oslash g$ and $\mathcal{W}$ are irreducible, so we cannot decompose these spaces into smaller $O(V)$ invariant subspaces.

Consider a Riemannian manifold (M, g). The Riemannian metric g turns each tangent space T_pM into a Euclidean linear space and the Riemannian curvature tensor R yields an algebraic curvature tensor on T_pM for all $p \in M$. Applying the above decomposition of algebraic curvature tensors pointwise, we obtain a decomposition of the Riemannian curvature tensor into three components, some of which automatically vanish in dimensions less then or equal to 3.

In dimension 2, the Riemannian curvature tensor can be written as

$$R = -\frac{s}{4} g \oslash g = -\frac{K}{2} g \oslash g,$$

where K assigns to each point $p \in M$ the sectional curvature in the direction T_pM. In particular,

$$\begin{aligned} R(X,Y,Z,W) &= \frac{s}{2}(\langle X,W\rangle\langle Y,Z\rangle - \langle X,Z\rangle\langle Y,W\rangle) \\ &= K(\langle X,W\rangle\langle Y,Z\rangle - \langle X,Z\rangle\langle Y,W\rangle). \end{aligned}$$

In dimension 3, the Riemannian curvature tensor is uniquely determined by the Ricci tensor according to the formula

$$R = -\frac{s}{12} g \oslash g - \left(\mathrm{Ric} - \frac{s}{3} g\right) \oslash g = \frac{s}{4} g \oslash g + \mathrm{Ric} \oslash g,$$

where $s = \operatorname{tr} \mathrm{Ric}$ is the scalar curvature function.

From dimensions greater than 3, the pointwise Weyl component appears and defines a tensor field $C = W(R)$ of type $(4,0)$ on M. This tensor field is called the *type* $(4,0)$ *conformal Weyl tensor field*. The adjective "conformal" and the related standard notation C for the Weyl tensor field refers to the invariance property of this tensor field under conformal changes of the Riemannian metric. This property will be discussed in the next section.

Some important classes of Riemannian manifolds can be defined by restrictions on the components of the curvature tensor.
For example, as we saw in the proof of Schur's theorem, a space has constant curvature $K \in \mathbb{R}$ if and only if its curvature tensor has the form $-\frac{K}{2} g \oslash g$. This means that a constant curvature space has vanishing trace free Ricci component and Weyl component. Schur's theorem can be interpreted as a converse of this statement: If the curvature tensor of a connected Riemannian manifold of dimension at least 3 has vanishing trace free Ricci component and Weyl component, then it has constant sectional curvature.

Definition 4.10.16. A Riemannian manifold is called an *Einstein manifold* if the Ricci curvature $r(X_p)$ equals the same constant $r \in \mathbb{R}$ for all $p \in M$ and $X_p \in T_pM$. ※

For an Einstein manifold with constant Ricci curvature r, the Ricci tensor Ric is equal to rg, so the trace free Ricci component $(\mathrm{Ric} - rg)$ is equal to 0. The following theorem is an analog of Schur's theorem.

Theorem 4.10.17. *A connected Riemannian manifold of dimension at least* 3 *is an Einstein manifold if and only if the trace free Ricci component of its curvature tensor vanishes.*

Proof. The trace free part of the Ricci tensor vanishes if and only if $\mathrm{Ric} = rg$ for a smooth function r on the manifold M. Our goal is to show that in this case r must be constant.
Since M is connected, it is enough to show that r has vanishing derivative with respect to any vector field Z. To check $Z(r) = 0$ at a point $p \in M$, choose a local coordinate system ϕ around p and apply the Gram–Schmidt orthogonalization process to the vector fields $\partial_1^\phi, \dots, \partial_n^\phi$ to obtain orthonormal vector fields $E_1, \dots, E_n$ in a neighborhood of p. Write $\nabla_Z E_i$ as a linear combination $\nabla_Z E_i = \sum \omega_i^j E_j$. The coefficients ω_i^j can be computed as $\omega_i^j = \langle \nabla_Z E_i, E_j \rangle$. Differentiating the equation $\langle E_i, E_j \rangle = \delta_{ij}$, we obtain that $\omega_i^j = -\omega_j^i$.
As a corollary of this skew symmetry, we obtain that for any two vector fields X, Y, we have

$$\begin{aligned}
&\sum_{i=1}^n (R(X, \nabla_Z E_i, E_i, Y) + R(X, E_i, \nabla_Z E_i, Y)) \\
&\quad = \sum_{i=1}^n (R(\nabla_Z E_i, X, Y, E_i) + R(E_i, X, Y, \nabla_Z E_i)) \\
&\quad = \sum_{i,j=1}^n (R(\omega_i^j E_j, X, Y, E_i) + R(E_i, X, Y, \omega_i^j E_j))
\end{aligned}$$

$$= \sum_{i,j=1}^{n} (\omega_i^j + \omega_j^i) R(E_j, X, Y, E_i) = 0. \tag{4.18}$$

Differentiating the defining equation $s = \sum_{i,j=1}^{n} R(E_i, E_j, E_j, E_i)$ of the scalar curvature with respect to Z using (4.18) and the second Bianchi identity,

$$\begin{aligned} Z(s) &= \sum_{i,j=1}^{n} [(\nabla_Z R)(E_i, E_j, E_j, E_i) - R(\nabla_Z E_i, E_j, E_j, E_i) \\ &\qquad - R(E_i, \nabla_Z E_j, E_j, E_i) - R(E_i, E_j, \nabla_Z E_j, E_i) \\ &\qquad\quad - R(E_i, E_j, E_j, \nabla_Z E_i)] \\ &= \sum_{i,j=1}^{n} (\nabla_Z R)(E_i, E_j, E_j, E_i) \qquad (4.19) \\ &= -\sum_{i,j=1}^{n} [(\nabla_{E_i} R)(E_j, Z, E_j, E_i) + (\nabla_{E_j} R)(Z, E_i, E_j, E_i)] \\ &= 2 \sum_{i,j=1}^{n} (\nabla_{E_j} R)(Z, E_i, E_i, E_j). \end{aligned}$$

Differentiate now the equation $r\langle Z, E_j\rangle = \mathrm{Ric}(Z, E_j) = \sum_{i=1}^{n} R(Z, E_i, E_i, E_j)$ with respect to E_j to get

$$\begin{aligned} &E_j(r)\langle Z, E_j\rangle + r\langle \nabla_{E_j} Z, E_j\rangle + r\langle Z, \nabla_{E_j} E_j\rangle \\ &= \sum_{i=1}^{n} \big[(\nabla_{E_j} R)(Z, E_i, E_i, E_j) + R(\nabla_{E_j} Z, E_i, E_i, E_j) + \\ &\quad + R(Z, \nabla_{E_j} E_i, E_i, E_j) + R(Z, E_i, \nabla_{E_j} E_i, E_j) + R(Z, E_i, E_i, \nabla_{E_j} E_j)\big] \\ &= \sum_{i=1}^{n} \big[(\nabla_{E_j} R)(Z, E_i, E_i, E_j) + R(\nabla_{E_j} Z, E_i, E_i, E_j) + R(Z, E_i, E_i, \nabla_{E_j} E_j)\big]. \end{aligned}$$

Since

$$\begin{aligned} r\langle \nabla_{E_j} Z, E_j\rangle + r\langle Z, \nabla_{E_j} E_j\rangle &\\ &= \mathrm{Ric}(\nabla_{E_j} Z, E_j) + \mathrm{Ric}(Z, \nabla_{E_j} E_j) \\ &= \sum_{i=1}^{n} \big[R(\nabla_{E_j} Z, E_i, E_i, E_j) + R(Z, E_i, E_i, \nabla_{E_j} E_j)\big], \end{aligned}$$

we obtain

$$E_j(r)\langle Z, E_j\rangle = \sum_{i=1}^{n} (\nabla_{E_j} R)(Z, E_i, E_i, E_j).$$

Summing for j and comparing the result to (4.19) we get

$$Z(r) = \sum_{j=1}^{n} (\langle Z, E_j \rangle E_j)(r) = \sum_{i=1,j}^{n} (\nabla_{E_j} R)(Z, E_i, E_i, E_j) = \frac{Z(s)}{2}.$$

On the other hand, taking the trace of the equation $\mathrm{Ric} = rg$, we obtain $s = r \dim M$, in particular $\dim M \cdot Z(r) = Z(s)$. As $\dim M \geq 3$, this implies that $Z(r) = Z(s) = 0$. □

We remark that 3-dimensional Einstein manifolds are spaces of constant curvature, since in dimension 3 the Weyl tensor vanishes automatically.

4.11 Conformal Invariance of the Weyl Tensor

The goal of this section is to compute how the curvature tensor of a Riemannian manifold is transformed when the Riemannian metric is replaced by another one, proportional to the original metric.

Definition 4.11.1. Two Riemannian metrics g and $\tilde{g}$ are called *conformally equivalent* if there is a smooth function λ on M such that $\tilde{g} = \lambda g$. Since both g and $\tilde{g}$ are positive definite, the conformal factor λ must be positive everywhere. Thus, it can be written as $\lambda = e^{2f}$ for some smooth function f. ※

To understand how the curvature tensors of conformally equivalent metrics are related to one another, first we compute the transformation rule between their Levi-Cività connections.

Definition 4.11.2. The *gradient vector field* $\operatorname{grad} f$ *of a smooth function* f defined on a Riemannian manifold M is the unique vector field satisfying $Xf = \langle X, \operatorname{grad} f \rangle$ for any vector field $X \in \mathfrak{X}(M)$. ※

Let us compute the gradient vector field of a function in local coordinates. Let $\phi = (x^1, \dots, x^n)\colon U \to \mathbb{R}^n$ be a chart and $\partial_1 = \partial_1^\phi, \dots, \partial_n = \partial_n^\phi$ be the basis vector fields induced by the chart. Denote by $\mathcal{G} = (g_{ij}) = (\langle \partial_i, \partial_j \rangle)$ the matrix of the Riemannian metric with respect to these basis vectors, and by $\mathcal{G}^{-1} = (g^{ij})$ the inverse matrix. We look for $\operatorname{grad} f$ as a linear combination $\operatorname{grad} f = \sum_{i=1}^n G^i \partial_i$. By the definition of the gradient, for any vector field $X = \sum_{i=1}^n X^i \partial_i$ we should have

$$Xf = \sum_{i=1}^{n} X^i \partial_i(f) = \langle X, \operatorname{grad} f \rangle = \sum_{i,j=1}^{n} X^i G^j g_{ij}.$$

Thus, the unknown coefficients G^i can be found by solving the linear system of equations

$$\partial_i(f) = \sum_{j=1}^{n} G^j g_{ij}, \qquad (i = 1, \dots, n).$$

Multiplying the ith equation by g^{ik}, and summing up for i we obtain

$$\sum_{i=1}^{n} \partial_i(f) g^{ik} = \sum_{i,j=1}^{n} G^j g_{ij} g^{ik} = \sum_{j=1}^{n} G^j \delta_{jk} = G^k,$$

hence

$$\operatorname{grad} f = \sum_{i,j=1}^{n} \partial_i(f) g^{ij} \partial_j.$$

Suppose that the Levi-Cività connection of g is ∇, and that of $\tilde{g} = e^{2f} g$ is $\tilde{\nabla}$. Let us compute the difference tensor $\nabla_X Y - \tilde{\nabla}_X Y$.
Applying the Koszul formula (4.12)

$$\begin{aligned} 2\tilde{g}(\tilde{\nabla}_X Y, Z) = {} & X(\tilde{g}(Y,Z)) + Y(\tilde{g}(X,Z)) - Z(\tilde{g}(X,Y)) \\ & + \tilde{g}([X,Y],Z) - \tilde{g}([X,Z],Y) - \tilde{g}([Y,Z],X), \end{aligned}$$

thus

$$\begin{aligned} 2e^{2f} g(\tilde{\nabla}_X Y, Z) = {} & e^{2f} X(g(Y,Z)) + 2X(f)e^{2f} g(Y,Z) + \\ & + e^{2f} Y(g(X,Z)) + 2Y(f)e^{2f} g(X,Z) - \\ & - e^{2f} Z(g(X,Y)) - 2Z(f)e^{2f} g(X,Y) + \\ & + e^{2f}(g([X,Y],Z) - g([X,Z],Y) - g([Y,Z],X)). \end{aligned}$$

Dividing by e^{2f} and taking the Koszul formula

$$\begin{aligned} 2g(\nabla_X Y, Z) = {} & X(g(Y,Z)) + Y(g(X,Z)) - Z(g(X,Y)) + \\ & + g([X,Y],Z) - g([X,Z],Y) - g([Y,Z],X). \end{aligned}$$

into consideration we conclude the formula

$$\tilde{\nabla}_X Y = \nabla_X Y + X(f)Y + Y(f)X - g(X,Y)\operatorname{grad}(f). \tag{4.20}$$

Definition 4.11.3. Let f be a smooth function on a manifold M endowed with a connection ∇. The *Hesse form of f with respect to the connection ∇* is the $(2,0)$ type tensor field $\operatorname{Hesse}(f)$ the evaluation of which on the vector fields X, Y is ※

$$\operatorname{Hesse}(f)(X,Y) = X(Y(f)) - (\nabla_X Y)(f).$$

Proposition 4.11.4. *The Hesse form is a tensor field indeed, that is,* $\mathrm{Hesse}(f)(X,Y)$ *is a bilinear function of* X *and* Y over smooth functions. *If* ∇ *is torsion free, then the Hesse form is symmetric. At a critical point* p *of* f*, i.e., at a point where the differential* $(df)_p$ *vanishes,* $\mathrm{Hesse}(f)_p$ *does not depend on the connection.*

Proof. Additivity in X and Y is obvious. If g is a smooth function on M, then

$$\begin{aligned}\mathrm{Hesse}(f)(gX,Y) &= gX(Y(f)) - (\nabla_{gX}Y)(f) = gX(Y(f)) - g(\nabla_X Y)(f)\\ &= g\,\mathrm{Hesse}(f)(X,Y),\end{aligned}$$

and

$$\begin{aligned}\mathrm{Hesse}(f)(X,gY) &= X(gY(f)) - (\nabla_X gY)(f)\\ &= gX(Y(f)) + X(g)Y(f) - g(\nabla_X Y)(f) - X(g)Y(f)\\ &= g\,\mathrm{Hesse}(f)(X,Y).\end{aligned}$$

The equation

$$\begin{aligned}\mathrm{Hesse}(f)(X,Y) - \mathrm{Hesse}(f)(Y,X) &= (X\circ Y - Y\circ X - (\nabla_X Y - \nabla_Y X))(f)\\ &= T(X,Y)(f)\end{aligned}$$

shows the symmetry for vanishing torsion.
Finally, if p is a critical point of f, then $(\nabla_X Y)_p(f) = (df)_p(\nabla_X Y) = 0$, so

$$\mathrm{Hesse}(f)_p(X_p,Y_p) = X_p(Y(f))$$

does not depend on ∇. (In the last equation, $X_p, Y_p \in T_pM$, Y is an arbitrary vector field the value of which at p is Y_p.) □

Now we are ready to compute the relation between the curvature tensor $\tilde{R}$ of the metric $\tilde{g} = e^{2f}g$ and the curvature tensor R of g.
From equation (4.20)

$$\begin{aligned}\tilde{\nabla}_X\tilde{\nabla}_Y Z &= \tilde{\nabla}_X(\nabla_Y Z + Y(f)Z + Z(f)Y - g(Y,Z)\,\mathrm{grad}(f))\\ &= \nabla_X(\nabla_Y Z + Y(f)Z + Z(f)Y - g(Y,Z)\,\mathrm{grad}(f)+)\\ &\quad + X(f)(\nabla_Y Z + Y(f)Z + Z(f)Y - g(Y,Z)\,\mathrm{grad}(f))+\\ &\quad + ((\nabla_Y Z)(f) + Y(f)Z(f) + Z(f)Y(f) - g(Y,Z)||\,\mathrm{grad}(f)||^2)X-\\ &\quad - (g(X,\nabla_Y Z) + Y(f)g(X,Z) + Z(f)g(X,Y) - g(Y,Z)X(f))\,\mathrm{grad}(f)\end{aligned}$$

Expanding the first line of the last expression we get

$$
\begin{aligned}
\tilde{\nabla}_X\tilde{\nabla}_Y Z &= \nabla_X\nabla_Y Z + X(Y(f))Z + \boxed{Y(f)\nabla_X Z}^{\circledcirc} + X(Z(f))Y \\
&+ Z(f)\nabla_X Y - g(\nabla_X Y, Z)\operatorname{grad}(f)\boxed{-g(Y,\nabla_X Z)\operatorname{grad}(f)}^{\times} \\
&- g(Y,Z)\nabla_X(\operatorname{grad}(f)) + \boxed{X(f)\nabla_Y Z}^{\circledcirc} + \boxed{X(f)Y(f)Z} \\
&+ \boxed{X(f)Z(f)Y}^{\ominus}\boxed{-X(f)g(Y,Z)\operatorname{grad}(f)}^{\circ} + \\
&+ (\nabla_Y Z)(f)X + \boxed{Y(f)Z(f)X}^{\ominus} + Z(f)Y(f)X - g(Y,Z)||\operatorname{grad}(f)||^2 X \\
&+ (\boxed{-g(X,\nabla_Y Z)}^{\times\cdot} - Y(f)g(X,Z) - \boxed{Z(f)g(X,Y)} \\
&+ \boxed{g(Y,Z)X(f)}^{\circ\cdot})\operatorname{grad}(f)
\end{aligned}
$$

To work with expressions antisymmetric in two variables, it is convenient to introduce the antisymmetrizing operator. If $F(X,Y,\dots)$ is an expression involving the variables X and Y and maybe other ones, then set $\Theta_{XY}(F(X,Y,\dots)) = F(X,Y,\dots) - F(Y,X,\dots)$. The *antisymmetrizer* $\Theta_{X,Y}$ makes the expression to which it is applied antisymmetric in X and Y.
The terms in the expansion of $\tilde{\nabla}_X\tilde{\nabla}_Y Z$ that are in boxes with the same labels give a symmetric expression in X and Y, therefore, they disappear when we apply the antisymmetrizer $\Theta_{X,Y}$ to both sides, so

$$
\begin{aligned}
[\tilde{\nabla}_X,\tilde{\nabla}_Y]Z &= \tilde{\nabla}_X\tilde{\nabla}_Y Z - \tilde{\nabla}_Y\tilde{\nabla}_X Z = \Theta_{XY}(\tilde{\nabla}_X\tilde{\nabla}_Y Z) \\
&= [\nabla_X,\nabla_Y]Z + [X,Y](f))Z + \Theta_{XY}(X(Z(f))Y) + Z(f)[X,Y] - \\
&\quad - g([X,Y],Z)\operatorname{grad}(f) + \Theta_{XY}(g(X,Z)\nabla_Y(\operatorname{grad}(f))) + \\
&\quad + \Theta_{XY}\big((\nabla_Y Z)(f)X + Z(f)Y(f)X + ||\operatorname{grad}(f)||^2 g(X,Z)Y\big) + \\
&\quad + \Theta_{XY}(X(f)g(Y,Z)\operatorname{grad}(f)).
\end{aligned}
$$

We know also that

$$
\tilde{\nabla}_{[X,Y]}Z = \nabla_{[X,Y]}Z + [X,Y](f)Z + Z(f)[X,Y] - g([X,Y],Z)\operatorname{grad}(f).
$$

Subtracting this equation from the previous one we obtain

$$
\begin{aligned}
\tilde{R}(X,Y)Z = R(X,Y)Z + \Theta_{XY}\Big(&X(Z(f))Y + g(X,Z)\nabla_Y(\operatorname{grad}(f)) + \\
&+ (\nabla_Y Z)(f)X + Z(f)Y(f)X + ||\operatorname{grad}(f)||^2 g(X,Z)Y + \\
&+ X(f)g(Y,Z)\operatorname{grad}(f)\Big).
\end{aligned}
$$

This is the transformation rule of the curvature tensor under conformal transformations of the metric. To produce from this a transformation rule for the

Riemannian curvature tensor, take the inner product of both sides with a fourth vector field W with respect to g. This produces

$$\begin{aligned} e^{-2f}\tilde{R}(X,Y,Z,W) &= R(X,Y,Z,W)+ \\ &+\Theta_{XY}\Big(X(Z(f))g(Y,W)+\boxed{g(X,Z)g(\nabla_Y(\mathrm{grad}(f)),W)} \\ &\quad+(\nabla_Y Z)(f)g(X,W)+Z(f)Y(f)g(X,W) \\ &\qquad+||\,\mathrm{grad}(f)||^2 g(X,Z)g(Y,W)+X(f)W(f)g(Y,Z)\Big). \end{aligned}$$

The expression in the box can be rewritten using the identity

$$\begin{aligned} g(\nabla_Y(\mathrm{grad}(f)),W) &= Y g(\mathrm{grad}(f),W)-g(\mathrm{grad}(f),\nabla_Y W) \\ &= Y(W(f))-(\nabla_Y W)(f). \end{aligned}$$

Using this a rearrangement gives

$$\begin{aligned} &e^{-2f}\tilde{R}(X,Y,Z,W) \\ &= R(X,Y,Z,W)+\Theta_{XY}\Big(X(Z(f))g(Y,W)+g(X,Z)Y(W(f))- \\ &\qquad -g(X,Z)(\nabla_Y W)(f)+ \\ &\qquad +(\nabla_Y Z)(f)g(X,W)+Z(f)Y(f)g(X,W)+ \\ &\qquad +||\,\mathrm{grad}(f)||^2 g(X,Z)g(Y,W)+ \\ &\qquad +X(f)W(f)g(Y,Z)\Big). \end{aligned}$$

Let us flip the role of X and Y in those terms, where there is no X in the inner product $g(.,.)$. Inside the antisymmetrizer Θ_{XY}, this can be done with a sign change. This results in the formula

$$\begin{aligned} &e^{-2f}\tilde{R}(X,Y,Z,W) \\ &= R(X,Y,Z,W)+\Theta_{XY}\Big(-Y(Z(f))g(X,W)+g(X,Z)Y(W(f))- \\ &\qquad -g(X,Z)(\nabla_Y W)(f)+ \\ &\qquad +(\nabla_Y Z)(f)g(X,W)+Z(f)Y(f)g(X,W)+ \\ &\qquad +||\,\mathrm{grad}(f)||^2\frac{g(X,Z)g(Y,W)-g(Y,Z)g(X,W)}{2}- \\ &\qquad -Y(f)W(f)g(X,Z)\Big). \end{aligned}$$

Observe that the expression to which Θ_{XY} is applied is antisymmetric in Z and W as well, so we can write the right-hand side in a more compressed form

$$
\begin{aligned}
e^{-2f}\tilde{R}(X,Y,Z,W)) \\
&= R(X,Y,Z,W)+ \\
&\quad + \Theta_{ZW}\Theta_{XY}\Big(-Y(Z(f))g(X,W) + (\nabla_Y Z)(f)g(X,W)+ \\
&\qquad + Z(f)Y(f)g(X,W) - \frac{\|\operatorname{grad}(f)\|^2}{2} g(Y,Z)g(X,W)\Big) \\
&= R(X,Y,Z,W) \\
&\quad + \left(\Big(df\otimes df - \operatorname{Hesse}(f) - \frac{\|\operatorname{grad}(f)\|^2}{2} g\Big) \oslash g\right)(X,Y,Z,W).
\end{aligned}
$$

Thus we obtained the formula

$$e^{-2f}\tilde{R} = R + \Big(df\otimes df - \operatorname{Hesse}(f) - \frac{\|\operatorname{grad}(f)\|^2}{2} g\Big) \oslash g.$$

The main corollary of this formula is that the difference $e^{-2f}\tilde{R} - R$ is the Kulkarni-Nomizu product of a symmetric tensor of type $(2,0)$ with g, consequently, its Weyl component is 0. This gives the following theorem.

Theorem 4.11.5. *If C and $\tilde{C}$ are the $(4,0)$ type conformal Weyl tensors of g and $\tilde{g} = e^{2f}g$, then $\tilde{C} = e^{2f}C$.*

The decomposition of the Riemannian curvature tensor induces a decomposition of the type $(3,1)$ curvature tensor as well. In this decomposition, the Weyl component will be a tensor field $\hat{C}$ of type $(3,1)$, which is related to the type $(4,0)$ Weyl tensor C by the identity

$$g(\hat{C}(X,Y,Z),W) = C(X,Y,Z,W).$$

$\hat{C}$ is called the *type* $(3,1)$ *conformal Weyl tensor* of the Riemannian manifold. As a corollary of the previous theorem, we see that the type $(3,1)$ Weyl tensor fields of conformally equivalent metrics are equal.

Definition 4.11.6. A Riemannian manifold is called *flat* if its curvature tensor vanishes. A Riemannian manifold is *locally conformally flat* if each point has an open neighborhood over which the metric is conformally equivalent to a flat one. ※

It can be proved that

(1) a flat Riemannian manifold is locally isometric to a Euclidean space;

(2) a space is locally conformally flat if it has an atlas of angle preserving charts.

Angle preserving charts, i.e., charts with respect to which the matrix of the Riemannian metric is a multiple of the unit matrix at each point are called *isothermal local coordinate systems.*
According to the conformal invariance of the Weyl tensor, if a Riemannian manifold is locally conformally flat, then its Weyl tensor is 0. It is a remarkable fact that this condition is not just necessary but also sufficient if the dimension of the manifold is not 3.

Theorem 4.11.7. *Let n be the dimension of a Riemannian manifold M.*

- *If $n = 2$, then M is locally conformally flat.*
- *If $n = 3$, then M is locally conformally flat if and only if its* Cotton tensor *defined by*

$$C(X,Y,Z) = (\nabla_Z \operatorname{Ric})(X,Y) - (\nabla_Y \operatorname{Ric})(X,Z) - \frac{Z(s)g(X,Y) - Y(s)g(X,Z)}{2(n-1)}$$

  vanishes.
- *If $n \geq 4$, then M is locally conformally flat if and only if its Weyl tensor vanishes.*

The proof of this theorem goes beyond the scope of this textbook.

4.12 Geodesics

We define the *length* of a smooth curve $\gamma\colon [a,b] \to M$ lying on a Riemannian manifold $(M, \langle\,,\rangle)$ to be the integral

$$l(\gamma) = \int_a^b \sqrt{\langle \gamma'(t), \gamma'(t)\rangle} dt.$$

It is worth mentioning that the classical definition of length as the limit of the lengths of inscribed broken lines does not make sense, since the distance of points is not directly defined. The situation is just the opposite. We can define first the length of curves as a primary concept and derive from it a so called *intrinsic metric* $d(p,q)$, at least for connected Riemannian manifolds as the infimum of the lengths of all curves joining p to q. The metric enables us to define the length of "broken lines" given just by a sequence of vertices $P_1, \dots, P_N$ to be the sum of the distances between consecutive vertices. There is a theorem saying that the length of a smooth curve $\gamma\colon [a,b] \to M$ is equal to the limit of the lengths of inscribed broken lines $\gamma(t_0), \gamma(t_1), \dots, \gamma(t_N)$, $a = t_0 < t_1 < \dots < t_N = b$ as the maximum of the distances $|t_i - t_{i-1}|$ tends to zero.

To find the analog of straight lines in the intrinsic geometry of a Riemannian manifold we have to characterize straight lines in a way that makes sense for Riemannian manifolds as well. Since the length of curves is one of the most fundamental concepts of Riemannian geometry, we can take the following characterization: a curve is a straight line if and only if for any two points on the curve, the segment of the curve bounded by the points is the shortest among curves joining the two points. A slight modification of this property could be used to distinguish a class of curves, but it is not clear at first glance whether such curves exist at all on a general Riemannian manifold.

For a physicist a straight line is the trajectory of a particle with zero acceleration or that of a light beam. This observation can also give rise to a definition. We only have to find a proper generalization of "acceleration" for curves lying in a Riemannian manifold. It seems quite natural to proceed as follows. The speed vectors of a curve yield a vector field along the curve. On the other hand, by the fundamental theorem of Riemannian geometry, the Riemannian metric determines a unique affine connection on the manifold which is symmetric and compatible with the metric. In particular, one can differentiate the speed vector field with respect to the curve parameter and may call the result $\gamma'' = \nabla_{\gamma'}\gamma'$ the acceleration vector field along the curve.

Definition 4.12.1. Let M be a Riemannian manifold, γ be a curve on it. We say that γ is a *geodesic* if

$$\nabla_{\gamma'}\gamma' = 0.$$

※

Remark. More generally, if (M, ∇) is a manifold with an affine connection, then the curves satisfying $\nabla_{\gamma'}\gamma' = 0$ are said to be *autoparallel*. Geodesics are autoparallel curves for the Levi-Cività connection.

Proposition 4.12.2. *The length of the speed vector of a geodesic is constant.*

Proof. By the compatibility of the connection with the metric, parallel transport preserves length and angles between vectors. The definition of geodesics implies that the speed vector field is parallel along the curve, consequently consists of vectors of the same length. □

The proposition follows also from the equality

$$\langle \gamma', \gamma' \rangle' = \langle \nabla_{\gamma'}\gamma', \gamma' \rangle + \langle \gamma', \nabla_{\gamma'}\gamma' \rangle = 0.$$

As a consequence, we get that the property of "being geodesic" is not invariant under reparameterization. The parameter t of a regular geodesic is always related to the natural parameter s through an affine linear transformation i.e. $t = as + b$ for some $a, b \in \mathbb{R}$. This motivates the following definition.

Definition 4.12.3. A regular curve on a Riemannian manifold is a *pre-geodesic* if its natural reparameterization is geodesic. ※

In terms of a local coordinate system with coordinates $(x^1, \ldots, x^n)$ a curve γ in the domain of the chart determines (and is determined by) n smooth functions $\gamma^i = x^i \circ \gamma$ $(1 \le i \le n)$. The equation $\nabla_{\gamma'}\gamma' = 0$ then takes the form

$$\gamma^{k\prime\prime} + \sum_{i,j=1}^{n} \Gamma_{ij}^k \circ \gamma \cdot \gamma^{i\prime} \cdot \gamma^{j\prime} = 0 \text{ for all } 1 \le k \le n.$$

The existence of geodesics depends, therefore, on the solutions of a certain system of second order differential equations.
Introducing the new functions $v^i = \gamma^{i\prime}$ this system of n second order differential equations becomes a system of $2n$ first order equations

$$\begin{cases} \gamma^{k\prime} = v^k, \\ v^{k\prime} = \displaystyle\sum_{i,j=1}^{n} \Gamma_{ij}^k \circ \gamma \cdot v^i \cdot v^j, \end{cases} \qquad \text{for all } 1 \le k \le n.$$

Applying the existence and uniqueness theorem for ordinary differential equations one obtains the following.

Proposition 4.12.4. *For any point p on a Riemannian manifold M and for any tangent vector $X \in T_pM$, there exists a unique maximal geodesic γ defined on an interval containing 0 such that $\gamma(0) = p$ and $\gamma'(0) = X$.*

If the maximal geodesic through a point p with initial velocity X is defined on an interval containing $[-\varepsilon, \varepsilon]$ then there is a neighborhood U of X in the tangent bundle such that every maximal geodesic started from a point q with initial velocity $Y \in T_qM$ is defined on $[-\varepsilon, \varepsilon]$.
Since a geodesic with zero initial speed can be defined on the whole real straight line, for each point p on the manifold one can find a positive δ such that for every tangent vector $X \in T_pM$ with $\|X\| < \delta$, the geodesic defined by the conditions $\gamma(0) = p$, $\gamma'(0) = X$ can be extended to the interval $[0, 1]$.
The following notation will be convenient. Let $X \in T_pM$ be a tangent vector and suppose that there exists a geodesic $\gamma\colon [0, 1] \to M$ satisfying the conditions $\gamma(0) = p$, $\gamma'(0) = X$. Then the point $\gamma(1) \in M$ will be denoted by $\exp_p(X)$ and called the *exponential* of the tangent vector X.
Using the fact that for any $c \in \mathbb{R}$, the curve $t \mapsto \gamma(ct)$ is also a geodesic we see that the geodesic γ is described by the formula

$$\gamma(t) = \exp_p(tX).$$

As we have observed, $\exp_p(X)$ is defined provided that $\|X\|$ is small enough. In general however, $\exp_p(X)$ is not defined for large vectors X. This motivates the following.

Definition 4.12.5. A Riemannian manifold is *geodesically complete* if for all $p \in M$, $\exp_p(X)$ is defined for all vectors $X \in T_pM$. ※

This is clearly equivalent to the requirement that maximal geodesics are defined on the whole real line $\mathbb{R}$.

Proposition 4.12.6. *For a fixed point $p \in M$, the exponential map $\exp_p$ is a smooth map from an open neighborhood of $0 \in T_pM$ into the manifold. Furthermore, the restriction of it onto a (possibly even smaller) open neighborhood of $0 \in T_pM$ is a diffeomorphism.*

Proof. Differentiability of the exponential mapping follows from the theorem on the differentiable dependence on the initial point for solutions of a system of ordinary differential equations. To show that $\exp_p$ is a local diffeomorphism, we only have to show that its derivative at the point $0 \in T_pM$ is a non-singular linear mapping (see Inverse Function Theorem). Since T_pM is a linear space, its tangent space $T_0(T_pM)$ at 0 can be identified with the vector space T_pM itself. Through this identification, the derivative of the exponential map at 0 maps $T_pM \cong T_0(T_pM)$ into T_pM. We show that this derivative is just the identity map of T_pM, hence non-singular.
Let X be an element of the tangent space $T_pM \cong T_0(T_pM)$. To determine where X is taken by the derivative of the exponential mapping, we represent X as the speed vector of the curve $t \mapsto \varphi(t) = tX$ at $t = 0$. The exponential mapping takes this curve to the geodesic curve $\gamma = \exp_p \circ \varphi$, $\gamma(t) = \exp_p(tX)$, the speed vector of which at $t = 0$ is X, so the derivative of the exponential map sends X to itself and this is what we claimed. □

By the proposition, we can introduce a local coordinate system, based on geodesics, about each point of the manifold as follows. We fix an *orthonormal* basis in the tangent space T_pM, which gives us an isomorphism $\iota\colon T_pM \to \mathbb{R}^n$ that assigns to each tangent vector its components with respect to the basis, and then take $\iota \circ \exp_p^{-1}$. The map $\iota \circ \exp_p^{-1}$ is a diffeomorphism between an open neighborhood of p and that of the origin in $\mathbb{R}^n$, therefore, it is a smooth chart on M. Coordinate systems obtained this way are called *normal coordinate systems*, while we shall call the inverse of them *normal parameterizations*.
For a Riemannian manifold M, we can define the sphere of radius r centered at $p \in M$ as the set of points $q \in M$ such that $d(p,q) = r$, where $d(p,q)$ denotes the intrinsic distance of p and q. When the radius of the sphere is increasing, the topological type of the sphere changes at certain critical values

of the radius. For small radii however, the intrinsic spheres are diffeomorphic to the ordinary spheres in $\mathbb{R}^n$, and what is more, we have the following.

Theorem 4.12.7. *The normal parameterization of a manifold about a point p maps the sphere about the origin with radius r, provided that it is contained in the domain of the parameterization, diffeomorphically onto the intrinsic sphere centered at p with radius r.*

We prove this theorem later.

Formula for the First Variation of the Length

Definition 4.12.8. A *variation* of a smooth curve $\gamma\colon [a,b] \to M$ is a smooth mapping γ_* from the rectangular domain $[-\delta,\delta] \times [a,b]$ into M such that $\gamma_*(0,t) = \gamma(t)$ for all $t \in [a,b]$. ※

Given a variation of a curve we may introduce a one parameter family of curves γ_ε, $\varepsilon \in [-\delta,\delta]$ by setting $\gamma_\varepsilon(t) = \gamma_*(\varepsilon,t)$. By our assumption, these curves yield a deformation of the curve $\gamma_0 = \gamma$.

Theorem 4.12.9. *Let γ_* be a variation of a geodesic γ. Let $l(\varepsilon)$ denote the length of the curve γ_ε. Then the following formula holds*

$$l'(0) = \left\langle \partial_1\gamma_*(0,b), \frac{\gamma'(b)}{\|\gamma'(b)\|} \right\rangle - \left\langle \partial_1\gamma_*(0,a), \frac{\gamma'(a)}{\|\gamma'(a)\|} \right\rangle.$$

Proof. By the definition of the length of a curve, one has

$$\begin{aligned} l'(0) &= \frac{d}{d\varepsilon}\int_a^b \|\partial_2\gamma_*(\varepsilon,\tau)\|\,d\tau\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon}\int_a^b \sqrt{\langle \partial_2\gamma_*(\varepsilon,\tau), \partial_2\gamma_*(\varepsilon,\tau)\rangle}d\tau\Big|_{\varepsilon=0} \\ &= \int_a^b \frac{d}{d\varepsilon}\sqrt{\langle \partial_2\gamma_*(\varepsilon,\tau), \partial_2\gamma_*(\varepsilon,\tau)\rangle}\Big|_{\varepsilon=0} d\tau \\ &= \int_a^b \frac{\frac{d}{d\varepsilon}\,\langle \partial_2\gamma_*(\varepsilon,\tau), \partial_2\gamma_*(\varepsilon,\tau)\rangle|_{\varepsilon=0}}{2\sqrt{\langle \partial_2\gamma_*(0,\tau), \partial_2\gamma_*(0,\tau)\rangle}} d\tau. \end{aligned}$$

With the help of the covariant differentiation induced by the Levi-Cività connection this expression can be written as follows.

$$\int_a^b \frac{\langle \nabla_1\partial_2\gamma_*(0,\tau), \partial_2\gamma_*(0,\tau)\rangle}{\|\gamma'(\tau)\|} d\tau = \int_a^b \left\langle \nabla_1\partial_2\gamma_*(0,\tau), \frac{\gamma'(\tau)}{\|\gamma'(\tau)\|} \right\rangle d\tau.$$

By the symmetry of the connection, this is equal to

$$\int_a^b \left\langle \nabla_2\partial_1\gamma_*(0,\tau), \frac{\gamma'(\tau)}{\|\gamma'(\tau)\|} \right\rangle d\tau$$

Observe, that the function $t \mapsto \left\langle \partial_1\gamma_*(0,t), \frac{\gamma'(t)}{\|\gamma'(t)\|} \right\rangle$ is a primitive function (antiderivative) of the function to be integrated. Indeed, the derivative of this function is

$$\begin{aligned}\frac{d}{dt}\left\langle \partial_1\gamma_*(0,t), \frac{\gamma'(t)}{\|\gamma'(t)\|} \right\rangle &= \left\langle \nabla_2\partial_1\gamma_*(0,t), \frac{\gamma'(t)}{\|\gamma'(t)\|} \right\rangle \\ &\quad + \left\langle \partial_1\gamma_*(0,t), \nabla_{\gamma'}\left(\frac{\gamma'}{\|\gamma'\|}\right)(t) \right\rangle,\end{aligned}$$

but the second term on the right-hand side is zero since γ is geodesic. Consequently,

$$\begin{aligned}l'(0) &= \int_a^b \left\langle \nabla_2\partial_1\gamma_*(0,t), \frac{\gamma'(t)}{\|\gamma'(t)\|} \right\rangle \\ &= \left\langle \partial_1\gamma_*(0,b), \frac{\gamma'(b)}{\|\gamma'(b)\|} \right\rangle - \left\langle \partial_1\gamma_*(0,a), \frac{\gamma'(a)}{\|\gamma'(a)\|} \right\rangle.\end{aligned}$$

□

Theorem 4.12.10 (Gauss Lemma)**.** *Let M be a Riemannian manifold, $p \in M$, and denote by S_r the sphere of radius r in T_pM centered at the zero tangent vector. Assume r is chosen to be so small that the exponential mapping is a diffeomorphism on a ball containing S_r and denote the exponential image of S_r by $\tilde{S}_r$. Then for any $X \in S_r$ the radial geodesic $t \mapsto \exp_p(tX)$ is perpendicular to $\tilde{S}_r$.*

Proof. Every tangent vector of $\tilde{S}_p$ can be obtained as the speed vector of a curve $\exp_p \circ \beta$ where β is a curve in S_r passing through $\beta(0) = X$. Given such a curve, let us define a variation of the geodesic $\gamma\colon t \mapsto \exp_p(tX)$ in the following way

$$\gamma_*(\varepsilon, t) := \exp_p(t\beta(\varepsilon)).$$

For a fixed ε, the curve γ_ε is a geodesic of length r so $l(\varepsilon)$ is constant. Thus, the previous theorem implies that

$$0 = l'(0) = \left\langle \partial_1\gamma_*(0,1), \frac{\gamma'(1)}{\|\gamma'(1)\|} \right\rangle - \left\langle \partial_1\gamma_*(0,0), \frac{\gamma'(0)}{\|\gamma'(0)\|} \right\rangle.$$

Since $\gamma_*(\varepsilon,0) = \exp_p(0\cdot\beta(\varepsilon)) = p$ and $\gamma_*(\varepsilon,1) = \exp_p(\beta(\varepsilon))$, we have $\partial_1\gamma_*(0,0) = \mathbf{0}$ and $\partial_1\gamma_*(0,1) = (\exp_p \circ \beta)'(0)$, therefore, we get

$$0 = \left\langle (\exp_p \circ \beta)'(0), \frac{\gamma'(1)}{|\gamma'(1)|} \right\rangle,$$

showing that γ intersects $\tilde{S}_r$ orthogonally. □

Now we are ready to prove the theorem saying that $\tilde{S}_r$ is a sphere in the intrinsic geometry of the manifold. It is clear that $d(p,q) \leq r$ for any point q on $\tilde{S}_r$, since the radial geodesic from p to r has length r, so all we need is the following.

Theorem 4.12.11. *If $\tilde{\gamma}\colon [a,b] \to M$ is an arbitrary curve connecting p to a point of $\tilde{S}_r$, then its length is $\geq r$.*

Proof. We may suppose without loss of generality that $\tilde{\gamma}(b)$ is the only intersection point of the curve with $\tilde{S}_r$ and $\tilde{\gamma}(t) \neq p$ for $t > a$. Then there is a unique curve γ in the tangent space T_pM such that $\tilde{\gamma} = \exp_p \circ \gamma$. Let $\mathbf{N}$ denote the vector field on $T_pM \setminus \{\mathbf{0}\}$ that is the gradient vector field of the function $f\colon X \mapsto \|X\|$ on T_pM, and therefore consists of unit vectors perpendicular to the spheres centered at the origin. The theorem above shows that the derivative of the exponential map takes $\mathbf{N}$ into a unit vector field $\tilde{\mathbf{N}}$ on M, perpendicular to the sets $\tilde{S}_t$.
We can estimate the length of a curve as follows

$$l(\tilde{\gamma}) = \int_a^b |\tilde{\gamma}'(\tau)| d\tau \geq \int_a^b \langle \tilde{\gamma}'(\tau), \tilde{\mathbf{N}}(\tilde{\gamma}(\tau)) \rangle d\tau.$$

Since $\langle \tilde{\gamma}'(\tau), \tilde{\mathbf{N}}(\tilde{\gamma}(\tau)\rangle$ is the component parallel to $\tilde{\mathbf{N}}(q)$ of the speed vector $\tilde{\gamma}'(\tau)$ with respect to the splitting $T_qM = \mathbb{R}\tilde{\mathbf{N}}(q) \oplus T_q\tilde{S}_*$ at $q = \tilde{\gamma}(\tau)$, it is equal to the component parallel to $N(X)$ of the speed vector $\gamma(\tau)$ with respect to the splitting

$$T_X(T_pM) = \mathbb{R}\mathbf{N}(X) \oplus T_X S_*$$

at $X = \gamma(\tau)$. Therefore,

$$\langle \tilde{\gamma}'(\tau), \tilde{\mathbf{N}}(\tilde{\gamma}(\tau)\rangle = \langle \gamma'(\tau), \mathbf{N}(\gamma(\tau)\rangle = \langle \gamma'(\tau), \operatorname{grad} f(\gamma(\tau))\rangle = (f \circ \gamma)'(\tau),$$

and

$$\int_a^b \langle \tilde{\gamma}'(\tau), \tilde{\mathbf{N}}(\tilde{\gamma}(\tau)\rangle d\tau = \int_a^b (f \circ \gamma)'(\tau) d\tau = \|\gamma(b)\| - \|\gamma(a)\| = r. \qquad \square$$

The proof also shows that the equality $l(\gamma) = r$ holds only for curves perpendicular to the spheres $\tilde{S}_*$.

Exercise 4.12.12. Show that such curves are pre-geodesics. ¤

Theorem 4.12.13. *A smooth curve $\gamma\colon [a,b] \to M$ parameterized by the natural parameter in a Riemannian manifold is geodesic if and only if there is a positive ε such that for any two values $t_1, t_2 \in [a,b]$ such that $|t_1 - t_2| < \varepsilon$, the restriction of γ onto $[t_1, t_2]$ is a curve of minimal length among curves joining $\gamma(t_1)$ to $\gamma(t_2)$.*

Remark. It is not true in general, that a geodesic curve is the a curve of minimal length among curves joining the same endpoints. To see this, it is enough to consider a long arc on a great circle on the sphere.

Exercise 4.12.14. Show that a regular curve in a hypersurface $M \subset \mathbb{R}^{n+1}$ is a geodesic if and only if its ordinary acceleration $\gamma''(t)$ is perpendicular to $T_{\gamma(t)}M$ for every t. It is a pre-geodesic if and only if $\gamma''(t)$ is contained in the plane spanned by $\gamma'(t)$ and the normal vector of M at $\gamma(t)$. ¤

Exercise 4.12.15. Show that great circles on the sphere and helices on a cylinder are pre-geodesics. ¤

Exercise 4.12.16. Find a regular pre-geodesic on the cone $x^2 + y^2 = z^2$, different from straight lines. ¤

Exercise 4.12.17. Show that straight lines on a hypersurface are pre-geodesic curves. ¤

Exercise 4.12.18. Show that symmetry planes of a surface in $\mathbb{R}^3$ intersect the surface in pre-geodesic lines. ¤

Exercise 4.12.19. Write the differential equation of geodesics on a surface of revolution with respect to the usual parameterization. Derive from the equations Clairaut's theorem: For a pre-geodesic curve on a surface of revolution the quantity $d\cos\beta$ is constant, where d denotes the distance of the curve point from the axis of symmetry, β is the angle between the speed vector of the curve and the circle of rotation passing through the curve point. ¤

4.13 Applications to Hypersurface Theory

4.13.1 Geodesic Curves on Hypersurfaces

The Frenet theory of curves in Euclidean spaces can be extended to curves lying in oriented Riemannian manifolds. If (M, g) is a Riemannian manifold and $\gamma\colon I \to M$ is a smooth parameterized curve, then using the Levi-Cività connection ∇ of M, we can define higher order derivatives of γ. There is no need for connection to define the speed vector field γ'. However, since γ' is a vector field along γ, to define higher order derivatives of γ as vector fields along γ, we need a connection. Let us define the vector field $\nabla^k\gamma$ recursively by $\nabla^1\gamma = \gamma'$ and $\nabla^{k+1}\gamma = \nabla_{\gamma'}(\nabla^k\gamma)$.

We say that a curve $\gamma\colon I \to M$ *is of general type in* M if the first $(n-1)$ covariant derivatives $\nabla^1\gamma(t), \dots, \nabla^{n-1}\gamma(t)$ are linearly independent for all $t \in I$, where $n = \dim M$.

For a curve of general type the distinguished Frenet vector fields $\mathbf{t}_i\colon I \to TM$ can be defined and computed in the same way as in $\mathbb{R}^n$. The first $(n-1)$ of them is the result of the Gram–Schmidt orthogonalization process applied to $\nabla^1\gamma, \dots, \nabla^{n-1}\gamma$, the last one is the unique vector which extends the first $(n-1)$ to a positively oriented orthonormal basis.

Once we defined the distinguished Frenet frame, the curvature functions of γ can be defined as the functions $\kappa_1, \dots, \kappa_{n-1}$ appearing in the Riemannian version of the Frenet formulae

$$\begin{aligned}
\frac{1}{\|\gamma'\|}\nabla_{\gamma'}\mathbf{t}_1 &= \kappa_1\mathbf{t}_2,\\
&\vdots\\
\frac{1}{\|\gamma'\|}\nabla_{\gamma'}\mathbf{t}_i &= -\kappa_{i-1}\mathbf{t}_{i-1} + \kappa_i\mathbf{t}_{i+1},\\
&\vdots\\
\frac{1}{\|\gamma'\|}\nabla_{\gamma'}\mathbf{t}_n &= -\kappa_{n-1}\mathbf{t}_{n-1}.
\end{aligned}$$

We can also prove the following extension of Theorem 2.5.14 for curves of general type in Riemannian manifolds.

Theorem 4.13.1. *Let $\gamma\colon I \to M$ be a curve of general type in M, $1 \le k \le n-1$. Then the curvature functions of γ can be computed by the equations*

$$\kappa_1 = \frac{\Delta_2}{v^3} \quad \textit{and} \quad \kappa_k = \frac{\Delta_{k+1}\Delta_{k-1}}{v\Delta_k^2} \text{ for } k \ge 2,$$

where $v = \|\gamma'\|$, and the numbers Δ_k are given by

$$\Delta_k = \sqrt{\det\begin{pmatrix} \langle\nabla^1\gamma, \nabla^1\gamma\rangle & \dots & \langle\nabla^1\gamma, \nabla^k\gamma\rangle \\ \vdots & \ddots & \vdots \\ \langle\nabla^k\gamma, \nabla^1\gamma\rangle & \dots & \langle\nabla^k\gamma, \nabla^k\gamma\rangle \end{pmatrix}}$$

$$\text{for } k < n \text{ and } \Delta_n = \det\begin{pmatrix} \nabla^1\gamma \\ \nabla^2\gamma \\ \vdots \\ \nabla^n\gamma \end{pmatrix},$$

where the last matrix is a matrix whose rows are the coordinates of the derivatives $\nabla^i\gamma$ with respect to an arbitrary positively oriented orthonormal frame.

(The matrix depends on the choice of the frame, but its determinant does not.)

Although up to the introduction of the Frenet frame, the curvature function, the Frenet formulae, and the computation of the curvatures everything goes word by word as in the Euclidean space, it is not true that all the theorems that were proved for curves in the Euclidean space are valid for curves in Riemannian manifolds.

Consider now a hypersurface M in $\mathbb{R}^n$, that is an $(n-1)$-dimensional submanifold of $\mathbb{R}^n$. Then a smooth curve $\gamma\colon I \to M$ in M lives a double life. It is both a curve on M and a curve in $\mathbb{R}^n$. If γ happens to be of general type in both spaces, then we should distinguish the Frenet frame and curvatures coming from the intrinsic geometry of M, from the Frenet frame and curvatures coming from $\mathbb{R}^n$.

Let us compare the *first* intrinsic and extrinsic curvatures of a curve in a hypersurface. Recall that the for a curve in $\mathbb{R}^n$, the first curvature function can be defined for any regular parameterized curve $\gamma\colon I \to \mathbb{R}^n$, being not necessarily of general type in $\mathbb{R}^n$. When $n \geq 3$, it is simply the non-negative function

$$\kappa_1 = \frac{\|\gamma' \wedge \gamma''\|}{\|\gamma'\|^3},$$

for $n = 2$, it is the signed function

$$\kappa_1 = \frac{\det\begin{pmatrix} x' & y' \\ x'' & y'' \end{pmatrix}}{\|\gamma'\|^3} = \pm\frac{\|\gamma' \wedge \gamma''\|}{\|\gamma'\|^3},$$

where $(x(t), y(t))$ are the coordinates of $\gamma(t)$ with respect to the standard basis. The intrinsic counterpart of κ_1 is what we call geodesic curvature.

Definition 4.13.2. Let $\gamma\colon I \to M$ be a regular curve in a hypersurface $M \subset \mathbb{R}^n$ (or, more generally, on a Riemannian manifold M). If $\dim M \geq 3$, then the *geodesic curvature* κ_g of γ is the function

$$\kappa_g = \frac{\|\gamma' \wedge \nabla^2\gamma\|}{\|\gamma'\|^3},$$

where ∇ is the Levi-Cività connection of M, $\nabla^2\gamma = \nabla_{\gamma'}\gamma'$. If $\dim M = 2$, and M is not orientable, or orientable, but not oriented, then we can use the same formula to define the geodesic curvature of γ.

However, if $\dim M = 2$ and M is oriented, then the geodesic curvature of γ is given an orientation depending sign as follows. We set

$$\kappa_g(t) = \begin{cases} \dfrac{\|\gamma'(t) \wedge \nabla^2\gamma(t)\|}{\|\gamma'(t)\|^3}, & \text{if } (\gamma'(t), \nabla^2\gamma(t)) \text{ is a positively oriented basis of } T_{\gamma(t)}M \\ -\dfrac{\|\gamma'(t) \wedge \nabla^2\gamma(t)\|}{\|\gamma'(t)\|^3}, & \text{otherwise.} \end{cases}, \quad ※$$

Proposition 4.13.3. *If $\gamma\colon I \to M \subset \mathbb{R}^n$ $(n \geq 3)$ is a regular curve on a hypersurface M, κ_1 is its first curvature in $\mathbb{R}^n$, κ_g is its geodesic curvature in M, and $k_{\gamma(t)}$ is the normal curvature function of M at $\gamma(t)$, then*

$$\kappa_1(t)^2 = \kappa_g(t)^2 + k_{\gamma(t)}(\gamma'(t))^2.$$

Proof. According to equation (4.14),

$$\gamma''(t) = \nabla^2\gamma(t) + II_{\gamma(t)}(\gamma'(t), \gamma'(t))\mathbf{N}_{\gamma(t)},$$

thus,

$$\gamma'(t) \wedge \gamma''(t) = \gamma'(t) \wedge \nabla^2\gamma(t) + II_{\gamma(t)}(\gamma'(t), \gamma'(t))\gamma'(t) \wedge \mathbf{N}_{\gamma(t)}.$$

Since $\mathbf{N}_{\gamma(t)}$ is orthogonal to both $\gamma'(t)$ and $\nabla^2\gamma(t)$, the bivector $\gamma'(t) \wedge \mathbf{N}_{\gamma(t)}$ is orthogonal to $\gamma'(t) \wedge \nabla^2\gamma(t)$ by formula (1.9). Thus, by the Pythagorean theorem,

$$\|\gamma'(t) \wedge \gamma''(t)\|^2 = \|\gamma'(t) \wedge \nabla^2\gamma(t)\|^2 + (II_{\gamma(t)}(\gamma'(t), \gamma'(t)))^2 \cdot \|\gamma'(t) \wedge \mathbf{N}_{\gamma(t)}\|^2. \tag{4.21}$$

Using formula (1.9) again to compute $\|\gamma'(t) \wedge \mathbf{N}_{\gamma(t)}\|^2$ we obtain

$$\|\gamma'(t) \wedge \mathbf{N}_{\gamma(t)}\|^2 = \det \begin{pmatrix} \|\gamma'(t)\|^2 & 0 \\ 0 & 1 \end{pmatrix} = \|\gamma'(t)\|^2.$$

Substituting back into (4.21) and dividing by $\|\gamma'(t)\|^6$, we get

$$\left(\frac{\|\gamma'(t) \wedge \gamma''(t)\|}{\|\gamma'(t)\|^3}\right)^2 = \left(\frac{\|\gamma'(t) \wedge \nabla^2\gamma(t)\|}{\|\gamma'(t)\|^3}\right)^2 + \left(\frac{II_{\gamma(t)}(\gamma'(t), \gamma'(t))}{\|\gamma'(t)\|^2}\right)^2,$$

which is exactly what we wanted to show. □

Proposition 4.13.4. *The following properties of a regular curve $\gamma\colon I \to M$ lying on a hypersurface M of $\mathbb{R}^{n-1}$ are equivalent.*

(1) *γ is a pre-geodesic curve.*

(2) *The geodesic curvature of* γ *is constant* 0*.*

(3) *The curvature of* γ *equals the normal curvature of the hypersurface in the direction of* γ' *at each point of the curve.*

(4) *The acceleration vector* $\gamma''(t)$ *is in the plane spanned by* $\gamma'(t)$ *and the normal vector* $\mathbf{N}_{\gamma(t)}$ *for all* $t \in I$*.*

(5) *At each point where the osculating plane exists, it contains the normal line of the hypersurface at that point.*

(6) *For any constant speed reparameterization* $\tilde{\gamma}$ *of* γ*, the acceleration vector field* $\tilde{\gamma}''$ *is orthogonal to the hypersurface.*

Proof. (1) $\iff$ (6). A curve γ is pre-geodesic if and only if any of its constant speed reparameterizations $\tilde{\gamma}$ is geodesic. $\tilde{\gamma}$ is geodesic if and only if $\nabla_{\tilde{\gamma}'}\tilde{\gamma}' = 0$. Since $\nabla_{\tilde{\gamma}'}\tilde{\gamma}'(t)$ is the orthogonal projection of $\gamma''(t)$ onto the tangent space of the hypersurface at $\gamma(t)$, $\nabla_{\tilde{\gamma}'}\tilde{\gamma}'(t) = 0$ if and only if $\gamma''(t)$ is orthogonal to the hypersurface at $\gamma(t)$.
If $\tilde{\gamma} \circ h = \gamma$, then the osculating plane of γ at t is the same as the osculating plane of $\tilde{\gamma}$ at $h(t)$.
(2) $\iff$ (3) follows from Proposition 4.13.3. □

4.13.2 Clairaut's Theorem

Alexis Claude Clairaut (1713-1765) was a French mathematician, astronomer and geophysicist. He considered great circles on a sphere centered at the origin. He observed that if we denote by r the distance of a point P on a great circle C from the z-axis and by θ the angle at which the great circle intersects the latitudinal circle at P, then $r\cos(\theta)$ is constant along C. Clairaut's relation was later extended to geodesic curves on surfaces of revolution.
We prove below a generalization of Clairaut's relation to Riemannian manifolds. The theorem presented here is still not the most general version of the assertion as it follows from a more general principle of Hamiltonian and Lagrangian mechanics known as Noether's theorem, and saying that symmetries of Lagrangian or Hamiltonian mechanical systems generate invariants of motion.

Definition 4.13.5. A vector field X on a Riemannian manifold (M, g) is a *Killing vector field* if and only if $\mathcal{L}_X g = 0$, or equivalently, if the flow Φ_t generated by X preserves the Riemannian metric in the sense that $\Phi_t^*(g) = g$ for all $t \in \mathbb{R}$.

Theorem 4.13.6. *Let* X *be a Killing field on a Riemannian manifold* (M, g)*. Then for any geodesic curve* $\gamma\colon I \to M$*, the function* $t \mapsto \langle X_{\gamma(t)}, \gamma'(t)\rangle$ *is constant.*

Proof. The statement is obvious if γ is constant. If γ is not constant, then the length $v = \|\gamma'\|$ of its speed vector is a nonzero constant.
Take two points $a, b \in I$, and consider the variation

$$\gamma_*\colon [-\delta, \delta] \times [a, b] \to M, \qquad \gamma_*(s,t) = \gamma_s(t) = \Phi_s(\gamma(t))$$

of γ, where $\{\Phi_t\}_{t\in\mathbb{R}}$ is the flow generated by X. The variation will be defined if $\delta > 0$ is small enough. As X is a Killing field, Φ_t is an isometry for all t, hence $\gamma_s|_{[a,b]}$ has the same length $l(s)$ as $\gamma|_{[a,b]}$ for any $|s| \leq \delta$. Application of the first variation formula for the length (Theorem 4.12.9) gives

$$\begin{aligned} 0 = l'(0) &= \left\langle \partial_1\gamma_*(0,b), \frac{\gamma'(b)}{\|\gamma'(b)\|} \right\rangle - \left\langle \partial_1\gamma_*(0,a), \frac{\gamma'(a)}{\|\gamma'(a)\|} \right\rangle \\ &= \left\langle X_{\gamma(b)}, \frac{\gamma'(b)}{v} \right\rangle - \left\langle X_{\gamma(a)}, \frac{\gamma'(a)}{v} \right\rangle, \end{aligned}$$

which proves the theorem. □

Consider now a surface of revolution in $\mathbb{R}^3$. If its symmetry axis is the z-axis, then it has a parameterization of the form

$$\mathbf{r}(u,v) = (x(u)\cos v, x(u)\sin v, z(u)),$$

where $u \mapsto (x(u), 0, z(u))$ is the generatrix in the (x,z)-plane. The group of rotations about the z axis acts on the surface of revolution by isometries. This isometry group is the flow of the vector field

$$\mathbf{r}_v(u,v) = (-x(u)\sin v, x(u)\cos v, 0),$$

therefore, $\mathbf{r}_v$ is a Killing vector field on the surface.
Assume now that $\gamma\colon I \to M = \operatorname{im}\mathbf{r}$, $\gamma(t) = \mathbf{r}(u(t), v(t))$ is a geodesic curve on the surface. Then the quantity

$$\langle \mathbf{r}_v(u(t), v(t)), \gamma'(t)\rangle = \|\mathbf{r}_v(u(t), v(t))\| \cdot \|\gamma'(t)\| \cdot \cos(\theta(t)),$$

where $\theta(t)$ is the angle between $\mathbf{r}_v(u(t), v(t))$ and $\gamma'(t)$, is constant. As $\mathbf{r}_v$ is tangent to the circles of latitude, $\theta(t)$ is the angle at which the geodesic curve γ intersects the circle of latitude through $\gamma(t)$. It is also clear from the formula for $\mathbf{r}_v$ that $\|\mathbf{r}_v(u,v)\| = |x(u)|$ is the distance of the point $\mathbf{r}(u,v)$ from the z-axis. Since $\|\gamma'\|$ is constant for any geodesic curve, we obtain the following corollary of the theorem for surfaces of revolution.

Corollary 4.13.7 (Clairaut Relation for Surfaces of Revolution)**.** *If γ is a non-constant geodesic curve on a surface of revolution, $r(t)$ is the distance of $\gamma(t)$ from the axis of the surface, $\theta(t)$ is the angle at which γ crosses the circle of latitude through $\gamma(t)$, then the function $t \mapsto r(t)\cos(\theta(t))$ is constant.*

4.13.3 Moving Orthonormal Frames Along a Hypersurface

Let $M \subset \mathbb{R}^n$ be a smooth hypersurface, $\mathbf{x}\colon M \to \mathbb{R}^n$ be its embedding map $\mathbf{e}_1, \dots, \mathbf{e}_n\colon M \to \mathbb{R}^n$ be a moving orthonormal frame along M, such that $\mathbf{e}_n$ is the unit normal vector field of M. We consider the maps $\mathbf{x}, \mathbf{e}_1, \dots, \mathbf{e}_n$ to be vector-valued functions on M. Then their differentials are vector-valued differential 1-forms, that can be written as a linear combination of the basis fields $\mathbf{e}_i$ with real valued 1-forms as coefficients

$$d\mathbf{x} = \sum_{i=1}^{n} \sigma^i \mathbf{e}_i, \tag{4.22}$$

$$d\mathbf{e}_i = \sum_{i=1}^{n} \omega_i^j \mathbf{e}_j. \tag{4.23}$$

Proposition 4.13.8. *If $\mathbf{v} \in T_pM$ is a tangent vector of M, then $d\mathbf{x}_p(\mathbf{v}) = \mathbf{v}$ and $\sigma_p^i(\mathbf{v}) = \langle \mathbf{v}, \mathbf{e}_i \rangle$. In particular, $\sigma^n \equiv 0$, since $\mathbf{e}_n(p)$ is orthogonal to T_pM.*

Proof. If $\gamma\colon (-\epsilon, \epsilon) \to M$ is a smooth curve that represents $\mathbf{v}$, which means that $\gamma(0) = p$ and $\gamma'(0) = \mathbf{v}$, then by the definition of the differential of functions we have

$$(d\mathbf{x})_p(\mathbf{v}) = (\mathbf{x} \circ \gamma)'(0) = \gamma'(0) = \mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{e}_i(p) \rangle \mathbf{e}_i(p) = \sum_{i=1}^{n-1} \langle \mathbf{v}, \mathbf{e}_i(p) \rangle \mathbf{e}_i(p).$$

□

We can arrange the 1-forms ω_i^j into a matrix ω using i as the row index and j as the column index. ω can be thought of as a matrix valued 1-form as well, which assigns to a tangent vector $\mathbf{v} \in T_pM$ the matrix $\big(\omega_{i,p}^j(\mathbf{v})\big)$.

Proposition 4.13.9. *ω is skew-symmetric, i.e., $\omega_i^j + \omega_j^i \equiv 0$ for all $1 \le i, j \le n$.*

Proof. Since the frame $\mathbf{e}_1, \dots, \mathbf{e}_n$ is orthonormal, $\langle \mathbf{e}_i, \mathbf{e}_j \rangle \equiv \delta_{ij}$. Differentiating we obtain $0 = \langle d\mathbf{e}_i, \mathbf{e}_j \rangle + \langle \mathbf{e}_i, d\mathbf{e}_j \rangle = \omega_i^j + \omega_j^i$. □

The one forms σ^i and ω_i^j are not independent of one another. The identity $d \circ d = 0$ for the exterior differentiation of differential forms yields some compatibility relations between them. Considering $dd\mathbf{x}$ we obtain

$$\begin{aligned} 0 = dd\mathbf{x} &= \sum_{i=1}^{n-1}\left(d\sigma^i \mathbf{e}_i - \sum_{j=1}^{n} \sigma^i \wedge \omega_i^j \mathbf{e}_j\right) \\ &= \sum_{j=1}^{n-1}\left(d\sigma^j - \sum_{i=1}^{n-1} \sigma^i \wedge \omega_i^j\right)\mathbf{e}_j - \left(\sum_{i=1}^{n-1}\sigma^i \wedge \omega_i^n\right)\mathbf{e}_n, \end{aligned}$$

which gives

$$d\sigma^j - \sum_{i=1}^{n-1} \sigma^i \wedge \omega_i^j = 0 \text{ for } 1 \le j \le n-1, \text{ and } \sum_{i=1}^{n-1}\sigma^i \wedge \omega_i^n = 0. \tag{4.24}$$

Vanishing of $dd\mathbf{e}_i$ gives

$$\begin{aligned} 0 = dd\mathbf{e}_i = d\left(\sum_{j=1}^{n} \mathbf{e}_j \omega_i^j\right) &= \sum_{j=1}^{n}(d\mathbf{e}_j \wedge \omega_i^j + \mathbf{e}_j d\omega_i^j) \\ &= \sum_{k=1}^{n} d\mathbf{e}_k \wedge \omega_i^k + \sum_{j=1}^{n}\sum_{k=1}^{n} \mathbf{e}_k \omega_j^k \wedge \omega_i^j. \end{aligned}$$

This means that

$$d\omega_i^k + \sum_{j=1}^{n} \omega_j^k \wedge \omega_i^j = 0, \text{ for all } 1 \le i, j \le n, \text{ or in matrix form, } d\omega - \omega \wedge \omega = 0.$$

4.13.4 Relation to Earlier Formulae for Parameterized Hypersurfaces

Recall that for a regular parameterized hypersurface $\mathbf{r}\colon \Omega \to \mathbb{R}^n$, the shape of the hypersurface and the way it is parameterized by $\mathbf{r}$ is encoded in the matrix valued functions $\mathcal{G}\colon \Omega \to \mathbb{R}^{(n-1)\times(n-1)}$ and $\mathcal{B}\colon \Omega \to \mathbb{R}^{(n-1)\times(n-1)}$ obtained form the first and second fundamental forms. From the viewpoint of manifold theory, if $\mathbf{r}$ is injective, then $M = \mathbf{r}(\Omega)$ is an injectively immersed submanifold of $\mathbb{R}^n$, the inverse of $\mathbf{r}$ is a chart on M. The first fundamental form is the Riemannian metric inherited from the metric of $\mathbf{R}^n$, and $\mathcal{G}$ is the matrix of this Riemannian metric with respect to the chart $\mathbf{r}^{-1}$.

If $\mathbf{e}_i$ is a moving orthonormal frame along M as above, then the first fundamental form is given by the fact that $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ is an orthonormal frame in the tangent space. For two tangent vectors $\mathbf{v}, \mathbf{w} \in T_pM$ we have

$$I_p(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \mathbf{w}\rangle = \sum_{i=1}^{n-1} \sigma_p^i(\mathbf{v})\sigma_p^i(\mathbf{w}).$$

Thus, the first fundamental form can be expressed by the differential forms σ^i.

To find the relation between the forms ω_i^j and the previously introduced notions for parameterized hypersurfaces, let us start with the equation

$$\partial_X Y = \nabla_X Y + II(X,Y)\mathbf{N},$$

where X and Y are tangential vector fields of M, ∂ and ∇ are the Levi-Cività connections of $\mathbb{R}^n$ and M respectively, II is the second fundamental form. It is clear that $\mathbf{e}_n = \pm\mathbf{N}$, assume $\mathbf{e}_n$ is chosen to be equal to $\mathbf{N}$. Applying this formula for $X = \mathbf{e}_i$ and $Y = \mathbf{e}_j$ for $1 \le i, j \le n-1$, we obtain

$$d\mathbf{e}_i(\mathbf{e}_j) = \partial_{\mathbf{e}_j}\mathbf{e}_i = \nabla_{\mathbf{e}_j}\mathbf{e}_i + II(\mathbf{e}_i, \mathbf{e}_j)\mathbf{e}_n.$$

Comparing this equation with (4.23), we get

$$\nabla_{\mathbf{e}_j}\mathbf{e}_i = \sum_{k=1}^{n-1} \omega_i^k(\mathbf{e}_j)\mathbf{e}_k,$$

and

$$II(\mathbf{e}_i, \mathbf{e}_j) = \omega_i^n(\mathbf{e}_j). \tag{4.25}$$

It is known that the Levi-Cività connection is determined by the first fundamental form. Thus, there should be a formula expressing ω_i^j for $1 \le i, j \le n-1$ in terms of the forms σ^k $(1 \le k \le n-1)$. Let us find this formula. The forms σ_p^k form a basis in the dual space T_p^*M, so we can write the 1-forms ω_i^j as their linear combinations $\omega_i^j = \sum_{k=1}^{n-1} b_{ik}^j \sigma^k$, where the coefficients $b_{ik}^j = \omega_i^j(\mathbf{e}_k)$ are smooth functions on M. Substituting into (4.24) we obtain a system of linear equations

$$d\sigma^j = \sum_{1 \le i < k \le n-1} (b_{ik}^j - b_{ki}^j)\sigma^i \wedge \sigma^k \text{ for } 1 \le j \le n-1 \tag{4.26}$$

and

$$0 = \sum_{1 \le i < k \le n-1} (b_{ik}^n - b_{ki}^n)\sigma^i \wedge \sigma^k.$$

The second equation is equivalent to the symmetry of the second fundamental form

$$II(\mathbf{e}_i, \mathbf{e}_k) = b_{ik}^n = b_{ki}^n = II(\mathbf{e}_k, \mathbf{e}_i).$$

Evaluating (4.26) on the basis vectors $\mathbf{e}_i$ and $\mathbf{e}_k$, taking the antisymmetry $b_{ki}^j = -b_{ji}^k$ into account and permuting the rôle of i, j and k we obtain the equations

$$d\sigma^j(\mathbf{e}_i, \mathbf{e}_k) = b^j_{ik} + b^k_{ji},$$
$$d\sigma^k(\mathbf{e}_j, \mathbf{e}_i) = b^k_{ji} + b^i_{kj},$$
$$d\sigma^i(\mathbf{e}_k, \mathbf{e}_j) = b^i_{kj} + b^j_{ik}.$$

Solving this system of equations for b^j_{ik}, we get

$$b^j_{ik} = \frac{1}{2}(d\sigma^i(\mathbf{e}_k, \mathbf{e}_j) + d\sigma^k(\mathbf{e}_j, \mathbf{e}_i) - d\sigma^j(\mathbf{e}_i, \mathbf{e}_k)),$$

therefore,

$$\omega^j_i = \sum_{k=1}^{n-1} \frac{1}{2}(d\sigma^i(\mathbf{e}_k, \mathbf{e}_j) + d\sigma^k(\mathbf{e}_j, \mathbf{e}_i) - d\sigma^j(\mathbf{e}_i, \mathbf{e}_k))\sigma^k \text{ for } 1 \le i, j \le n-1.$$

4.13.5 The Gauss–Bonnet Formula

Consider a surface $M \subset \mathbb{R}^3$ in $\mathbb{R}^3$. Define a local orthonormal frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ over an open subset U of M. Define the one forms σ^i and ω^j_i in the usual way. Then the Levi-Cività connection ∇ of M is determined by the form $\omega^2_1 = -\omega^1_2$ through the relations

$$\nabla_X \mathbf{e}_1 = \omega^2_1(X)\mathbf{e}_2,$$
$$\nabla_X \mathbf{e}_2 = \omega^1_2(X)\mathbf{e}_1.$$

The structure equations for $d\omega^2_1$ yield the equation

$$d\omega^2_1 = \omega^1_1 \wedge \omega^2_1 + \omega^2_1 \wedge \omega^2_2 + \omega^3_1 \wedge \omega^2_3 = -\omega^3_1 \wedge \omega^3_2.$$

The form $\omega^3_1 \wedge \omega^3_2$ is a multiple of the area form $\sigma^1 \wedge \sigma^2$ of the surface M. The multiplier f in the equation

$$\omega^3_1 \wedge \omega^3_2 = f\sigma^1 \wedge \sigma^2$$

can be obtained by evaluating both sides on the orthonormal system $\mathbf{e}_1, \mathbf{e}_2$. This gives

$$f = \omega^3_1 \wedge \omega^3_2(\mathbf{e}_1, \mathbf{e}_2) = \det \begin{pmatrix} \omega^3_1(\mathbf{e}_1) & \omega^3_1(\mathbf{e}_2) \\ \omega^3_2(\mathbf{e}_1) & \omega^3_2(\mathbf{e}_2) \end{pmatrix}.$$

According to equation (4.25), the matrix on the right-hand side is the matrix of the second fundamental form of M with respect to the orthonormal basis

$\mathbf{e}_1, \mathbf{e}_2$, consequently, its determinant is the Gaussian curvature K of M. We conclude that

$$d\omega_2^1 = \omega_1^3 \wedge \omega_2^3 = K\sigma^1 \wedge \sigma^2.$$

Applying Stokes' theorem this formula implies that fixing an orientation of U, and a regular domain $D \subset U$,

$$\int_D K\sigma^1 \wedge \sigma^2 = \int_{\partial D} \omega_2^1.$$

Variants of this formula together with the geometrical interpretations of the second integral are called *local Gauss–Bonnet formulae.*

4.13.6 Steiner's Formula

Steiner's formula, named after the Swiss geometer Jacob Steiner (1796-1863), asserts that for a compact convex subset K of $\mathbb{R}^n$, the volume of the parallel body

$$B(K,r) = \{q \in \mathbb{R}^n \mid d(q,K) \le r\}$$

is a polynomial

$$\lambda_n(B(K,r)) = \binom{n}{0} W_0(K) + \binom{n}{1} W_1(K) r + \cdots + \binom{n}{n} W_n(K) r^n$$

of degree n of r. The constant $W_j(K)$ appearing in the coefficient of r^j is a geometrical invariant of the convex body K, called its *jth quermassintegral.* The number $V_j(K)$ given by the equation $\omega_{n-j} V_j(K) = \binom{n}{n-j} W_{n-j}(K)$, where ω_{n-j} denotes the volume of the $(n-j)$-dimensional Euclidean unit ball, is the *jth intrinsic volume of K.*

For example, the constant term $W_0(K) = V_n(K)$ is the volume of K. The coefficient $2V_{n-1}(K) = nW_1(K)$ of r is the surface volume of the boundary of K. The coefficient of r^{n-1} is the average width of K multiplied with a constant depending only on the dimension. The coefficient $\omega_n V_0(K) = W_n(K)$ of r^n does not depend on K, it is the volume of the n-dimensional unit ball for all K.

In this part we compute a smooth analogue of Steiner's formula. A practical formulation of the problem we want to solve is "How much paint we need to cover one side of a surface with a coat of paint of thickness r?". As the layer of paint has positive thickness, it fills a part of space between the original surface M and a parallel surface of M lying at distance r from M.

Definition 4.13.10. Let M be a hypersurface in $\mathbb{R}^n$, with unit normal vector field $\mathbf{N}$ and moving orthonormal frame $\mathbf{e}_1, \dots, \mathbf{e}_n = \mathbf{N}$. For a real number $r \in \mathbb{R}$, we define the *parallel surface M_r of M lying at distance* r as the image of the map $\mathbf{x}_r \colon M \to M_r$, $\mathbf{x}_r = \mathbf{x} + r\mathbf{e}_n$. ※

The following proposition summarizes some basic facts on the geometry of the parallel surfaces.

Proposition 4.13.11. *The differential of the map $\mathbf{x}_r : M \to M_r$ has maximal rank $(n-1)$ at $p \in M$ if and only if $1/r$ is not a principal curvature of M at p. If p is such a point, then*

- *the tangent spaces T_pM and $T_{\mathbf{x}_r(p)}M_r$ are parallel;*
- *the principal directions of M at p are also principal directions of M_r at $\mathbf{x}_r(p)$. If the principal curvatures of M and M_r are computed with respect to the unit normal vector field $\mathbf{e}_n$, then if the principal curvature of M corresponding to the principal direction $\mathbf{v} \in T_pM$ is κ_i, then the principal curvature of M_r corresponding to the same principal direction $\mathbf{v} \in T_{\mathbf{x}_r(p)}M_r$ is $\kappa_i/(1 - r\kappa_i)$.*

Proof. The differential of the map $\mathbf{x}_r$ is $d\mathbf{x}_r = d\mathbf{x} + rd\mathbf{e}_n = \sum_{i=1}^{n-1} \mathbf{e}_i(\sigma^i + r\omega_n^i)$. In a coordinate free way, $d\mathbf{x}_r$ takes the tangent vector $\mathbf{v} \in T_pM$ to the vector $\mathbf{v} - rL_p(\mathbf{v})$, where L_p is the Weingarten map of M at $p \in M$. This means that the map $d\mathbf{x}_r$ has maximal rank at $p \in M$ if and only if $1/r$ is not a principal curvature of M at p. Assuming that $\mathbf{x}_r$ has maximal rank, the tangent space of M_r at $\mathbf{x}_r(p)$ is parallel to $\operatorname{im}(I_{T_pM} - rL_p) = T_pM$. Thus, we can choose $\mathbf{e}_1 \circ \mathbf{x}_r^{-1}, \dots, \mathbf{e}_n \circ \mathbf{x}_r^{-1}$ as an orthonormal frame on M_r. By the chain rule, the Weingarten map of M_r at $\mathbf{x}_r(p)$ is $-d(\mathbf{e}_n \circ \mathbf{x}_r^{-1})(\mathbf{x}_r(p)) = L_p \circ (I_{T_pM} - rL_p)^{-1}$. This proves that if $\mathbf{v}$ is a principal direction of M with principal curvature κ_i, then $\mathbf{v}$ is also a principal direction of M_r with principal curvature $\kappa_i/(1 - r\kappa_i)$. $\square$

Proposition 4.13.12. *If $D \subset M$ is a compact connected regular domain in M such that $1/r$ is not a principal curvature of M at any point of D, then the volume measure of $D_r = \mathbf{x}_r(D) \subset M_r$ can be expressed by the formula*

$$\mu_{n-1}(D_r) = \left| \sum_{i=0}^{n-1} (-1)^i \left(\int_D K_i(p)dp \right) r^i \right|,$$

as the absolute value of a degree $n-1$ polynomial of r, where $K_i(p)$ denotes the ith elementary symmetric polynomial

$$\begin{aligned} K_0(p) &= 1, \\ K_1(p) &= \kappa_1(p) + \dots + \kappa_{n-1}(p), \\ &\vdots \\ K_i(p) &= \sum_{1 \le j_i < j_2 < \dots < j_i \le n-1} \kappa_{j_1}(p) \cdot \kappa_{j_2}(p) \cdots \kappa_{j_i}(p), \end{aligned} \tag{4.27}$$

$$\vdots$$

$$K_{n-1}(p) = \kappa_1(p) \cdot \kappa_2(p) \cdots \kappa_{n-1}(p)$$

of the principal curvatures $\kappa_1(p), \dots, \kappa_{n-1}(p)$ at p.

Proof. As the derivative of $\mathbf{x}_r$ at p is the linear map $(I_{T_pM} - rL_p)$, the pull-back of the volume form of M_r to M is $\det(I_{T_{\mathbf{x}(.)}M} - rL_{\mathbf{x}(.)})\sigma_1 \wedge \cdots \wedge \sigma^{n-1}$. For this reason,

$$\mu_{n-1}(D_r) = \int_D |\det(I_{T_pM} - rL_p)| d\sigma(p),$$

where the integral is taken with respect to the volume measure σ of M. The integrand can be expressed with the help of the characteristic polynomial of L_p as

$$\begin{aligned} \det(I_{T_pM} - rL_p)) &= r^{n-1} \cdot \det\left(\frac{1}{r} I_{T_pM} - L_p)\right) = r^{n-1} \prod_{i=1}^{n-1} \left(\frac{1}{r} - \kappa_i(p)\right) \\ &= r^{n-1} \sum_{i=0}^{n-1} (-1)^i K_i(p) \frac{1}{r^{n-1-i}} = \sum_{i=0}^{n-1} (-1)^i K_i(p) r^i. \end{aligned}$$

Since $1/r$ is not a principal curvature at any point of D, $\det(I_{T_pM} - rL_p)$ does not vanish on D, therefore, as D is connected, it has constant sign. As a corollary we obtain that the integral of its absolute value is the absolute value of its integral. Combining these observations,

$$\mu_{n-1}(D_r) = \left| \int_D \det(I_{T_pM} - rL_p) d\sigma(p) \right| = \left| \sum_{i=0}^{n-1} (-1)^i \left(\int_D K_i(p) dp \right) r^i \right|,$$

as we wanted to prove. □

The layer of thickness r over the regular domain $D \subset M$, lying between D and D_r is parameterized by the map

$$h \colon M \times [0, r] \to \mathbb{R}^n, \qquad h(p, s) = p + s\mathbf{e}_n(p).$$

We claim that if D is compact and $r > 0$ is sufficiently small, then h is a diffeomorphism. Using the natural decomposition $T_{(p,s)}(M \times [0, r]) \cong T_pM \oplus T_s[0, r]$, the tangent space $T_{(p,s)}(M \times [0, r])$ is spanned by the vectors $\mathbf{e}_1(p), \dots, \mathbf{e}_{n-1}(p)$ and $\mathbf{d}(s)$, where the vector field $\mathbf{d}$ on $\mathbb{R}$ is the unit vector field corresponding to the derivation of functions with respect to their single variable.

The images of these tangent vectors under the derivative map of h are $\mathbf{e}_1(p) - sL_p(\mathbf{e}_1(p)), \dots, \mathbf{e}_{n-1}(p) - sL_p(\mathbf{e}_{n-1}(p))$ and $\mathbf{e}_n(p)$. These vectors are linearly independent if $1/s$ is not a principal curvature of M at p. If M has a positive principal curvature at a point p of D, then let

$$\kappa_+ = \max_{p\in D} \max_{1\leq i\leq n-1} \kappa_i(p)$$

be the maximal value of principal curvatures at all points of D, otherwise set $\kappa_+ = 0$. From the previous arguments, if $0 < r < 1/\kappa_+$, then the derivative of h has maximal rank n at each point of $D \times [0, r]$, hence h is a local diffeomorphism by the inverse function theorem. Suppose that h is not a diffeomorphism for any $r > 0$. Then, since h is a local diffeomorphism for small values of r, the reason why it is not a diffeomorphism for these r's, is that h is not injective. If h is not injective for the $r = 1/2, 1/3, \dots, 1/k, \dots,$ then we can find two sequences of pairs $(p_k, s_k) \neq (p_k^*, s_k^*)$ such that $p_k, p_k^* \in M$, $s_k, s_k^* \in (0, 1/k)$, and $p_k + s_k\mathbf{e}_n(p_k) = p_k^* + s_k^*\mathbf{e}_n(p_k^*)$. Since D is compact, there is a convergent subsequence of the sequence p_k, say p_{k_i} which tends to p as i tends to infinity. Then the pairs (p_{k_i}, s_{k_i}) and $(p_{k_i}^*, s_{k_i}^*)$ tend to the pair $(p, 0)$ But then $(p, 0)$ would not have a neighborhood on which h is injective. This contradicts that h is a local diffeomorphism.

Proposition 4.13.13. *Using the above notation, if h is a diffeomorphism the volume of the layer of thickness r over D parameterized by h is equal to the polynomial*

$$\sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \left(\int_D K_{k-1}(p) d\sigma(p) \right) r^k.$$

Proof. Introduce on $M \times [0, r]$ the Riemannian metric, in which the basis $\mathbf{e}_1(p), \dots, \mathbf{e}_{n-1}(p)$, $\mathbf{d}(s)$ is orthonormal. The pull-back of the volume form of $\mathbb{R}^n$ by h is the form

$$\pm \det(I_{T_pM} - sL_p)) d\sigma^1 \wedge \dots \wedge d\sigma^{n-1} \wedge ds,$$

where the sign depends on the orientations chosen. Thus,

$$\lambda_n \left(\bigcup_{s\in[0,r]} D_s \right) = \int_0^r \int_D |\det(I_{T_pM} - sL_p)| d\sigma(p) ds.$$

As h is diffeomorphism, it does not vanish on $D \times [0, r]$. It is also clear that $\det(I_{T_pM} - sL_p) = 1$ for $s = 0$, therefore $\det(I_{T_pM} - sL_p) > 0$ on $D \times [0, r]$, and the absolute value can be omitted. We conclude

$$\lambda_n\left(\bigcup_{s\in[0,r]} D_s\right) = \int_0^r \left(\sum_{i=0}^{n-1}(-1)^i \left(\int_D K_i(p)dp\right) s^i\right) ds$$
$$= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \left(\int_D K_{k-1}(p)d\sigma(p)\right) r^k,$$

as it was to be proven. □

Exercise 4.13.14. Prove that the above proposition implies Steiner's formula for the volume of parallel bodies of a convex body K, the boundary $M = \partial K$ of which is a smooth hypersurface in $\mathbb{R}^n$. Show that we get the following explicit expression for the coefficients of the polynomial:

$$\lambda_n(B(K,r)) = \lambda_n(K) + \sum_{k=1}^{n} \frac{1}{k} \left(\int_D |K_{k-1}(p)|d\sigma(p)\right) r^k. \quad ¤$$

Exercise 4.13.15. Show that if K is a convex body in $\mathbb{R}^n$ bounded by a smooth hypersurface M with Gauss–Kronecker curvature function $K = K_{n-1}$, then

$$\int_M |K(p)|d\sigma(p) = n\omega_n,$$

where ω_n is the volume of the n-dimensional unit ball (consequently, $n\omega_n$ is the surface volume of the unit sphere $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$). ¤

4.13.7 Minkowski's Formula

Being important geometrical invariants of a convex body, the coefficients of the polynomial in Steiner's formula are studied thoroughly. In this part, we focus on the coefficient r^{n-1}, which is, up to some constant multiplier, the integral of the $(n-1)$st elementary symmetric polynomial of the principal curvatures of the boundary hypersurface of K. Hermann Minkowski (1864-1909) proved that if K is a compact convex set in $\mathbb{R}^n$, then this coefficient is the average width of K up to some constant, depending only on the dimension n.

We prove below a formula which is true for any compact hypersurface M in $\mathbb{R}^n$ which, in the special case when M is the boundary of a convex set reduces to Minkowski's formula.

Proposition 4.13.16 (Minkowki's Formula)**.** *Let M be a smooth compact hypersurface in $\mathbb{R}^n$, $\mathbf{e}_n\colon M \to \mathbb{R}^n$ be a unit normal vector field on M. Let $p_n\colon M \to \mathbb{R}$, $p_n(p) = \langle \mathbf{e}_n, p\rangle$ denote the signed distance of the origin from* *the affine tangent space of M at p. Then denoting by σ the surface volume measure on M, we have*

$$\frac{-1}{n-1}\int_M K_{n-2}d\sigma = \int_M p_n K_{n-1} d\sigma,$$

where the elementary symmetric polynomials K_i of the principal curvatures of M taken with respect to the unit normal vector field $\mathbf{e}_n$ are defined by (4.27).

Proof. Let $\mathbf{x}\colon M \to \mathbb{R}^n$ be the inclusion map, and choose a positively oriented orthonormal frame $\mathbf{e}_1, \dots, \mathbf{e}_n$ over an open subset U of M and orient M so that the vector fields $(\mathbf{e}_1, \dots, \mathbf{e}_{n-1})$ give positively oriented bases at points of U. Consider the functions $p_i\colon U \to \mathbb{R}$, $p_i(p) = \langle \mathbf{e}_i, p\rangle$. Using the fundamental equations of hypersurfaces, we can express the differentials of the functions p_i by the equations

$$\begin{aligned} dp_i &= \langle d\mathbf{e}_i, \mathbf{x}\rangle + \langle \mathbf{e}_i, d\mathbf{x}\rangle = \left\langle \sum_{j=1}^{n} \mathbf{e}_j \omega_i^j, \mathbf{x} \right\rangle + \left\langle \mathbf{e}_i, \sum_{j=1}^{n-1} \mathbf{e}_j \sigma^j \right\rangle \\ &= \begin{cases} \sum_{j=1}^n p_j \omega_i^j + \sigma^i, & \text{if} \quad i \le n-1, \\ \sum_{j=1}^n p_j \omega_n^j, & \text{if} \quad i = n. \end{cases} \end{aligned}$$

Define the differential $(n-2)$-form η on U as

$$\eta = \sum_{i=1}^{n-1} (-1)^{i+1} p_i \omega_1^n \wedge \dots \wedge \widehat{\omega_i^n} \wedge \dots \wedge \omega_{n-1}^n,$$

where the hat above ω_i^n means that ω_i^n is omitted. Compute the differential of η. Clearly,

$$\begin{aligned} d\eta = &\sum_{i=1}^{n-1} (-1)^{i+1} dp_i \wedge \omega_1^n \wedge \dots \wedge \widehat{\omega_i^n} \wedge \dots \wedge \omega_{n-1}^n + \\ &+ \sum_{i=1}^{n-1} \sum_{1 \le j < i} (-1)^{i+j} p_i \omega_1^n \wedge \dots \wedge d\omega_j^n \wedge \dots \wedge \widehat{\omega_i^n} \wedge \dots \wedge \omega_{n-1}^n + \\ &+ \sum_{i=1}^{n-1} \sum_{i < j \le n-1} (-1)^{i+j-1} p_i \omega_1^n \wedge \dots \wedge \widehat{\omega_i^n} \wedge \dots \wedge d\omega_j^n \wedge \dots \wedge \omega_{n-1}^n, \end{aligned}$$

hence

$$
\begin{aligned}
d\eta = &\sum_{i=1}^{n-1}\sum_{j=1}^{n}(-1)^{i+1}p_j\omega_i^j \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\
&+ \sum_{i=1}^{n-1}(-1)^{i+1}\sigma^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\
&+ \sum_{i=1}^{n-1}\sum_{1\le j<i}\sum_{k=1}^{n}(-1)^{i+j}p_i\omega_1^n \wedge \cdots \wedge (\omega_j^k \wedge \omega_k^n) \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\
&+ \sum_{i=1}^{n-1}\sum_{i<j\le n-1}\sum_{k=1}^{n}(-1)^{i+j-1}p_i\omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge (\omega_j^k \wedge \omega_k^n) \wedge \cdots \wedge \omega_{n-1}^n.
\end{aligned}
$$

The sum of the last two terms can be simplified as

$$
\begin{aligned}
&\sum_{i=1}^{n-1}\sum_{1\le j<i}(-1)^{i+j}p_i\omega_1^n \wedge \cdots \wedge (\omega_j^i \wedge \omega_i^n) \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\
&\qquad + \sum_{i=1}^{n-1}\sum_{i<j\le n-1}(-1)^{i+j-1}p_i\omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge (\omega_j^i \wedge \omega_i^n) \wedge \cdots \wedge \omega_{n-1}^n \\
&= \sum_{i=1}^{n-1}\sum_{1\le j<i}(-1)^{(i+j)+(j-1)+(i-j-1)}p_i\omega_j^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_j^n} \wedge \cdots \wedge \omega_{n-1}^n + \\
&\quad + \sum_{i=1}^{n-1}\sum_{i<j\le n-1}(-1)^{(i+j-1)+(j-2)+(j-i-1)}p_i\omega_j^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_j^n} \wedge \cdots \wedge \omega_{n-1}^n \\
&= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}(-1)^j p_i\omega_j^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_j^n} \wedge \cdots \wedge \omega_{n-1}^n.
\end{aligned}
$$

Substituting back into the last formula for $d\eta$ we obtain

$$
\begin{aligned}
d\eta =&p_n\sum_{i=1}^{n-1}(-1)^{i+1}\omega_i^n \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\
&\qquad + \sum_{i=1}^{n-1}(-1)^{i+1}\sigma^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n \\
=&(n-1)p_n\omega_1^n \wedge \cdots \wedge \omega_{n-1}^n + \sum_{i=1}^{n-1}(-1)^{i+1}\sigma^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n.
\end{aligned}
\tag{4.28}
$$

The forms η, $p_n \cdot \omega_1^n \wedge \cdots \wedge \omega_{n-1}^n$ and $\sum_{i=1}^{n-1}(-1)^{i+1}\sigma^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n$ are natural forms on M in the sense that they do not depend on the choice of the orthonormal frame. Indeed, if $\mathbf{v}_1, \ldots, \mathbf{v}_{n-2} \in T_pM$, then $\eta_p(\mathbf{v}_1, \ldots, \mathbf{v}_{n-2})$ is the signed volume of the parallelepiped spanned by the orthogonal projection $\pi(p) = \sum_{i=1}^{n-1} p_i(p)\mathbf{e}_i(p)$ of p onto T_pM and $L_p(\mathbf{v}_1), \ldots, L_p(\mathbf{v}_{n-2})$, where L_p is the Weingarten map at p. The differential form $p_n \wedge \omega_1^n \wedge \cdots \wedge \omega_{n-1}^n$ is the volume form of M multiplied by the product of the Gauss–Kronecker curvature K_{n-1} and the signed distance p_n of the origin from the tangent plane. Naturality of the third form follows from equation (4.28) and naturality of the first two forms. Computing the third form using a frame, whose vectors point in principal directions at a point, we see easily that

$$\sum_{i=1}^{n-1}(-1)^{i+1}\sigma^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n = K_{n-2}\sigma^1 \wedge \cdots \wedge \sigma^{n-1}.$$

The importance of naturality is that although the forms η, $p_n \cdot \omega_1^n \wedge \cdots \wedge \omega_{n-1}^n$ and $\sum_{i=1}^{n-1}(-1)^{i+1}\sigma^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n$ were defined locally, on an open set, on which a local orthonormal frame $\mathbf{e}_1, \ldots, \mathbf{e}_n$ can be defined, these forms are globally defined on M and relation (4.28) also holds globally on M. If M is a compact hypersurface (with no boundary), then by Stokes' theorem

$$\int_M d\eta = \int_{\partial M} \eta = 0,$$

thus, integrating (4.28) we obtain the formula

$$\frac{-1}{n-1}\int_M K_{n-2}d\sigma = \int_M p_n K_{n-1} d\sigma$$

we wanted to show. □

Exercise 4.13.17. Prove that for a compact hypersurface $M \subset \mathbb{R}^n$,

$$\int_M \mathbf{e}_n K_{n-1} d\sigma = \mathbf{0}.$$ ¤

Hint: Apply Minkowski's formula with different choices of the origin of the coordinate system.
In the special case, when M is strictly convex and the Gauss map $\mathbf{e}_n \colon M \to \mathbb{S}^{n-1}$ is a diffeomorphism between M and $\mathbb{S}^{n-1}$, the second integral can be rewritten as an integral on the sphere.

Since the derivative map $T_p\mathbf{e}_n\colon T_pM \to T_{\mathbf{e}_n(p)}\mathbb{S}^{n-1}$ is $-L_p$, the pull-back of the volume form of $\mathbb{S}^{n-1}$ by $\mathbf{e}_n$ is $(-1)^{n-1}K_{n-1}\sigma^1\wedge\cdots\wedge\sigma^{n-1}$. Consequently,

$$\begin{aligned}\int_M (-1)^n p_n K_{n-1} d\sigma &= \int_{\mathbb{S}^{n-1}} p_n(\mathbf{e}_n^{-1}(\mathbf{u}))d\mathbf{u}\\ &= \frac{1}{2}\int_{\mathbb{S}^{n-1}} (p_n(\mathbf{e}_n^{-1}(\mathbf{u})) + p_n(\mathbf{e}_n^{-1}(-\mathbf{u})))d\mathbf{u}.\end{aligned}$$

Definition 4.13.18. The *width* $w_K(\mathbf{u})$ *of a compact set* $K \subset \mathbb{R}^n$ *in the direction* $\mathbf{u} \in \mathbb{S}^{n-1}$ is the width of the narrowest slab bounded by hyperplanes orthogonal to $\mathbf{u}$ that contains K. The width is given by the formula

$$w_K(\mathbf{u}) = \max_{\mathbf{x}\in K}\langle \mathbf{u}, \mathbf{x}\rangle - \min_{\mathbf{x}\in K}\langle \mathbf{u}, \mathbf{x}\rangle. \quad ※$$

Definition 4.13.19. The *mean width* $\bar{w}(K)$ of a compact set $K \subset \mathbb{R}^n$ is the average of its widths in different directions, that is

$$\bar{w}(K) = \frac{1}{n\omega_n}\int_{\mathbb{S}^{n-1}} w_K(\mathbf{u})d\mathbf{u}.$$

If K is a compact convex regular domain with boundary $M = \partial K$, and $\mathbf{e}_n\colon M \to \mathbb{R}^n$ is the exterior unit normal vector field along M, then for $\mathbf{u} \in \mathbb{S}^{n-1}$, $\langle \mathbf{u}, \mathbf{x}\rangle$ attains its maximum (or minimum respectively) on K at the points $\mathbf{x} \in M$ of the boundary, at which $\mathbf{e}_n(\mathbf{x}) = \mathbf{u}$ (or $\mathbf{e}_n(\mathbf{x}) = -\mathbf{u}$ respectively). At such a point $\langle \mathbf{u}, \mathbf{x}\rangle = p_n(\mathbf{x})$ (or $\langle \mathbf{u}, \mathbf{x}\rangle = -p_n(\mathbf{x})$ respectively). In the special case, when the Gauss map is a diffeomorphism between M and $\mathbb{S}^{n-1}$, we get

$$\int_{\mathbb{S}^{n-1}} (p_n(\mathbf{e}_n^{-1}(\mathbf{u})) + p_n(\mathbf{e}_n^{-1}(-\mathbf{u})))d\mathbf{u}. = \int_{\mathbb{S}^{n-1}} w_K(\mathbf{u})d\mathbf{u} = n\omega_n\bar{w}(K).$$

Thus, we get the following corollary of Proposition 4.13.16, also known as Minkowski's Formula.

Corollary 4.13.20 (Minkowski's Formula on the Mean Width)**.** *Let* $K \subset \mathbb{R}^n$ *be a compact convex regular domain with boundary* $M = \partial K$*, such that the exterior unit normal vector field* $\mathbf{e}_n\colon M \to \mathbb{R}^n$ *along* M *is a diffeomorphism between* M *and* $\mathbb{S}^{n-1}$*. Then the coefficient of* r^{n-1} *in the polynomial expressing the volume of* $B(K, r)$ *is equal to the following quantities*

$$nW_{n-1}(K) = \omega_{n-1}V_1(K) = \frac{(-1)^{n-2}}{n-1}\int_M K_{n-2}d\sigma = \frac{n\omega_n}{2}\bar{w}(K).$$

4.13.8 Rigidity of Convex Surfaces

Cauchy's rigidity theorem for convex polytopes says that the shape of a 3-dimensional convex polytope is uniquely determined by the shapes of the facets and the combinatorial structure describing which are the common edges of the neighboring facets. More formally, if we have two convex 3-dimensional polytopes P_1 and P_2 and a bijection $\Phi\colon \partial P_1 \to \partial P_2$, which maps each facet of P_1 isometrically onto a facet of P_2, then Φ extends to an isometry of the whole space $\mathbb{R}^3$.

We want to prove a smooth analogue of Cauchy's rigidity theorem here. Instead of two convex polytopes, we shall consider two convex compact regular domains, K_1 and K_2 in $\mathbb{R}^3$, and require that the bijection $\Phi\colon \partial K_1 \to \partial K_2$ between their boundaries be a bending. This means that Φ should preserve the lengths of curves lying on the boundary of K_1.

Let M be a smooth hypersurface in $\mathbb{R}^n$, $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a local orthonormal frame on an open subset $U \subset M$, $h\colon M \to \tilde{M}$ be a bending of M. Then the images of $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ under the derivative of h yield an orthonormal tangential frame $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_{n-1}$ along $h(U)$, which can be extended uniquely to an orthonormal frame $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ having the same orientation as $\mathbf{e}_1, \dots, \mathbf{e}_n$. Using this frame on $\tilde{M}$, we can introduce the differential 1-forms $\tilde{\sigma}^i$, $\tilde{\omega}_i^j$ and the functions p_i, $\tilde{p}_i$ in the usual way. Instead of working with these differential forms and functions, we prefer to work with their pull-back forms $\bar{\sigma}^i = h^*\tilde{\sigma}^i$, $\bar{\omega}_i^j = h^*\tilde{\omega}_i^j$ and the functions $\bar{p}_i = \tilde{p}_i \circ h$. Since h is an isometry of Riemannian manifolds, $\sigma^i = \bar{\sigma}^i$. The Riemannian metric also determines the Levi-Cività connection, which is encoded by the one forms ω_i^j for $1 \leq i, j \leq n-1$, so we also have $\omega_i^j = \bar{\omega}_i^j$ for $1 \leq i, j \leq n-1$.

As the following proposition claims, the exterior shapes of U and $h(U)$ are determined by the forms ω_i^n and $\bar{\omega}_i^n$.

Proposition 4.13.21. *If U is connected, then $h|_U$ extends to an orientation-preserving isometry of the whole space if and only if $\omega_i^n = \bar{\omega}_i^n$ for $1 \leq i \leq n-1$.*

Proof. The easier part of the statement is that if the bending extends to an orientation preserving isometry of the space, then $\omega_i^n = {\omega'}_i^n$. We leave the details of this direction to the reader.

Assume now that $\omega_i^n = \bar{\omega}_i^n$ for all $1 \leq i \leq n-1$. Choose a point $p \in U$ and an orientation preserving isometry Φ of $\mathbb{R}^n$ such that $\Phi(p) = h(p)$ and $T_p\Phi(\mathbf{e}_i(p)) = \bar{\mathbf{e}}_i(h(p))$ for $1 \leq i \leq n-1$. Then $\hat{h} = \Phi^{-1} \circ h$ is a bending of M which fixes p and the frame $\mathbf{e}_1(p), \dots, \mathbf{e}_{n-1}(p)$. Denote by $\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n$ the frame induced on $\hat{h}(U)$ in the same way as the frame $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ was obtained on $h(U)$.

Let $\gamma\colon [0,1] \to U$ be a smooth curve starting from $\gamma(0) = p$. Consider the collection of the vector valued functions $(\mathbf{e}_1 \circ \gamma, \dots, \mathbf{e}_n \circ \gamma)\colon [0,1] \to \mathbb{R}^{n^2}$. It satisfies the system of differential equations

$$\frac{d}{dt}(\mathbf{e}_i \circ \gamma)(\tau) = \sum_{j=1}^{n} \omega_i^j(\gamma'(\tau))(\mathbf{e}_j \circ \gamma)(\tau), \qquad (1 \leq i \leq n).$$

The same system of differential equations is satisfied also by the vector valued function $(\hat{\mathbf{e}}_1 \circ \hat{h} \circ \gamma, \dots, \hat{\mathbf{e}}_n \circ \hat{h} \circ \gamma)\colon [0,1] \to \mathbb{R}^{n^2}$. Both solutions start from the same initial vectors at 0, therefore, by the uniqueness of the solutions of ordinary differential equations with given initial value, $\mathbf{e}_i \circ \gamma = \hat{\mathbf{e}}_i \circ \hat{h} \circ \gamma$ for $i = 1, \dots, n$. From this we obtain

$$\gamma'(\tau) = \sum_{i=1}^{n-1} \sigma^i(\gamma'(\tau))\mathbf{e}_i(\gamma(\tau)) = \sum_{i=1}^{n-1} \sigma^i(\gamma'(\tau))\hat{\mathbf{e}}_i(\hat{h}(\gamma(\tau))) = (\hat{h} \circ \gamma)'(\tau),$$

which implies by $\gamma(0) = \hat{h}(\gamma(0)) = p$ that $\gamma = \hat{h} \circ \gamma$. Since U is connected, $\gamma(1)$ can be any point in U, consequently $\hat{h}\Phi^{-1} \circ h = \mathrm{id}_M$, that is, Φ extends $h|_U$ to an isometry of the whole space. □

The main tool of proving the rigidity of convex surfaces in $\mathbb{R}^3$ is a generalization of Minkowski's formula due to Gustav Herglotz (1881-1953). This formula has the same form as Minkowski's formula, but some of the "ingredients" of the formula are taken not from the hypersurface M but from a bending $\tilde{M}$ of M. This way, Herglotz's formula relates the geometries of M and $\tilde{M}$ to each other, and in the special case, when $M = \tilde{M}$, it returns Minkowski's formula.

Proposition 4.13.22 (Herglotz's Formula)**.** *Let M be a compact hypersurface in $\mathbb{R}^n$, $h\colon M \to \mathbb{R}^n$ be a bending, $h(M) = \tilde{M}$. Starting with a local orthonormal frame $\mathbf{e}_1, \dots, \mathbf{e}_n$ over an open subset U of M define the forms ω_i^j, $\bar{\omega}_i^j$ and the functions K_{n-2}, p_i and $\bar{p}_i$ as above. Then the forms $\bar{p}_n\left(\sum_{i=1}^{n-1} \omega_1^n \wedge \dots \wedge \bar{\omega}_i^n \wedge \dots \wedge \omega_{n-1}^n\right)$ and $K_{n-2} \cdot \sigma^1 \wedge \dots \wedge \sigma^{n-1}$ do not depend on the choice of the frame, consequently they are properly defined over M, and*

$$-\int_M \bar{p}_n\left(\sum_{i=1}^{n-1} \omega_1^n \wedge \dots \wedge \bar{\omega}_i^n \wedge \dots \wedge \omega_{n-1}^n\right) = \int_M K_{n-2} \cdot \sigma^1 \wedge \dots \wedge \sigma^{n-1}.$$

Proof. The proof follows the same line of computation as the proof of Minkowski's formula, the only difference is that some terms are marked with a bar to indicate that they come from the bent surface $\tilde{M}$. Due to the similarity of the two proofs, we skip some details. Define the differential $(n-2)$-form

η on M by the equation

$$\eta = \sum_{i=1}^{n-1} (-1)^{i+1} \bar{p}_i \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n,$$

and compute its differential as

$$\begin{aligned} d\eta = &\sum_{i=1}^{n-1} (-1)^{i+1} d\bar{p}_i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\ &+ \sum_{i=1}^{n-1} \sum_{1 \le j < i} (-1)^{i+j} \bar{p}_i \omega_1^n \wedge \cdots \wedge d\omega_j^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\ &+ \sum_{i=1}^{n-1} \sum_{i < j \le n-1} (-1)^{i+j-1} \bar{p}_i \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge d\omega_j^n \wedge \cdots \wedge \omega_{n-1}^n, \end{aligned}$$

hence

$$\begin{aligned} d\eta = &\sum_{i=1}^{n-1} \sum_{j=1}^{n} (-1)^{i+1} \bar{p}_j \bar{\omega}_i^j \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\ + &\sum_{i=1}^{n-1} (-1)^{i+1} \sigma^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\ + &\sum_{i=1}^{n-1} \sum_{1 \le j < i} \sum_{k=1}^{n} (-1)^{i+j} \bar{p}_i \omega_1^n \wedge \cdots \wedge (\omega_j^k \wedge \omega_k^n) \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge \omega_{n-1}^n + \\ + &\sum_{i=1}^{n-1} \sum_{i < j \le n-1} \sum_{k=1}^{n} (-1)^{i+j-1} \bar{p}_i \omega_1^n \wedge \cdots \wedge \widehat{\omega_i^n} \wedge \cdots \wedge (\omega_j^k \wedge \omega_k^n) \wedge \cdots \wedge \omega_{n-1}^n \end{aligned}$$

The sum of the last two terms can be simplified to

$$\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} (-1)^j \bar{p}_i \omega_j^i \wedge \omega_1^n \wedge \cdots \wedge \widehat{\omega_j^n} \wedge \cdots \wedge \omega_{n-1}^n.$$

Substituting back we obtain

$$d\eta = \bar{p}_n \left(\sum_{i=1}^{n-1} \omega_1^n \wedge \cdots \wedge \bar{\omega}_i^n \wedge \cdots \wedge \omega_{n-1}^n \right) + K_{n-2} \cdot \sigma^1 \wedge \cdots \wedge \sigma^{n-1}. \quad (4.29)$$

The forms η, $\bar{p}_n \sum_{i=1}^{n-1} \omega_1^n \wedge \cdots \wedge \bar{\omega}_i^n \wedge \cdots \wedge \omega_{n-1}^n$ and $K_{n-2}\sigma^1 \wedge \cdots \wedge \sigma^{n-1}$ do not depend on the choice of the orthonormal frame $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$, hence they are

defined globally. For example, if $\mathbf{v}_1, \dots, \mathbf{v}_{n-2} \in T_pM$, then $\eta_p(\mathbf{v}_1, \dots, \mathbf{v}_{n-2})$ is the signed volume of the parallelepiped spanned by $(T_ph)^{-1}(\pi(h(p)))$ and $L_p(\mathbf{v}_1), \dots, L_p(\mathbf{v}_{n-2})$, where $\pi(h(p))$ is the orthogonal projection of $h(p)$ onto $T_{h(p)}\tilde{M}$, L_p is the Weingarten map at p. Naturality of the third form was obtained during the proof of the Minkowski formula, that of the second form follows form the relation (4.29) between the three forms.
If M is a compact hypersurface with no boundary, then integrating (4.29) we obtain Herglotz's formula. □

Theorem 4.13.23. *Let M and $\tilde{M}$ be the boundary surfaces of the compact convex regular domains C and $\tilde{C}$ in $\mathbb{R}^3$. Assume that M has positive Gaussian curvature $K_2 > 0$ and that there is a bending $h\colon M \to \tilde{M}$. Then h extends to an isometry Φ of the whole space. In particular, C and $\tilde{C}$ are congruent.*

Proof. Orient M and $\tilde{M}$ so that augmenting a positively oriented basis of one of the tangent spaces by the exterior unit normal of C or $\tilde{C}$ respectively give a positively oriented basis of the space $\mathbb{R}^3$. We may assume without loss of generality the $h\colon M \to \tilde{M}$ is orientation preserving, otherwise consider instead of h a composition of h with a reflection in a plane.
Let us use the same notations as in Minkowski's and Herglotz's formulae.
Translating C and $\tilde{C}$ we may assume that the origin is in the interior of C and $\tilde{C}$. Then $p_3 > 0$ and $\bar{p}_3 > 0$.
The integral of the Minkowski curvature $H = K_1/2$ of M over M can be expressed both by Minkowski's and Herglotz's formulae as

$$-\int_M H d\sigma = \int_M p_3 \omega_1^3 \wedge \omega_2^3 = \frac{1}{2}\int_M \bar{p}_3(\bar{\omega}_1^3 \wedge \omega_2^3 + \omega_1^3 \wedge \bar{\omega}_2^3).$$

Changing the role of M and $\tilde{M}$ yields

$$-\int_{\tilde{M}} \tilde{H} d\tilde{\sigma} = \int_M \bar{p}_3 \bar{\omega}_1^3 \wedge \bar{\omega}_2^3 = \frac{1}{2}\int_M p_3(\bar{\omega}_1^3 \wedge \omega_2^3 + \omega_1^3 \wedge \bar{\omega}_2^3).$$

Taking into account that M and $\tilde{M}$ has the same Gaussian curvature at corresponding points by Theorema Egregium, we get

$$\omega_1^3 \wedge \omega_2^3 = K_2 \sigma^1 \wedge \sigma^2 = \bar{\omega}_1^3 \wedge \bar{\omega}_2^3,$$

and

$$\int_{\tilde{M}} \tilde{H} d\tilde{\sigma} - \int_M H d\sigma = \frac{1}{2}\int_M p_3(\omega_1^3 - \bar{\omega}_1^3) \wedge (\omega_2^3 - \bar{\omega}_2^3).$$

The differential form $(\omega_1^3 - \bar{\omega}_1^3) \wedge (\omega_2^3 - \bar{\omega}_2^3)$ can be written as a multiple

$$f\sigma^1 \wedge \sigma^2 = (\omega_1^3 - \bar{\omega}_1^3) \wedge (\omega_2^3 - \bar{\omega}_2^3)$$

of the volume form $\sigma^1 \wedge \sigma^2$. We show that the function f is non-positive.
Fix a point $q \in M$ and choose a positively oriented orthonormal basis $\mathbf{v}_1, \mathbf{v}_2$ in T_qM then

$$\mathcal{B} = \begin{pmatrix} \omega_1^3(\mathbf{v}_1) & \omega_1^3(\mathbf{v}_2) \\ \omega_2^3(\mathbf{v}_1) & \omega_2^3(\mathbf{v}_2) \end{pmatrix}$$

is the matrix of the second fundamental form and the Weingarten map of M at q with respect to the basis $(\mathbf{v}_1, \mathbf{v}_2)$. For this reason, $\mathcal{B}$ is negative definite, and has determinant $K_2(q)$. Similarly,

$$\bar{\mathcal{B}} = \begin{pmatrix} \bar{\omega}_1^3(\mathbf{v}_1) & \bar{\omega}_1^3(\mathbf{v}_2) \\ \bar{\omega}_2^3(\mathbf{v}_1) & \bar{\omega}_2^3(\mathbf{v}_2) \end{pmatrix}$$

is the matrix of the second fundamental form and the Weingarten map of $\tilde{M}$ at $h(q)$ with respect to the basis $(T_qh(\mathbf{v}_1), T_qh(\mathbf{v}_2))$, and it is also negative definite, and has determinant $K_2(q)$.
Consider the quadratic polynomial

$$D(c) = \det(\mathcal{B} - c\bar{\mathcal{B}}) = K_2(q)(c^2 - \operatorname{tr}(\mathcal{B}\bar{\mathcal{B}}^{-1})c + 1).$$

Since $\bar{\mathcal{B}}$ is negative definite, $\mathcal{B} - c\bar{\mathcal{B}}$ is positive definite for large c, consequently, D must have a positive root $c_0 > 0$. Since the product of the two roots is 1, the second root is $1/c_0$. Then $D(c) = K_2(q)(c - c_0)(c - 1/c_0)$ and

$$f(q) = D(1) = -K_2(q)\frac{(1 - c_0)^2}{c_0} \leq 0.$$

Equality $f(q) = 0$ holds if and only if $\operatorname{tr}(\mathcal{B}\bar{\mathcal{B}}^{-1}) = 2$. But we also know that $\det(\mathcal{B}\bar{\mathcal{B}}^{-1}) = 1$, hence $f(q) = 0$ if and only if both eigenvalues of $\mathcal{B}\bar{\mathcal{B}}^{-1}$ are equal to 1. Since both $\mathcal{B}$ and $\bar{\mathcal{B}}$ are symmetric, and $\bar{\mathcal{B}}$ is negative definite, $\mathcal{B}\bar{\mathcal{B}}^{-1}$ is diagonalizable because of the principal axis theorem. This way, the eigenvalues of $\mathcal{B}\bar{\mathcal{B}}^{-1}$ are equal to 1 if and only if $\mathcal{B}\bar{\mathcal{B}}^{-1} = I$ and $\mathcal{B} = \bar{\mathcal{B}}$.
As a corollary, we obtain

$$\int_{\tilde{M}} \tilde{H} d\tilde{\sigma} - \int_M H d\sigma = \frac{1}{2}\int_M p_3(\omega_1^3 - \bar{\omega}_1^3) \wedge (\omega_2^3 - \bar{\omega}_2^3) = \frac{1}{2}\int_M p_3 f \sigma^1 \wedge \sigma^2 \leq 0,$$

which implies

$$\int_{\tilde{M}} \tilde{H} d\tilde{\sigma} \leq \int_M H d\sigma.$$

However, the role of M and $\tilde{M}$ is symmetric, so the reversed inequality

$$\int_{\tilde{M}} \tilde{H} d\tilde{\sigma} \geq \int_M H d\sigma$$

must also be true, therefore we must have equality in both inequalities. Equality implies $f \equiv 0$, which can hold only if the matrices $\mathcal{B}$ and $\bar{\mathcal{B}}$ are equal for each $q \in M$ and each choice of $\mathbf{v}_1 \mathbf{v}_2 \in T_q M$, that is, if $\omega_1^3 = \bar{\omega}_1^3$ and $\omega_2^3 = \bar{\omega}_2^3$. However this condition is equivalent to the extendability of h to an orientation preserving isometry of the space $\mathbb{R}^3$. $\square$

Bibliography

[1] W. Blaschke, H. Reichardt, *Einführung in die Differentalgeometrie.* Zweite aufl., Springer-Verlag Berlin-Gottingen-Heidelberg, 1960.

[2] W.M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press, Inc. Orlando 1986.

[3] M.P. do Carmo, *Differential Geometry of Curves and Surfaces.* Prentice-Hall, Inc., Englewood Cliffs New Jersey 1976.

[4] M.P. do Carmo, *Riemannian Geometry.* Birkhäuser Boston 1992.

[5] D.L. Cohn, *Measure Theory.* Birkhäuser, Boston-Basel-Stuttgart, 1980.

[6] A. Hatcher, *Algebraic Topology.* Cambridge University Press 2002.

[7] S.Mac Lane, *Categories for the Working Mathematician.* Graduate Texts in Mathematics, Springer Verlag New York Heidelberg Berlin 1998.

[8] J.M. Lee, *Riemannian Manifolds: An Introduction to Curvature.* Graduate Texts in Mathematics, Springer Verlag New York 1997.

[9] J.R. Munkres, *Topology.* 2nd ed. Prentice Hall, Inc. Upper Saddle River, 2000.

[10] J.J. Stoker, *Differential Geometry.* Wiley Classics Library, John Wiley & Sons, Inc. 1989.

[11] F.W. Warner, *Foundations of Differentiable Manifolds and Lie Groups.* Graduate Texts in Mathematics, Springer-Verlag Berlin-Heidelberg-New York 1983.

Index