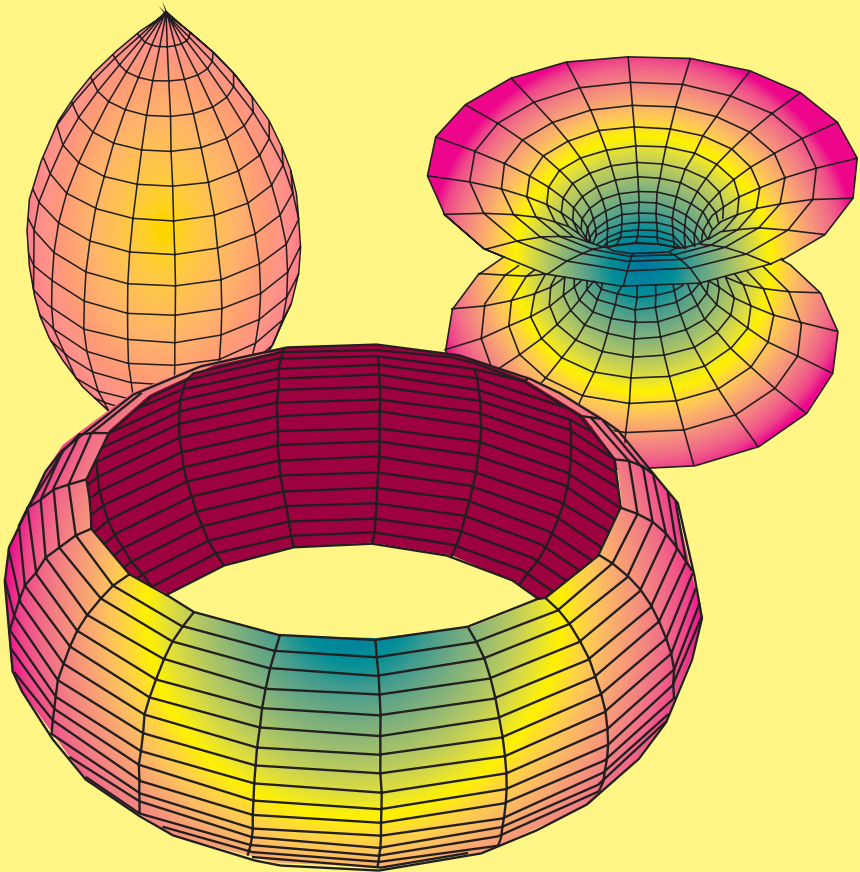


Paulo Ventura Araújo

Differential Geometry



 Springer

Differential Geometry

Paulo Ventura Araújo

Differential Geometry

 Springer

Paulo Ventura Araújo
Department of Mathematics
Faculdade de Ciências da Universidade do Porto
Porto, Portugal

Translation from the Portuguese language edition: “Geometria Diferencial” by Paulo Ventura Araújo, © Paulo Ventura Araújo 1998. Published by IMPA. All Rights Reserved.

ISBN 978-3-031-62383-7 ISBN 978-3-031-62384-4 (eBook)
<https://doi.org/10.1007/978-3-031-62384-4>

Mathematics Subject Classification (2020): 53-XX, 51-XX

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cover illustration: Image created from two figured to place above the title.

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

If disposing of this product, please recycle the paper.

Preface

This book is based on the lecture notes of the course Differential Geometry taught at the Faculty of Sciences of the University of Porto in the academic years 1992–93 and 1993–94. Students from different courses and with different mathematical backgrounds attended the course and, consequently, its prerequisites were reduced to Linear Algebra, Calculus (of one and several variables), and the study of curves up to the Frenet trihedron. Furthermore, we avoided the introduction of certain technical apparatus, such as Tensor Calculus, in order to insist instead on results with accessible geometric content whose proofs, although possibly long, used more elementary means.

That said, one understands why we have restricted ourselves to the study of curves and surfaces in Euclidean space. But, in our opinion, even for students pursuing a scientific career, this is the right approach for a first study of Differential Geometry, grounding intuition and motivating the problems that arise in higher dimensions.

Although the idea was to reproduce, in order and content, the course notes, the notes grew and included subjects not discussed in the lectures. There is a risk, when teaching Differential Geometry at an introductory level, that the harvest of interesting results will not compensate for the work spent in digesting definitions and assimilating techniques. The digressions in this text may lead the student to discover some of the richness of Differential Geometry which, by the imperative of bureaucratic “realism”, is so often absent from the classroom.

The exercises included in the text are rarely routine, although few are really difficult, and were chosen on the assumption that a good exercise, with a medium level of difficulty, should reward the students’ effort by teaching them something.

Among the books we consulted, Manfredo do Carmo’s [6] deserves to be highlighted: some of the exercises and the structuring of some subjects come from there. But several exercises are original, and the selection of themes and the composition of the proofs reflect personal taste and work.

We now give some hints on how to use the book: Sections 1.1 – 1.3 cover subjects probably already known to the student, and may be omitted in well-prepared classes. Chapters 2–4 cover a basic course in Differential Geometry. Sections 3.3, 4.4 may be omitted if time is tight. If time permits, a choice of topics from Chapter 1 (sections

1.4 to 1.8) and Chapter 5 can be made; the interdependence between the various sections of these chapters is indicated at the beginning of each chapter.

Porto, July 1996

Paulo Ventura Araújo.

Contents

1	Differentiable Curves	1
1.1	Velocity and Arc Length	1
1.2	Acceleration, Curvature and the Frenet Trihedron	5
1.3	Planar Curves	9
1.4	Contact of Curves	12
1.5	Convex Curves	15
1.6	Curves of Constant Width	19
1.7	Theorem of the Four Vertices	25
1.8	The Isoperimetric Inequality	26
2	Regular Surfaces	31
2.1	Definition and Examples	31
2.2	Change of Parameters, Level Surfaces	37
2.3	Differentiable Functions on Surfaces, Tangent Space	42
2.4	Orientability	46
2.5	Areas, Lengths, and Angles: The First Fundamental Form	51
3	The Geometry of the Gauss Map	59
3.1	The Gauss Map and its Derivative	59
3.2	The Second Fundamental Form	67
3.3	Vector Fields	76
4	The Intrinsic Geometry of Surfaces	83
4.1	Conformal Mappings and Isometries	83
4.2	Gauss's Theorema Egregium	90
4.3	Covariant Derivative, Parallel Transport, Geodesic Curvature	92
4.4	The Divergence Theorem. First Variation of Area	101
4.5	The Gauss-Bonnet Theorem	110
4.6	Minimizing Properties of Geodesics	121
	Appendix: Rotation Index	133

5	The Global Geometry of Surfaces	139
5.1	Complete Surfaces	139
5.2	Coverings	144
5.3	Complete Surfaces of Non-Positive Curvature	149
5.4	Ovals (First Part): The Rigidity of the Sphere	152
5.5	Ovals: Areas and Volumes; Surfaces of Constant Width	157
5.6	Abstract Surfaces. The Hyperbolic Plane	163
5.7	Complete Surfaces of Constant Curvature	172
References		181
Index		183



Chapter 1

Differentiable Curves

The first three sections of this chapter contain the basics on curves, and, because of their brevity, are rather a review of concepts and results, gathered in a form we will use later; in the last five sections we will deal with subjects whose inclusion in the course is optional. Section 1.5 should be read before 1.6 and 1.7, but otherwise, and except for a few exercises, sections 1.4 – 1.8 are independent of each other.

1.1 Velocity and Arc Length

In the space $\mathbb{R}^n$ we will denote vectors by the symbols $\mathbf{v}$, $\mathbf{w}$ and points by lowercase consonants p , q . This space is equipped with a canonical Euclidean structure in which the *inner product* of two vectors is the sum of the products of their components of equal index, i.e. $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + \dots + v_n w_n$. The *norm* or *length* of a vector is given by $|\mathbf{v}| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$, and the *angle* between the nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ is the only number $\theta \in [0, \pi]$ such that

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

The *distance* between the points p and q is defined as the length of the vector $p - q$.

A *parametrized curve* is a continuous function $\alpha: I \rightarrow \mathbb{R}^n$ on an interval of $\mathbb{R}$ in Euclidean n -dimensional space, and its *trace* is the image of that function. Writing $\alpha(t) = (x_1(t), \dots, x_n(t))$, the functions x_i are the *component functions* of α . We say that α is of *class* C^∞ if each of its component functions has continuous derivatives of all orders (if α is defined, for example, on a closed interval $[a, b]$, then we require the existence of all right-hand side derivatives at a and left-hand side derivatives at b).

The *velocity vector* of the curve is $\alpha'(t) = (x'_1(t), \dots, x'_n(t))$ and, when nonzero, points in the direction tangent to the curve at time t . *Regular curves* are those whose velocity vector never vanishes and therefore have a well-defined tangent direction at each instant.

From now on, unless otherwise stated, by *curve* we mean a regular parametrized curve of class C^∞ .

The simplest example of a curve is a straight line $p + t\mathbf{v}$, $t \in \mathbb{R}$, parametrized with nonzero constant velocity $\mathbf{v}$. Other examples are the circle $(\cos t, \sin t)$ and, shown in Fig. 1.1, the cloverleaf $(\cos 3t \cos t, \cos 3t \sin t)$ and the helix $(\cos t, \sin t, t)$.

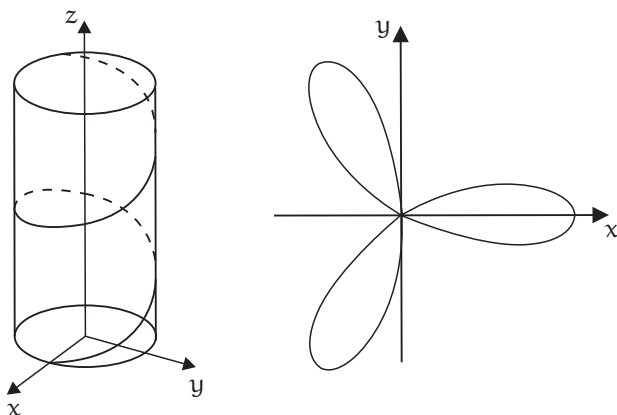


Figure 1.1

One of the first questions to ask about a curve is how to compute its length. Assume that the curve α is defined on a closed interval $[a, b]$. Take an arbitrary partition $a = t_0 < t_1 < \dots < t_m = b$ of that interval. The sum

$$\sum_{i=1}^m |\alpha(t_i) - \alpha(t_{i-1})|$$

gives the length of the polygonal line obtained by replacing, for each $1 \leq i \leq m$, the trace of the curve in the interval $[t_{i-1}, t_i]$ by the line segment joining $\alpha(t_{i-1})$ with $\alpha(t_i)$ (see Fig. 1.2). The narrower and more numerous are the intervals of the partition, the better the sum should approximate the length of the curve. In Exercise 3 of this section we show that the limit of these sums, as the maximum of the differences $t_i - t_{i-1}$ tends to zero, is given by the integral $\int_a^b |\alpha'(t)| dt$, and this is how the *length* $l(\alpha)$ of the curve α is defined.

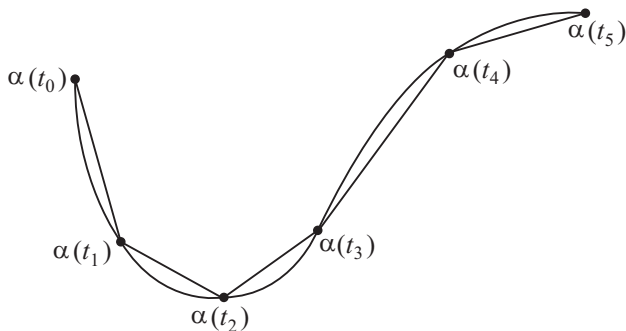


Figure 1.2

Our first result reassures us as to the correctness of the definition just given, by showing that the line is the shortest path between two points:

Proposition 1.1.1 *Let $\alpha: [a, b] \rightarrow \mathbb{R}^n$ be a curve. Then $l(\alpha) \geq |\alpha(b) - \alpha(a)|$, and equality holds if and only if the trace of α is a line segment.*

Proof We assume that $\alpha(b) \neq \alpha(a)$, otherwise there is nothing to be shown. We write $\alpha(b) - \alpha(a) = \int_a^b \alpha'(t) dt$, an equality in which the right-hand side is the vector whose coordinates are the integrals of the components of $\alpha'(t)$. Forming the inner product of both sides with the vector $\alpha(b) - \alpha(a)$, we obtain

$$\begin{aligned} |\alpha(b) - \alpha(a)|^2 &= \int_a^b \langle \alpha'(t), \alpha(b) - \alpha(a) \rangle dt \\ &\leq \int_a^b |\alpha'(t)| |\alpha(b) - \alpha(a)| dt = l(\alpha) |\alpha(b) - \alpha(a)|, \end{aligned}$$

from which, simplifying, we obtain the desired inequality. For equality to hold, the Cauchy-Schwarz inequality

$$\langle \alpha'(t), \alpha(b) - \alpha(a) \rangle \leq |\alpha'(t)| |\alpha(b) - \alpha(a)|$$

will have to reduce to an equality for all $t \in [a, b]$, which happens if and only if $\alpha'(t)$ is a positive multiple of $\alpha(b) - \alpha(a)$ — i.e., if and only if the trace of α is the line segment joining $\alpha(a)$ to $\alpha(b)$. $\square$

The value $v(t) = |\alpha'(t)|$ is the *scalar velocity* of the curve α at time t , and

$$S(t) = \int_a^t v(r) dr$$

is the *arc length* function. When the scalar velocity is constant, the arc length is proportional to time. We recall that the curve $\tilde{\alpha}$ is a reparametrization of α if $\tilde{\alpha} = \alpha \circ h$, for some *increasing* diffeomorphism $h: J \rightarrow I$ between intervals of $\mathbb{R}$. (If h were

decreasing, then we would obtain a curve with the same trace as α but with reversed orientation — that is, traversed in the opposite direction.) Let us now show, by defining a suitable function h , that it is *possible to reparameterize α so that $\alpha \circ h$ has constant scalar velocity, equal to 1*.

Since $S'(t) = v(t) > 0$, the function S is increasing and sends $[a, b]$ diffeomorphically onto the interval $[0, l(\alpha)]$. Defining h as the inverse of S , and putting $\tilde{\alpha}(s) = \alpha \circ h(s)$, we have

$$\tilde{\alpha}'(s) = h'(s)\alpha'(h(s)) = \frac{1}{v(t)}\alpha'(t),$$

where $t = h(s)$, and from this we see that $|\tilde{\alpha}'(s)| = 1$ for all s in $[0, l(\alpha)]$.

We say that a curve is *parametrized by arc length* when it is traversed with constant scalar velocity equal to 1, regardless of whether the time at which the parameterization starts is zero.

Exercises

1. Let $\alpha: [a, b] \rightarrow \mathbb{R}^n$ and $\beta: [c, d] \rightarrow \mathbb{R}^n$ be two regular, injective curves with the same trace. Show that the function $\beta^{-1} \circ \alpha: [a, b] \rightarrow [c, d]$ is differentiable and its derivative never vanishes.

2. Let $\alpha(t) = (e^{bt} \cos t, e^{bt} \sin t)$, where b is a negative constant and $t \in \mathbb{R}$.

(a) Sketch the trace of α .

(b) Check that α has finite length on $[t_0, +\infty[$ and compute it.

3. Let $\alpha(t) = (x_1(t), \dots, x_n(t))$, $t \in [a, b]$ be a curve of class C^1 . Show that

(a) given $\varepsilon > 0$, there exists $\delta > 0$ such that for all $1 \leq k \leq n$,

$$|t - s| < \delta \Rightarrow \left| \frac{x_k(t) - x_k(s)}{t - s} - x'_k(t) \right| < \varepsilon;$$

(b) for the ε and δ just obtained, if $a = t_0 < t_1 < \dots < t_m = b$ is a partition of $[a, b]$ such that $t_i - t_{i-1} < \delta$ for all $1 \leq i \leq m$, then

$$\left| \sum_{i=1}^m |\alpha(t_i) - \alpha(t_{i-1})| - \sum_{i=1}^m |\alpha'(t_i)|(t_i - t_{i-1}) \right| < \sqrt{n}(b - a)\varepsilon;$$

(c) the limit of sums

$$\sum_{i=1}^m |\alpha(t_i) - \alpha(t_{i-1})|, \quad \text{as } \max_{1 \leq i \leq m} (t_i - t_{i-1}) \rightarrow 0,$$

$$\text{is } \int_a^b |\alpha'(t)| dt.$$

4. A curve $\alpha: [a, b] \rightarrow \mathbb{R}^n$ is called *rectifiable* if the supremum of the sums

$$\sum_{i=1}^m |\alpha(t_i) - \alpha(t_{i-1})|,$$

where $t_0 < t_1 < \dots < t_m$ is some partition of $[a, b]$, is finite. We call this supremum the length of α .

(a) Show that, for curves of class C^1 , this new definition of length is equivalent to the previous one.

(b) Does a rectifiable curve necessarily have derivatives at all points?

(c) Is it true that the graph of any continuous and monotone function $[a, b] \rightarrow \mathbb{R}$ is a rectifiable curve?

(d) Consider the Weierstrass example of a continuous function $[0, 1] \rightarrow \mathbb{R}$ that is nowhere differentiable (see [24], chapter 23, Theorem 5), and find out whether the graph of this function is rectifiable.

1.2 Acceleration, Curvature and the Frenet Trihedron

In this section we only consider curves in three dimensional Euclidean space. In the previous section we learned how to compute the length of a curve with the help of its velocity or first derivative. This calculation does not exhaust the analysis of a curve, since it tells us nothing about the shape it can take, and does not distinguish a line from a circle. To proceed we will also have to take into account the second derivative.

Given a curve $\alpha: I \rightarrow \mathbb{R}^3$, the *tangent unit vector* to α is $\tau(t) = \frac{1}{v(t)}\alpha'(t)$. If $\tilde{\alpha}(r) = \alpha \circ h(r)$ is a reparametrization of α then the unit vector tangent to $\tilde{\alpha}$ is given by $\tilde{\tau} = \tau \circ h$. Hence, defining $\tilde{v}(r) = |\tilde{\alpha}'(r)|$, we have

$$\begin{aligned}\tilde{v}(r) &= (v \circ h(r))h'(r), \\ |\tilde{\tau}'(r)| &= |\tau'(h(r))|h'(r).\end{aligned}$$

If we put

$$k(t) = \frac{1}{v(t)}|\tau'(t)|$$

and denote by $\tilde{k}$ the analogous function for $\tilde{\alpha}$, then we see that $\tilde{k} = k \circ h$. The quantity $k(t)$ is the *curvature* of α at the point $\alpha(t)$. The preceding calculations show that the curvature does not depend on the way the curve is traversed but only on the point at which it is computed, and does not change even when we reverse its orientation. We can thus assume, whenever convenient, that the parameter of the curve is the arc length. In this case we simply have $k(s) = |\alpha''(s)|$.

A line, for example, has zero constant curvature: in fact, when parametrized with constant scalar velocity, its second derivative vanishes. But the converse is also true, since the condition $k(s) \equiv 0$ implies that $\alpha''(s) \equiv 0$, which shows that there exist constants p and $\mathbf{v}$ such that $\alpha(s) \equiv p + s\mathbf{v}$.

Since $\tau(t)$ has constant norm, it is orthogonal to its derivative. In fact, by differentiating the equality $\langle \tau(t), \tau(t) \rangle = 1$, we obtain $2\langle \tau'(t), \tau(t) \rangle = 0$. Thus, when

$k(t) \neq 0$, the vector $\tau'(t)$ points in a direction normal to the curve, the so-called *principal normal*. In these cases we can define the unit vector

$$\mathbf{n}(t) = \frac{1}{|\tau'(t)|} \tau'(t)$$

and the *center of curvature*, which is the point

$$\alpha(t) + \frac{1}{k(t)} \mathbf{n}(t).$$

The value $1/k(t)$ is the *curvature radius* at the point $\alpha(t)$. The *osculating plane* is the plane parallel to $\tau(t)$ and $\mathbf{n}(t)$ that passes through $\alpha(t)$.

The curvature measures the variation of the direction of the curve, but it does not determine the form of the curve: both the circumference and the helix, for instance, have constant curvature, that of the former being equal to the inverse of the radius; and while the circumference is a planar curve, in the helix the osculating plane varies from point to point. What we lack then is to measure the variation of the osculating plane — or, to put it differently, how far a curve deviates from being planar.

Continuing to assume that $k(t) \neq 0$, the bi-normal vector $\mathbf{b}(t)$ is defined as the only vector such that $(\tau(t), \mathbf{n}(t), \mathbf{b}(t))$ is a direct orthonormal trihedron — that is, an ordered triplet of unit vectors, orthogonal to each other, such that the 3×3 -matrix whose columns are these vectors in the same order has positive determinant. Even simpler, we have $\mathbf{b}(t) = \tau(t) \times \mathbf{n}(t)$, where $\times$ denotes the vector product on $\mathbb{R}^3$.

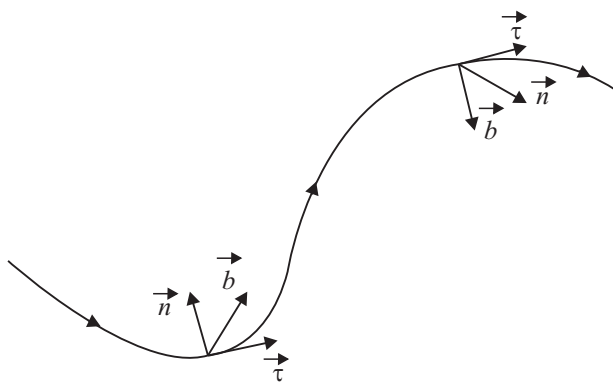


Figure 1.3

The trihedron $(\tau(t), \mathbf{n}(t), \mathbf{b}(t))$ is the *Frenet trihedron*. Each of the vectors of the trihedron is orthogonal to its derivative, so that each derivative is expressible as a linear combination of the other two vectors. We will next study the coefficients of this linear combination. We have

$$\begin{aligned}
\mathbf{b}'(t) &= \frac{d}{dt}(\boldsymbol{\tau}(t) \times \mathbf{n}(t)) \\
&= \boldsymbol{\tau}'(t) \times \mathbf{n}(t) + \boldsymbol{\tau}(t) \times \mathbf{n}'(t) \\
&= \boldsymbol{\tau}(t) \times \mathbf{n}'(t)
\end{aligned}$$

(because $\boldsymbol{\tau}'$ and $\mathbf{n}$ are collinear for each t). From this equality it follows that $\mathbf{b}'$, which we already knew to be orthogonal to $\mathbf{b}$, is also orthogonal to $\boldsymbol{\tau}$, and is therefore a multiple of $\mathbf{n}$. The *torsion* of α at the point $\alpha(t)$ is the value $\nu(t)$ defined by the equality

$$\frac{1}{\nu(t)} \mathbf{b}'(t) = \nu(t) \mathbf{n}(t);$$

the factor $1/\nu(t)$ again ensures that the torsion is independent of the parameterization. Note that torsion can assume negative values — although, like curvature, it does not depend on the orientation of the curve.

Differentiating the product $\mathbf{n}(t) = \mathbf{b}(t) \times \boldsymbol{\tau}(t)$, and using the formulas seen above for $\mathbf{b}'(t)$ and $\boldsymbol{\tau}'(t)$, we find that

$$\frac{1}{\nu(t)} \mathbf{n}'(t) = -k(t) \boldsymbol{\tau}(t) - \nu(t) \mathbf{b}(t).$$

We thus establish the equalities

$$\begin{aligned}
\frac{1}{\nu} \boldsymbol{\tau}' &= k \mathbf{n} \\
\frac{1}{\nu} \mathbf{n}' &= -k \boldsymbol{\tau} - \nu \mathbf{b} \\
\frac{1}{\nu} \mathbf{b}' &= \nu \mathbf{n},
\end{aligned}$$

which are known as *Frenet formulas*. When the curve is parametrized by arc length the factor $1/\nu$ (which is then equal to 1) is suppressed.

The next proposition gives us the geometric meaning of torsion being identically zero.

Proposition 1.2.1 *Assume that the curvature of α never vanishes. Then α is a planar curve if and only if it has zero constant torsion.*

Proof If α is all contained in the plane S , the vectors $\boldsymbol{\tau}(t)$ and $\mathbf{n}(t)$ are parallel to S and linearly independent, and the product $\mathbf{b}(t) = \boldsymbol{\tau}(t) \times \mathbf{n}(t)$ is a unit vector orthogonal to S ; therefore $\mathbf{b}(t)$ can only take two distinct values and, since it varies continuously, it must be constant — from which it follows that $\nu \equiv 0$.

Conversely, if $\nu \equiv 0$ then $\mathbf{b}(t)$ is a nonzero constant vector $\mathbf{b}$ and for all t holds $\langle \mathbf{b}, \alpha'(t) \rangle = 0$. Therefore there exists a constant c such that $\langle \mathbf{b}, \alpha(t) \rangle \equiv c$, and this equality shows that the curve α is planar. $\square$

Let us now consider a curve $\alpha(s)$, where the parameter $s \in [a, b]$ is the arc length, and whose curvature $k(s)$ never vanishes. We ask ourselves to what extent the

functions $k(s)$ and $\nu(s)$ determine the curve α . They do not do so completely, since any rigid movement of $\mathbb{R}^3$ transforms α into another curve with the same curvature and torsion functions. (By *rigid movement* we mean the composition of a translation with a linear mapping that transforms the canonical basis of $\mathbb{R}^3$ into some direct orthonormal trihedron.) But this is the only freedom:

Theorem 1.2.2 *Let $\alpha, \tilde{\alpha}: [a, b] \rightarrow \mathbb{R}^3$ be curves parametrized by arc length and with nonzero curvature at all points. If we have $k(s) = \tilde{k}(s)$ and $\nu(s) = \tilde{\nu}(s)$ for all s on $[a, b]$, then there exists a rigid movement $L: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ such that $\alpha \equiv L \circ \tilde{\alpha}$.*

Proof Let L_1 be the linear mapping that transforms the trihedron $(\tilde{\tau}(a), \tilde{\mathbf{n}}(a), \tilde{\mathbf{b}}(a))$ into the trihedron $(\tau(a), \mathbf{n}(a), \mathbf{b}(a))$, L_2 the translation that maps $L_1(\tilde{\alpha}(a))$ into $\alpha(a)$, and $L = L_2 \circ L_1$. Then the curve $\alpha_0 = L \circ \tilde{\alpha}$ has at the start time $s = a$ the same Frenet trihedron as α and satisfies $\alpha_0(a) = \alpha(a)$. Identifying by the subindex 0 the vectors and quantities concerning the curve α_0 , we define the function

$$\delta(s) = \frac{1}{2} (|\tau(s) - \tau_0(s)|^2 + |\mathbf{n}(s) - \mathbf{n}_0(s)|^2 + |\mathbf{b}(s) - \mathbf{b}_0(s)|^2).$$

We know that $\delta(a) = 0$; furthermore, we have

$$\delta' = \langle \tau' - \tau'_0, \tau - \tau_0 \rangle + \langle \mathbf{n}' - \mathbf{n}'_0, \mathbf{n} - \mathbf{n}_0 \rangle + \langle \mathbf{b}' - \mathbf{b}'_0, \mathbf{b} - \mathbf{b}_0 \rangle.$$

From this, using Frenet's formulas and the fact that $k_0 = k$ and $\nu_0 = \nu$, we easily obtain $\delta' \equiv 0$, therefore also $\delta \equiv 0$. In particular we have $\alpha'_0 = \alpha'$ — and, since $\alpha_0(a) = \alpha(a)$, we conclude that $\alpha_0 = \alpha$. $\square$

Note 1.2.3 To complement Theorem 1.2.2, we will now show the following result: *for any differentiable functions $k, \nu: [a, b] \rightarrow \mathbb{R}$ (with k strictly positive), there exists a curve $\alpha(s)$, parametrized by arc length, whose curvature and torsion functions are precisely $k(s)$ and $\nu(s)$.* We make use of the Theorem of Existence and uniqueness of solutions of differential equations that we will state in Section 3.3; it deserves mention that this approach provides another proof of Theorem 1.2.2. We keep the above proof though as it is more elementary.

Once functions $k(s)$ and $\nu(s)$ are fixed, Frénet's formulas can be viewed as a non-autonomous equation of the form $X = \mathbf{v}(s, X)$, where $X = (\tau, \mathbf{n}, \mathbf{b}) \in \mathbb{R}^9$ and where $\mathbf{v}: [a, b] \times \mathbb{R}^9 \rightarrow \mathbb{R}^9$ is differentiable. Take any direct orthonormal trihedron $(\tau(a), \mathbf{n}(a), \mathbf{b}(a))$. Then there exists $\varepsilon > 0$ such that the solution $X(s) = (\tau(s), \mathbf{n}(s), \mathbf{b}(s))$ with this initial condition is defined for $[a, a + \varepsilon]$. But since

$$\frac{d}{dt} \{ |\tau|^2 + |\mathbf{n}|^2 + |\mathbf{b}|^2 \} = 2 (\langle \tau, k\mathbf{n} \rangle + \langle \mathbf{n}, -k\tau - \nu\mathbf{b} \rangle + \langle \mathbf{b}, \nu\mathbf{n} \rangle) = 0,$$

we see that $X(s)$ stays in the compact set $\{X \in \mathbb{R}^9: |X| = \sqrt{3}\}$, and is therefore defined on the entire interval $[a, b]$ (see Theorem 3 on p. 17 of [23]). By differentiating, we obtain the equalities

$$\begin{aligned}
\langle \boldsymbol{\tau}, \mathbf{n} \rangle' &= k|\mathbf{n}|^2 - k|\boldsymbol{\tau}|^2 - \nu \langle \boldsymbol{\tau}, \mathbf{b} \rangle \\
\langle \mathbf{n}, \mathbf{b} \rangle' &= \nu|\mathbf{n}|^2 - \nu|\mathbf{b}|^2 - k \langle \boldsymbol{\tau}, \mathbf{b} \rangle \\
\langle \boldsymbol{\tau}, \mathbf{b} \rangle' &= k \langle \mathbf{n}, \mathbf{b} \rangle - \nu \langle \boldsymbol{\tau}, \mathbf{n} \rangle \\
(|\boldsymbol{\tau}|^2)' &= 2k \langle \boldsymbol{\tau}, \mathbf{n} \rangle \\
(|\mathbf{n}|^2)' &= -2k \langle \boldsymbol{\tau}, \mathbf{n} \rangle - 2\nu \langle \mathbf{n}, \mathbf{b} \rangle \\
(|\mathbf{b}|^2)' &= 2\nu \langle \mathbf{n}, \mathbf{b} \rangle
\end{aligned}$$

— from which it follows that $\langle \boldsymbol{\tau}, \mathbf{n} \rangle$, $\langle \mathbf{n}, \mathbf{b} \rangle$, $\langle \boldsymbol{\tau}, \mathbf{b} \rangle$, $|\boldsymbol{\tau}|^2$, $|\mathbf{n}|^2$ and $|\mathbf{b}|^2$ are constant functions, equal to 0, 0, 0, 1, 1 and 1 respectively, since, as it is easily seen, these constants constitute a solution, with the same initial condition, of the same differential equation defined in $\mathbb{R}^6$. This proves that $(\boldsymbol{\tau}(s), \mathbf{n}(s), \mathbf{b}(s))$ is an orthonormal, necessarily direct trihedron for all $s \in [a, b]$. To conclude we take for $\alpha(s)$ any primitive of $\boldsymbol{\tau}(s)$, e.g. $\alpha(s) = \int_a^s \boldsymbol{\tau}(t) dt$: we verify without difficulty that $(\boldsymbol{\tau}(s), \mathbf{n}(s), \mathbf{b}(s))$ is the Frenet trihedron of $\alpha(s)$, and that $k(s)$ and $\nu(s)$ are its curvature and torsion. $\square$

Exercise

5. Show that if we permit curves whose curvature vanishes at some point then the conclusion of Theorem 1.2.2 holds — that is, there exists a pair of curves $\alpha, \tilde{\alpha}: [a, b] \rightarrow \mathbb{R}^3$ parametrized by arc length such that their curvature and torsion functions coincide whenever they are defined, but which cannot be transformed into each other by a rigid movement.

Hint: look for an example that also shows that in Proposition 1.2.1 the assumption that the curvature is positive is essential.

1.3 Planar Curves

We will deal in this section with planar curves, more specifically with curves in $\mathbb{R}^2$. To simplify the calculations, we only consider curves parametrized by arc length: the formulas we obtain are easily adapted to any other parameterizations.

Consider a curve $\alpha: I \rightarrow \mathbb{R}^2$. Then there exists, for each $s \in I$, a single vector $\mathbf{n}(s)$ such that the pair $(\boldsymbol{\tau}(s), \mathbf{n}(s))$ forms a direct orthonormal or positively oriented dihedron: if $\boldsymbol{\tau}(s) = (x'(s), y'(s))$ then $\mathbf{n}(s) = (-y'(s), x'(s))$. As before, we know that the vectors $\boldsymbol{\tau}'(s)$ and $\boldsymbol{\tau}(s)$ are orthogonal; but in this case we can conclude that $\boldsymbol{\tau}'(s)$ is a multiple of $\mathbf{n}(s)$. The *curvature* of α at the point $\alpha(s)$ is the number $K(s)$ such that $\boldsymbol{\tau}'(s) = k(s)\mathbf{n}(s)$.

This curvature can take negative values and is therefore sometimes called *signed curvature*; its absolute value is equal to the curvature defined in the previous section. Whenever we talk about the curvature of a curve in $\mathbb{R}^2$ we will be referring to the signed curvature.

Frenet's formulas in this case boil down to

$$\begin{aligned}\tau' &= k\mathbf{n} \\ \mathbf{n}' &= -k\tau.\end{aligned}$$

Since $\tau(s)$ is a unit vector, it describes the position vector of a point on $\mathbb{S}^1$ (circle with radius 1 centered at the origin). Denoting by $\varphi(s)$ the oriented angle that $\tau(s)$ makes with the positive part of the x -axis, we have $\tau(s) = (\cos \varphi(s), \sin \varphi(s))$ and $\mathbf{n}(s) = (-\sin \varphi(s), \cos \varphi(s))$. From this we obtain

$$\begin{aligned}\tau'(s) &= \varphi'(s)(-\sin \varphi(s), \cos \varphi(s)), \\ k(s) &= \langle \tau'(s), \mathbf{n}(s) \rangle = \varphi'(s).\end{aligned}\tag{1.1}$$

This last formula gives a geometric interpretation of curvature, showing that it measures the variation of the angle φ that the tangent line to the curve makes with a fixed oriented line; and that the curvature is positive or negative according to whether the curve turns left or right (see Fig. 1.4).

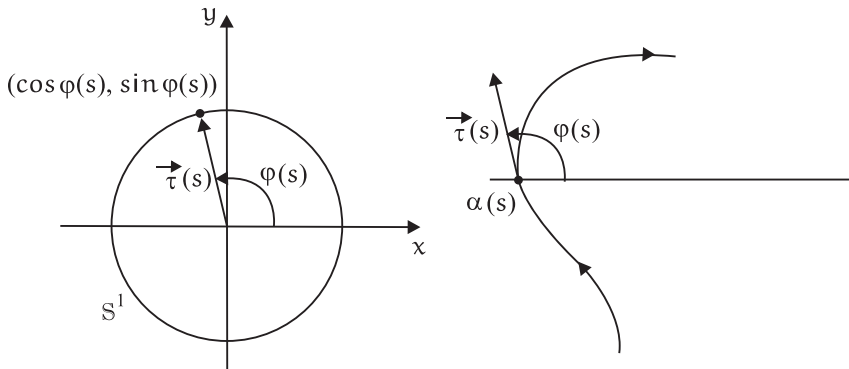


Figure 1.4

Note 1.3.1 We must make one caveat: the value $\varphi(s)$ of the angle is only determined up to an integer multiple of 2π , and it is not clear that we can make a choice for each s such that the resulting function is differentiable or even continuous. The way to solve this problem is to take advantage of the formula (1.1) to define $\varphi(s)$: fixing $s_0 \in I$ and a number φ_0 such that $\tau(s_0) = (\cos \varphi_0, \sin \varphi_0)$, we put

$$\varphi(s) = \varphi_0 + \int_{s_0}^s k(t) dt.$$

The function φ is obviously differentiable and all that remains to be shown is that

$$\tau(s) = (\cos \varphi(s), \sin \varphi(s))$$

– or, which is the same, that the function

$$\delta(s) = \frac{1}{2} |\boldsymbol{\tau}(s) - (\cos \varphi(s), \sin \varphi(s))|^2$$

is identically zero. Let us put $\boldsymbol{\tau}(s) = (x'(s), y'(s))$: by Frenet's formulas, we have $x''(s) = -k(s)y'(s)$ and $y''(s) = k(s)x'(s)$. Therefore

$$\begin{aligned} \delta'(s) &= \langle \boldsymbol{\tau}'(s) - \varphi'(s)(-\sin \varphi(s), \cos \varphi(s)), \boldsymbol{\tau}(s) - (\cos \varphi(s), \sin \varphi(s)) \rangle \\ &= (k(s) - \varphi'(s))(y'(s) \cos \varphi(s) - x'(s) \sin \varphi(s)) = 0, \end{aligned}$$

since $\varphi'(s) = k(s)$. From this, and since $\delta(s_0) = 0$, we conclude that $\delta \equiv 0$.

We further stress that any other continuous choice of the angle between $\boldsymbol{\tau}(s)$ and the x -axis is of the form $\varphi(s) + 2n\pi$, for some constant $n \in \mathbb{Z}$ (since the difference between two choices is continuous and only takes values in the discrete set $\{2n\pi : n \in \mathbb{Z}\}$).

We end this section with a result that is the version for planar curves of Theorem 1.2.2 and the note 1.2.3; its proof is kept as an exercise (it is possible to prove directly, without using the theorem on solutions of differential equations, the existence of α , and for uniqueness adapt the proof of Theorem 1.2.2)

Theorem 1.3.2 *Given a differentiable function $k: [a, b] \rightarrow \mathbb{R}$, there exists some curve $\alpha: [a, b] \rightarrow \mathbb{R}^2$ whose curvature at $\alpha(s)$ is $k(s)$. Any other curve with the same curvature function is the composite of α with some rigid plane movement.*

It follows in particular from Theorem 1.3.2 that the only planar curves with nonzero constant curvature are circles or arcs of circumference. A direct proof of this fact is certainly possible and is an exercise well worth to be done.

Exercises

6. (a) Consider a curve $\alpha(t) = (x(t), y(t))$ not necessarily parametrized by arc length, and put, as usual, $v(t) = |\alpha'(t)|$ and $\boldsymbol{\tau}(t) = \frac{1}{v(t)} \alpha'(t)$. Prove each of the following formulas for the curvature of α :

$$\begin{aligned} k &= \frac{1}{v} \langle \boldsymbol{\tau}', \mathbf{n} \rangle = \frac{1}{v^2} \langle \alpha'' \mathbf{n} \rangle = \frac{1}{v^3} \langle \alpha'', v \mathbf{n} \rangle \\ &= \frac{x'y'' - x''y'}{((x')^2 + (y')^2)^{3/2}}. \end{aligned}$$

(b) Show that the curvature of the ellipse given by the equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, parametrized by $\alpha(t) = (a \cos t, b \sin t)$, is given by

$$k(t) = \frac{ab}{(a^2 \sin^2 t + b^2 \cos^2 t)^{3/2}}.$$

7. Let $\mathbf{v}: [a, b] \rightarrow \mathbb{S}^1$ be a differentiable mapping. Show that there exists a differentiable function $\varphi: [a, b] \rightarrow \mathbb{R}$ such that $\mathbf{v}(t) = (\cos \varphi(t), \sin \varphi(t))$ for all t on $[a, b]$.

8. Let $\alpha:]a, b[\rightarrow \mathbb{R}^2$ be a regular curve parametrized by arc length. Given $s_0 \in]a, b[$ and $p \in \mathbb{R}^2 \setminus \{\alpha(s_0)\}$, let C be the circle with center p and radius $|\alpha(s_0) - p|$.

(a) Show that C is tangent to α at the point $\alpha(s_0)$ if and only if p is a point of the normal to α at $\alpha(s_0)$.

(b) Assume $p = \alpha(s_0) + \lambda \mathbf{n}(s_0)$ and consider the function $\rho(s) = |\alpha(s) - p|^2$. Show that if $\lambda k(s_0) > 1$ then s_0 is a strict local maximum of ρ and if $\lambda k(s_0) < 1$ then s_0 is a strict local minimum.

(c) Let $\mathcal{D} \subseteq \mathbb{R}^2$ be a circle with radius R such that $\alpha(]a, b[)$ is contained in the closed disk bounded by $\mathcal{D}$. Conclude that at the instants s_0 at which $\alpha(s_0) \in \mathcal{D}$, one has $|k(s_0)| \geq \frac{1}{R}$.

9. Let $\alpha: I \rightarrow \mathbb{R}^2$ be a regular curve of always nonzero curvature. The curve $\beta(t) = \alpha(t) + \frac{1}{k(t)} \mathbf{n}(t)$ ($t \in I$) traversed by the center of curvature of α is called the *evolute* of α .

(a) Show that, if it is defined, the tangent line to its evolute at time t coincides with the normal to α at the same instant.

(b) Assume that the curve $\tilde{\beta}(t) = \alpha(t) + \lambda(t) \mathbf{n}(t)$ has the property described in (a). Show that $\lambda(t) = \frac{1}{k(t)}$.

(c) Consider the normals to α at two nearby points $\alpha(t_0)$ and $\alpha(t_0 + h)$. Show that as $h \rightarrow 0$, the point of intersection of the two normals tends to $\beta(t_0)$.

(d) Study the evolute of the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

1.4 Contact of Curves

We continue our study of curves by considering their Taylor polynomial expansions: given s_0 inside the interval I and $n \geq 1$, we can write

$$\alpha(s_0 + s) = \alpha(s_0) + s\alpha'(s_0) + \frac{1}{2}s^2\alpha''(s_0) + \cdots + \frac{1}{n!}s^n\alpha^{(n)}(s_0) + o(s^n),$$

where the remainder $o(s^n)$ is a vector such that $\lim_{s \rightarrow 0} \frac{1}{s^n} |o(s^n)| = 0$. This expression of the curve can be used to detect its local properties: below we consider the contact theory of planar curves; in the exercises we obtain the geometric meaning of the osculating plane and the sign of torsion.

We now introduce a concept that measures the degree of closeness of two planar curves in the neighborhood of an intersection point. We say that the two curves α and $\tilde{\alpha}$ have *n-order contact* at the point $\alpha(s_0) = \tilde{\alpha}(s_0)$ if

$$\lim_{s \rightarrow 0} \frac{1}{s^n} |\tilde{\alpha}(s_0 + s) - \alpha(s_0 + s)| = 0.$$

Considering the Taylor polynomial expansion, it is easily seen that this condition is equivalent to the condition that the derivatives of α and $\tilde{\alpha}$ up to order n coincide at s_0 . In particular, since $\alpha'(s_0) = \tau(s_0)$ and $\alpha''(s_0) = k(s_0)\mathbf{n}(s_0)$, we conclude that *two planar curves have second-order contact at a point p if and only if they are tangent to p and had equal curvature there* (of course when we speak of tangency here we require that the velocity vectors of the two curves are identical — i.e., point in the same direction); when the curvature is nonzero, this is equivalent to these two curves having the same center of curvature at p . From this we conclude that *the only circle which has second-order contact with α at a point of nonzero curvature is the one with center at the center of curvature of α at that point and radius equal to the radius of curvature* (of course there is no circle with second-order contact at a point of zero curvature).

The definition of n -order contact we have given is perhaps not the most natural nor the easiest to handle, since it depends on a special parameterization of the two curves. To improve the situation, we start by defining $\Delta(t)$ as the distance between the points $\alpha(s_0 + s)$ and $\tilde{\alpha}(s_0 + \tilde{s})$ given by the condition

$$\langle \alpha(s_0 + s) - \alpha(s_0), \tau(s_0) \rangle = t = \langle \tilde{\alpha}(s_0 + \tilde{s}) - \alpha(s_0), \tau(s_0) \rangle;$$

$\Delta(t)$ is thus the length of the line segment bounded by the intersections with α and $\tilde{\alpha}$ of the line orthogonal to $\tau(s_0)$ and the (oriented) distance t from $\alpha(s_0)$ (see Fig. 1.5).

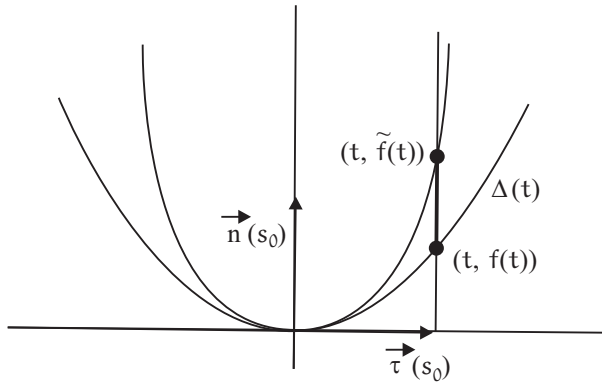


Figure 1.5

Proposition 1.4.1 *The curves α and $\tilde{\alpha}$ have n -order contact at the point $\alpha(s_0) = \tilde{\alpha}(s_0)$ if and only if $\lim_{t \rightarrow 0} \frac{1}{t^n} \Delta(t) = 0$.*

Proof We first assume that α and $\tilde{\alpha}$ have n -order contact — i.e., that $\alpha(s_0) = \tilde{\alpha}(s_0)$ and $\alpha^{(i)}(s_0) = \tilde{\alpha}^{(i)}(s_0)$ for $1 \leq i \leq n$, — and we define functions g and $\tilde{g}$ by

$$\begin{aligned} g(s) &= \langle \alpha(s_0 + s) - \alpha(s_0), \tau(s_0) \rangle \\ \tilde{g}(s) &= \langle \tilde{\alpha}(s_0 + s) - \alpha(s_0), \tau(s_0) \rangle. \end{aligned}$$

Note that $g(0) = \tilde{g}(0) = 0$ and $g^{(i)}(0) = \tilde{g}^{(i)}(0)$ for $1 \leq i \leq n$, and that, since the first derivative at 0 of these functions is nonzero, both have local inverses in a neighborhood of 0, which we will denote by g^{-1} and $\tilde{g}^{-1}$. Let us now put

$$\begin{aligned} F(s) &= \langle \alpha(s_0 + s) - \alpha(s_0), \mathbf{n}(s_0) \rangle, \\ \tilde{F}(s) &= \langle \tilde{\alpha}(s_0 + s) - \alpha(s_0), \mathbf{n}(s_0) \rangle, \\ f(t) &= F \circ g^{-1}(t), \tilde{f}(t) = \tilde{F} \circ \tilde{g}^{-1}(t), \end{aligned}$$

and note that, for t near 0, one has $\Delta(t) = |f(t) - \tilde{f}(t)|$ (see Fig. 1.5.). Moreover, the first n derivatives of f and $\tilde{f}$ at 0 coincide, since at this point also the first n derived from F are identical to those from $\tilde{F}$, and those from g^{-1} to those from $\tilde{g}^{-1}$. It follows, as desired, that $\lim_{t \rightarrow 0} \frac{1}{t^n} \Delta(t) = 0$.

We now prove the converse implication. To simplify the notation, we assume that $s_0 = 0$, take $\alpha(0) = \tilde{\alpha}(0)$ for the origin of the coordinates, and further assume that the tangent and normal vectors at this point are $(1, 0)$ and $(0, 1)$ respectively.

With the above notation, let $\beta(t) = \alpha \circ g^{-1}(t)$, $\tilde{\beta}(t) = \tilde{\alpha} \circ \tilde{g}^{-1}(t)$: we then have $\beta(t) = (t, f(t))$ and $\tilde{\beta}(t) = (t, \tilde{f}(t))$, and by hypothesis f and $\tilde{f}$ are functions whose derivatives up to the n th order coincide at 0. Since α and $\tilde{\alpha}$ are parametrized by arc length, we have

$$\begin{aligned} g^{-1}(t) &= \int_0^t |\beta'(r)| dr = \int_0^t \sqrt{1 + [f'(r)]^2} dr, \\ \tilde{g}^{-1}(t) &= \int_0^t |\tilde{\beta}'(r)| dr = \int_0^t \sqrt{1 + [\tilde{f}'(r)]^2} dr, \end{aligned}$$

and these formulas show that the derivatives of g^{-1} and $\tilde{g}^{-1}$ at the point 0 coincide at least to the same order as the derivatives of f and $\tilde{f}$ at the same point — that is, to the order n . Since $\alpha = \beta \circ g$ and $\tilde{\alpha} = \tilde{\beta} \circ \tilde{g}$, we conclude that $\alpha^{(i)}(0) = \tilde{\alpha}^{(i)}(0)$ for $1 \leq i \leq n$, which means that α and $\tilde{\alpha}$ have n -order contact in $\alpha(0) = \tilde{\alpha}(0)$.

□

Exercises

10. Assume that two planar curves α and $\tilde{\alpha}$, not necessarily parametrized by arc length, touch at the point $p = \alpha(0) = \tilde{\alpha}(0)$. Show that if $\lim_{t \rightarrow 0} \frac{1}{t^n} |\tilde{\alpha}(t) - \alpha(t)| = 0$, then the curves have n -order contact at p .

11. With the same terminology as in Proposition 1.4.1, but under the assumption that $s_0 = 0$ and $\varphi(s_0) = 0$, show that $f'(t) = tg\varphi(s)$ and $f''(t) = \frac{k(s)}{\cos^3 \varphi(s)}$, where $s = g^{-1}(t)$. Conclude that if $k(0) > 0$ then there exists $\varepsilon > 0$ such that, on an appropriate coordinate system, the trace of $\alpha|_{[-\varepsilon, \varepsilon]}$ is the graph of a convex function.

12. Let I be an open interval such that $o \in I$ and let $\alpha: I \rightarrow \mathbb{R}^3$ be a curve parametrized by arc length whose curvature at 0 is nonzero.

(a) Show that

$$\alpha(s) = \alpha(0) + \left(s - \frac{1}{6}k^2s^3\right)\boldsymbol{\tau} + \left(\frac{1}{2}ks^2 + \frac{1}{6}k's^3\right)\mathbf{n} - \frac{1}{6}k\nu s^3\mathbf{b} + o(s^3),$$

where the quantities k , ν and the vectors $\boldsymbol{\tau}$, $\mathbf{n}$, $\mathbf{b}$ are computed at 0.

(b) Conclude that if $\nu(0) > 0$ then, when s reaches the instant 0, the curve crosses the osculating plane at $\alpha(0)$ from top to bottom (the “top part” is the one pointed to by $\mathbf{b}(0)$).

13. Using the notation and assumptions from Exercise 12, show that the plane that contains the points $\alpha(0)$, $\alpha(h_0)$ and $\alpha(h_1)$, for $h_1 < 0 < h_0$, converges to the osculating plane at $\alpha(0)$ when $|h_0| + |h_1| \rightarrow 0$.

Hint: Using the Taylor expansion — just up to the second order — show that

$$\mathbf{v}(h_0, h_1) = \frac{(\alpha(h_0) - \alpha(0)) \times (\alpha(h_1) - \alpha(0))}{|(\alpha(h_0) - \alpha(0)) \times (\alpha(h_1) - \alpha(0))|}$$

has limit $\mathbf{b}(0)$ when $|h_0| + |h_1| \rightarrow 0$.

1.5 Convex Curves

Continuing with planar curves parametrized by arc length, let us talk about simple closed curves and characterize those that are convex. We say that a curve $\alpha: [a, b] \rightarrow \mathbb{R}^2$ is *closed* if $\alpha(a) = \alpha(b)$; if its periodic extension, defined by $\alpha(s + n(b - a)) = \alpha(s)$ for $s \in [a, b]$ and $n \in \mathbb{Z}$, is differentiable (i.e., C^∞), the curve is *closed regular*; and if the curve has no self-intersections — that is, if its restriction on $[a, b[$ is injective — we say it is *simple*.

In this section all closed curves are regular; and, where necessary, we consider them defined in $\mathbb{R}$ by periodic extension.

We recall from Section 1.3 that there is a continuous choice $\varphi(s)$ of the angle that the vector $\boldsymbol{\tau}(s)$ makes with the positive part of the x -axis. Since $\boldsymbol{\tau}(b) = \boldsymbol{\tau}(a)$, the difference $\varphi(b) - \varphi(a)$ is an integer multiple of 2π , which, by note 1.3.1, does not depend on the choice of $\varphi(s)$. We call *rotation index* of the closed curve α the integer $\Re(\alpha) = \frac{1}{2\pi}(\varphi(b) - \varphi(a))$; $\Re(\alpha)$ thus counts the number of turns that its tangent vector $\boldsymbol{\tau}(s)$ makes in the unit circle when the point $\alpha(s)$ completes one turn around the curve. Since $\varphi'(s) = k(s)$, the rotation index can be given in integral form

$$\Re(\alpha) = \frac{1}{2\pi} \int_a^b k(s) ds.$$

A closed curve $\alpha: [a, b] \rightarrow \mathbb{R}^2$ is called *convex* if, for every $s_0 \in [a, b]$, the curve is all on the same side of the tangent line to α at the point $\alpha(s_0)$ — that is, if the function

$h(s) = \langle \alpha(s) - \alpha(s_0), \mathbf{n}(s_0) \rangle$ does not change sign. Our next result characterizes these curves.

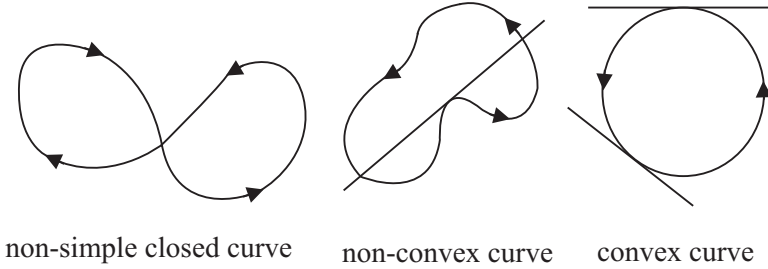


Figure 1.6

Theorem 1.5.1 *A closed curve is convex if and only if its curvature does not change sign and its rotation index is ± 1 . Any convex curve is simple.*

Proof (i) We begin by showing that any closed curve with non-negative curvature at all its points and rotation index 1 is convex. Let $\alpha: [a, b] \rightarrow \mathbb{R}^2$ be such a curve. Then the function $\varphi(s)$ is non-decreasing and $\varphi(b) = \varphi(a) + 2\pi$.

Given $s_0 \in [a, b]$, we want to prove that the function $h(s) = \langle \alpha(s) - \alpha(s_0), \mathbf{n}(s_0) \rangle$ does not change sign. Otherwise $h(s)$ reaches a positive maximum and a negative minimum at points $s_1, s_2 \in [a, b] \setminus \{s_0\}$, and at each of these points the tangent line is parallel to the tangent line at $\alpha(s_0)$. Hence, there exist $i \neq j \in \{0, 1, 2\}$ such that $\varphi(s_i) = \varphi(s_j)$ and, φ being non-decreasing, this is only possible if φ is constant in the interval between s_i and s_j . This means that the curve contains the line segment from $\alpha(s_i)$ to $\alpha(s_j)$, and therefore the tangent lines at these points coincide — which is absurd given the way they were chosen. Therefore h does not change sign and α is convex.

(ii) We assume now and until the end of the proof that $\alpha: [a, b] \rightarrow \mathbb{R}^2$ is a convex curve. Let us first see that $k(s)$ does not change sign. Consider the function of two variables $(s, t) \in [a, b] \times [a, b]$ defined by

$$H(s, t) = \langle \alpha(s) - \alpha(t), \mathbf{n}(t) \rangle.$$

By hypothesis, for each $t \in [a, b]$ the function h_t given by $h_t(s) = H(s, t)$ has constant sign — that is, the restriction of H to each horizontal line segment $[a, b] \times \{t\}$ is either non-negative or non-positive. What we want for now is to prove that the function H itself does not change sign.

Assume, instead, that there exist $(s_0, t_0), (s_1, t_1) \in [a, b] \times [a, b]$ such that $H(s_0, t_0) < 0 < H(s_1, t_1)$; and let us agree that $t_0 < t_1$. Consider the set

$$A = \{\hat{t} \geq t_0 : H(s, t) \leq 0 \quad \forall (s, t) \in [a, b] \times [t_0, \hat{t}]\};$$

A is a non-empty interval (because $t_0 \in A$) and closed (because H is continuous). By continuity of H , there exists $\delta > 0$ such that $H(s_0, t) < 0$ whenever $t_0 \leq t \leq t_0 + \delta$,

and therefore A contains the interval $[t_0, t_0 + \delta]$. Since $t_1 \notin A$, $t_2 = \sup(A)$ lies in $]t_0, t_1[$. Since $t_1 \in A$, we have $H(s, t_1) \leq 0$ for all $s \in [a, b]$. If for all s this inequality were strict, then the same argument as above would show that there exists $\delta > 0$ such that $[t_2, t_2 + \delta] \subset A$, which contradicts the definition of t_2 . We then conclude that $H(s, t_2) = 0$ for all $s \in [a, b]$ — which is absurd because it means that the curve α is all contained in a straight line.

So we have proved that H does not change sign. That the curvature does not either is now immediate: when $k(t_0) > 0$, the function $h_{t_0}(s)$ has a strict local minimum for $s = t_0$ (since $h'_{t_0}(t_0) = 0$ and $h''_{t_0}(t_0) = k(t_0)$) and therefore $H(s, t_0) > 0$ for s near t_0 ; and, conversely, when $k(t_0) < 0$ one has $H(s, t_0) < 0$ for s near t_0 .

(iii) Let us now prove that α is simple. Let us assume, to the contrary, that it has some self-intersection which, changing if necessary the initial point, we suppose to take place at $\alpha(a)$. Then there exists $x \in]a, b[$ such that $\alpha(c) = \alpha(a)$.

The function $H(s, t)$ vanishes for $(s, t) = (a, c)$ and, by (ii), reaches at that point a local extremum. Thus $\langle \tau(a), \mathbf{n}(c) \rangle = \frac{\partial H}{\partial s} \Big|_{(a,c)} = 0$, and therefore $\tau(a) = \pm \tau(c)$. If it were $\tau(a) = -\tau(c)$, we would have $H(s, a) = -H(s, c)$ for all $s \in [a, b]$, which is impossible by (ii). Therefore we have $\tau(a) = \tau(c)$.

To simplify the notation, we assume that $\alpha(a) = \alpha(c) = (0, 0)$. Let us put, (as in the proof of 1.4.1) for $s \geq 0$,

$$\begin{aligned} g(s) &= \langle \alpha(a+s), \tau(a) \rangle, & \tilde{g}(s) &= \langle \alpha(c+s), \tau(a) \rangle, \\ F(s) &= \langle \alpha(a+s), \mathbf{n}(a) \rangle, & \tilde{F}(s) &= \langle \alpha(c+s), \mathbf{n}(a) \rangle, \\ f(t) &= F \circ g^{-1}(t), & \tilde{f}(t) &= \tilde{F} \circ \tilde{g}^{-1}(t) \end{aligned}$$

— where the functions f and $\tilde{f}$ are defined on some interval $[0, \varepsilon]$, $\varepsilon > 0$. The graphs of f and $\tilde{f}$ are portions of the trace of α : in fact, putting $s = g^{-1}(t)$, $\tilde{s} = \tilde{g}^{-1}(t)$, we can write

$$\begin{aligned} \alpha(a+s) &= t\tau(a) + f(t)\mathbf{n}(a), \\ \alpha(c+\tilde{s}) &= t\tilde{\tau}(a) + \tilde{f}(t)\mathbf{n}(a). \end{aligned}$$

From these formulas we obtain

$$\begin{aligned} H(a+s, c+\tilde{s}) &= \{f(t) - \tilde{f}(t)\} \langle \mathbf{n}(a), \mathbf{n}(c+\tilde{s}) \rangle, \\ H(c+\tilde{s}, a+s) &= \{\tilde{f}(t) - f(t)\} \langle \mathbf{n}(t), \mathbf{n}(a+s) \rangle. \end{aligned}$$

In each of these products, and since $\mathbf{n}(a) = \mathbf{n}(c)$, the second factor is positive for $s, \tilde{s}$ sufficiently small; hence, if it were $f(t) \neq \tilde{f}(t)$ for some $t \in [0, \varepsilon]$, $H(a+s, c+\tilde{s})$ and $H(c+\tilde{s}, a+s)$ would have opposite signs, in contradiction to (ii). We thus have $f(t) = \tilde{f}(t)$ for all $t \in [0, \varepsilon]$ — and from this, since α is parametrized by arc length, we conclude that there exists $\delta > 0$ such that $\alpha(a+s) = \alpha(c+s)$ for all $s \in [0, \delta]$. A

trivial argument now proves that for all $s \geq 0$ one has $\alpha(a+s) = \alpha(c+s)$, and this says that when the curve returns to the starting point, it repeats the same path from then on. The given hypothesis thus leads us to conclude that $\alpha|_{[a,b]}$ gives more than one turn to the same closed curve. Assuming this does not happen, such a $c \in]a, b[$ with $\alpha(c) = \alpha(a)$ does not exist and the curve is simple.

(iv) Let us now show that $\Re(\alpha) = \pm 1$. Assuming that α is positively oriented, and putting $\varphi(b) = \varphi(a) + 2n\pi$, we have $n = \Re(\alpha) \geq 1$, and we want to see that $n = 1$. We can assume, without loss of generality, that $k(a) > 0$. Take $c \in [a, b]$ such that $\varphi(c) = \varphi(a) + 2\pi$: then $H(a, c) = -H(c, a)$, and it follows by (ii) that $H(a, c) = 0$, which means that the tangent lines at $\alpha(a)$ and $\alpha(c)$ coincide. The function $\lambda(s) = H(c, a+s)$ reaches a minimum at 0, and so

$$0 = \lambda'(0) = -k(a)\langle \alpha(c) - \alpha(a), \tau(a) \rangle,$$

whence it follows, since $k(a) > 0$ and the points $\alpha(c)$ and $\alpha(a)$ lie on a straight line parallel to $\tau(a)$, that $\alpha(c) = \alpha(a)$. Since α is simple, we must have $c = b$ and therefore $\Re(\alpha) = 1$. $\square$

It is important to note that any simple curve, whether convex or not, has rotation index ± 1 : this is what the *rotation index theorem* says, the proof of which we give in the Appendix to Chap. 4, but of which a special case is given in Exercise 16 below. If we already had this result, the proof of 1.5.1 would be somewhat simplified; another simplification would be to suppress step (iii) if, as some authors do, we already required in the definition that a convex curve be simple.

One result we will not prove, but of which we will make important use, not always explicit, is the *Jordan curve theorem*. This theorem states that any simple closed curve divides the plane into two disjoint connected open subsets of which it is a common boundary. (For a proof of the theorem in the differentiable case, and its generalization to higher dimensions, see [15]; for the topological version we suggest [17], which also includes a proof of Schönflies' theorem: *the region bounded by a simple closed curve is homeomorphic to an open disc*).

To finish this section we mention that a convex curve of nonzero curvature at all its points is usually called *strictly convex*. In this case $\varphi(s)$ is strictly monotone and therefore every tangent line touches the curve at a single point.

Exercises

14. Show that if a line intersects a closed convex curve then one and only one of the following cases occurs: either the line is tangent to the curve, or it intersects the curve at exactly two points.

15. Let α be a closed, simple, regular curve, Ω be the open set bounded by α , and $\overline{\Omega} = \Omega \cup \alpha$ be the closure of α . Show that the following conditions are equivalent:

- (i) $\overline{\Omega}$ is a convex set (i.e., $p, q \in \overline{\Omega} \Rightarrow$ the line segment $[p, q] \subseteq \overline{\Omega}$);

(ii) α is a convex curve.

(Suggestion for (i) $\Rightarrow$ (ii): Show that for each s_0 , the image of the function s ($s \neq s_0$) $\mapsto (\alpha(s) - \alpha(s_0))/|\alpha(s) - \alpha(s_0)|$ is contained in a semicircle.)

16. (a) Let $\mathcal{H}: [0, 1] \times [a, b] \rightarrow \mathbb{R}^2$ be a differentiable mapping such that every $\alpha_s = \mathcal{H}(s, \cdot)$ is a regular closed curve. Show that $\mathfrak{R}(\alpha_s)$ is constant. (Exercise 7 may be helpful.)

(b) Let $\alpha: [a, b] \rightarrow \mathbb{R}^2$ be a regular closed curve and $p \in \mathbb{R}^2 \setminus \alpha([a, b])$ such that each half-line with origin at p intersects α exactly once, and this intersection is transverse (i.e., the half-line is **not** tangent to α at the point of intersection). Prove that $\mathfrak{R}(\alpha) = \pm 1$.

(c) Now assume only that all intersections of α with half-lines r of origin p are transversal. Show that the cardinal of the set $T(r) = \{t \in [a, b] : \alpha(t) \in r\}$ is the same for all such half-lines.

1.6 Curves of Constant Width

In this section, we explore the varying width of a planar curve. The width of the curve in any given direction is the narrowest distance between two lines perpendicular to that direction that can contain the curve. This means that the width of a curve is not necessarily the same in all directions. Remarkably, besides the circle, there are other convex curves that have a constant width regardless of direction; and the perimeter of such curves is equal to that of the circle of the same width.

We will deal in this section with the width of a planar curve. The width of a curve in a given direction is the minimal width among the strips that contain the curve and are bounded by lines orthogonal to that direction.

Given a closed curve $\alpha: [a, b] \rightarrow \mathbb{R}^n$ we define, for $\mathbf{v} \in \mathbf{S}^1$, $h(\mathbf{v}) = \max_{a \leq s \leq b} \langle \alpha(s), \mathbf{v} \rangle$.

Since the maximum of $\langle \alpha(s), \mathbf{v} \rangle$ is only reached at points s such that $\langle \boldsymbol{\tau}(s), \mathbf{v} \rangle = 0$, $h(\mathbf{v})$ is the maximum among the (oriented) distances from the origin to the tangent lines to α that are orthogonal to $\mathbf{v}$. The *width of α in the direction of $\mathbf{v}$* is $\mathcal{L}(\mathbf{v}) = h(\mathbf{v}) + h(-\mathbf{v})$.

If α is a convex curve, then there are exactly two tangent lines to α that are orthogonal to $\mathbf{v}$ (although each of them may be tangent to α at more than one point), and $\mathcal{L}(\mathbf{v})$ is the distance between these lines (see Fig. 1.7). For example, the width of a circle is, in all directions, equal to its diameter.

Proposition 1.6.1 *In any closed curve the diameter and maximum width are equal.*

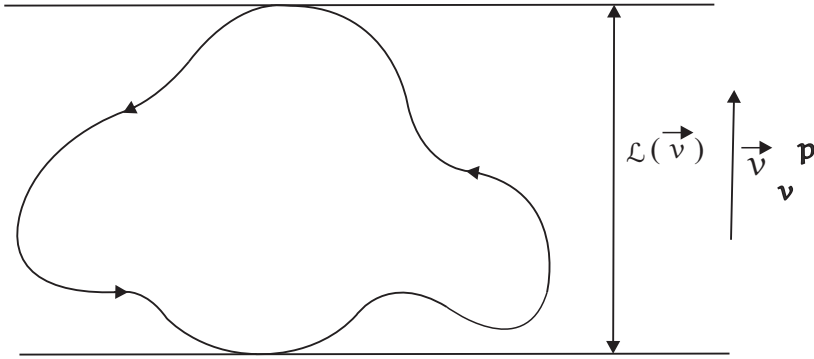


Figure 1.7

Proof We denote the diameter of α by $D = \max\{|\alpha(s) - \alpha(t)| : s, t \in [a, b]\}$, and by $\mathfrak{L} = \max\{\mathcal{L}(\mathbf{v}) : \mathbf{v} \in \mathbf{S}^1\}$ the maximum width of α .

Let us check that $D \leq \mathfrak{L}$. Consider the function $\mathcal{D}(s, t) = |\alpha(s) - \alpha(t)|$. This function reaches its maximum D on a pair of points (s_0, t_0) such that

$$\left. \frac{\partial \mathcal{D}}{\partial s} \right|_{(s_0, t_0)} = \left. \frac{\partial \mathcal{D}}{\partial t} \right|_{(s_0, t_0)} = 0,$$

conditions which are translated by the equalities

$$\langle \tau(s_0), \alpha(s_0) - \alpha(t_0) \rangle = \langle \tau(t_0), \alpha(s_0) - \alpha(t_0) \rangle = 0.$$

This shows that the tangent lines at $\alpha(s_0)$ and $\alpha(t_0)$ are parallel, both being orthogonal to the line segment joining $\alpha(s_0)$ to $\alpha(t_0)$. Moreover the curve is completely contained in the strip between these tangent lines, otherwise the maximum distance between distinct points of α would exceed D . Thus the width of the strip, which is D , is also equal to $\mathcal{L}(\mathbf{n}(s_0))$, and thus $D \leq \mathfrak{L}$.

Let us now deal with the opposite inequality. Given $\mathbf{v} \in \mathbf{S}^1$, let $s_0, t_0 \in [a, b]$ be such that $h(\mathbf{v}) = \langle \alpha(s_0), \mathbf{v} \rangle$ and $h(-\mathbf{v}) = \langle \alpha(t_0), -\mathbf{v} \rangle$. Then

$$|\alpha(s_0) - \alpha(t_0)| \geq \langle \alpha(s_0) - \alpha(t_0), \mathbf{v} \rangle = \mathcal{L}(\mathbf{v}),$$

and therefore $D \geq \mathcal{L}(\mathbf{v})$. Since this inequality holds for all $\mathbf{v}$, it follows that $D \geq \mathfrak{L}$. $\square$

Let us now assume that the curve α is convex and has constant width $\mathfrak{L}$. By 1.6.1, also its diameter is equal to $\mathfrak{L}$. Let us now see that this diameter is realized by many pairs of points on the curve.

Fixing $s_0 \in [a, b]$, let $\alpha(s_1)$ be a point such that $\tau(s_1) = -\tau(s_0)$. We know from the analysis done in Section 1.5 that the curve is contained in the strip bounded by tangent lines at $\alpha(s_0)$ and $\alpha(s_1)$. Hence these tangent lines are at a distance $\mathfrak{L}$ from each other, and therefore $|\alpha(s_0) - \alpha(s_1)| \geq \mathfrak{L}$. Since the diameter of α is $\mathfrak{L}$, it must be $|\alpha(s_0) - \alpha(s_1)| = \mathfrak{L}$, an equality that is only possible if the line segment between

$\alpha(s_0)$ and $\alpha(x_1)$ is orthogonal to the tangent lines to α at these points. Furthermore, there is no other point $\alpha(\tilde{s}_1)$ such that $|\alpha(s_0) - \alpha(\tilde{s}_1)| = \mathfrak{L}$, for the proof of 1.6.1 shows that $\alpha(\tilde{s}_1)$ would also be on the normal to α at $\alpha(s_0)$.

We conclude that *for every point p of a convex curve of constant width $\mathfrak{L}$, there is a single point $\tilde{p}$ of the curve at the maximum distance $\mathfrak{L}$ from p , and $\tilde{p}$ is situated on the normal to α at p . We call this point $\tilde{p}$ the antipode of p . Assuming that the curve is positively oriented (and therefore has non-negative curvature at all points), our conclusion translates into $\tilde{p} = p + \mathfrak{L}\mathbf{n}(p)$. We stress that “being antipodal to” is a reflexive relation, and that two tangent lines to α that are parallel and distinct meet α at points that are antipodes of each other.*

Consider now the circle $\mathcal{C}$ with center $\tilde{p}$ and radius $\mathfrak{L}$. Such a circle is tangent to α at the point p ; and all other points of α are contained in the interior of the disk bounded by $\mathcal{C}$. Exercise 8 then says that the absolute value of the curvature of α at p is greater than or equal to $1/\mathfrak{L}$. We thus conclude that *any convex curve of constant width is strictly convex*.

Example 1.6.2 At this point it is good to wonder about the existence of constant-width convex curves that are not circles. The above discussion suggests that such a curve is determined by knowing the arc between two antipodal points p and $\tilde{p}$: the remaining segment is found by marking, from each point of this arc, a distance of $\mathfrak{L}$ along the normal.

To construct an example where $\mathfrak{L} = 2$, we consider a curve $\alpha: [0, c] \rightarrow \mathbb{R}^2$ with the following properties:

- (i) α is parametrized by arc length;
- (ii) $\alpha(0) = (1, 0)$, $\alpha(c) = (-1, 0)$, and there exists $\varepsilon > 0$ such that $\alpha([0, \varepsilon] \cup [c - \varepsilon, c]) \subseteq \mathbf{S}_+^1 = \{(x, y) \in \mathbf{S}^1: y \geq 0\}$;
- (iii) the trace of α is not contained in $\mathbf{S}_+^1$;
- (iv) the tangent vector $\tau(s)$, $s \in [0, c]$, describes a semicircle;
- (v) $k(s) > \frac{1}{2}$ for all $s \in [0, c]$.

Such a curve can be obtained by considering $\mathbf{S}_+^1$ as the graph of a function $[-1, 1] \rightarrow \mathbb{R}$ and adding to that function a non-constant function of class C^∞ that is identically zero in the intervals $[-1, -1 + \delta]$ and $[1 - \delta, 1]$, for some $\delta > 0$. (See Exercise 17 for the existence of functions with these properties.) Reparameterizing such that the graph of the resulting function starts at $(1, 0)$, we obtain a curve α that verifies conditions (i)-(iv). If the added function and its first and second derivatives are close to zero then the curvature of α is close to that of $\mathbf{S}_+^1$, which guarantees (v).

We define $\beta: [0, 2c] \rightarrow \mathbb{R}^n$ by $\beta(t) = \alpha(t)$ for $0 \leq t \leq c$, and $\beta(t) = \alpha(t - c) + 2\mathbf{n}(t - c)$ for $c \leq t \leq 2c$. Condition (ii) guarantees that β is well-defined at $t = c$ and that $\beta(2c) = \beta(0)$. For $c \leq t \leq 2c$ we have $\beta'(t) = \{1 - 2k(t - c)\}\tau(t - c)$ and, by (v), this vector never vanishes. Furthermore conditions (ii) and (i) imply that, in neighborhoods of the “gluing points” $(1, 0)$ and $(-1, 0)$, the curve β runs through arcs of $\mathbf{S}^1$ and is parametrized by arc length. Therefore β is a regular closed curve.

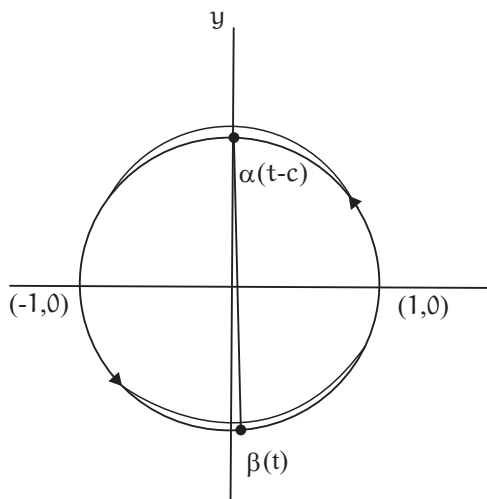


Figure 1.8

Noting that we have

$$k(t) = \frac{k(t-c)}{2k(t-c)-1},$$

for $c \leq t \leq 2c$, if $v(t) = 2k(t-c) - 1$ and $\tau(t) = -\tau(t-c)$ and this equality shows that the curvature of β is positive at all points. Finally, by (iv), the vector $\tau(t)$ turns exactly once around the circumference when t runs through $[0, 2c]$, and therefore $\Re(\beta) = 1$. By Theorem 1.5.1, β is convex. That β has constant width equal to 2 is now an immediate exercise. $\square$

We now state the most important result of this section, originally proved by E. Barbier in the 19th century using probabilistic methods (see [1]; Barbier's proof is also reproduced in [5], pp. 161-163).

Theorem 1.6.3 *The perimeter of any constant-width curve $\mathfrak{L}$ is equal to $\pi\mathfrak{L}$.*

Proof We fix a curve $\alpha: [0, L] \rightarrow \mathbb{R}^2$ that is convex and positively oriented and has constant width $\mathfrak{L}$. We assume that the parameter of the curve is given by arc length, so that its perimeter is L , and we consider α defined on $\mathbb{R}$ through its periodic extension. We further denote by $\varphi(s)$ a differentiable choice of the angle of $\tau(s)$ with the positive x -half-axis. The rotation index of α is 1, so $\varphi(s+L) = \varphi(s) + 2\pi$ for all $s \in \mathbb{R}$; and, since α is strictly convex, $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ is strictly increasing and has differentiable inverse.

For each $s \in \mathbb{R}$, we denote the antipode of $\alpha(s)$ by $\tilde{\alpha}(s)$. This function $s \mapsto \tilde{\alpha}(s)$ is also periodic with period equal to L ; and is differentiable, since we can write $\tilde{\alpha}(s) = \alpha(s) + \mathfrak{L}\mathbf{n}(s)$.

Lemma 1.6.4 *There exists a differentiable function $f: \mathbb{R} \rightarrow \mathbb{R}$ such that $\tilde{\alpha}(s) = \alpha(f(s))$ for all $s \in \mathbb{R}$. This function satisfies $f(s+L) = f(s) + L$, and its derivative is strictly positive at all points.*

Proof Assume that a certain function f satisfies the equation $\varphi \circ f(s) = \varphi(s) + \pi$. Then $\tau(f(s)) = (\cos(\varphi \circ f(s)), \sin(\varphi \circ f(s))) = -\tau(s)$, and therefore the points $\alpha(s)$ and $\alpha(f(s))$ are antipodes of each other — that is, $\tilde{\alpha}(s) = \alpha(f(s))$ just as we intend.

This means that we just have to find f such that $\varphi \circ f(s) = \varphi(s) + \pi$. Such a function is given by $f(s) = \varphi^{-1}(\varphi(s) + \pi)$, which is differentiable and has positive derivative. Furthermore, we have $f(s+L) = \varphi^{-1}(\varphi(s+L) + \pi) = \varphi^{-1}(\{\varphi(s) + \pi\} + 2\pi) = f(s) + L$, as we want. $\square$

We now finish the proof of 1.6.3. Differentiating the equality $\alpha(s) + \mathfrak{L}\mathbf{n}(s) = \alpha(f(s))$, we obtain $\{1 - \mathfrak{L}k(s)\} \tau(s) = f'(s)\tau(f(s))$ — and from this, as $\tau(f(s)) = -\tau(s)$, yields $f'(s) = -1 + \mathfrak{L}k(s)$. Finally, using 1.6.4, and since the rotation index of α is 1, we have

$$L = f(L) - f(0) = \int_0^L f'(s) ds = -L + \mathfrak{L} \int_0^L k(s) ds = -L + 2\pi\mathfrak{L}.$$

Note 1.6.5 We cannot omit the simplest example of a non-circular curve of constant width: *Reuleaux's triangle*, which is formed by three arcs, each centered at one of the vertices of an equilateral triangle ABC and radius equal to the side of the triangle. Its perimeter L and width $\mathfrak{L}$ are also related by $L = \pi\mathfrak{L}$, but the proof of 1.6.3 does not cover this case: the antipode of each point of the arc $\widehat{BC}$ (resp. $\widehat{CA}$, $\widehat{AB}$) is the point A (resp. B , C).

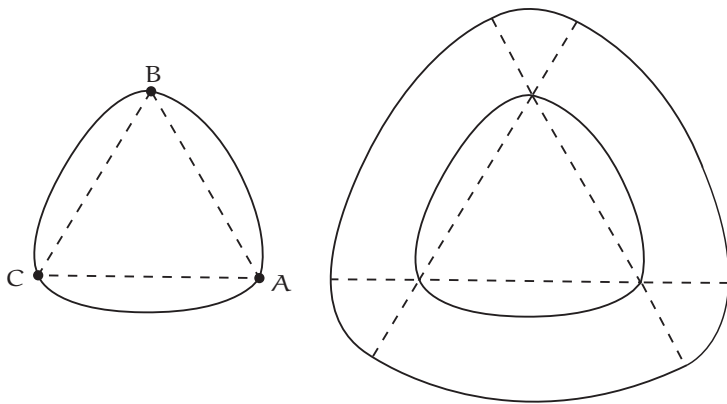


Figure 1.9

The Reuleaux triangle is a *piecewise regular* curve; we say that $\alpha: [a, b] \rightarrow \mathbb{R}^2$ is such a curve if there exists a partition $a = t_0 < t_1 < \dots < t_k = b$ of $[a, b]$ such that each restriction $\alpha|_{(t_{i-1}, t_i]}$ is regular. We now sketch how to extend 1.6.3 to piecewise regular convex closed curves, provided that each regular segment is at least of class C^2 .

Given $d > 0$, let α_d be the curve which surrounds α at a constant distance from it equal to d . The curve α_d is called *parallel* to α , and if α has constant width $\mathfrak{L}$, α_d has

constant width $\mathfrak{L} + 2d$. In Fig. 1.9 we show a curve parallel to the Reuleaux triangle. Each “corner” of α is replaced by an arc of a circle at α_d . Therefore α_d already has a well-defined tangent vector $\tau_d(s)$ at each point; the angle $\varphi_d(s)$ that $\tau_d(s)$ makes with the x -axis is a continuous, strictly increasing function, which is piecewise C^1 ; and the antipode mapping is already a bijection of the curve onto itself. The proof of 1.6.3 can easily be adapted to show that the perimeter of α_d is

$$l(\alpha_d) = \pi(\mathfrak{L} + 2d)$$

– and, letting $d \rightarrow 0$, we obtain $l(\alpha) = \pi\mathfrak{L}$, as desired. $\square$

There are numerous results on constant width curves: [7] contains a careful discussion of the topic (and its generalization to higher dimensions) and an extensive bibliography. In Section 5.5 we give some results on constant width surfaces.

Exercises

17. (a) Check that the function $f(x) = e^{-1/x(1-x)}$ if $0 < x < 1$, $f(x) = 0$ otherwise it is C^∞ , and that $F(x) = \int_0^x f(t) dt / \int_0^1 f(t) dt$ satisfies the conditions: $F(x) = 0$ for $x < 0$, F is strictly increasing on $[0, 1]$, $F(x) = 1$ for $x \geq 1$.

(b) Given $a < b$ and $y_1, y_2 \in \mathbb{R}$, show that there exists a nonconstant $g: \mathbb{R} \rightarrow \mathbb{R}$ of class C^∞ such that $g(x) = y_1 \quad \forall x \in [-\infty, a]$, and $g(x) = y_2 \quad \forall x \in [b, +\infty[$.

18. Let $\alpha: [a, b] \rightarrow \mathbb{R}^2$ be a closed curve and $\mathcal{L}(\mathbf{v})$ be the corresponding “width function”. Show that:

(a) $\mathcal{L}$ is continuous and therefore there exists $\max_{\mathbf{v} \in S} \mathcal{L}(\mathbf{v})$;

(b) if α is regular and strictly convex then $\mathcal{L}$ is differentiable (i.e., the function $\theta \rightarrow \mathcal{L}(\cos \theta, \sin \theta)$ is differentiable).

19. Show that the function of Lemma 1.6.4 is unique but for the addition of a constant.

20. Let α be a regular convex curve of constant width $\mathfrak{L}$. Show that:

(a) $k(p) > 1/\mathfrak{L}$ for any point p of α (assume that α has positive orientation);

(b) if p and $\tilde{p}$ are antipodal points then

$$\frac{1}{k(p)} + \frac{1}{k(\tilde{p})} = \mathfrak{L};$$

(c) if each pair of antipodal points divide α into two arcs of equal length then α is a circle.

21. Convex curves of constant width $\mathfrak{L}$ are characterized by the fact that all the rectangles that circumscribe them are squares of side $\mathfrak{L}$. In this exercise we prove a generalization of Barbier’s Theorem: *if α is a regular curve strictly convex such that all rectangles that surround it have perimeter $4\mathfrak{L}$, then the perimeter of α is $\pi\mathfrak{L}$.*

Using the same notation as in the proof of 1.6.3, denote by L the perimeter of the curve and by $\tilde{\alpha}(s)$ the only point of α at which the tangent vector is $-\tau(s)$. Show that:

(a) there exists a differentiable function $f: \mathbb{R} \rightarrow \mathbb{R}$ such that $\tilde{\alpha}(s) = \alpha(f(s))$ and $f(s+L) = f(s) + L$;

(b) there exist differentiable functions λ, η such that

$$\tilde{\alpha}(s) = \alpha(s) + \lambda(s)\tau(s) + \eta(s)\mathbf{n}(s);$$

(c) λ and η are periodic functions of period L ;

(d) $f'(s) = -1 - \lambda'(s) + \eta(s)k(s)$;

(e) $2L = \int_0^L \eta(s)k(s) ds$.

Now make the change of variable $\theta = \varphi(s)$ to prove that

$$\int_0^L \eta(s)k(s) ds = \int_0^{2\pi} \eta(\theta) d\theta,$$

where $\eta(\theta)$ denotes $\eta(\varphi^{-1}(\theta))$. Note that $\eta(\theta)$ is periodic of period 2π and that the assumption about α translates to $\eta(\theta) + \eta\left(\theta + \frac{\pi}{2}\right) = 2\mathfrak{L}$. Finally, we can write

$$\begin{aligned} 4L &= \int_0^{2\pi} \eta(\theta) d\theta + \int_0^{2\pi} \eta\left(\theta + \frac{\pi}{2}\right) d\theta \\ &= \int_0^{2\pi} \left\{ \eta(\theta) + \eta\left(\theta + \frac{\pi}{2}\right) \right\} d\theta = 4\pi\mathfrak{L}. \end{aligned}$$

22. Modify 1.6.2 to give examples of curves that satisfy the hypothesis of Exercise 21 but do not have constant width.

1.7 Theorem of the Four Vertices

We now give the four-vertex theorem, which states that the curvature of any closed convex curve has at least four critical points (this result is also valid for nonconvex closed planar curves, but we will not prove it in such generality). This result is best possible: a non-circular ellipse has exactly four vertices, which are its points of intersection with the axes (see Exercise 6).

Let $\alpha: [a, b] \rightarrow \mathbb{R}^2$ be a regular closed curve, and $k(s)$ its curvature function. A *vertex* of α is a point $\alpha(s_0)$ such that $k'(s_0) = 0$. This definition does not depend on the parameterization and so we assume that s is the arc length.

Theorem 1.7.1 *Any closed convex curve has at least four vertices.*

Proof We can assume that $k(s)$ has a finite number of critical points, because otherwise there is nothing to be shown. The function $k(s)$ attains some maximum

and some minimum — which, changing the starting point if necessary, we suppose happen at the points $s = a$ and $s = s_0 \in]a, b[$. By applying a rotation or translation we can ensure that both points $\alpha(a)$ and $\alpha(s_0)$ are on the x -axis.

Let us check that there are no other points of α on the x -axis: for if there were another one — let it be p — then the tangent line to α at the one of the three points $\alpha(a)$, $\alpha(s_0)$ and p which lies between the other two is the x -axis; otherwise there would be points of α on opposite sides of this tangent line, in contradiction to the convexity of α . It follows that the tangent line at $\alpha(a)$ and $\alpha(s_0)$ is also the horizontal axis, and that (as in the proof of 1.5.1) the trace of $\alpha|_{[a, s_0]}$ is a line segment. This however contradicts our assumption that α has a finite number of vertices.

Putting $\alpha(s) = (x(s), y(s))$, we then have that $y(s)$ never vanishes on the intervals $]a, s_0[$ and $]s_0, b[$, taking on the second interval a sign opposite to the one it takes on the first; and the same is true of $k'(s)$ if we assume that α has at most two vertices. Under this assumption the function $k'(s)y(s)$ then has constant sign, vanishing only at a , s_0 and b , and so $\int_a^b k'(s)y(s) ds \neq 0$. But, writing $(x'(s), y'(s)) = (\cos \varphi(s), \sin \varphi(s))$, and using integration by parts and the equality $\varphi'(s) = k(s)$, we have

$$\begin{aligned} \int_a^b k'(s)y(s) ds &= k(s)y(s)\Big|_a^b - \int_a^b k(s)y'(s) ds = - \int_a^b k(s)y'(s) ds \\ &= - \int_a^b \varphi'(s) \sin \varphi(s) ds = \cos \varphi(s)\Big|_a^b = 0. \end{aligned}$$

This contradiction shows that $k'(s)$ changes sign on some intervals $]a, s_0[$ and $]s_0, b[$. Since in each of them $k'(s)$ has the same sign near the endpoints, we conclude that $k'(s)$ changes sign at least twice in such an interval, which proves the theorem. $\square$

The four-vertex theorem is still valid for non-convex curves. For a very elegant geometric proof that also covers this generalization, we suggest [21].

1.8 The Isoperimetric Inequality

The isoperimetric inequality states that, among all planar curves with a given perimeter, the circumference encompasses the largest area. The proof we give (by A. Hurwitz, 1902) makes essential use of the theory of Fourier series (see [11] for an introduction to this theory). An elementary proof appears in [18], and [9] contains a generalization of the isoperimetric inequality for convex bodies in dimensions greater than two.

Lemma 1.8.1 (Wirtinger) Let f be a function of class C^1 , periodic of period 2π , such that $\int_0^{2\pi} f(t) dt = 0$. Then

$$\int_0^{2\pi} f'(t)^2 dt \geq \int_0^{2\pi} f(t)^2 dt,$$

and equality holds if and only if there exist a and b such that $f(t) = a \cos t + b \sin t$.

Proof Be

$$f(t) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt)$$

the Fourier series expansion of f . Since $f'(t)$ is continuous, its expansion is obtained from that of $f(t)$ by term-by-term differentiation, thus

$$f'(t) \sim \sum_{n=1}^{\infty} (n b_n \cos nt - n a_n \sin nt).$$

Since $\int_0^{2\pi} f(t) dt = \pi a_0$, our hypothesis yields $a_0 = 0$. Using Parseval's formula, we have

$$\begin{aligned} \frac{1}{\pi} \int_0^{2\pi} f(t)^2 dt &= \sum_{n=1}^{\infty} (a_n^2 + b_n^2), \\ \frac{1}{\pi} \int_0^{2\pi} f'(t)^2 dt &= \sum_{n=1}^{\infty} n^2 (a_n^2 + b_n^2). \end{aligned}$$

It follows that

$$\int_0^{2\pi} f'(t)^2 dt - \int_0^{2\pi} f(t)^2 dt = \sum_{n=1}^{\infty} \pi(n^2 - 1)(a_n^2 + b_n^2) \geq 0,$$

and equality only holds if $a_n = b_n = 0$ for all $n > 1$. Since continuous functions are determined by their Fourier expansion, this is equivalent to $f(t) = a_1 \cos t + b_1 \sin t$. $\square$

Theorem 1.8.2 (Isoperimetric inequality) Let α be a simple regular closed curve of perimeter L , bounding a region Ω of area A . Then

$$A \leq \frac{L^2}{4\pi},$$

and equality holds only when α is a circle.

We can rescale the figure using a homothety, so there is no loss of generality if we suppose that $L = 2\pi$, and therefore $\alpha(s) = (x(s), y(s))$, $s \in [0, 2\pi]$. With a translation, we can achieve $\int_0^{2\pi} x(s) ds = 0$. Furthermore, we assume that $\alpha(s)$ runs through the boundary of Ω in the counterclockwise direction. Applying Green's theorem

$$\left(\int_{\partial\Omega} P dx + Q dy = \iint_{\Omega} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy \right)$$

to the vector field $(P, Q) = (0, x)$, we obtain

$$A = \int_0^{2\pi} x y' ds,$$

and on the other hand

$$L = 2\pi = \int_0^{2\pi} (x'^2 + y'^2) ds.$$

We can then write

$$2(\pi - A) = \int_0^{2\pi} (x'^2 - x^2) ds + \int_0^{2\pi} (x - y')^2 ds.$$

The second integral of this sum is non-negative and, by Wirtinger's lemma, so is the first. We thus conclude, as desired, that $A \leq \pi$. To achieve equality both integrals have to be zero, so $y'(s) = x(s)$ and $x(s) = a \cos s + b \sin s$. We thus have

$$\begin{aligned} x(s) &= a \cos s + b \sin s, \\ y(s) &= a \sin s - b \cos s + c, \end{aligned}$$

and α is therefore the unit circle with center at $(0, c)$. □

Exercises

23. The coefficients of the Fourier series

$$f(t) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt)$$

of a periodic function of period 2π , integrable on $[0, 2\pi]$ (a class that includes bounded functions with a finite number of discontinuities), are defined by

$$\begin{aligned} a_0 &= \frac{1}{\pi} \int_0^{2\pi} f(t) dt, \\ a_n &= \frac{1}{\pi} \int_0^{2\pi} f(t) \cos nt dt, \\ b_n &= \frac{1}{\pi} \int_0^{2\pi} f(t) \sin nt dt. \end{aligned}$$

Show that:

(a) if f is of class C^1 and $(a'_n)_{n=0}^{+\infty}$ and $(b'_n)_{n=1}^{+\infty}$ are the Fourier coefficients of f' , then $a'_0 = 0$ and $a'_n = n b_n$ and $b'_n = -n a_n$ for $n \geq 1$ (use integration by parts);

(b) the result of (a) is still true if f is only piecewise C^1 ;

(c) the isoperimetric inequality is valid for piecewise C^1 curves.

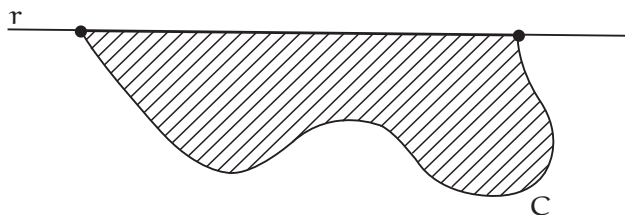


Figure 1.10

24. Consider a straight line r in the plane and a flexible string C of length L . By placing C in the plane so that its ends are on r , we obtain a figure bounded by r and by C and whose area depends on the shape we give the string (see figure above). Show that the figure of maximum area among all those so obtained is a semicircle based on r .

25. Given two points p and q in the plane and a flexible string C of length $L > |p - q|$, determine the figure of largest area among those bounded by C and by the line segment $\overline{pq}$.

26. Let α be a convex closed curve, piecewise C^1 , of perimeter L , bounding a region Ω of area A . Let r_1 and r_2 be two parallel lines at a distance d from each other such that both touch α , and α is contained in the strip bounded by the them. Consider an orthonormal Cartesian coordinate system whose vertical axis is r_1 and whose origin is the midpoint of the line segment $r_1 \cap \alpha$ (a line segment which may contain a single point). There thus exist functions piecewise C^1 $f, g: [0, d] \rightarrow \mathbb{R}$ such that:

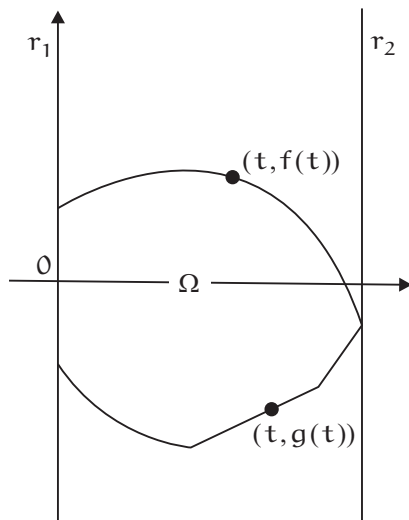


Figure 1.11

1. $f(t) \geq g(t) \quad \forall t \in [0, d], \quad f(0) = -g(0) \geq 0$;
2. the trace of α consists of the graphs of f and g and the vertical segments $\{0\} \times [g(0), f(0)]$ and $\{d\} \times [g(d), f(d)]$;
3. f is a concave function and g is a convex function.

(a) Define $h = \frac{1}{2}(f - g)$. Show that h is a concave function piecewise C^1 , and conclude that the region $\tilde{\Omega}$ bounded by r_1 and r_2 and by the graphs of the functions h and $-h$ is convex.

(b) Let $\tilde{L}$ and $\tilde{A}$ be the perimeter and area of $\tilde{\Omega}$. Show that $\tilde{A} = A$ and $\tilde{L} \leq L$, and that the inequality is an equality only in the case where $f = -g$.

(c) Assume that α has minimal perimeter among all convex piecewise C^1 curves encompassing a fixed area A . Conclude, without using the isoperimetric inequality, that α is a circle.

Hint: α has an axis of symmetry in each direction. Show that all these axes pass through the same point.

27. Let α be a regular convex curve of constant width $\mathfrak{L}$ that bounds a region of area

A. Show that $A \leq \frac{\pi \mathfrak{L}^2}{4}$, with equality only if α is a circle.



Chapter 2

Regular Surfaces

In this chapter we introduce regular surfaces, the object of all our further study, defining them as those subsets of $\mathbb{R}^3$ that can be described locally by two independent parameters. We introduce notions such as tangent space, differentiable function and diffeomorphism, and consider the problems of orientability and the measurement of quantities (areas, lengths . . .) on surfaces.

2.1 Definition and Examples

We all have an intuitive notion of what a surface is, and any attempt to describe that notion would inevitably fall into redundancy. We accept, however, that the plane is the simplest surface of all, and that a good way to construct models of others is by gluing together various pieces of paper. Our definition of surface is the mathematical elaboration of this idea.

A subset S of $\mathbb{R}^3$ is called a *regular surface* if, for each $p \in S$, there exist an open neighborhood $V \subseteq \mathbb{R}^3$ of p , an open subset $U \subseteq \mathbb{R}^2$, and a bijection $\Phi: U \rightarrow V \cap S$ with the following properties:

- i. Φ is of class C^∞ ;
- ii. Φ is a homeomorphism (i.e., its inverse $\Phi^{-1}: V \cap S \rightarrow U$ is continuous);
- iii. for all $q \in U$ the Jacobian matrix $J\Phi(q)$ has rank two.

A mapping Φ with these three properties is named *parameterization* or *system of (local) coordinates* of S . We usually denote the points of U by (u, v) , so that u and v are *local parameters* of S , and the partial derivatives of Φ are denoted by Φ_u and Φ_v . These vectors describe the velocities of the *coordinate curves*, which are the curves obtained by fixing one of the parameters and varying the other. Moreover the columns of the matrix $J\Phi(u, v)$ are precisely Φ_u and Φ_v , so that condition iii. above expresses that, for each $(u, v) \in U$, Φ_u and Φ_v are linearly independent.

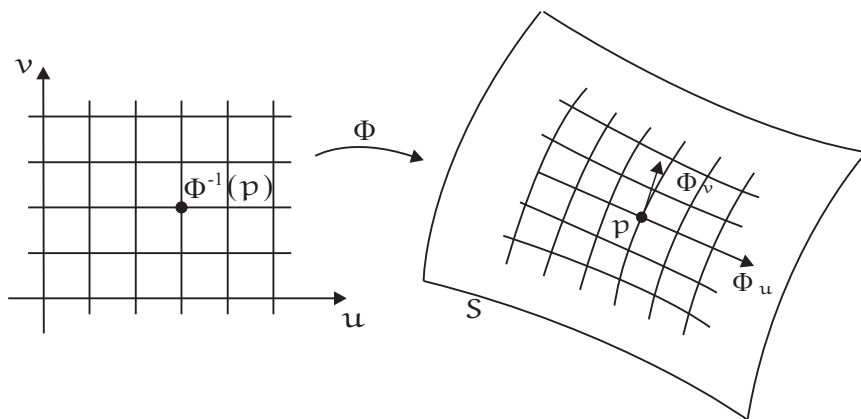


Figure 2.1

A subset of the surface S is called *open* if it is of the form $V \cap S$, where V is an open subset of $\mathbb{R}^3$. An important observation is that any point of S has an open neighborhood (in S) homeomorphic to a disk, and this neighborhood can be taken as small as we wish: indeed, if (U, Φ) is a parameterization in the neighborhood of p , there exists some open disk D (of arbitrarily small radius) containing $\Phi^{-1}(p)$ and contained in U , and $\Phi(D)$ is the sought neighborhood.

Examples 2.1.1 A. Any plane Π in $\mathbb{R}^3$ is a surface: in fact Π admits a description of the form $\phi(u, v) = p_0 + u\mathbf{w}_1 + v\mathbf{w}_2$, where $(u, v) \in \mathbb{R}^2$, $p_0 \in \Pi$ and $\mathbf{w}_1, \mathbf{w}_2$ are linearly independent vectors. The conditions i. and iii. are trivially verified, and ii. follows from the fact that the solution of the equations $\phi(u, v) = p$, for $p = (a, b, c) \in \Pi$, is a first degree function at a, b, c , hence continuous. This means that Π is all covered by a single (called *global*) parameterization.

B. If $f: U \rightarrow \mathbb{R}^2$ is a differentiable function defined on an open subset $\mathbb{R}^2$, its graph $\{(u, v, f(u, v)): (u, v) \in U\} \subseteq \mathbb{R}^3$ is a surface admitting the global parameterization $\Phi(u, v) = (u, v, f(u, v))$, $(u, v) \in U$.

C. The union $S = \Pi_1 \cup \Pi_2$ of two non-parallel planes is not a surface, since the points of $\Pi_1 \cap \Pi_2$ have on S no neighborhood homeomorphic to a disk.

D. A parameterization of the sphere $\mathbf{S}^2 = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 = 1\}$ covering the northern hemisphere is

$$\Phi(u, v) = \left(u, v, \sqrt{1 - (u^2 + v^2)} \right),$$

defined on the disk $\{(u, v): u^2 + v^2 < 1\}$. With a few more analogous parameterizations (how many are needed?) we can cover the whole sphere, which therefore is a surface.

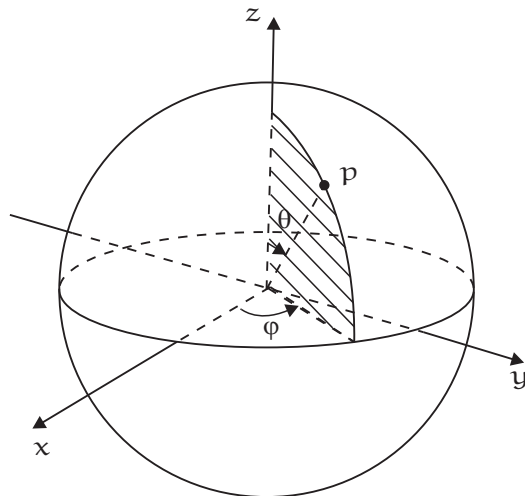


Figure 2.2

Another type of parameterization of S^2 , which excludes only one meridian, is given by the spherical coordinates, which are the colatitude $\theta \in]0, \pi[$ and the longitude $\varphi \in]-\pi, \pi[$ (Fig. 2.2), defining the point

$$\Psi(\varphi, \theta) = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta).$$

The tangent vectors to the coordinate curves are

$$\Psi_\varphi(\varphi, \theta) = (-\sin \theta \sin \varphi, \sin \theta \cos \varphi, 0),$$

$$\Psi_\theta(\varphi, \theta) = (\cos \theta \cos \varphi, \cos \theta \sin \varphi, -\sin \theta),$$

whose vector product

$$\Psi_\varphi \times \Psi_\theta = -\sin \theta (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$$

has length $\sin \theta$, and so is nonzero: therefore Ψ_φ and Ψ_θ are linearly independent.

We mention that some authors use, in spherical coordinates, the latitude $\tilde{\theta} = \frac{\pi}{2} - \theta$ ($\tilde{\theta} \in]-\frac{\pi}{2}, \frac{\pi}{2}[$) instead of the colatitude, thus obtaining

$$\tilde{\Psi}(\varphi, \tilde{\theta}) = (\cos \tilde{\theta} \cos \varphi, \cos \tilde{\theta} \sin \varphi, \sin \tilde{\theta}).$$

E. The sphere is a special case of a surface of revolution, which is obtained by rotating a planar curve around an axis contained in the plane of the curve. Assuming that the curve $\alpha(v) = (\rho(v), 0, z(v))$ is defined on an open interval I , is a homeomorphism onto its image, and that $\rho(v) > 0$ for all $v \in I$, the mapping

$$\Phi(u, v) = (\rho(v) \cos u, \rho(v) \sin u, z(v)),$$

where $(u, v) \in]-\pi, \pi[\times I$, is a parameterization of the surface obtained by rotating α around the z -axis. To show that Φ^{-1} is continuous, we make use of the formula

$$\operatorname{tg} \frac{u}{2} = \frac{\sin u}{1 + \cos u}$$

and of the fact that $f(u) = \operatorname{tg} \frac{u}{2}$ is a diffeomorphism of $]-\pi, \pi[$ and $\mathbb{R}$. The equality $\Phi(u, v) = (x, y, z)$ is then equivalent to the combination of the two equalities

$$u = f^{-1} \left(\frac{y}{x + \sqrt{x^2 + y^2}} \right), \quad v = \alpha^{-1}(\sqrt{x^2 + y^2}, O, z),$$

which proves that Φ^{-1} is continuous.

F. By analogy with the definition of a parametrized curve, we define *parametrized surface* as a differentiable mapping Φ of a connected open subset of $\mathbb{R}^2$ into $\mathbb{R}^3$ whose Jacobian has rank two at all points. This definition does not require that Φ be injective, and so its image, being allowed to have self-intersections, is not necessarily a regular surface; but it may not be so even if Φ is injective, as the following example shows:

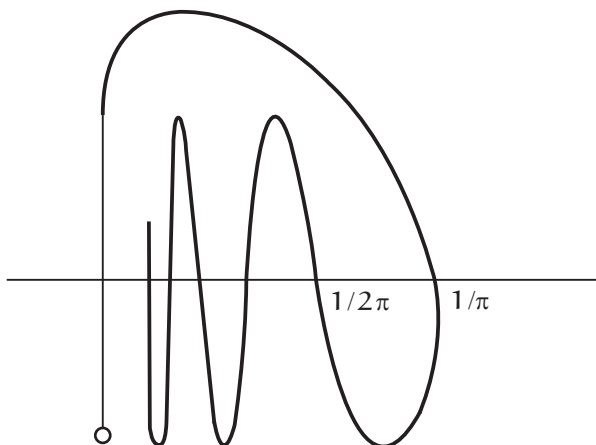


Figure 2.3

Let $\alpha(u) = (x(u), y(u))$, $u > 0$, be a simple regular curve of class C^∞ that includes the vertical line segment $\{0\} \times]1, 1]$ and the graph of the function $\sin \frac{1}{x}$, $x \in]0, \frac{1}{\pi}]$; this curve accumulates in the neighborhood of $\{0\} \times]-1, 1[$ (see Fig. 2.3). The trace S of the parametrized surface $\Phi(u, v) = (x(u), y(u), v)$ is **not** a surface: if it were, each $p \in S$ would have arbitrarily small neighborhoods V in $\mathbb{R}^3$ with $V \cap S$ homeomorphic to disks; but that does not happen if $p \in \{0\} \times]-1, 1[\times \mathbb{R}$, because $V \cap S$ has infinitely many connected components for all sufficiently small V .

This means that not all parametrized surfaces define regular surfaces. But it is also not easy, and in some cases not even possible, to describe regular surfaces as

parametrized surfaces (i.e., as the image of **one single** function Φ). There is in general no reason to privilege a particular (even global) parameterization on a given surface.

In conclusion: surface for us means regular surface, and only in the exercises we will mention parametrized surfaces.

Exercises

28. For each $a \in \mathbb{R}$, the polar coordinates $\Phi(\rho, \varphi) = (\rho \cos \varphi, \rho \sin \varphi)$, with $\rho > 0$ and $\varphi \in]a - \pi, a + \pi[$, define a parameterization of $\mathbb{R}^2$ that excludes a half-line. (We consider $\mathbb{R}^2$ as a surface by identifying it with the plane $\mathbb{R}^2 \times \{0\} \subseteq \mathbb{R}^3$.)

29. Consider a helix parametrized by $(\cos t, \sin t, t)$ ($t \in \mathbb{R}$). The *helicoid* is the set formed by all (horizontal) lines connecting each point of the z -axis with the point of the helix at the same height (see Fig. 2.4). Show that the helicoid is a regular surface.

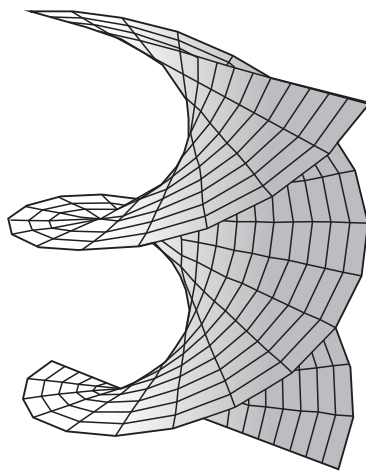


Figure 2.4

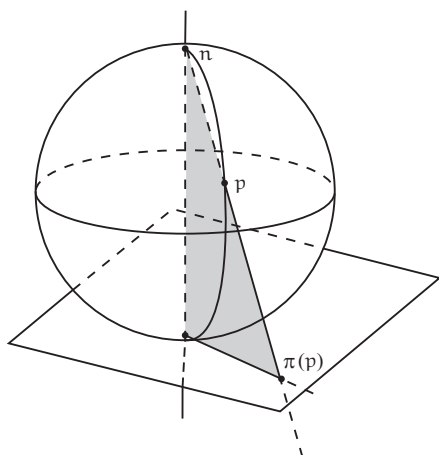


Figure 2.5

30. Consider the sphere $\mathbf{S}^2 = \{(x, y, z): x^2 + y^2 + z^2 = 1\}$. The *stereographic projection* is the mapping $\pi: \mathbf{S}^2 \setminus \{(0, 0, 1)\} \rightarrow \mathbb{R}^2$ defined as follows: $(\pi(p), -1)$ is the intersection point of the plane $z = -1$ with the line that contains the points $(0, 0, 1)$ and p (see Fig. 2.5).

(a) Obtain an explicit formula for π . Show that π is a bijection and that $(\mathbb{R}^2, \pi^{-1})$ is a parameterization of $\mathbf{S}^2$.

(b) Conclude that there are two parameterizations of $\mathbf{S}^2$ whose union covers the sphere.

(c) Is there any global parameterization of $\mathbf{S}^2$ (i.e., whose image is $\mathbf{S}^2$)?

31. Consider in $\mathbb{R}^3$ the circle $C = \{(x, y, z): (y - 2)^2 + z^2 = 1, x = 0\}$. Show that the set $\mathbb{T}^2$ which is obtained by rotating C around the z -axis is a regular surface (the *torus*) — a parameterization is given by $\Phi(u, v) = ((2 + \cos v) \cos u, (2 + \cos v) \sin u, \sin v)$, where $(u, v) \in]-\pi, \pi[\times]-\pi, \pi[$ (see Fig. 2.6).

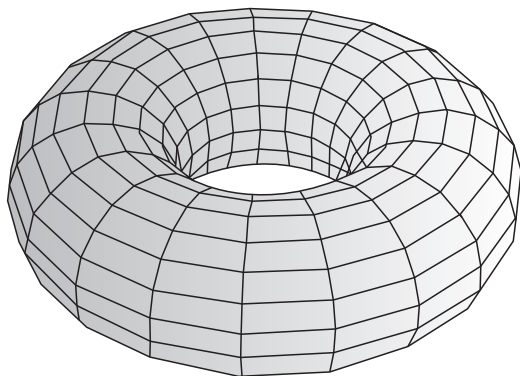


Figure 2.6

32. The *tractrix* is the planar curve obtained as follows: let us fix a line (let it be the z -axis); the distance from any point p on the curve to the point of intersection p' of the tangent line to the curve at p with the fixed line is constant, equal to $C > 0$. Parametrize the tractrix using the angle t in Fig. 2.7 as a parameter (note that $t \in]\frac{\pi}{2}, \pi[$). [The surface of revolution obtained from the tractrix around the z -axis is the *pseudosphere*].

2.2 Change of Parameters, Level Surfaces

In this section we gather a number of technical results, such as the change of parameters on a surface, level surfaces, and the fact that any surface is locally the graph of a function. We make systematic use of the inverse function theorem, and the proofs are largely routine — so it seems more instructive (and less monotonous) if, instead of reading all these proofs, the reader tries to reconstruct some of them by herself.

In our further study we will make use of local coordinates to express certain concepts, and our first caution is that such concepts should not depend on the coordinate system used, but only on the surface. Assume then that (U, Φ) and $(\tilde{U}, \Psi)$ are two parameterizations of the surface S , and that the open set $W = \Phi(U) \cap \Psi(\tilde{U})$ is non-empty. Under these assumptions (see Fig. 2.8) we have the following result.

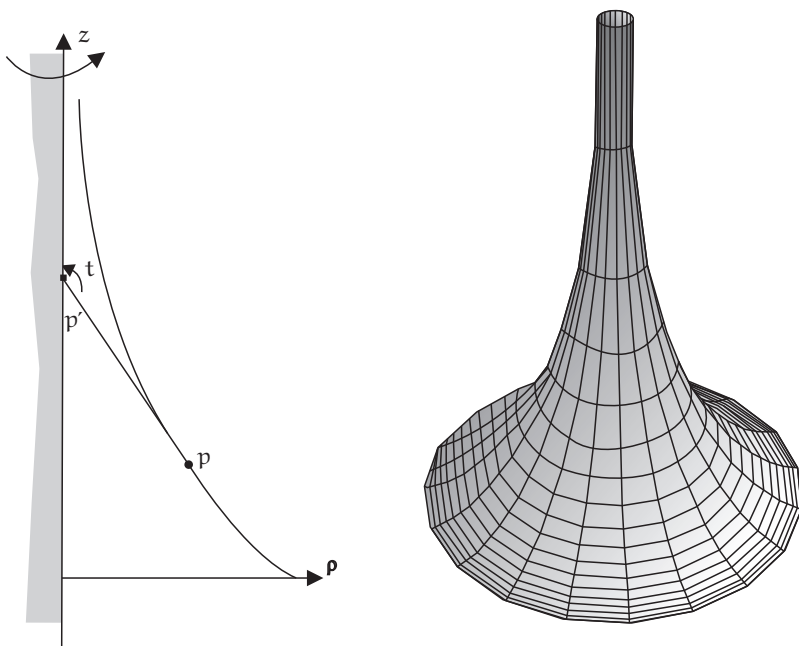


Figure 2.7

Proposition 2.2.1 *The coordinate change $\Phi^{-1} \circ \Psi: \Phi^{-1}(W) \rightarrow \Psi^{-1}(W)$ is a diffeomorphism.*

Proof It suffices to show that $\Phi^{-1} \circ \Psi$ is differentiable, because the same argument proves the differentiability of its inverse $\Psi^{-1} \circ \Phi$. We write

$$\Phi(u, v) = (x(u, v), y(u, v), z(u, v)),$$

$$\Psi(\tilde{u}, \tilde{v}) = (x(\tilde{u}, \tilde{v}), y(\tilde{u}, \tilde{v}), z(\tilde{u}, \tilde{v}));$$

and, given $(\tilde{u}_0, \tilde{v}_0) \in \Psi^{-1}(W)$, let us show that $\Phi^{-1} \circ \Psi$ is differentiable at $(\tilde{u}_0, \tilde{v}_0)$.

Putting $(u_0, v_0) = \Phi^{-1} \circ \Psi(\tilde{u}_0, \tilde{v}_0)$, some 2×2 submatrix of $J\Phi(u_0, v_0)$ has nonzero determinant, and we assume that it is the one formed by the first two rows (whose determinant is usually denoted by $\frac{\partial(x, y)}{\partial(u, v)}$). By the inverse mapping theorem,

there exists some open neighborhood $D \subseteq \Phi^{-1}(W)$ of (u_0, v_0) such that the restriction of $f(u, v) = (x(u, v), y(u, v))$ to D is a diffeomorphism onto the image. It follows that $\Phi^{-1} \circ \Psi|_{\Psi^{-1} \circ \Phi(D)}$ is differentiable because it is a composition of differentiable mappings:

$$(\tilde{u}, \tilde{v}) \mapsto \Psi(\tilde{u}, \tilde{v}) = (x, y, z) \mapsto (x, y) \xrightarrow{f^{-1}} (u, v) = \Phi^{-1} \circ \Psi(\tilde{u}, \tilde{v}). \square$$

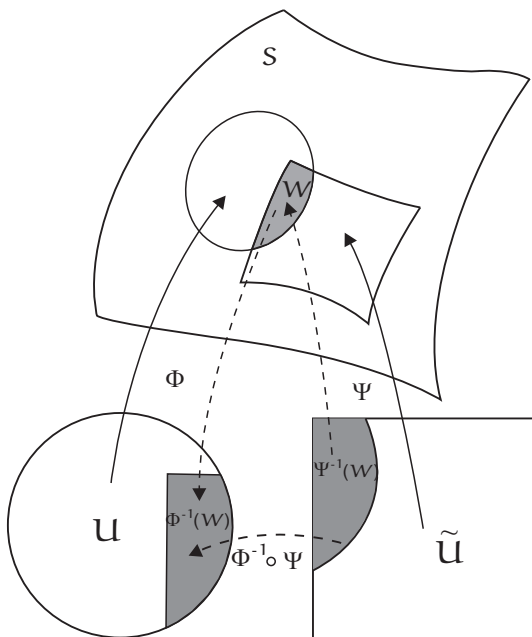


Figure 2.8

Let us point out that the foregoing proof establishes a more general fact than 2.2.1: if α is a differentiable function defined on an open subset of $\mathbb{R}^n$ (a curve, for example)

whose image is contained in a surface S , then, for any parameterization Φ of S , $\Phi^{-1} \circ \alpha$ is differentiable at all points where it is defined.

In the previous section (example D) we noted that the graph of any differentiable function is a surface; the next proposition says that locally this example is as general as possible.

Proposition 2.2.2 Any point p of a regular surface S has in S a neighborhood W of one of the following three forms:

$$W = \{(x, y, h(x, y)) : (x, y) \in R\},$$

$$W = \{(x, h(x, z), z) : (x, z) \in R\},$$

$$W = \{(h(y, z), y, z) : (y, z) \in R\}$$

— with, in all three cases, R is an open subset of $\mathbb{R}^2$ and h is a differentiable function.

Proof Let (U, Φ) be a parameterization in the neighborhood of p . One of the three determinants

$$\frac{\partial(x, y)}{\partial(u, v)}, \quad \frac{\partial(x, z)}{\partial(u, v)} \quad \text{and} \quad \frac{\partial(y, z)}{\partial(u, v)},$$

say the first one, is nonzero when computed at $\Phi^{-1}(p)$. The inverse mapping theorem then guarantees that there exists an open neighborhood $D \subseteq U$ of $\Phi^{-1}(p)$ restricted to which $f(u, v) = (x(u, v), y(u, v))$ is a diffeomorphism onto the image. Now $W = \Phi(D)$ is the sought neighborhood, since $R = f(D)$ is an open subset of $\mathbb{R}^2$, $h(x, y) = z \circ f^{-1}(x, y)$ is differentiable, and $W = \Phi \circ f^{-1}(R) = \{(x, y, h(x, y)) : (x, y) \in R\}$. $\square$

Example 2.2.3 Proposition 2.2.2 gives us a criterion to show that certain sets are not surfaces, which we illustrate with the cone

$$\mathcal{C} = \{(x, y, z) \in \mathbb{R}^3 : z = \sqrt{x^2 + y^2}\}.$$

If $\mathcal{C}$ is a surface, there exists an open subset V of $\mathbb{R}^3$ containing the point $(0, 0, 0)$ such that $V \cap \mathcal{C}$ is the graph of a differentiable function h . But h can only be a function of (x, y) , because none of the projections of $\mathcal{C}$ on the other coordinate planes contains a neighborhood of $(0, 0)$. Thus $h(x, y) = \sqrt{x^2 + y^2}$ and this function is not differentiable at $(0, 0)$. Therefore, $\mathcal{C}$ is not a surface. [But the origin is the only problematic point: $\mathcal{C} \setminus \{(0, 0, 0)\}$ is a surface]. $\square$

A ready way to define a surface is by an equation of the form $f(x, y, z) = a$, where $f: V \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}$ is a differentiable function. Not always, of course, does such an equation define a surface: we have to impose on f a certain degree of non-degeneracy, which we are going to describe.

A point $p \in V$ is called *regular* (for the function f) if the gradient vector

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right),$$

computed at p , is a nonzero vector; and $a \in \mathbb{R}$ is a *regular value* of f if $f^{-1}(\{a\})$ is non-empty and contains only regular points.

The condition $\frac{\partial f}{\partial z}(p) \neq 0$ guarantees that in a neighborhood U of $p \in f^{-1}(\{a\})$, f is strictly monotone along vertical segments, and therefore each of these segments intersects $f^{-1}(\{a\})$ at most one point. In fact, the proof of the next proposition consists *essentially* in showing that under these circumstances $f^{-1}(\{a\}) \cap U$ is the graph of a differentiable function of (x, y) .

Proposition 2.2.4 *If a is a regular value of $f: V \rightarrow \mathbb{R}$ then $f^{-1}(\{a\})$ is a regular surface.*

Proof Assuming that $\frac{\partial f}{\partial z}(p) \neq 0$, we define

$$F(x, y, z) = (x, y, f(x, y, z)).$$

Since $\det JF(p) = \frac{\partial f}{\partial z}(p) \neq 0$, the function F is invertible in a neighborhood of p : so there exist open subsets $U, W \subseteq \mathbb{R}^3$ such that $p \in U \subseteq V$, and F sends U diffeomorphically onto W . Let us now note that the inverse $G: W \rightarrow U$ of $F|_U$ has the form $G(x, y, z) = (x, y, g(x, y, z))$ and that, for $(x, y, z) \in U$, all the following equalities are equivalent:

$$\begin{aligned} f(x, y, z) &= a, \\ F(x, y, z) &= (x, y, a), \\ (x, y, z) &= G(x, y, a), \\ z &= g(x, y, a). \end{aligned}$$

The equivalence between the first and last subset of these equalities shows that $U \cap f^{-1}(\{a\})$ is the graph of the differentiable function $h(x, y) = g(x, y, a)$, whose domain is the open subset $R = \{(x, y) \in \mathbb{R}^2: (x, y, a) \in W\}$ — and this concludes the proof that $f^{-1}(\{a\})$ is a surface. $\square$

The sets $f^{-1}(\{a\})$ are the *level sets* of f and, when a is a regular value, they are also called *level surfaces*. Taking for example

$$f(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2},$$

we see that the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

is a regular surface, since $\nabla f(x, y, z)$ is nonzero for all $(x, y, z) \neq (0, 0, 0)$. In general, any non-degenerate quadric in $\mathbb{R}^3$ is a regular surface, since, for an appropriate orthonormal basis, it has equation $x + \varepsilon_2 y^2 + \varepsilon_3 z^2 = 0$ if it is a paraboloid, or $\varepsilon_1 x^2 + \varepsilon_2 y^2 + \varepsilon_3 z^2 = 1$, with $(\varepsilon_1, \varepsilon_2, \varepsilon_3) \neq (0, 0, 0)$, if it is a hyperboloid or an ellipsoid.

An important caveat is that *the condition that a be a regular value is by no means necessary for $f^{-1}(\{a\})$ to be a surface*. A simple example is given by $f(x, y, z) = x^2$: the set $f^{-1}(\{0\})$ is a surface, even though it consists only of singular points of f .

Note 2.2.5 A level surface is not necessarily connected, as shown by the two-leaf hyperboloid defined by the equation $z^2 - x^2 - y^2 = 1$. For the benefit of the reader unfamiliar with the concept, we give here the definition of connectedness and a brief discussion: a set $A \subseteq \mathbb{R}^n$ is called *connected* if it cannot be *split*, i.e., if **there are no** disjoint open subsets V and W of $\mathbb{R}^n$ such that $A \cap V \neq \emptyset \neq A \cap W$ and $A \subseteq V \cup W$. The hyperboloid above, for example, is not connected because it admits the splitting $V = \{(x, y, z) \in \mathbb{R}^3 : z > 0\}$ and $W = \{(x, y, z) \in \mathbb{R}^3 : z < 0\}$. The connected subsets of $\mathbb{R}$ are the intervals; the balls in $\mathbb{R}^n$ (disks in $\mathbb{R}^2$) are connected. Connectedness is a topological property, in the sense that the image of a connected set under a continuous function is still connected.

To put it suggestively, a surface is connected when it is made up of a single chunk. A useful connectivity criterion for surfaces (and for open subsets of $\mathbb{R}^n$) is the following: *S is connected if and only if, for every p and q on S , there exists a piecewise differentiable curve $\alpha: [a, b] \rightarrow S$ such that $\alpha(a) = p$ and $\alpha(b) = q$* . **[Proof:** if V and W split S , then there is no curve in S that joins $p \in S \cap V$ to $q \in S \cap W$, because the trace of a curve, being a continuous image of an interval, is connected. On the other hand, if S is connected and $p \in S$, consider the set $R = \{q \in S : \text{there exists a curve in } S \text{ from } p \text{ to } q\}$. Given $q \in S$, let (D, Φ) be a parameterization in the neighborhood of q , where $D \subseteq \mathbb{R}^2$ is an open disk. Any $r \in \Phi(D)$ can be joined to q by a curve in S : the image under Φ of the line segment $[\Phi^{-1}(q), \Phi^{-1}(r)]$. Thus, if $q \in R$ (resp. $q \in S \setminus R$) then $\Phi(D) \subseteq R$ (resp. $\Phi(D) \subseteq S \setminus R$). Therefore R and $S \setminus R$ are open subsets of S and, since S is connected, one of them, necessarily $S \setminus R$, is empty. Therefore $S = R$, which proves what we wanted.] $\square$

Now that we have a method for establishing that a set is a surface without using any parameterization, our next proposition states that if S is a surface, anything that appears to be a parameterization of S is indeed so.

Proposition 2.2.6 *Let S be a surface, U be an open subset of $\mathbb{R}^2$, and $\Phi: U \rightarrow S$ be a differentiable mapping. If Φ is injective and $J\Phi(u, v)$ has rank two for all (u, v) on U , then Φ is a parameterization of S .*

Proof One just has to check the continuity of the inverse $\Phi^{-1}: \Phi(U) \rightarrow U$. Given $(u_0, v_0) \in U$, the point $\Phi(u_0, v_0)$ has, by 2.2.2, an open neighborhood V in $\mathbb{R}^3$ such that $V \cap S$ is the graph of a function that we assume to depend on (x, y) . Thus, $V \cap S = \{(x, y, h(x, y)) : (x, y) \in R\}$, where R is an open subset of $\mathbb{R}^2$; and, taking an open disk $D \subseteq U$ centered at (u_0, v_0) and such that $\Phi(D) \subseteq V$, the restriction of Φ to D can be written as $\Phi(u, v) = (x(u, v), y(u, v), h(x(u, v), y(u, v)))$. We then have

$$\begin{aligned}
\Phi_u &= \frac{\partial x}{\partial u} \left(1, 0, \frac{\partial h}{\partial x}\right) + \frac{\partial y}{\partial u} \left(0, 1, \frac{\partial h}{\partial y}\right), \\
\Phi_v &= \frac{\partial x}{\partial v} \left(1, 0, \frac{\partial h}{\partial x}\right) + \frac{\partial y}{\partial v} \left(0, 1, \frac{\partial h}{\partial y}\right), \\
\Phi_u \times \Phi_v &= \left\{ \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \right\} \left(1, 0, \frac{\partial h}{\partial x}\right) \times \left(0, 1, \frac{\partial h}{\partial y}\right) \\
&= \frac{\partial(x, y)}{\partial(u, v)} \left(-\frac{\partial h}{\partial x}, -\frac{\partial h}{\partial y}, 1\right)
\end{aligned}$$

– and from this, since $\Phi_u \times \Phi_v$ is nonzero, it follows that $\frac{\partial(x, y)}{\partial(u, v)} \neq 0$. We can therefore assume, shrinking D if necessary, that $\pi \circ \Phi|_D$ [where $\pi: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ is the projection on the first two coordinates] is a diffeomorphism onto its image, which is then an open subset E of $\mathbb{R}^2$. Thus, $\Phi(D) = \{(x, y, h(x, y)): (x, y) \in E\}$ is an open neighborhood of $\Phi(u_0, v_0)$ in S , and the restriction $\Phi^{-1}|_{\Phi(D)}$ is continuous, because it is given by the composite $(\pi \circ \Phi|_D)^{-1} \circ \pi$ of continuous functions. Thus Φ^{-1} is continuous on $\Phi(u_0, v_0)$. $\square$

Exercises

33. Show that any surface is locally a level surface. Given $p \in S$, there exist an open neighborhood V of p in $\mathbb{R}^3$ and a differentiable function $f: V \rightarrow \mathbb{R}$ such that $S \cap V = f^{-1}(\{0\})$ and 0 is a regular value of f .

34. Show that if two surfaces S_1 and S_2 intersect transversely at p then there exists an open neighborhood V of p (in $\mathbb{R}^3$) such that $S_1 \cap S_2 \cap V$ is the trace of a regular curve. (We say that S_1 and S_2 intersect transversely at p if $T_p S_1 \neq T_p S_2$.)

2.3 Differentiable Functions on Surfaces, Tangent Space

The results of the previous section prepared the setting to do Differential Calculus on surfaces; and we can now, in this section, explain what is a differentiable function in such a context. The derivatives of such functions are defined not on the surface but on its tangent spaces, a concept that we also introduce here.

Let S_1 and S_2 be two surfaces. A mapping $f: S_1 \rightarrow S_2$ is called *differentiable* if its expression in local coordinates is differentiable: more precisely, if there exist, for each $p \in S_1$, parameterizations (U, Φ) of S_1 and (V, Ψ) of S_2 in the neighborhoods of p and $f(p)$, respectively, such that $\Psi^{-1} \circ f \circ \Phi$ is differentiable. Similarly, a function $f: S_1 \rightarrow \mathbb{R}$ is called *differentiable* if every point of S_1 has a parametrized neighborhood (U, Φ) such that $f \circ \Phi$ is differentiable. A *diffeomorphism* is a differentiable bijection $f: S_1 \rightarrow S_2$ whose inverse is also differentiable.

Observations and Examples 2.3.1 A. Proposition 2.2.1 guarantees that if $f: S_1 \rightarrow S_2$ is differentiable, then, for all parameterizations Φ and Ψ , the mapping $\Psi^{-1} \circ f \circ \Phi$ is differentiable. This means that our definition does not depend on any choice of parameterization.

B. If $f: S_1 \rightarrow S_2$ is differentiable and (U, Φ) is a parameterization of S_1 then $f \circ \Phi: U \rightarrow \mathbb{R}^3$ is differentiable, because locally we can write $f \circ \Phi = \Psi \circ (\Psi^{-1} \circ f \circ \Phi)$, using appropriate local coordinates (V, Ψ) in S_2 . But the converse is also true: if $f \circ \Phi$ is differentiable for any parameterization (U, Φ) of S_1 then $f: S_1 \rightarrow S_2$ is differentiable. This is a consequence of the observation we make following the proof of 2.2.1.

C. Let $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be a differentiable mapping such that $f(S_1) \subseteq S_2$. Then $f|_{S_1}: S_1 \rightarrow S_2$ is differentiable, and an analogous observation can be made for functions $g: \mathbb{R}^3 \rightarrow \mathbb{R}$. As examples, we have $g_1(p) = \langle p, \mathbf{v} \rangle$, where $\mathbf{v}$ is a unit vector of $\mathbb{R}^3$ [g_1 measures the “oriented height”, in the direction of $\mathbf{v}$, of p relative to the origin $(0, 0, 0)$], and $g_2(p) = |p - p_0|^2$, which measures the square of the distance from p to a fixed point p_0 . Both these functions, restricted to any surface S , are differentiable.

D. Let S be a surface of revolution around an axis r and let R_θ be the rotation of angle θ around r . The mapping $R_\theta|_S: S \rightarrow S$ is a diffeomorphism: its inverse is $R_{-\theta}|_S$. For a more interesting example, consider the torus of revolution $\mathbb{T}^2$ parametrized by $\Phi(u, v) = ((2 + \cos v) \cos u, (2 + \cos v) \sin u, \sin v)$. The restriction of Φ to any square of the form $]a - \pi, a + \pi[\times]b - \pi, b + \pi[$ is injective, and is therefore a parameterization of $\mathbb{T}^2$. Fixing $(u_0, v_0) \in \mathbb{R}^2$, we define a mapping $f: \mathbb{T}^2 \rightarrow \mathbb{T}^2$ by the condition:

- if $p = \Phi(u, v)$ then $f(p) = \Phi(u + u_0, v + v_0)$.

Let us show that f is differentiable: in fact, given $q \in \mathbb{T}$, $q = \Phi(u_1, v_1)$, the mappings

$$\begin{aligned}\tilde{\Phi} &= \Phi|_{]u_1 - \pi, u_1 + \pi[\times]v_1 - \pi, v_1 + \pi[} \\ \tilde{\Psi} &= \Phi|_{]u_1 + u_0 - \pi, u_1 + u_0 + \pi[\times]v_1 + v_0 - \pi, v_1 + v_0 + \pi[}\end{aligned}$$

are parameterizations of $\mathbb{T}^2$ in the neighborhoods of q and $f(q)$, respectively; and $\tilde{\Psi}^{-1} \circ f \circ \tilde{\Phi}(u, v) = (u + u_0, v + v_0)$ is obviously differentiable, which proves that f is differentiable in a neighborhood of q . The same argument proves that the inverse is differentiable, and therefore f is a diffeomorphism. It follows in particular that, given any two points $p, q \in \mathbb{T}^2$, there exists some diffeomorphism $f: \mathbb{T}^2 \rightarrow \mathbb{T}^2$ such that $f(p) = q$. (See also Exercise 40 in this section.) $\square$

We now deal with the *tangent space* to a surface S at a point p , which we denote by $T_p S$. We define $T_p S$ to be the set of velocity vectors, at the point p , of the curves whose graph is in S :

$$T_p S = \{\alpha'(0) \mid \alpha:]-\varepsilon, \varepsilon[\rightarrow S \text{ is } C^\infty \text{ and } \alpha(0) = p\}.$$

Our next proposition shows that $T_p S$ is a vector subspace of $\mathbb{R}^3$ of dimension two, which justifies calling $p + T_p S$ the *tangent plane* to S at p .

Proposition 2.3.2 *If (U, Φ) is a parameterization of S in the neighborhood of p then $T_p S = D\Phi_{\Phi^{-1}(p)}(\mathbb{R}^2)$.*

Proof Let us prove the inclusion $T_p S \subseteq D\Phi_{\Phi^{-1}(p)}(\mathbb{R}^2)$. Given $\alpha'(0) \in T_p S$, we can assume that $\alpha(]-\varepsilon, \varepsilon[) \subseteq \Phi(U)$. Then the curve $\beta = \Phi^{-1} \circ \alpha:]-\varepsilon, \varepsilon[\rightarrow U$ is differentiable; writing $\alpha = \Phi \circ \beta$, we obtain, by the chain rule, $\alpha'(0) = D\Phi_{\Phi^{-1}(p)}(\beta'(0))$.

Let us treat the opposite inclusion. Given a vector $\mathbf{w} \in \mathbb{R}^2$, let us take $\varepsilon > 0$ so that the line segment $\beta(t) = \Phi^{-1}(p) + t\mathbf{w}$, $|t| < \varepsilon$, is contained in U . Putting $\alpha = \Phi \circ \beta$, we have $D\Phi_{\Phi^{-1}(p)}(\mathbf{w}) = D\Phi_{\Phi^{-1}(p)}(\beta'(0)) = \alpha'(0) \in T_p S$. $\square$

In practice, what we did was to write the curve α in local coordinates: if $\alpha(t) = \Phi(u(t), v(t))$ then α is differentiable if and only if so are both functions $u(t)$ and $v(t)$; and the chain rule provides the equality $\alpha'(t) = u'(t)\Phi_u + v'(t)\Phi_v$, which shows that at each point of $\Phi(U)$ the tangent space is generated by the vectors Φ_u and Φ_v .

Example 2.3.3 The tangent space to the level surface $S_a = f^{-1}(\{a\})$ at point p is the orthogonal complement of the line generated by $\nabla f(p)$. In fact, if $\alpha:]-\varepsilon, \varepsilon[\rightarrow S_a$ satisfies $\alpha(0) = p$ then $f \circ \alpha(t) = a$ for all $t \in]-\varepsilon, \varepsilon[$, so that $\langle \nabla f(p), \alpha'(0) \rangle = (f \circ \alpha)'(0) = 0$ — which shows that $\nabla f(p)$ is orthogonal to $T_p S_a$. $\square$

Let $f: S_1 \rightarrow S_2$ be a differentiable mapping at $p \in S_1$. The *derivative* of f at p is the mapping $Df_p: T_p S_1 \rightarrow T_{f(p)} S_2$ defined as follows: if $\alpha'(0) \in T_p S_1$ then $Df_p(\alpha'(0)) = (f \circ \alpha)'(0)$. That is, Df_p sends the velocity vector at p of a given curve to the velocity vector at $f(p)$ of the transform of that curve by f . Of course, the same vector represents the velocity vector at p of many different curves, but we will see below that Df_p is well-defined. Let us take local coordinates $\Phi(u, v)$ and $\Psi(\tilde{u}, \tilde{v})$ at p and $f(p)$, and let us put $\tilde{f} = \Psi^{-1} \circ f \circ \Phi$: with this notation we have the following result.

Proposition 2.3.4 *$Df_p: T_p S_1 \rightarrow T_{f(p)} S_2$ is a linear mapping whose matrix with respect to the bases (Φ_u, Φ_v) of $T_p S_1$ and $(\Psi_{\tilde{u}}, \Psi_{\tilde{v}})$ of $T_{f(p)} S_2$ is the Jacobian of $\tilde{f}$ on $\Phi^{-1}(p)$.*

Proof Writing $\alpha(t) = \Phi(u(t), v(t))$, the curve $\beta = f \circ \alpha$ is given by $\beta(t) = \Psi(\tilde{u}(t), \tilde{v}(t))$, where $(\tilde{u}(t), \tilde{v}(t)) = \tilde{f}(u(t), v(t))$. Differentiating the last equality, we obtain

$$\begin{pmatrix} \tilde{u}'(0) \\ \tilde{v}'(0) \end{pmatrix} = J\tilde{f}_{(u(0), v(0))} \begin{pmatrix} u'(0) \\ v'(0) \end{pmatrix} = J\tilde{f}_{\Phi^{-1}(p)} \begin{pmatrix} u'(0) \\ v'(0) \end{pmatrix} \quad (*)$$

The equality $Df_p(\alpha'(0)) = (f \circ \alpha)'(0)$, which defines Df_p , can be rewritten as

$$Df_p(u'(0)\Phi_u + v'(0)\Phi_v) = \tilde{u}'(0)\Psi_{\tilde{u}} + \tilde{v}'(0)\Psi_{\tilde{v}}. \quad (**)$$

From (*) and (**) it follows that $Df_p(\alpha'(0))$ is well-defined, not depending on α but only on $\alpha'(0)$; and that furthermore Df_p is linear and has matrix $J\tilde{f}_{\Phi^{-1}(p)}$ with respect to the given bases. $\square$

We end this section with two results which are the transpositions of the inverse mapping theorem and the chain rule to the context of surfaces. The proof of the first one is left as an (easy) exercise.

Proposition 2.3.5 *Let $f: S_1 \rightarrow S_2$ be a differentiable mapping and $p \in S_1$ such that $Df_p: T_p S_1 \rightarrow T_{f(p)} S_2$ is a linear isomorphism. Then there exists an open neighborhood U of p in S_1 and an open subset V of S_2 such that $f|_U: U \rightarrow V$ is a diffeomorphism.*

Proposition 2.3.6 *If $f: S_1 \rightarrow S_2$ and $g: S_2 \rightarrow S_3$ are differentiable then $g \circ f$ is differentiable and, for all $p \in S_1$, we have $D(g \circ f)_p = Dg_{f(p)} \circ Df_p$.*

Proof The verification that $g \circ f$ is differentiable is left to the reader. As for the second statement, let us take $\mathbf{u} \in T_p S_1$ and a curve α such that $\alpha'(0) = \mathbf{u}$, and let us put $\beta = f \circ \alpha$, $\gamma = g \circ f \circ \alpha$, $\mathbf{v} = \beta'(0)$, $\mathbf{w} = \gamma'(0)$. We then have $Df_p(\mathbf{u}) = \mathbf{v}$ (because $\beta = f \circ \alpha$), $Dg_{f(p)}(\mathbf{v}) = \mathbf{w}$ (because $\gamma = g \circ \beta$), $D(g \circ f)_p(\mathbf{u}) = \mathbf{w}$ (because $\gamma = (g \circ f) \circ \alpha$), and therefore $D(g \circ f)_p(\mathbf{u}) = (Dg_{f(p)} \circ Df_p)(\mathbf{u})$. $\square$

Exercises

35. Consider the function $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ given by $f(x, y, z) = 2x^2 - y^2 - z^2$. Determine the equations of the planes which are tangent to the surface $f^{-1}(\{1\})$ and parallel to the plane given by the equation $2\sqrt{2}x + y + z = 0$.

36. (a) Show that the paraboloid $z = x^2 + y^2$ is diffeomorphic to the plane.
(b) Show that the sphere S^2 and the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

are diffeomorphic.

37. Let V be a neighborhood of the origin in $\mathbb{R}^2$ and let $\Phi: V \rightarrow \mathbb{R}^3$ given by $\Phi(u, v) = f(u) + g(v)$ be a parameterization of a regular surface S . Show that the tangent planes to S along the curve $\Phi(u, 0)$ are all parallel to the same line.

38. A differentiable mapping $f: S_1 \rightarrow S_2$ is called a *local diffeomorphism* if each point $p \in S_1$ has a neighborhood W in S_1 such that $f|_W: W \rightarrow f(W)$ is a diffeomorphism. Show that if f is a local diffeomorphism then Df_p is a linear isomorphism for all $p \in S_1$.

39. Show that if all normal lines to a connected surface pass through the same point, then that surface is contained in a sphere.

40. (a) Given $0 < r_1 < r_2$ and φ_0 , consider a function $g: \mathbb{R} \rightarrow \mathbb{R}$ that is C^∞ , monotone and such that $g(x) = \varphi_0$ for $x \leq r_1$, and $g(x) = 0$ for $x \geq r_2$ (see ex. 17). Let $h: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the mapping that sends the point with polar coordinates (ρ, φ) to the point with coordinates $(\rho, \varphi + g(\rho))$. Show that h is a C^∞ diffeomorphism. How does h behave in $\{p \in \mathbb{R}^2: |p| \leq r_1\}$ and $\{p \in \mathbb{R}^2: |p| \geq r_2\}$?

(b) Let (U, Φ) be a parameterization of S such that U contains the closed disk with radius r_2 centered at the origin. Show that $\Phi \circ h \circ \Phi^{-1}: \Phi(U) \rightarrow \Phi(U)$ extends to a diffeomorphism of S .

(c) Show that if S is connected then for any two points of S there exists a diffeomorphism of S that sends one of these points to the other one.

41. Define explicitly a differentiable mapping $\mathbb{T}^2 \rightarrow \mathbf{S}^2$ that is surjective.

42. Let $S = \{(x, y, z) \in \mathbb{R}^3: x \neq 0, z = xf(y/x)\}$, where $f: \mathbb{R} \rightarrow \mathbb{R}$ is a C^∞ function. Show that S is a regular surface, and that all tangent planes to S pass through the origin.

43. Consider the mapping

$$\Phi(u, v) = \left(\frac{a(uv + 1)}{u + v}, \frac{b(u - v)}{u + v}, \frac{c(uv - 1)}{u + v} \right),$$

where $a, b, c \neq 0$ and $u + v \neq 0$. Find an equation that implicitly defines the image of Φ , and conclude that it is a surface. Compute the normal vector and the tangent plane at each point.

2.4 Orientability

A surface is orientable when it is possible to distinguish its top from the bottom, so that an observer placed on it can distinguish left from right. This approach works when the observer is three-dimensional and has an idea of the position of the surface in space; it is more intricate to explain how two-dimensional beings whose universe is the surface will know whether it is orientable or not.

Given two linearly independent vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$, the trihedron $(\mathbf{v}, \mathbf{w}, N)$, where

$$N = \frac{1}{|\mathbf{v} \times \mathbf{w}|} (\mathbf{v} \times \mathbf{w}),$$

forms a positively oriented basis of $\mathbb{R}^3$, meaning that the matrix whose columns are (in the same order) these vectors has positive determinant. The unit vector N is orthogonal to the plane Π generated by $\mathbf{v}$ and $\mathbf{w}$, introducing an orientation in Π as follows: a basis $(\mathbf{v}_1, \mathbf{w}_1)$ of Π is called *positively oriented* if the triplet $(\mathbf{v}_1, \mathbf{w}_1, N)$ is a positively oriented basis of $\mathbb{R}^3$; in other words, if

$$N = \frac{1}{|\mathbf{v}_1 \times \mathbf{w}_1|} (\mathbf{v}_1 \times \mathbf{w}_1).$$

We thus recognize that Π has exactly two orientations, one induced by N and the other one by $-N$.

We say that the surface S is *orientable* if it is possible to choose, for each $p \in S$, an orientation on $T_p S$ that varies continuously with p — more precisely, if there exists a continuous function $N: S \rightarrow \mathbb{S}^2$ such that, for each p , $N(p)$ is orthogonal to $T_p S$. We call such a *field of normal vectors* N an *orientation* of S .

For example, *level surfaces are orientable*, because the vector field

$$N(p) = \frac{1}{|\nabla f(p)|} \nabla f(p)$$

is an orientation of $f^{-1}(\{a\})$ (see example 2.3.3).

Surfaces that admit a global parameterization $\Phi(u, v)$ are also orientable, because on them we can define

$$N(p) = \frac{1}{|\Phi_u \times \Phi_v|} \Phi_u \times \Phi_v|_{\Phi^{-1}(p)}.$$

More generally, each parameterization (U, Φ) of S induces, by the preceding formula, an orientation in the open $\Phi(U) \subseteq S$; the problem is to “glue” together the various local orientations to obtain an orientation of the whole surface.

Proposition 2.4.1 *Any orientable connected surface has exactly two distinct orientations.*

Proof Given two orientations N and $\tilde{N}$ of S , we have, for each p in S , $N(p) = \tilde{N}(p)$ or $N(p) = -\tilde{N}(p)$, since these two unit vectors are orthogonal to the same plane $T_p S$. Thus, the function $\sigma: S \rightarrow \mathbb{R}$ defined by $\sigma(p) = \langle N(p), \tilde{N}(p) \rangle$ is continuous and only takes the values 1 or -1 . Since S is connected, its image $\sigma(S) \subseteq \{-1, 1\}$ is also connected, and is therefore reduced to only one element. Hence, we have $N(p) = \tilde{N}(p)$ or $N(p) = -\tilde{N}(p)$ for all $p \in S$. $\square$

Example 2.4.2 The Möbius strip $\mathbb{M}$ is the surface obtained by gluing the two ends of a paper strip so that their opposite vertices coincide. We will now see that this surface is non-orientable:

To obtain a parameterization of $\mathbb{M}$, we fix a circumference and consider a line segment that intersects, at its midpoint, orthogonally the circumference. We let the line segment travel around the whole circumference, letting it rotate around its midpoint and return to the starting point with reversed endpoints. We put

$$\Phi(\theta, t) = \left(\left(2 - t \sin \frac{\theta}{2} \right) \cos \theta, \left(2 - t \sin \frac{\theta}{2} \right) \sin \theta, t \cos \frac{\theta}{2} \right),$$

where $(\theta, t) \in \mathbb{R} \times]-1, 1[$. We note that every restriction of θ to some interval of length 2π yields a different parameterization of $\mathbb{M}$. The curves $\theta = c^{\text{te}}$ represent the various positions of the generating line segment of $\mathbb{M}$ along the circumference $t = 0$. We also note that $\Phi(\theta + 2\pi, t) = \Phi(\theta, -t)$. To simplify the calculations, we introduce the moving orthonormal trihedron

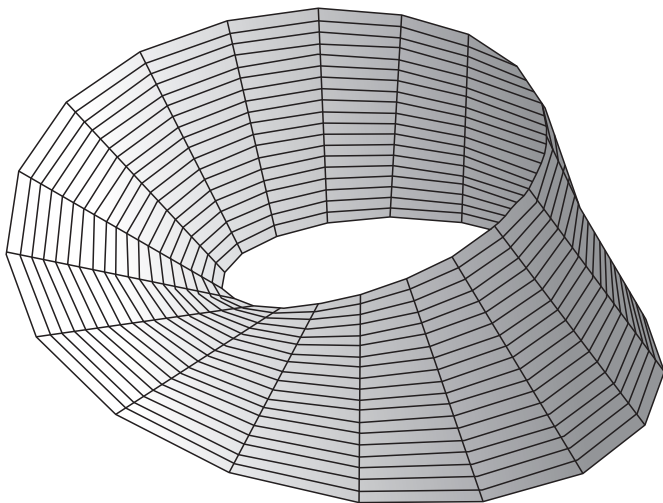


Figure 2.9

given by the three vectors

$$\begin{aligned}\mathbf{e}_1(\theta) &= (\cos \theta, \sin \theta, 0) \\ \mathbf{e}_2(\theta) &= (-\sin \theta, \cos \theta, 0) \\ \mathbf{e}_3(\theta) &= (0, 0, 1),\end{aligned}$$

which satisfies the relations $\mathbf{e}_1'(\theta) = \mathbf{e}_2(\theta)$, $\mathbf{e}_2'(\theta) = -\mathbf{e}_1(\theta)$, $\mathbf{e}_1(\theta) \times \mathbf{e}_2(\theta) = \mathbf{e}_3(\theta)$, $\mathbf{e}_2(\theta) \times \mathbf{e}_3(\theta) = \mathbf{e}_1(\theta)$, $\mathbf{e}_3(\theta) \times \mathbf{e}_1(\theta) = \mathbf{e}_2(\theta)$. Writing

$$\Phi(\theta, t) = \left(2 - t \sin \frac{\theta}{2}\right) \mathbf{e}_1(\theta) + t \cos \frac{\theta}{2} \mathbf{e}_3(\theta),$$

we easily conclude that

$$\Phi_\theta \times \Phi_t = \left(2 - t \sin \frac{\theta}{2}\right) \left[\cos \frac{\theta}{2} \mathbf{e}_1(\theta) + \sin \frac{\theta}{2} \mathbf{e}_3(\theta) \right] + \frac{t}{2} \mathbf{e}_2(\theta).$$

Let us now assume $\mathbb{M}$ orientable with orientation given by $N: \mathbb{M} \rightarrow \mathbf{S}^2$. The parameterization $\Phi|_{]0, 2\pi[\times]-1, 1[}$ induces the orientation $\tilde{N} = \frac{1}{|\Phi_\theta \times \Phi_t|} \Phi_\theta \times \Phi_t$ on $W = \mathbb{M} \setminus \{(2, 0, t): t \in]-1, 1[\}$; and, by 2.4.1, we have $\tilde{N} = N|_W$ or $\tilde{N} = -N|_W$. We shall now see that $\tilde{N}$ has no continuous extension to $\mathbb{M}$, which proves the non-existence of N .

In fact, if there was such an extension, then there would exist $\lim_{p \rightarrow (2, 0, 0)} \tilde{N}(p)$, but

$$\begin{aligned}\lim_{\theta \rightarrow 0^+} \tilde{N} \circ \Phi(\theta, 0) &= \lim_{\theta \rightarrow 0^+} \left[\cos \frac{\theta}{2} \mathbf{e}_1(\theta) + \sin \frac{\theta}{2} \mathbf{e}_2(\theta) \right] = (1, 0, 0) \neq \\ \lim_{\theta \rightarrow 2\pi^-} \tilde{N} \circ \Phi(\theta, 0) &= \lim_{\theta \rightarrow 2\pi^-} \left[\cos \frac{\theta}{2} \mathbf{e}_1(\theta) + \sin \frac{\theta}{2} \mathbf{e}_2(\theta) \right] = (-1, 0, 0).\end{aligned}$$

To sum up, what we have done shows that if a normal vector makes one complete turn around $\mathbb{M}$, then it returns to the starting position pointing in the opposite direction. It is thus possible to go “from up to down” by walking on the Möbius strip, which justifies the statement that it has only one side. $\square$

Our definition of orientability has the disadvantage of obscuring that this concept is invariant under diffeomorphisms. For example, we saw that a Möbius strip, corresponding to a certain $\mathbb{M}$ of $\mathbb{R}^3$, is non-orientable; but can we from this draw the same conclusion for all Möbius strips (i.e., for all surfaces that are diffeomorphic to $\mathbb{M}$)? There is another orientability criterion that allows one to more easily justify the (affirmative) answer to this question.

An *oriented atlas* of a surface S is a collection $(U_\alpha, \Phi_\alpha)_{\alpha \in \mathcal{I}}$ of parameterizations of S such that:

(i) the parameterizations cover S , i.e., $S = \bigcup_{\alpha \in \mathcal{I}} \Phi_\alpha(U_\alpha)$;

(ii) for all $\alpha, \beta \in \mathcal{I}$, the Jacobian of $\Phi_\beta^{-1} \circ \Phi_\alpha$ has positive determinant at all points where it is defined.

Before we move on, consider two parameterizations $\Phi(u, v)$ and $\Psi(\tilde{u}, \tilde{v})$ that intersect on an open W of S . We thus have

$$\Phi(u, v) = \Psi(\tilde{u}, \tilde{v}) \quad (*)$$

for $(u, v) \in \Phi^{-1}(W)$ and $(\tilde{u}, \tilde{v}) = \Psi^{-1} \circ \Phi(u, v)$. By differentiation of (*) we obtain the two equalities

$$\Phi_u = \frac{\partial \tilde{u}}{\partial u} \Psi_{\tilde{u}} + \frac{\partial \tilde{v}}{\partial u} \Psi_{\tilde{v}}, \quad \Phi_v = \frac{\partial \tilde{u}}{\partial v} \Psi_{\tilde{u}} + \frac{\partial \tilde{v}}{\partial v} \Psi_{\tilde{v}},$$

from which

$$\Phi_u \times \Phi_v = \left(\frac{\partial \tilde{u}}{\partial u} \frac{\partial \tilde{v}}{\partial v} - \frac{\partial \tilde{u}}{\partial v} \frac{\partial \tilde{v}}{\partial u} \right) \Psi_{\tilde{u}} \times \Psi_{\tilde{v}} = (\det J(\Psi^{-1} \circ \Phi)_{(u,v)}) \Psi_{\tilde{u}} \times \Psi_{\tilde{v}}.$$

This latter formula allows us to conclude that the two conditions

$$\begin{aligned}\text{a)} \quad & \frac{1}{|\Phi_u \times \Phi_v|} \Phi_u \times \Phi_v = \frac{1}{|\Psi_{\tilde{u}} \times \Psi_{\tilde{v}}|} \Psi_{\tilde{u}} \times \Psi_{\tilde{v}}, \\ \text{b)} \quad & \det J(\Psi^{-1} \circ \Phi)_{(u,v)} > 0,\end{aligned} \quad (**)$$

are equivalent. We can now state the alternative orientability criterion.

Proposition 2.4.3 *S is orientable if and only if it has an orientable atlas.*

Proof Suppose that N is an orientation of S . Given $p_\alpha \in S$, let (U, Φ) , with U connected, be a parameterization in the neighborhood of p_α . Then, by 2.4.1, we

have $\frac{1}{|\Phi_u \times \Phi_v|} \Phi_u \times \Phi_v = N|_{\Phi(U)}$ or $\frac{1}{|\Phi_u \times \Phi_v|} \Phi_u \times \Phi_v = -N|_{\Phi(U)}$. In the first hypothesis, we take $(U_\alpha, \Phi_\alpha) = (U, \Phi)$; in the second one, we take $U_\alpha = \{(u, v) \in \mathbb{R}^2 : (u, -v) \in U\}$ and $\Phi_\alpha(u, v) = \Phi(u, -v)$; in each case Φ_α induces on $\Phi_\alpha(U_\alpha)$ the same orientation as N . The atlas $(U_\alpha, \Phi_\alpha)_{\alpha \in \mathcal{I}}$ obtained in this way covers S and, by the equivalence of conditions a) and b) above, is oriented.

Given now an oriented atlas $(U_\alpha, \Phi_\alpha)_{\alpha \in \mathcal{I}}$, we define an orientation $N: S \rightarrow \mathbf{S}^2$ by requiring that its restriction to each open subset $\Phi_\alpha(U_\alpha)$ is the orientation induced by Φ_α . By (**), there is no ambiguity in the definition of N ; and, since $N|_{\Phi_\alpha(U_\alpha)}$ is continuous and $(\Phi_\alpha(U_\alpha))_{\alpha \in \mathcal{I}}$ is a covering of S by open sets, N is continuous. $\square$

It follows from this proof that any orientation $N: S \rightarrow \mathbf{S}^2$ is a differentiable mapping, because N is expressed as a differentiable function of the local parameters.

Let now $f: S_1 \rightarrow S_2$ be a diffeomorphism between surfaces. If S_1 is oriented then f induces an orientation on S_2 as follows: given an oriented atlas $(U_\alpha, \Phi_\alpha)_{\alpha \in \mathcal{I}}$ of S_1 (which is compatible with the orientation of S_1), then $\Psi_\alpha = f \circ \Phi_\alpha$, $(U_\alpha, \Psi_\alpha)_{\alpha \in \mathcal{I}}$ is an oriented atlas of S_2 — since, by $\Psi_\beta^{-1} \circ \Psi_\alpha = \Phi_\alpha^{-1} \circ \Phi_\beta$, the coordinate changes in either atlas are precisely the same. We call the orientation defined by the atlas $(U_\alpha, \Psi_\alpha)_{\alpha \in \mathcal{I}}$ on S_2 the *orientation induced by the diffeomorphism f* (from the given orientation of S_1). In particular, *any two diffeomorphic surfaces are either both orientable or both non-orientable*.

In the case of a diffeomorphism $f: S \rightarrow S$ of an orientable *connected* surface onto itself, we say that f *preserves orientation* if the orientation induced by f on S from a given orientation of S is equal to the original one; if it is the opposite one, then we say that f *reverses orientation*.

Proposition 2.4.4 *Let $f: S \rightarrow S$ be a diffeomorphism of a connected orientable surface. Then:*

- (i) *Whether or not f preserves orientation only depends on f , not on the orientation of S .*
- (ii) *For each oriented atlas $\mathcal{A}$ of S , one and only one of the following statements holds:*

$$\det J(\Psi^{-1} \circ f \circ \Phi) > 0 \quad \text{for any } (U, \Phi), (V, \Psi) \in \mathcal{A};$$

or

$$\det J(\Psi^{-1} \circ f \circ \Phi) < 0 \quad \text{for any } (U, \Phi), (V, \Psi) \in \mathcal{A}.$$

In the first case f preserves orientation, while in the second case it inverts it.

Proof Let N and $-N$ be the two orientations of S . We define two oriented atlases $\mathcal{A}_1$ and $\mathcal{A}_2$ of S as follows: $\mathcal{A}_1$ (resp. $\mathcal{A}_2$) includes all parameterizations (U, Φ) of S such that Φ induces in $\Phi(U)$ the orientation $N|_{\Phi(U)}$ (resp. $-N|_{\Phi(U)}$). Thus, any oriented atlas is included either in $\mathcal{A}_1$ or in $\mathcal{A}_2$, so that we can assume $\mathcal{A} = \mathcal{A}_1$. Furthermore, $\det J(\Psi^{-1} \circ \Phi) < 0$ whenever $(U, \Phi) \in \mathcal{A}_1$ and $(V, \Psi) \in \mathcal{A}_2$.

The set $f(\mathcal{A}_1)$ of the parameterizations $(U, f \circ \Phi)$ such that $(U, \Phi) \in \mathcal{A}_1$ is an oriented atlas of S , so that it defines on S one of the orientations N or

$-N$. In the first hypothesis, $f(\mathcal{A}_1) \subseteq \mathcal{A}_1$ and therefore $\det J(\Psi^{-1} \circ f \circ \Phi) > 0$ whenever $(U, \Phi), (V, \Psi) \in \mathcal{A}_1$; in the second hypothesis, $f(\mathcal{A}_1) \subseteq \mathcal{A}_2$ and thence $\det J(\Psi^{-1} \circ f \circ \Phi) < 0$ for $(U, \Phi), (V, \Psi) \in \mathcal{A}_1$. This proves (ii).

To prove (i), we take (U, Φ) in $\mathcal{A}_1$ and set $V = \{(u, v) \in \mathbb{R}^2 : (u, -v) \in U\}$, $\Psi(u, v) = \Phi(u, -v)$. Then (V, Ψ) belongs to $\mathcal{A}_2$, since the Jacobian of $\Psi^{-1} \circ \Phi(u, v) = (u, -v)$ is negative; and, for the same reason, the parameterizations $(U, f \circ \Phi)$ and $(V, f \circ \Psi)$ cannot belong both to $\mathcal{A}_1$ or both to $\mathcal{A}_2$. Therefore $f(\mathcal{A}_1)$ and $f(\mathcal{A}_2)$ define distinct orientations, which proves (i). $\square$

Example 2.4.5 Consider the diffeomorphism $f: \mathbb{S}^2 \rightarrow \mathbb{S}^2$ given by $f(x, y, z) = (-x, y, z)$. The parameterization

$$\Phi(u, v) = (u, v, \sqrt{1 - (u^2 + v^2)})$$

belongs to some oriented atlas of $\mathbb{S}^2$, since its domain is connected; and, since the Jacobian of $\Phi^{-1} \circ f \circ \Phi(u, v) = (-u, v)$ is negative, we conclude that f reverses orientation.

Exercises

44. Find out whether the antipodal mapping $h: \mathbb{S}^2 \rightarrow \mathbb{S}^2$ given by $h(x, y, z) = (-x, -y, -z)$ preserves orientation or not.
45. Consider the Möbius strip $\mathbb{M}$ parametrized by $\Phi(\theta, t) = ((2 - t \sin \frac{\theta}{2}) \cos \theta, (2 - t \sin \frac{\theta}{2}) \sin \theta, t \cos \frac{\theta}{2})$. Show that if the circumference $t = 0$ is removed from $\mathbb{M}$, then the resulting surface is still connected but is then orientable.
46. Let $f: S_1 \rightarrow S_2$ be a local diffeomorphism. Check whether the following statements are true:
 - (a) if S_2 is orientable then S_1 is orientable;
 - (b) if S_1 is orientable and f is surjective then S_2 is orientable.
47. Let S be a connected orientable surface and let $f: S \rightarrow S$ be a diffeomorphism. Is it true that $f \circ f$ preserves orientation?

2.5 Areas, Lengths, and Angles: The First Fundamental Form

Any surface $S \subseteq \mathbb{R}^3$ inherits from the ambient space a notion of size that can be used to measure the area of regions and the length of curves in S . This metric structure, which we now introduce, enriches the concept of surface and enables a finer classification than that by diffeomorphisms.

The *first fundamental form* of S at $p \in S$ is the quadratic form $I_p: T_p S \rightarrow \mathbb{R}^+$ defined by $I_p(\mathbf{v}) = \langle \mathbf{v}, \mathbf{v} \rangle_p$, where $\langle \cdot, \cdot \rangle_p$ is the restriction to $T_p S$ of the usual inner product on $\mathbb{R}^3$.

If $\Phi(u, v)$ is a parameterization of S and $\alpha(t) = \Phi(u(t), v(t))$ is a differentiable curve, we have

$$\begin{aligned}
I_{\alpha(t)}(\alpha'(t)) &= \langle u'(t)\Phi_u + v'(t)\Phi_v, u'(t)\Phi_u + v'(t)\Phi_v \rangle_{\alpha(t)} \\
&= I_{\alpha(t)}(\Phi_u)u'(t)^2 + 2\langle \Phi_u, \Phi_v \rangle_{\alpha(t)}u'(t)v'(t) + I_{\alpha(t)}(\Phi_v)v'(t)^2 \\
&= E u'(t)^2 + 2F u'(t)v'(t) + G v'(t)^2,
\end{aligned}$$

where E , F and G are the so-called *coefficients of the first fundamental form* for the parameterization $\Phi(u, v)$, defined by $E(u, v) = I_{\Phi(u, v)}(\Phi_u)$, $F(u, v) = \langle \Phi_u, \Phi_v \rangle_{\Phi(u, v)}$, $G(u, v) = I_{\Phi(u, v)}(\Phi_v)$. The above calculations show that the length of $\alpha(t)$, $t \in [a, b]$, is given by

$$l(\alpha) = \int_a^b \sqrt{E u'(t)^2 + 2F u'(t)v'(t) + G v'(t)^2} dt.$$

Therefore it is possible to compute the length of any curve in S knowing only the first fundamental form (and hence its coefficients in any parameterization) without further reference to the ambient space.

We point out that the matrix of the quadratic form $L_{\Phi(u, v)}$ relative to the basis (Φ_u, Φ_v) of $T_{\Phi(u, v)}S$ is $M = \begin{bmatrix} E & F \\ F & G \end{bmatrix}$: so if $\mathbf{v} = a\Phi_u + b\Phi_v$ and $\mathbf{w} = c\Phi_u + d\Phi_v$, the inner product of $\mathbf{v}$ and $\mathbf{w}$ is given by the matrix product $[a, b] M \begin{bmatrix} c \\ d \end{bmatrix} = Eac + F(ad + bc) + Gbd$.

Examples 2.5.1 If $\mathbf{v}$ and $\mathbf{w}$ are orthonormal vectors and $p \in \mathbb{R}^3$ then the parameterization $\Psi(u, v) = p + u\mathbf{v} + v\mathbf{w}$ of the plane parallel to $\mathbf{v}$ and $\mathbf{w}$ which passes through p has coefficients $E = 1$, $F = 0$, $G = 1$. On the other hand, the coefficients of the parameterization $\Phi(u, v) = ((2 + \cos v) \cos u, (2 + \cos v) \sin u, \sin v)$ of $\mathbb{T}^2$ are $E = 1$, $F = 0$ and $G = (2 + \cos v)^2$. $\square$

The first fundamental form also allows one to compute the angle between two nonzero vectors $\mathbf{v}, \mathbf{w} \in T_p S$: this (non-oriented) angle is the only $\theta \in [0, \pi]$ such that

$$\cos \theta = \frac{1}{\sqrt{I_p(\mathbf{v})I_p(\mathbf{w})}} \langle \mathbf{v}, \mathbf{w} \rangle_p;$$

in local coordinates, writing $\mathbf{v} = a\Phi_u + b\Phi_v$ and $\mathbf{w} = c\Phi_u + d\Phi_v$, we have

$$\cos \theta = \frac{Eac + F(ad + bc) + Gbd}{\sqrt{(Ea^2 + 2Fab + Gb^2)(Ec^2 + 2Fcd + Gd^2)}}. \quad (*)$$

If the surface S is oriented and $\Phi(u, v)$ is compatible with the orientation, we can assign a sign to the angles: the oriented angle $\angle(\mathbf{v}, \mathbf{w})$ (from $\mathbf{v}$ to $\mathbf{w}$) is the only $\theta \in]-\pi, \pi]$ such that equality $(*)$ is satisfied and such that it is negative when $ad - bc < 0$, non-negative when $ad - bc \geq 0$. [Since the oriented angles are defined up to integer multiples of 2π , the representatives of the angle $\angle(\mathbf{v}, \mathbf{w})$ are thus all numbers of the form $\theta + 2k\pi$, $k \in \mathbb{Z}$.]

The angle between two curves $\alpha(t)$ and $\beta(s)$ in S at an intersection point $\alpha(t_0) = \beta(s_0)$ is, by definition, the angle between the velocity vectors $\alpha'(t_0)$ and

$\beta'(s_0)$. For instance, it follows from the formulas deduced above that the angle between the coordinate curves of the parameterization $\Phi(u, v)$ is $\arccos(F/\sqrt{EG}) \in]0, \pi[$. When $F = 0$ the coordinate curves intersect each other orthogonally; in this case we say that $\Phi(u, v)$ is an *orthogonal parameterization*. The above examples 2.5.1 are orthogonal parameterizations; in fact, as we showed in Section 3.3, any surface admits orthogonal parameterizations.

We finally deal with the measurement of areas. If $\Delta \subseteq S$ is a region contained in a single coordinate system (U, Φ) , its area is defined by the integral

$$\iint_{\Phi^{-1}(\Delta)} |\Phi_u \times \Phi_v| du dv$$

— if such an integral exists (and it certainly exists when Δ is open or closed and the closure of $\Phi^{-1}(\Delta)$ is a compact set contained in U). If Δ is not contained in a single parametrized neighborhood, we can write it as a disjoint, finite or countable union of regions Δ_n whose areas we can compute, and add up the results.

In Section 2.4 we deduced the formula $\Phi_u \times \Phi_v = (\det J(\psi^{-1} \circ \Phi)_{(u,v)}) \Psi_{\tilde{u}} \times \Psi_{\tilde{v}}$. Therefore, if we have $\Delta \subseteq \Phi(U) \cap \Psi(U)$, the equality

$$\iint_{\Phi^{-1}(\Delta)} |\Phi_u \times \Phi_v| du dv = \iint_{\Psi^{-1}(\Delta)} |\Psi_{\tilde{u}} \times \Psi_{\tilde{v}}| d\tilde{u} d\tilde{v}$$

is a consequence of the change of variables theorem for multiple integrals, and it follows that the area is well-defined, being independent of any parameterizations used to compute it.

To motivate the formula

$$\iint_{\Phi^{-1}(\Delta)} |\Phi_u \times \Phi_v| du dv$$

for the calculation of the area of Δ , let us cover $\Phi^{-1}(\Delta)$ with a fine lattice of horizontal and vertical lines, and let $R_{i,j} = [u_i, u_{i+1}] \times [v_j, v_{j+1}]$ be any rectangle of this lattice, whose intersection with $\Phi^{-1}(\Delta)$ be non-empty. Then the integral in question is the limit, as the maximum diameter of the $R_{i,j}$ tends to zero, of the sums

$$\begin{aligned} \sum_{i,j} (u_{i+1} - u_i)(v_{j+1} - v_j) |\Phi_u \times \Phi_v|_{(u_i, v_j)} &= \\ &= \sum_{i,j} |(u_{i+1} - u_i)\Phi_u \times (v_{j+1} - v_j)\Phi_v| \end{aligned}$$

— where each summand gives the area of the parallelogram of sides $(u_{i+1} - u_i)\Phi_u$ and $(v_{j+1} - v_j)\Phi_v$. The sides of this parallelogram are tangent to $\Phi(u_i, v_j)$ and have lengths approximating the sides of the “rectangle” $\Phi(R_{i,j})$.

We now want to express the area using the coefficients E , F and G . From the identity

$$|\Phi_u \times \Phi_v|^2 + \langle \Phi_u, \Phi_v \rangle^2 = |\Phi_u|^2 |\Phi_v|^2,$$

we obtain

$$|\Phi_u \times \Phi_v| = \sqrt{EG - F^2},$$

and therefore the area of Δ is given by the integral

$$\iint_{\Phi^{-1}(\Delta)} \sqrt{EG - F^2} \, du \, dv.$$

We again point out that it follows from this formula that the notion of area depends only on the knowledge of the first fundamental form.

Example 2.5.2 The coefficients of the first fundamental form of the spherical coordinates (φ, θ) in $\mathbf{S}^2$ (example 2.1.1 **D**) are $E = \sin^2 \theta$, $F = 0$ and $G = 1$. The area of $\mathbf{S}^2$ is then given by the integral

$$\int_{-\pi}^{\pi} \left(\int_0^{\pi} \sin \theta \, d\theta \right) d\varphi = 4\pi.$$

It is also interesting to note that the area of the *spindle* between the meridians $\varphi = 0$ and $\varphi = \varphi_0$ is equal to $2\varphi_0$; more generally, any spindle of amplitude φ_0 (bounded by two **maximal semicircles** of $\mathbf{S}^2$ that intersect at an angle $\varphi_0 \in]0, \pi[$) has area $2\varphi_0$. This allows us to deduce *Girard's formula*, which gives the area of a spherical triangle (which is the figure inside $\mathbf{S}^2$ bounded by three maximal circular arcs) as a function of its interior angles.

Assume that such a triangle $\mathcal{T}$ has vertices A, B, C and interior angles $\varphi_1, \varphi_2, \varphi_3$ and denote by α, β, γ the maximal circles containing respectively the pairs of points B and C , A and C , A and B . The antipodes $\tilde{A}, \tilde{B}$, and $\tilde{C}$ of the vertices of $\mathcal{T}$ form a triangle $\tilde{\mathcal{C}}$ bounded by arcs of the same maximal circles α, β and γ (see Fig. 2.10). Because they are antipodes of each other, $\mathcal{T}$ and $\tilde{\mathcal{T}}$ have the same area (see Exercise 51 in this section).

The two maximal circles β and γ define two spindles of amplitude φ_1 in $\mathbf{S}^2$; one of them contains $\mathcal{T}$ and the other one $\tilde{\mathcal{T}}$. We denote by Δ_1 the union of these two spindles, and define analogously (using the pairs α and γ , α and β) the regions Δ_2 and Δ_3 . The union of the Δ_i covers $\mathbf{S}^2$, but each point of $\mathcal{T} \cup \tilde{\mathcal{T}}$ is counted three times, since, for $i \neq j$, we have $\Delta_i \cap \Delta_j = \mathcal{T} \cup \tilde{\mathcal{T}}$. Thus,

$$\sum_{i=1}^3 \text{area}(\Delta_i) = \text{area}(\mathbf{S}^2) + 2[\text{area}(\mathcal{T}) + \text{area}(\tilde{\mathcal{T}})]$$

and therefore

$$\begin{aligned} \text{area}(\mathcal{T}) &= \frac{1}{4} \left\{ \sum_{i=1}^3 \text{area}(\Delta_i) - \text{area}(\mathbf{S}^2) \right\} \\ &= \varphi_1 + \varphi_2 + \varphi_3 - \pi. \end{aligned}$$

We can describe our conclusion by saying that the area of a spherical triangle is proportional to its *spherical excess* (with proportionality constant equal to the square

of the radius). This formula is generalized by the remarkable Gauss-Bonnet theorem, which we discuss later. $\square$

We end this section by defining what is meant by the integral of a real function defined on a surface: given a function $f: S \rightarrow \mathbb{R}$, a parameterization (U, Φ) of S , and a region $\Delta \subseteq \Phi(U)$, the *integral of f along Δ* is

$$\int_{\Delta} f d\sigma = \iint_{\Phi^{-1}(\Delta)} f \circ \Phi(u, v) \sqrt{EG - F^2} du dv.$$

For regions not contained in a single parametrized neighborhood, we partition them, as before, into smaller regions and add up the results.

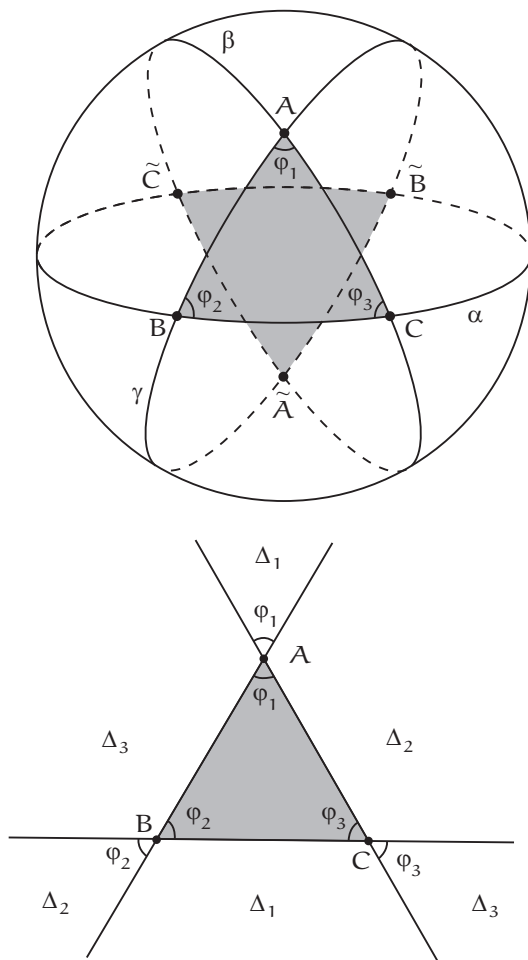


Figure 2.10

In particular, we have just defined what is meant by $\int_S f d\sigma$. Again it can be shown that these definitions do not depend on the parameterizations we use. The “quantity” $d\sigma$, which in local coordinates is written $\sqrt{EG - F^2} du dv$, is usually called *area element* of the surface.

Exercises

48. Obtain the coefficients of the first fundamental form for: (i) the helicoid (choose a parameterization); (ii) the sphere (parametrized by the inverse of the stereographic projection); (iii) a surface of revolution (parametrized as in example 2.1.1 E).

49. The coordinate curves of the parameterization $\Phi(u, v)$ constitute a Chebyshev net if the opposite sides of any quadrangle formed by them have the same lengths. Show that this happens if and only if $\frac{\partial E}{\partial v} \equiv \frac{\partial G}{\partial u} \equiv 0$.

50. Find all curves of the cylinder $\{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 = 1\}$ that intersect the generatrices (vertical lines) at a constant angle.

51. We say that a diffeomorphism of one surface onto another (or of an open subset of one surface onto an open subset of another surface) preserves area if the area of each open set is equal to that of its image.

(a) Let $\Psi: U \rightarrow S_1$ and $\Phi: U \rightarrow S_2$ be parameterizations of two surfaces, E, F, G and $\tilde{E}, \tilde{F}, \tilde{G}$ the coefficients of the first fundamental form for Ψ and Φ , respectively. Show that $\Psi \circ \Phi^{-1}$ preserves areas if and only if the functions $EG - F^2$ and $\tilde{E}\tilde{G} - \tilde{F}^2$ are identical.

(b) Show that the antipodal mapping $\mathbf{S}^2 \rightarrow \mathbf{S}^2$, $(x, y, z) \mapsto (-x, -y, -z)$, preserves areas.

(c) The *Archimedes projection* sends each point p of $\mathbf{S}^2$ (except the north and south poles) to the point of intersection of the circumscribing vertical cylinder with the half-line qp whose origin is the point q on the z -axis at the same height as p . Show that this mapping preserves areas.

(d) Find a mapping of an open subset of $\mathbb{T}^2$ into the plane that preserves areas.

52. Let U be a connected open subset of $\mathbb{R}^2$ and let $h: U \rightarrow \mathbb{R}$ be a differentiable function. Consider the surface $S = \{(x, y, z) \in \mathbb{R}^3: (x, y) \in U, z = h(x, y)\}$. Show that:

(a) the mapping $\pi: S \rightarrow U$ given by $\pi(x, y, z) = (x, y)$ is a diffeomorphism;

(b) π decreases area (for any open subset $W \subseteq S$ the area of $\pi(W)$ is $\leq$ to the area of W);

(c) if π preserves areas then S is contained in a horizontal plane.

53. The *gnomonic projection* $\mathcal{P}_\Pi: \mathbf{S}^2 \setminus \gamma_\Pi \rightarrow \Pi$ sends each point on the sphere (except the points on a certain maximal circle γ_Π) into a tangent plane Π by projecting it from the center of the sphere. Show that:

- (a) two points have the same image for $\mathcal{P}_\Pi$ if and only if they are antipodes, and the restriction $\mathcal{P}_\Pi$ to each of the hemispheres of $\mathbf{S}^2 \setminus \gamma_\Pi$ is a diffeomorphism;
- (b) the maximal circles of $\mathbf{S}^2$ are transformed into the lines of Π ;
- (c) the two drawings of [Fig. 2.10](#) are related by a gnomonic projection.



Chapter 3

The Geometry of the Gauss Map

In this chapter we deal with the extrinsic geometry of the surface, by defining quantities (curvatures) that express how the surface is embedded in its ambient space. The main tool for this study is the normal vector field to the surface; hence, we shall deal only with oriented surfaces.

3.1 The Gauss Map and its Derivative

Given an oriented surface S , the *Gauss map* is the field of normal vectors $N: S \rightarrow \mathbf{S}^2$ that defines the orientation of S . We noted in the previous chapter (following Proposition 2.4.3) that N is a differentiable mapping. By analogy with planar curves, it is to be expected that the study of the variation of N (i.e., of its derivative) will shed light on the local shape of S .

For $p \in S$ the tangent spaces $T_p S$ and $T_{N(p)} \mathbf{S}^2$ are the same subspace of $\mathbb{R}^3$, since both are the orthogonal complement of the line generated by $N(p)$. This means that the derivative DN_p is an endomorphism $T_p S \rightarrow T_p S$. Let us now take a parameterization $\Phi(u, v)$ of S and put $N(u, v) = N \circ \Phi(u, v)$. By definition of the derivative, we have

$$N_u = DN_{\Phi(u,v)}(\Phi_u), \quad N_v = DN_{\Phi(u,v)}(\Phi_v).$$

Differentiating the equalities $\langle \Phi_u, N \rangle = 0 = \langle \Phi_v, N \rangle$ with respect to v and u respectively, we obtain

$$\langle \Phi_{uv}, N \rangle + \langle \Phi_u, N_v \rangle = 0, \quad \langle \Phi_{vu}, N \rangle + \langle \Phi_v, N_u \rangle = 0,$$

and from this, subtracting term by term, and given that $\Phi_{uv} = \Phi_{vu}$, we get $\langle \Phi_u, N_v \rangle = \langle \Phi_v, N_u \rangle$. The latter equality can be rewritten in the form $\langle \Phi_u, DN_{\Phi(u,v)}(\Phi_v) \rangle = \langle \Phi_v, DN_{\Phi(u,v)}(\Phi_u) \rangle$. It follows that for all vectors $\mathbf{w}_1, \mathbf{w}_2 \in T_{\Phi(u,v)} S$, we have

$$\langle \mathbf{w}_1, DN_{\Phi(u,v)}(\mathbf{w}_2) \rangle = \langle \mathbf{w}_2, DN_{\Phi(u,v)}(\mathbf{w}_1) \rangle, \quad (3.1)$$

To verify the equality it suffices to express $\mathbf{w}_1$ and $\mathbf{w}_2$ as linear combinations of Φ_u and Φ_v .

Equality (3.1) means that DN_p is a symmetric linear mapping of $T_p S$ (with respect to the inner product $\langle \cdot, \cdot \rangle_p$ on $T_p S$). In general, a linear endomorphism $L: E \rightarrow E$ of a finite-dimensional real vector space equipped with an inner product $\langle \cdot, \cdot \rangle$, is called *symmetric* (or *self-adjoint*) if, for all $\mathbf{w}_1, \mathbf{w}_2 \in E$, we have

$$\langle \mathbf{w}_1, L(\mathbf{w}_2) \rangle = \langle L(\mathbf{w}_1), \mathbf{w}_2 \rangle. \quad (3.2)$$

The next proposition, whose complete proof can be found in numerous Linear Algebra texts, gathers the essentials about symmetric endomorphisms.

Proposition 3.1.1 *Let E be a space with inner product $\langle \cdot, \cdot \rangle$, $\mathcal{B} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ an orthonormal basis of E , and $L: E \rightarrow E$ an endomorphism. Then:*

- (i) *L is symmetric if and only if its matrix with respect to the basis $\mathcal{B}$ is symmetric;*
- (ii) *if L is symmetric, E has an orthonormal basis formed by eigenvectors of L .*

Proof (i) Since $\mathcal{B}$ is orthonormal, the matrix of L in this basis is $M = (a_{ij})_{1 \leq i, j \leq n}$ given by $a_{ij} = \langle \mathbf{e}_i, L(\mathbf{e}_j) \rangle$. We thus observe that if L is symmetric then $a_{ij} = a_{ji}$ for all i, j , a condition, that expresses the symmetry of M . Conversely, if M is symmetric then $\langle \mathbf{e}_i, L(\mathbf{e}_j) \rangle = \langle L(\mathbf{e}_i), \mathbf{e}_j \rangle$ for all $1 \leq i, j \leq n$, and it follows that the symmetry condition (3.2) is verified for any two vectors that are linear combinations of the $\mathbf{e}_i$; but every vector of E is such a linear combination, and therefore L is symmetric.

- (ii) We give the proof only in the case $n = 2$, the only one we will need. Let

$$M = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$$

be the (symmetric) matrix of L with respect to the orthonormal basis $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$. The eigenvalues of L are the roots of the characteristic polynomial $P(\lambda) = \lambda^2 - (a + b)\lambda + ab - c^2$, whose discriminant is $\Delta = (a - b)^2 + 4c^2 \geq 0$. If $\Delta = 0$ then $c = 0$ and $a = b$, which shows that L is a homothety, and so any orthonormal basis of E is formed by eigenvectors of L . If $\Delta > 0$, then L has two real eigenvalues $\lambda_1 < \lambda_2$, and we let $\mathbf{v}_1$ and $\mathbf{v}_2$ be the associated unit eigenvectors. These vectors constitute the promised basis, since

$$\lambda_1 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle L(\mathbf{v}_1), \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, L(\mathbf{v}_2) \rangle = \lambda_2 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle,$$

and therefore $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$. □

Corollary 3.1.2 *Let $\zeta: E \times E \rightarrow \mathbb{R}$ be a symmetric bilinear form on the Euclidean space E . Then there exists an orthonormal basis $\mathcal{C} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ of E such that $\zeta(\mathbf{v}_i, \mathbf{v}_j) = 0$ for all $1 \leq i < j < n$.*

Proof Consider the matrix $M = (a_{ij})_{1 \leq i, j \leq n}$ of the bilinear form ζ relative to an orthonormal basis $\mathcal{B} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ of E . This matrix, defined by $a_{ij} = \zeta(\mathbf{e}_i, \mathbf{e}_j)$, is

symmetric because of the symmetry of ζ . Let L be the linear mapping whose matrix with respect to $\mathcal{B}$ is M : by 3.1.1 (i), L is symmetric. A simple calculation shows that $\zeta(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, L(\mathbf{w}) \rangle$. By 3.1.1 there exists an orthonormal basis $\mathcal{C}$ of E formed by eigenvectors of L , and $\mathcal{C}$ is the sought basis. $\square$

The eigenvalues $k_1(p) \leq k_2(p)$ of the symmetric endomorphism $-DN_p$ (beware of the minus sign!) are called *principal curvatures* of S at the point p . If $k_1(p) < k_2(p)$, we call *principal directions* the two **orthogonal** directions defined in $T_p S$ by the eigenvectors of $-DN_p$.

To justify this terminology we introduce yet another definition. Let $\alpha:]a, b[\rightarrow S$ be a curve parametrized by arc length. The *normal curvature* of α at $\alpha(s)$ is the component of $\alpha''(s)$ in the direction of the normal to S at that point, and is given by $k_n(\alpha, s) = \langle \alpha''(s), N \circ \alpha(s) \rangle$. Note that this quantity does not depend on the orientation of the curve and that if the curve is not parametrized by arc length, the formula for computing the normal curvature is $k_n(\alpha, t) = \frac{1}{v(t)^2} \langle \alpha''(t), N \circ \alpha(t) \rangle$, where $v(t) = |\alpha'(t)|$.

Proposition 3.1.3 (i) *The normal curvature $k_n(\alpha, s)$ at $\alpha(s)$ depends only on the tangent direction to the curve at instant s : more precisely, if α and β are curves in S tangent to each other at $\alpha(s_0) = \beta(t_0) = p_0$ then $k_n(\alpha, s_0) = k_n(\beta, t_0)$.*

(ii) *The set of normal curvatures at p_0 is the interval $[k_1(p_0), k_2(p_0)]$. If $k_1(p_0) < k_2(p_0)$, then the minimum and maximum of these normal curvatures are the principal curvatures at p_0 , which occur precisely in the principal directions associated with $k_1(p_0)$ and $k_2(p_0)$.*

Proof Let us put $N(s) = N \circ \alpha(s)$. Differentiating the equality $\langle \alpha'(s), N(s) \rangle = 0$, we obtain $\langle \alpha''(s), N(s) \rangle + \langle \alpha'(s), N'(s) \rangle = 0$, and from this we get $k_n(\alpha, s) = \langle \alpha'(s), -N'(s) \rangle = \langle \alpha'(s), -DN_p(\alpha'(s)) \rangle$, where we let $p = \alpha(s)$. This equality shows that $k_n(\alpha, s)$ only depends on $\alpha'(s) \in T_p S$ and proves statement (i).

Let us now fix $p_0 = \alpha(s_0)$ and let $(\mathbf{v}_1, \mathbf{v}_2)$ be an orthonormal basis of $T_{p_0} S$ consisting of eigenvectors of $-DN_{p_0}$. Putting $\alpha'(s_0) = a\mathbf{v}_1 + b\mathbf{v}_2$, we have $a^2 + b^2 = 1$; furthermore,

$$\begin{aligned} k_n(\alpha, s_0) &= \langle a\mathbf{v}_1 + b\mathbf{v}_2, -DN_{p_0}(a\mathbf{v}_1 + b\mathbf{v}_2) \rangle \\ &= a^2 \langle \mathbf{v}_1, -DN_{p_0}(\mathbf{v}_1) \rangle + b^2 \langle \mathbf{v}_2, -DN_{p_0}(\mathbf{v}_2) \rangle \\ &= k_1 a^2 + k_2 b^2. \end{aligned}$$

Thus we obtain the inequalities

$$k_1 = k_1(a^2 + b^2) \leq k_n(\alpha, s_0) \leq k_2(a^2 + b^2) = k_2,$$

From which it follows that the normal curvatures cover the entire interval $[k_1, k_2]$ and that if $k_1 < k_2$ then the minimum is only reached for $\alpha'(s_0) = \pm \mathbf{v}_1$ and the maximum for $\alpha'(s_0) = \pm \mathbf{v}_2$. $\square$

As we have said, the normal curvature $k_n(\alpha, s)$ gives the component of the curvature vector $\alpha''(s_0)$ of α in the direction of the normal $N \circ \alpha(s)$ to the surface. If these vectors are collinear, i.e., if the principal normal to the curve α at instant s points in the direction of the normal to the surface at $\alpha(s)$, then the absolute value of $k_n(\alpha, s)$ is equal to the curvature of α at that point.

Given an arbitrary direction $\mathbf{v} \in T_p S$ with $|\mathbf{v}| = 1$, there exists at least one curve that passes through p with velocity $\mathbf{v}$, and whose principal normal at p points in the direction of $N(p)$: the intersection of S with the plane that passes through p and is parallel to the vectors $N(p)$ and $\mathbf{v}$ (see Fig. 3.1 and Exercise 34 in Section 2.2). A curve obtained this way is called a *normal section* of S at p .

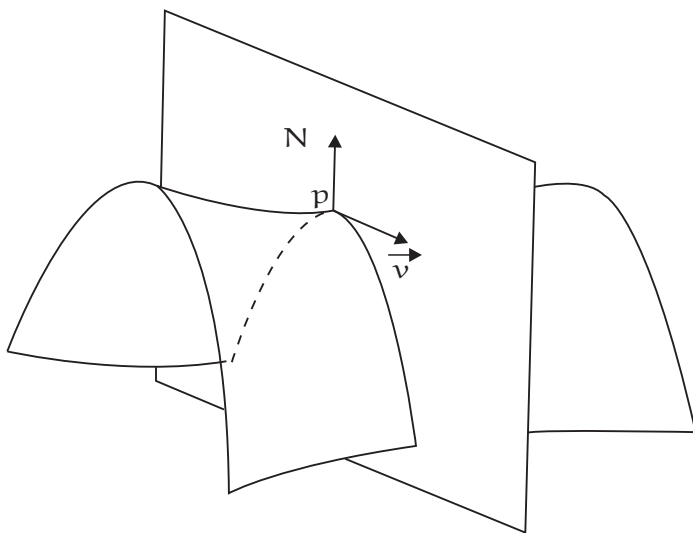


Figure 3.1

This means that to analyze the normal curvatures at $p \in S$ it suffices to study the curvatures of the normal sections. The sign of k_n will depend on whether the concavity at p of such a curve points in the direction of $N(p)$ or in the opposite direction.

The *Gaussian curvature* of S at p is defined by $K(p) = k_1(p)k_2(p)$, and the *mean curvature* is $H(p) = \frac{1}{2}(k_1(p) + k_2(p))$; equivalently, $K(p)$ and $H(p)$ are equal, respectively, to the determinant and the semi-trace of the linear mapping $-DN_p$. According to the value of these curvatures, a point $p \in S$ is called

- *elliptic* if $K(p) > 0$ (i.e., if $k_1(p)$ and $k_2(p)$ are both positive or both negative);
- *hyperbolic* if $K(p) < 0$ (the principal curvatures have opposite signs);
- *parabolic* if $K(p) = 0$ and $H(p) \neq 0$ (one of the principal curvatures is zero, the other one is nonzero);
- *planar* if $K(p) = 0 = H(p)$ (both principal curvatures are zero);
- *umbilical* if $k_1(p) = k_2(p)$ (this condition is equivalent to the equality $H(p)^2 - K(p) = 0$).

Note that any point on the surface belongs to one and only one of the first four classes, but that the umbilical points can be either elliptical or planar. We further note that although the sign of the principal curvatures depends on the orientation of S , the Gaussian curvature does not change when one changes the orientation and the classification of the points we gave above does not depend on the orientation chosen. Hence, since any surface is locally orientable (any parametrized neighborhood has an orientation induced by the parameterization), the classification given extends also to points on non-orientable surfaces.

In the next section we study the meaning of these definitions with the help of local coordinates. Now we give some examples.

Examples 3.1.4 A. All points in a plane are planar: in fact, the normal sections are straight lines, and therefore have zero curvature. This means that all normal curvatures (and hence both principal curvatures) of the plane are zero.

B. The normal sections of a sphere with radius r are maximal circles whose curvature is $1/r$. This means that at each point, the absolute values of all normal curvatures are equal to $1/r$, and therefore the two principal curvatures are equal (if they were $-1/r$ and $1/r$, some normal curvature would be zero, which is not the case) and have absolute value $1/r$. All points on the sphere are therefore umbilical, and their Gaussian curvature is constant and positive, equal to $1/r^2$.

Of course, the analysis of the signs of the normal curvatures could be replaced, in this example, by a simple calculation. But with this analysis we illustrate a useful principle: *at the point $p \in S$ there is some direction in which the normal curvature vanishes (called an asymptotic direction) if and only if $K(p) \leq 0$.*

C. Consider the hyperbolic paraboloid $\{(x, y, z): z = x^2 - y^2\}$ with the orientation induced by the parameterization $\Phi(u, v) = (u, v, u^2 - v^2)$, i.e.

$$N(u, v) = \frac{1}{\sqrt{1 + 4u^2 + 4v^2}} (-2u, 2v, 1).$$

At the point $O = \Phi(0, 0)$ we have

$$\begin{aligned} -N_u &= (2, 0, 0) = 2\Phi_u, \\ -N_v &= (0, -2, 0) = -2\Phi_v. \end{aligned}$$

The principal curvatures at O are thus 2 and -2 , the principal directions are those of the vectors $\Phi_u = (1, 0, 0)$ and $\Phi_v = (0, 1, 0)$, and the Gaussian curvature is negative, equal to -4 .

D. Let E be the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1,$$

where $a \geq b \geq c > 0$. We use the symmetric bilinear form $\zeta: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ given by

$$\zeta((x, y, z), (x', y', z')) = \frac{xx'}{a^2} + \frac{yy'}{b^2} + \frac{zz'}{c^2}$$

and we denote by $\mathcal{Q}$ the associated quadratic form:

$$\mathcal{Q}(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}.$$

Given $p_0 \in \mathcal{E}$ and a plane Π passing through the origin $(0, 0, 0)$, let us study the intersection of $\mathcal{E}$ with the plane $\Pi_0 = p_0 + \Pi$.

Note that, for every vector $\mathbf{w} \in \mathbb{R}^3$, we have

$$\frac{1}{a^2} |\mathbf{w}|^2 \leq \mathcal{Q}(\mathbf{w}) \leq \frac{1}{c^2} |\mathbf{w}|^2. \quad (3.3)$$

Given now $\mathbf{w} \in \Pi$ (so that $p_0 + \mathbf{w} \in \Pi_0$), we have

$$\begin{aligned} \mathcal{Q}(p_0 + \mathbf{w}) &= \mathcal{Q}(\mathbf{w}) + 2\zeta(p_0, \mathbf{w}) + \mathcal{Q}(p_0) \\ &= \mathcal{Q}(\mathbf{w}) + L(\mathbf{w}) + 1, \end{aligned}$$

where $L: \Pi \rightarrow \mathbb{R}$ is a linear form. Thus, the intersection $\mathcal{F} = \mathcal{E} \cap \Pi_0$ is the set of sums $p_0 + \mathbf{w}$, where $\mathbf{w} \in \Pi$ satisfies $\mathcal{Q}(\mathbf{w}) + L(\mathbf{w}) = 0$.

Consider the bilinear form $\tilde{\zeta} = \zeta|_{\Pi \times \Pi}$ and the associated quadratic form $\tilde{\mathcal{Q}} = \mathcal{Q}|_{\Pi}$. By corollary 3.1.2, there exists an orthonormal basis $(\mathbf{v}_1, \mathbf{v}_2)$ of Π such that $\tilde{\zeta}(\mathbf{v}_1, \mathbf{v}_2) = 0$. By (3.3), $\tilde{\mathcal{Q}}(\mathbf{v}_1)$ and $\tilde{\mathcal{Q}}(\mathbf{v}_2)$ are positive, belonging to the interval $\left[\frac{1}{a^2}, \frac{1}{c^2}\right]$. We then have

$$\tilde{\mathcal{Q}}(\tilde{x}\mathbf{v}_1 + \tilde{y}\mathbf{v}_2) = \frac{\tilde{x}^2}{\tilde{a}^2} + \frac{\tilde{y}^2}{\tilde{b}^2},$$

where we can assume that $a \geq \tilde{a} \geq \tilde{b} \geq c$. Using coordinates $\tilde{x}, \tilde{y}$ the equation of $-p_0 + \mathcal{F}$ is written

$$\frac{\tilde{x}^2}{\tilde{a}^2} + \frac{\tilde{y}^2}{\tilde{b}^2} + \lambda\tilde{x} + \mu\tilde{y} = 0,$$

or, “completing squares”,

$$\frac{(\tilde{x} - \tilde{x}_0)^2}{\tilde{a}^2} + \frac{(\tilde{y} - \tilde{y}_0)^2}{\tilde{b}^2} = r^2,$$

where $\lambda, \mu, \tilde{x}_0, \tilde{y}_0$ and r are certain constants. Assuming that $r > 0$ (otherwise $\lambda = \mu = 0$ and $\Pi_0 \cap \mathcal{F} = \{p_0\}$), the last equation is equivalent to

$$\frac{(\tilde{x} - \tilde{x}_0)^2}{(\tilde{a}r)^2} + \frac{(\tilde{y} - \tilde{y}_0)^2}{(\tilde{b}r)^2} = 1,$$

which describes an ellipse of semi-major axis $\tilde{a}r$ and semi-minor axis $\tilde{b}r$. Thus, the maximum and minimum curvatures of $\mathcal{F}$ are, respectively,

$$\frac{\tilde{a}r}{(\tilde{b}r)^2} = \frac{\tilde{a}}{\tilde{b}^2r} \quad \text{and} \quad \frac{\tilde{b}r}{(\tilde{a}r)^2} = \frac{\tilde{b}}{\tilde{a}^2r}$$

(see Exercise 6, in Section 1.3). Since $\mathcal{F} \subseteq \mathcal{E}$ and the diameter of $\mathcal{E}$ is $2a$, we have $\tilde{a}r \leq a$. Hence, the curvature of $\mathcal{F}$ is, at all points, greater than or equal to

$$\frac{\tilde{b}}{\tilde{a}^2r} \geq \frac{\tilde{b}}{a\tilde{a}} \geq \frac{c}{a^2}.$$

We thus prove that the normal curvatures of $\mathcal{E}$ are not less than $\frac{c}{a^2}$ in absolute value. Hence (see example **B**) *the Gaussian curvature of $\mathcal{E}$ is positive at all points, being bounded by c^2/a^4 .*

We now want to obtain an upper bound for the principal curvatures of $\mathcal{E}$, and for this we need a lower bound for the diameter of its normal sections. The normal line to $\mathcal{E}$ at the point $(x_0, y_0, z_0) \in \mathcal{E}$, given by

$$t \mapsto (x_0, y_0, z_0) + t \left(\frac{x_0}{a^2}, \frac{y_0}{b^2}, \frac{z_0}{c^2} \right),$$

intersects $\mathcal{E}$ for $t = 0$ and for $t = t_0$, where

$$t_0 = \frac{-2 \left(\frac{x_0^2}{a^4} + \frac{y_0^2}{b^4} + \frac{z_0^2}{c^4} \right)}{\frac{x_0^2}{a^6} + \frac{y_0^2}{b^6} + \frac{z_0^2}{c^6}}.$$

The length of the line segment between the two intersections is then

$$\begin{aligned} |t_0| \sqrt{\frac{x_0^2}{a^4} + \frac{y_0^2}{b^4} + \frac{z_0^2}{c^4}} &= \frac{2 \left(\frac{x_0^2}{a^4} + \frac{y_0^2}{b^4} + \frac{z_0^2}{c^4} \right)^{3/2}}{\frac{x_0^2}{a^6} + \frac{y_0^2}{b^6} + \frac{z_0^2}{c^6}} \\ &\geq 2c^2 \sqrt{\frac{x_0^2}{a^4} + \frac{y_0^2}{b^4} + \frac{z_0^2}{c^4}} \\ &\geq \frac{2c^2}{a} \sqrt{\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} + \frac{z_0^2}{c^2}} = \frac{2c^2}{a}. \end{aligned}$$

Thus, under the assumption that $\mathcal{F} = \mathcal{E} \cap \Pi_0$ is a normal section of $\mathcal{E}$, we have

$\tilde{a}r \geq \frac{c^2}{a}$, or $\frac{1}{r} \leq \frac{\tilde{a}a}{c^2}$. The curvature of $\mathcal{F}$ is then not greater than

$$\frac{\tilde{a}}{\tilde{b}^2r} \leq \frac{\tilde{a}^2a}{\tilde{b}^2c^2} \leq \frac{a^3}{c^4},$$

and so the absolute values of the normal curvatures of $\mathcal{E}$ are also not greater than a^3/c^4 . In conclusion: *at any point $p \in \mathcal{E}$ we have*

$$\frac{c^2}{a^4} \leq K(p) \leq \frac{a^6}{c^8}.$$

Note that when $\mathcal{E}$ is a sphere then $a = b = c$, and both these inequalities become the equality already seen in example B. It should however be made clear that it is possible, by more ingenious methods than ours, to obtain an explicit expression for the curvature of the points of $\mathcal{E}$. This example was intended to show that we can estimate (in this case obtain lower and upper bounds) the Gaussian curvature without computing it explicitly. $\square$

In examples 3.1.4 A, B we observed that all points on the plane and the sphere are umbilical; we end the section by showing that these are the only surfaces with such a property.

Proposition 3.1.5 *Let S be a connected surface whose points are all umbilical. Then S is contained in a sphere or in a plane.*

Proof Our hypothesis implies the existence of a function $\lambda: S \rightarrow \mathbb{R}$ such that, for every $p \in S$, DN_p is a homothety of ratio $\lambda(p)$. Let us take a parameterization (U, Φ) of S , with U connected, and put $N(u, v) = N \circ \Phi(u, v)$ and $\lambda(u, v) = \lambda \circ \Phi(u, v)$. We then have

$$\begin{aligned} N_u &= \lambda(u, v) \Phi_u \\ N_v &= \lambda(u, v) \Phi_v \end{aligned} \tag{3.4}$$

and these equalities ensure that $\lambda(u, v)$ is differentiable, for we deduce from them that

$$\lambda(u, v) = \frac{\langle N_u, \Phi_u \rangle}{|\Phi_u|^2} = \frac{\langle N_v, \Phi_v \rangle}{|\Phi_v|^2}.$$

By differentiation of (3.4) we obtain

$$\begin{aligned} N_{uv} &= \lambda_v \Phi_u + \lambda \Phi_{uv} \\ N_{vu} &= \lambda_u \Phi_v + \lambda \Phi_{vu} \end{aligned}$$

Whence, subtracting term by term,

$$\lambda_v \Phi_u - \lambda_u \Phi_v = 0,$$

an equality which is only possible when λ_v and λ_u are identically zero. Thus, the function $\lambda(u, v)$ is constant on U , and therefore $\lambda: S \rightarrow \mathbb{R}$ is locally constant, hence (since S is connected) constant, equal to $\lambda \in \mathbb{R}$.

If $\lambda = 0$ then by (3.4) the normal vector N is constant, and the function $p \mapsto \langle N, p \rangle$ is locally constant, hence constant, on S , which means that S is contained in a plane $\{p \in \mathbb{R}^3: \langle N, p \rangle = a\}$, for some $a \in \mathbb{R}$.

If $\lambda \neq 0$ then, again by (3.4), the mapping $S \rightarrow \mathbb{R}^3$ given by $p \mapsto p - \frac{1}{\lambda} N(p)$ is locally constant, hence constant. Denoting by q_0 this constant, we have, for all p in S , $|p - q_0| = \left| \frac{1}{\lambda} N(p) \right| = \frac{1}{|\lambda|}$, and therefore S is contained in the sphere with center q_0 and radius $\frac{1}{|\lambda|}$. $\square$

Exercises

54. What is the region on the sphere covered by the image of the Gauss map of the surface given by the equation: (i) $z = x^2 + y^2$; (ii) $x^2 + y^2 - z^2 = 1$; (iii) $x^2 + y^2 = \cosh^2 z$.

55.(a) Compute the principal curvatures at $(0, 0, 0)$ of each of the following surfaces: (i) $z = x^2 + y^3$; (ii) $z = x^2 + y^4$; (iii) $z = x^3 - 3xy^2$. Sketch the surface (iii), indicating the region that lies above the plane $z = 0$.

(b) Conclude that when p_0 is a parabolic (or planar) point of S the following two cases are possible: (i) there exists a neighborhood of p_0 in S that lies entirely on the same side of the tangent plane to S at p_0 ; (ii) any neighborhood of p_0 in S contains points on both sides of the tangent plane.

56. (a) Show that at a point on a surface, the arithmetic mean of the normal curvatures in two orthogonal directions is equal to the mean curvature at that point.

(b) Show that the mean curvature at $p \in S$ is given by $\frac{1}{\pi} \int_0^\pi k_n(\theta) d\theta$, where $k_n(\theta)$ is the normal curvature at p in the direction that makes an angle θ with a fixed principal direction.

57. Let S be an oriented regular surface, and suppose that $p_0 \in S$ is a maximum of the function $f: S \rightarrow \mathbb{R}$, $f(p) = |p|^2$. Show that:

(a) the line segment $[O, p_0]$ is orthogonal to S at p_0 ;

(b) the Gaussian curvature of S at p_0 is greater than or equal to $1/f(p_0)$ (use Exercise 8, in Section 1.3);

(c) if S is compact, then S has some point with positive Gaussian curvature.

58. Show that if a surface is tangent to a plane along a regular curve then the points on that curve are parabolic or planar.

59. Let p be a hyperbolic point of S , and assume there exists a neighborhood U of p in S such that $(p + T_p S) \cap U$ is the union of two regular curves that intersect at p . Show that the tangent line at p to each of these curves defines an asymptotic direction in $T_p S$.

3.2 The Second Fundamental Form

Using local coordinates, we will in this section continue the study of Gaussian curvature, obtaining explicit formulas to compute it and a better understanding of its geometric meaning. The tool is again a quadratic form, now related to the Gaussian normal mapping.

We observed in Section 3.1 that $DN_p: T_p S \rightarrow T_p S$ is a symmetric linear mapping with respect to the inner product $\langle \cdot, \cdot \rangle_p$ on $T_p S$, which means that the bilinear form $(\mathbf{v}, \mathbf{w}) \mapsto \langle \mathbf{v}, -DN_p(\mathbf{w}) \rangle$ is symmetric. The *second fundamental form* at $p \in S$ is the quadratic form associated with this symmetric bilinear form, i.e. $\Pi_p(\mathbf{v}) = \langle \mathbf{v}, -DN_p(\mathbf{v}) \rangle$.

From the proof of 3.1.3 it follows that the normal curvature at p in the direction of $\mathbf{v}$ is precisely $\Pi_p(\mathbf{v})$ when $|\mathbf{v}| = 1$ [if $\mathbf{v} \neq 0$ is not a unit vector, then that normal curvature is given by $\Pi_p(\mathbf{v})/|\mathbf{v}|^2 = \Pi_p(\mathbf{v})/I_p(\mathbf{v})$] and that the principal curvatures

are the maximum and minimum of the set $\{\Pi_p(\mathbf{v}): \mathbf{v} \in T_p S, |\mathbf{v}| = 1\}$. This means that at each point of the surface the second fundamental form gathers all information about normal curvatures, principal curvatures and Gaussian curvature.

Given a parameterization $\Phi(u, v)$, we want to determine the matrix $\begin{pmatrix} e & f \\ f & g \end{pmatrix}$ of $\Pi_{\Phi(u,v)}$ relative to the base (Φ_u, Φ_v) of $T_{\Phi(u,v)} S$. The entries e, f, g of this matrix, which are functions of (u, v) , are called the *coefficients of the second fundamental form* in the coordinates (u, v) , and are computed by the formulas

$$\begin{aligned} e &= \langle \Phi_u, -DN_{\Phi(u,v)}(\Phi_u) \rangle = \langle \Phi_u, -N_u \rangle \\ &= \langle \Phi_{uu}, N \rangle \end{aligned}$$

$$\begin{aligned} f &= \langle \Phi_u, -DN_{\Phi(u,v)}(\Phi_v) \rangle = \langle \Phi_v, -DN_{\Phi(u,v)}(\Phi_u) \rangle \\ &= \langle \Phi_u, -N_v \rangle = \langle \Phi_v, -N_u \rangle \\ &= \langle \Phi_{uv}, N \rangle \end{aligned}$$

$$\begin{aligned} g &= \langle \Phi_v, -DN_{\Phi(u,v)}(\Phi_v) \rangle = \langle \Phi_v, -N_v \rangle \\ &= \langle \Phi_{vv}, N \rangle. \end{aligned}$$

For computational purposes, formulas that **do not involve** the derivatives of N are in general easier to handle. Once we have computed e, f, g it is easy to compute the normal curvature of a curve $\alpha(t) = \Phi(u(t), v(t))$ expressed in local coordinates: writing $\alpha'(t) = u'(t)\Phi_u + v'(t)\Phi_v$, we have

$$\begin{aligned} \Pi_{\alpha(t)}(\alpha'(t)) &= \langle u'\Phi_u + v'\Phi_v, -DN_{\alpha(t)}(u'\Phi_u + v'\Phi_v) \rangle \\ &= u'^2 \langle \Phi_u, -DN_{\alpha(t)}(\Phi_u) \rangle + u'v' (\langle \Phi_u, -DN_{\alpha(t)}(\Phi_v) \rangle + \\ &\quad + \langle \Phi_v, -DN_{\alpha(t)}(\Phi_u) \rangle) + v'^2 \langle \Phi_v, -DN_{\alpha(t)}(\Phi_v) \rangle \\ &= eu'^2 + 2fu'v' + gv'^2, \end{aligned}$$

where e, f, g are computed at $(u(t), v(t))$; the normal curvature is then

$$\begin{aligned} k_n(t) &= \frac{eu'^2 + 2fu'v' + gv'^2}{|\alpha'(t)|^2} \\ &= \frac{eu'^2 + 2fu'v' + gv'^2}{Eu'^2 + 2Fu'v' + Gv'^2}. \end{aligned} \tag{3.5}$$

Let us now determine the matrix $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ of $-DN_{\Phi(u,v)}$ relative to the basis (Φ_u, Φ_v) : the entries of the matrix are determined by the equalities

$$\begin{aligned} -N_u &= a_{11}\Phi_u + a_{21}\Phi_v \\ -N_v &= a_{12}\Phi_u + a_{22}\Phi_v. \end{aligned}$$

Forming the inner product of each of these equalities with Φ_u and with Φ_v we obtain

$$\begin{aligned} e &= a_{11}E + a_{21}F & f &= a_{11}F + a_{21}G \\ f &= a_{12}E + a_{22}F & g &= a_{12}F + a_{22}G, \end{aligned}$$

and these equalities can be written in matrix form as

$$\begin{bmatrix} e & f \\ f & g \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

i.e.

$$\begin{aligned} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} &= \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} e & f \\ f & g \end{bmatrix} \\ &= \frac{1}{EG - F^2} \begin{bmatrix} G & -F \\ -F & E \end{bmatrix} \begin{bmatrix} e & f \\ f & g \end{bmatrix}. \end{aligned} \quad (3.6)$$

From this we derive explicit formulas for the Gaussian curvature and mean curvature:

$$K \circ \Phi(u, v) = \det(-DN_{\Phi(u, v)}) = \frac{eg - f^2}{EG - F^2} \quad (3.7)$$

$$\begin{aligned} H \circ \Phi(u, v) &= \frac{1}{2} \operatorname{tr}(-DN_{\Phi(u, v)}) = \frac{1}{2} (a_{11} + a_{22}) \\ &= \frac{Ge - 2Ff + Eg}{2(EG - F^2)}; \end{aligned} \quad (3.8)$$

and also for the principal curvatures k_1 and k_2 , which are the eigenvalues of the matrix $[a_{ij}]$:

$$k_1 = H - \sqrt{H^2 - K}, \quad k_2 = H + \sqrt{H^2 - K}. \quad (3.9)$$

Example 3.2.1 The surface of revolution

$$\Phi(u, v) = (\rho(v) \cos u, \rho(v) \sin u, z(v))$$

has coefficients

$$\begin{aligned} e &= \frac{-z \rho}{\sqrt{\rho^2 + z^2}}, & f &= 0, & g &= \frac{\rho z - \rho z}{\sqrt{\rho^2 + z^2}}, \\ E &= \rho^2, & F &= 0, & G &= \rho^2 + z^2. \end{aligned}$$

It then follows from (3.6) that the matrix $[a_{ij}]$ is diagonal. The principal directions at the non-umbilical points are thus those of Φ_u and Φ_v , the tangent lines to the meridians and the parallels. The principal curvatures are

$$\frac{e}{E} = \frac{-z}{\rho\sqrt{\rho^2 + z^2}} \quad \text{and} \quad \frac{g}{G} = \frac{\rho z - \rho z}{(\rho^2 + z^2)^{3/2}}$$

and the Gaussian curvature is

$$K = \frac{z(\rho z - \rho z)}{\rho(\rho^2 + z^2)^2}.$$

The expression of the curvature is simplified by assuming that the generating curve $\alpha(v) = (\rho(v), 0, z(v))$ is parametrized by arc length — that is, $\rho^2 + z^2 = 1$. Differentiating this equality we obtain $zz' = -\rho\rho'$, and replacing the left-hand side with the right-hand side in the expression of K yields

$$K = \frac{-\rho(\rho^2 + z^2)'}{\rho(\rho^2 + z^2)^2} = \frac{-\rho'}{\rho}.$$

Using this formula we will now determine the surfaces of revolution of constant curvature; to do this we simply solve the differential equation $\rho' + K\rho = 0$, with K constant. Leaving the cases $K = 0$ and $K < 0$ as an exercise, let us deal with the case $K > 0$; to simplify the formulas we let $K = 1$.

The general solution of the equation $\rho' + \rho = 0$ is of the form $\rho(v) = C \cos(v + B)$, where C and B are constants. We can choose $B = 0$, since the solutions we obtain with $B \neq 0$ correspond only to a translation of the domain; so we are reduced to the solutions $\rho_C(v) = C \cos v$. Since we want $\rho > 0$, we take $C > 0$ and restrict v to an interval of the form $] -v_0, v_0[$. Putting $z_C(0) = 0$ and integrating the equality $\rho_C^2 + z_C^2 = 1$, we obtain

$$z_C(v) = \int_0^v \sqrt{1 - C^2 \sin^2 t} \, dt$$

(the opposite solution gives another parameterization of the same surface).

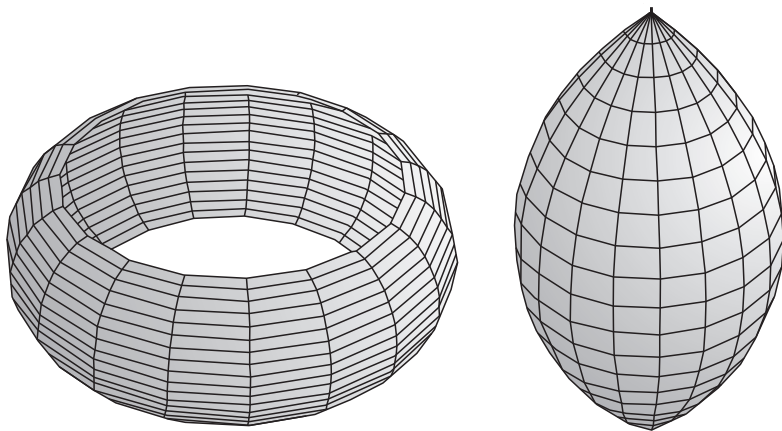


Figure 3.2

So necessarily $v_0 = \frac{\pi}{2}$ when $0 < C \leq 1$, and $v_0 = \arcsin(1/C) \in]0, \frac{\pi}{2}[$ when $C > 1$. The resulting surfaces S_C are symmetric with respect to the equator $z = 0$, whose radius is precisely C . When $C < 1$, then S_C has a cusp; when $C = 1$ it is the sphere with radius 1. When $C > 1$ it is a kind of bead of a rosary (flattened sphere pierced from one end to the other one). [See Figure 3.2]. $\square$

We can now specify the meaning of the sign of the Gaussian curvature:

Proposition 3.2.2 *Let p_0 be a point on the surface S . Then:*

- (i) *if $K(p_0) > 0$ then there exists a neighborhood U of p_0 in S which is entirely on the same side of the plane tangent to S at p_0 ;*
- (ii) *if $K(p_0) < 0$ then any neighborhood of p_0 in S contains points on either side of the plane tangent to S at p_0 .*

Proof It suffices to analyze the sign of the function $\tilde{h}: S \rightarrow \mathbb{R}$ given by $\tilde{h}(p) = \langle N(p_0), p - p_0 \rangle$ at points p near p_0 . Let us consider a parameterization $\Phi(u, v)$ such that $\Phi(0, 0) = p_0$, and let $h = \tilde{h} \circ \Phi$. The expansion of h into a Taylor polynomial at the point $(0, 0)$ gives

$$h(u, v) = h(0, 0) + \frac{\partial h}{\partial u}u + \frac{\partial h}{\partial v}v + \frac{1}{2} \left(\frac{\partial^2 h}{\partial u^2}u^2 + 2 \frac{\partial^2 h}{\partial u \partial v}uv + \frac{\partial^2 h}{\partial v^2}v^2 \right) + R(u, v),$$

where all partial derivatives are computed at $(0, 0)$ and

$$\lim_{(u,v) \rightarrow (0,0)} \frac{R(u, v)}{u^2 + v^2} = 0.$$

A quick calculation shows that $h(0, 0) = \frac{\partial h}{\partial u} = \frac{\partial h}{\partial v} = 0$ and that the second-order derivatives are nothing else but the coefficients at p_0 of the second fundamental form: i.e.,

$$\frac{\partial^2 h}{\partial u^2} = e, \quad \frac{\partial^2 h}{\partial u \partial v} = f, \quad \frac{\partial^2 h}{\partial v^2} = g.$$

This leaves us with

$$h(u, v) = \frac{1}{2} (eu^2 + 2fuv + gv^2) + R(u, v). \quad (3.10)$$

Let us put $Q(u, v) = eu^2 + 2fuv + gv^2$. By (3.7), $K(p_0)$ and $eg - f^2$ have the same sign. If $eg - f^2 > 0$ then e, g are both positive or both negative: let us now show that in this case there exists $\delta > 0$ such that $h(u, v)$ is nonzero and has constant sign (positive if $e, g > 0$, negative otherwise) whenever $0 < |(u, v)| < \delta$.

Assuming then that $eg - f^2 > 0$ and that e, g are positive, we can write

$$Q(u, v) = e \left(u + \frac{f}{e}v \right)^2 + \frac{eg - f^2}{e}v^2 \geq \frac{eg - f^2}{e}v^2,$$

and analogously,

$$Q(u, v) \geq \frac{eg - f^2}{g}u^2.$$

From these inequalities we obtain $Q(u, v) \geq 2M(u^2 + v^2)$, where

$$M = \frac{eg - f^2}{4} \min \left\{ \frac{1}{g}, \frac{1}{e} \right\} > 0.$$

Finally we have, by (3.10), that

$$\frac{h(u, v)}{u^2 + v^2} \geq M - \frac{R(u, v)}{u^2 + v^2} > 0$$

whenever $0 < |(u, v)| < \delta$, provided that $\delta > 0$ is chosen sufficiently small. The case where e, g are negative is treated analogously, and the first statement is thus proved.

If $eg - f^2 < 0$ then we can write $Q(u, v)$ as a product $\mathfrak{L}_1(u, v)\mathfrak{L}_2(u, v)$ of linearly independent linear forms $\mathfrak{L}_i: \mathbb{R}^2 \rightarrow \mathbb{R}$. The two lines $\mathfrak{L}_i(u, v) = 0$ divide the plane into four sectors: in two of these sectors $\mathfrak{L}_1$ and $\mathfrak{L}_2$ have equal signs, and in the other two they have opposite signs. There are therefore unit vectors $(u_0, v_0), (u_1, v_1) \in \mathbb{R}^2$ such that $a_0 = Q(u_0, v_0) < 0 < Q(u_1, v_1) = a_1$. Now observe that

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{h(tu_i, tv_i)}{t^2} &= \lim_{t \rightarrow 0} \left\{ \frac{Q(tu_i, tv_i)}{2t^2} + \frac{R(tu_i, tv_i)}{|(tu_i, tv_i)|^2} \right\} \\ &= \frac{a_i}{2} + \lim_{t \rightarrow 0} t \rightarrow 0 \frac{R(tu_i, tv_i)}{|(tu_i, tv_i)|^2} = \frac{a_i}{2} \end{aligned}$$

for $i = 0, 1$. This shows that $h(u, v)$ takes positive and negative values for (u, v) arbitrarily close to the origin and proves (ii). $\square$

Let us now study the contact between two surfaces. Assume that S_1 and S_2 are tangent at p_0 , and denote $\pi: \mathbb{R}^2 \rightarrow p_0 + T_{p_0} S_1$ the orthogonal projection on the

tangent plane to S at p_0 : thus, the vector $\pi(p) - p_0$ belongs to $T_{p_0} S_1$, and $\pi(p) - p$ is orthogonal to the same plane. Note that the derivative at p_0 of the restriction of π to each of the surfaces is the identity. Hence, by 2.3.5, there exist open neighborhoods U_1 and U_2 of p_0 in S_1 and S_2 , respectively, such that $\pi|_{U_1}$ and $\pi|_{U_2}$ are diffeomorphisms onto their image. Let us fix $\delta > 0$ such that

$$\{p \in p_0 + T_{p_0} S_1 : |p - p_0| < \delta\} \subseteq \pi(U_1) \cap \pi(U_2);$$

and, given an orthonormal basis $(\mathbf{w}_1, \mathbf{w}_2)$ of $T_{p_0} S_1$ and putting $N_0 = \mathbf{w}_1 \times \mathbf{w}_2$, define

$$\begin{aligned} \Phi(u, v) &= \pi|_{U_1}^{-1}(p_0 + u\mathbf{w}_1 + v\mathbf{w}_2) \\ &= p_0 + u\mathbf{w}_1 + v\mathbf{w}_2 + h_1(u, v)N_0, \\ \Psi(u, v) &= \pi|_{U_2}^{-1}(p_0 + u\mathbf{w}_1 + v\mathbf{w}_2) \\ &= p_0 + u\mathbf{w}_1 + v\mathbf{w}_2 + h_2(u, v)N_0, \end{aligned}$$

whenever $u^2 + v^2 < \delta^2$. These parameterizations correspond to regarding S_1 and S_2 locally as the graphs of the functions h_1 and h_2 defined at $p_0 + T_{p_0} S_1$.

We say that S_1 and S_2 have *contact order* ≥ 2 at p_0 if

$$\lim_{(u,v) \rightarrow (0,0)} \frac{h_1(u, v) - h_2(u, v)}{u^2 + v^2} = 0.$$

Expanding the functions h_i into Taylor polynomials, we see that this happens if and only h_1 and h_2 and their partial derivatives up to second order are equal at $(0, 0)$; but, since we have $h_i(0, 0) = 0$ and

$$\frac{\partial h_i}{\partial u} \Big|_{(0,0)} = \frac{\partial h_i}{\partial v} \Big|_{(0,0)} = 0$$

for $i = 1, 2$, this condition boils down to

$$\begin{aligned} \frac{\partial^2 h_1}{\partial u^2} \Big|_{(0,0)} &= \frac{\partial^2 h_2}{\partial u^2} \Big|_{(0,0)}, \\ \frac{\partial^2 h_1}{\partial u \partial v} \Big|_{(0,0)} &= \frac{\partial^2 h_2}{\partial u \partial v} \Big|_{(0,0)}, \\ \frac{\partial^2 h_1}{\partial v^2} \Big|_{(0,0)} &= \frac{\partial^2 h_2}{\partial v^2} \Big|_{(0,0)}. \end{aligned} \tag{3.11}$$

Now if S_1 and S_2 are oriented such that their normal vectors both coincide with N_0 at the point p_0 , then the second-order derivatives of h_1 and h_2 are the coefficients of the second fundamental form at p_0 of the parameterizations $\Phi(u, v)$ and $\Psi(u, v)$. Since $\Phi_u|_{p_0} = \Psi_u|_{p_0} = \mathbf{w}_1$ and $\Phi_v|_{p_0} = \Psi_v|_{p_0} = \mathbf{w}_2$, the equality of the coefficients implies that the second fundamental forms of the two surfaces coincide at the point p_0 . We thus proved half of the following proposition.

Proposition 3.2.3 *Let S_1 and S_2 be two oriented surfaces tangent at p_0 whose normal vectors coincide at that point. Then the following two conditions are equivalent:*

- (i) S_1 and S_2 have contact of order ≥ 2 at p_0 ;
- (ii) the restrictions on $T_{p_0} S_1 = T_{p_0} S_2$ of the second fundamental forms of S_1 and S_2 coincide.

It is left to the reader to prove that (ii) $\Rightarrow$ (i).

Given a point p_0 on a surface S_1 , the equalities 3.1.1, plus the fact that the function and its first derivatives vanish at $(0, 0)$, completely determine a polynomial $h_2(u, v)$ of degree **at most two**, given by $h_2(u, v) = \frac{1}{2}(eu^2 + 2fuv + gv^2)$ — where e, f, g are the coefficients of the second fundamental form of $\Phi(u, v)$ at p_0 . This means that *there is exactly one paraboloid which has contact of order ≥ 2 with S_1 on p_0* ; it is called the *osculating paraboloid*. It is an elliptic paraboloid when p_0 is an elliptic point, hyperbolic when p_0 is hyperbolic; it is a plane when p_0 is planar, and when p_0 is parabolic it is a straight paraboloid (a figure generated by a line perpendicular to a plane when its base point runs through a parabola).

We end this section by considering two special types of curves on surfaces; the question of the existence of such curves will be dealt with in the next section.

We say that a regular curve $\alpha(t)$ in S is a *line of curvature* if, for every t , the vector $\alpha'(t)$ defines one of the principal directions in $\alpha(t)$ — that is, if $\alpha'(t)$ is an eigenvector of $DN_{\alpha(t)}$. Using this formulation, we recognize that α is a line of curvature if and only if there exists a function $\lambda(t)$ such that $(N \circ \alpha)'(t) = \lambda(t)\alpha'(t)$ and, if such a function exists, it is differentiable because it is defined by

$$\lambda(t) = \frac{\langle (N \circ \alpha)'(t), \alpha'(t) \rangle}{|\alpha'(t)|^2}.$$

An *asymptotic line* is a regular curve whose velocity vector defines at each point an asymptotic direction — that is, it is a curve whose normal curvature is constant and equal to zero. It follows from the argument in example 3.1.4 **B** that at all points on an asymptotic line the Gaussian curvature is nonpositive. In local coordinates, formula (3.5) implies that a regular curve $\Psi(u(t), v(t))$ is asymptotic if and only if $e(u')^2 + 2fu'v' + g(v')^2 \equiv 0$.

Exercises

60. Define orientations for $S_1 = \{(x, y, z): z = x^2 - y^2\}$ and $S_2 = \{(x, y, z): z = x^3 - 3xy^2\}$, and then determine for each of these surfaces:

- (a) the Gaussian curvature and the mean curvature at each point;
- (b) the points where the mean curvature vanishes.
- (c) Do any of the answers to the above questions depend on the chosen orientations?

61. (a) Show that the Gaussian curvature of the Möbius strip $\mathbb{M}$ (with the parameterization given in 2.4.2) is given by

$$K(\theta, t) = -\left(\frac{1}{4}t^2 + \left(2 - t \sin \frac{\theta}{2}\right)^2\right)^{-2}.$$

(b) Show that if S is a surface of strictly positive curvature at all its points, then S is orientable. Show further that it is possible to choose this orientation such that all principal curvatures are positive. **Hint:** For every point p_0 of S , let U be the neighborhood given by 3.2.2 (i) and $N(p_0)$ the normal vector pointing to the side where U lies; then $N: S \rightarrow \mathbf{S}^2$ is an orientation of S .

62. Show that the pseudosphere (Exercise 32) has constant negative curvature.

63. (a) Show that the only surfaces of revolution with zero constant curvature are the cylinder, cone and plane.

(b) Show that any surface of revolution of constant curvature $K = -1$ is, up to reparametrization, of the form $\Phi(u, v) = (\rho(v) \cos u, \rho(v) \sin u, z(v))$, where the generating curve $v \mapsto (\rho(v), 0, z(v))$ is of one of the following three types:

- (i) $\rho(v) = A \cosh v$, $z(v) = \int_0^v \sqrt{1 - A^2 \sinh^2 t} dt$, $A > 0$;
- (ii) $\rho(v) = Ae^{-v}$, $z(v) = \int_0^v \sqrt{1 - A^2 e^{-2t}} dt$, $A > 0$;
- (iii) $\rho(v) = A \sinh v$, $z(v) = \int_0^v \sqrt{1 - A^2 \cosh^2 t} dt$, $0 < A < 1$;

In each of the cases find the domain of v and sketch the generating curve.

(c) Show that the surface of type (ii) is the pseudosphere.

64. Let $p_0 \in S$ be such that $K(p_0) \neq 0$, and let (U, Φ) be a parametrized neighborhood of p_0 where K has constant sign. Show that:

(a) $N_u \times N_v = K(u, v)(\Phi_u \times \Phi_v)$;

(b) if $V \subseteq \Phi(U)$, then the area of $N(V) \subseteq \mathbf{S}^2$ is given by $\int_V |K| d\sigma$;

(c) $|K(p_0)|$ is the limit of the ratio $\frac{\text{area of } N(V)}{\text{area of } V}$ as the diameter of its neighborhood V tends to zero.

65. Complete the proof of Proposition 3.2.3.

66. Let p_0 be a point shared by the surfaces S_1 and S_2 . Show that the following assertions are equivalent:

(i) S_1 and S_2 have contact of order ≥ 2 at p_0 .

(ii) There exist parameterizations $\Phi(u, v)$ and $\Psi(u, v)$ of S_1 and S_2 such that $p_0 = \Phi(0, 0) = \Psi(0, 0)$ and

$$\lim_{(u,v) \rightarrow (0,0)} \frac{|\Phi(u, v) - \Psi(u, v)|}{u^2 + v^2} = 0.$$

67. A regular surface S is a *ruled surface* if $S = \{\alpha(t) + \lambda \mathbf{v}(t) : t \in J, \lambda \in \mathbb{R}\}$, where J is an interval, $\alpha, \mathbf{v} : J \rightarrow \mathbb{R}^3$ are C^∞ , and $\alpha'(t), \mathbf{v}(t)$ are linearly independent vectors for all $t \in J$. Note that then $S = \bigcup_{t \in J} r_t$, where r_t is the line $\{\alpha(t) + \lambda \mathbf{v}(t) : \lambda \in \mathbb{R}\}$.

(a) Show that the hyperbolic paraboloid $z = x^2 - y^2$ and the hyperboloid $x^2 + y^2 - z^2 = 1$ are ruled surfaces, and find out whether the equation $e^x = z + y^2$ defines a ruled surface;

(b) Show that the tangent planes to the ruled surface S intersect S along a line, which is an asymptotic line; conclude that the curvature of S is ≤ 0 at all points.

68. With the notation of Exercise 67, let S be a ruled surface without planar points. Show that the following conditions are equivalent:

- i. S has curvature 0 at all its points.
- ii. The lines $\lambda \mapsto \alpha(t) + \lambda \mathbf{v}(t)$ are lines of curvature of S .
- iii. For every $t \in I$, the vector $\mathbf{v}'(t)$ is a linear combination of $\mathbf{v}(t)$ and $\alpha'(t)$.

69. Given a point p on a compact surface $S \subseteq \mathbb{R}^3$, we define the diameter of S at p by $d(p) = \max \{|q - p| : q \in S\}$. Assume that S has constant diameter d — that is, that $d(p) = d$ for every point p of S . Show that:

- (a) for every p in S there exists a single point $f(p)$ in S such that $|f(p) - p| = d$;
- (b) f is an involutive diffeomorphism of S (i.e., $f \circ f = \text{id}$);
- (c) S has strictly positive curvature, and therefore [Exercise 61-b)] admits an orientation for which all principal curvatures are positive;
- (d) if $k_2(p) \geq k_1(p)$ denote the principal curvatures of S then

$$\frac{1}{k_1(p)} + \frac{1}{k_2(f(p))} = d.$$

3.3 Vector Fields

A vector field assigns to each point in a surface a vector in the tangent space of the surface to that point. If this assignment is made in a sufficiently regular way, the vector field can be interpreted as a velocity field, and so it determines certain curves (*trajectories*) on the surface. With this approach we can establish in this section the existence of curves and parameterizations satisfying certain requirements (such as asymptotic lines, lines of curvature, and orthogonal parameterizations).

A *vector field* of class C^k ($k \geq 1$) in an open subset $V \subseteq \mathbb{R}^n$ is a mapping $\mathbf{v}: V \rightarrow \mathbb{R}^n$ of class C^k , and a *trajectory* (or *integral curve*) of $\mathbf{v}$ is a curve $\varphi: I \rightarrow V$ such that $\varphi'(t) = \mathbf{v}(\varphi(t))$. In other words, a trajectory is a curve whose velocity at each point is the vector $\mathbf{v}$ assigned to that point. The fundamental theorem of differential equations, which we now state, asserts the existence and uniqueness of the trajectory passing through each $p \in V$ at a given instant; for its proof we suggest [23].

Theorem 3.3.1 *Given a vector field $\mathbf{v}: V \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ of class C^k ($k \geq 1$) and $p_0 \in V$, there exist $\varepsilon > 0$, an open neighborhood $U \subseteq V$ of p_0 , and a C^k mapping $\varphi:]-\varepsilon, \varepsilon[\times U \rightarrow V$, such that, for every $p \in U$, the curve $t \mapsto \varphi(t, p)$ is the only trajectory of $\mathbf{v}$ with initial condition $\varphi(0, p) = p$ (in the sense that any other trajectory with the same initial condition coincides with this one in the intersection of their domains).*

At a point p where the vector field $\mathbf{v}$ vanishes, one trajectory of $\mathbf{v}$ that passes through p is the constant curve $\varphi(t) = p$; the theorem guarantees that this is the only trajectory that passes through p . This means that the *singularities* of the vector field

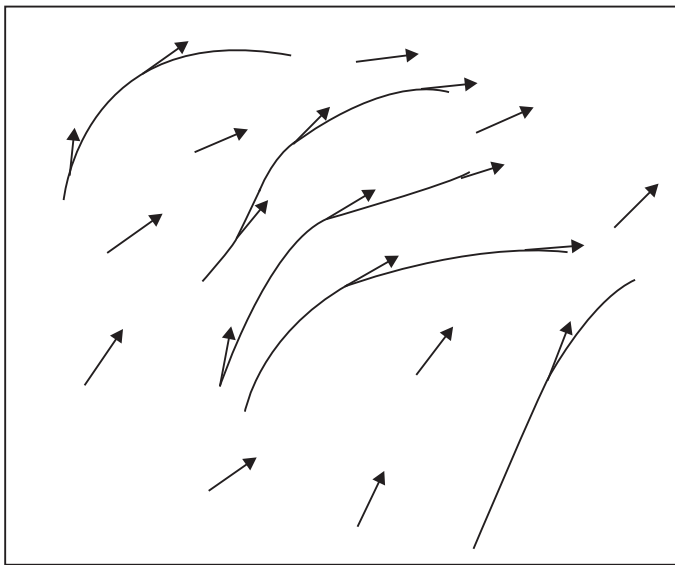


Figure 3.3

$\mathbf{v}$ correspond to the constant trajectories, and the trajectories of the other points are regular curves: for if the derivative of $\varphi(t)$ vanishes at t_0 , then $\varphi(t_0)$ is a singularity of $\mathbf{v}$ and the only trajectory through $\varphi(t_0)$ is the constant one: so a non-constant trajectory cannot have points of zero velocity.

The mapping $\varphi(t, p)$ in 3.3.1 is called the *local flow* of the vector field $\mathbf{v}$. A frequent way to write the differential equation associated with a given field is in the form

$$X' = \mathbf{v}(X)$$

where $X = (x_1, \dots, x_n)$ is a point of V . A differential equation of this type is called *autonomous*. A *non-autonomous* equation is one where the right-hand side depends explicitly on the independent variable t , i.e. of the form

$$X' = \mathbf{v}(t, X), \quad (3.12)$$

where $\mathbf{v}: I \times V \subseteq \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ is of class C^k . A simple example is given by the equation $X' = tX$: a curve $\alpha(t)$ in $\mathbb{R}^n$ is a solution of this equation if and only if $\alpha'(t) = t\alpha(t)$.

For non-autonomous equations there is also a theorem of existence, uniqueness and differentiable dependence on initial conditions. In this instance, given $(t_0, p_0) \in I \times V$, there exists a mapping $\varphi:]t_0 - \varepsilon, t_0 + \varepsilon[\times U \rightarrow V$ of class C^k (where U is an open

neighborhood of p_0) such that, for every $p \in U$, the function $t \mapsto \varphi(t, p)$ is the only solution of 3.1.2 with $\varphi(t_0, p) = p$ defined in the interval $]t_0 - \varepsilon, t_0 + \varepsilon[$.

This result for non-autonomous equations, although seemingly more general than 3.3.1, can in fact be easily deduced from it. To this end we consider the autonomous equation

$$Y = \mathbf{w}(Y), \quad (3.13)$$

where we put $Y = (t, X) \in I \times V$, and where the vector field $\mathbf{w}: I \times V \subseteq \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$ is given by $\mathbf{w}(t, X) = (1, \mathbf{v}(t, X))$. If $\psi(s)$ is the solution of (3.13) with $\psi(0) = (t_0, p)$, then we have $\psi(s) = (t_0 + s, \tilde{\varphi}(s))$, and therefore $t = t_0 + s$. Putting $\varphi(t) = \tilde{\varphi}(t - t_0)$, we recognize that $\varphi(t)$ is the solution of (3.12) with initial condition $\varphi(t_0) = p$ if and only if $\psi(s) = (t_0 + s, \tilde{\varphi}(s))$ is a solution of (3.13) with initial condition $\psi(0) = (t_0, p)$.

Returning to surfaces, a *vector field* in an open subset $V \subseteq S$ is a mapping $\mathbf{v}$ such that $\mathbf{v}(p) \in T_p S$ for every $p \in V$. Using local coordinates, we can write $\mathbf{v} \circ \Phi(u, v) = \alpha(u, v)\Phi_u + \beta(u, v)\Phi_v$ for certain functions α and β . We say that $\mathbf{v}$ is of class C^k if every point of V has a parametrized neighborhood $\Phi(u, v)$ in which the coordinate functions α and β of $\mathbf{v}$ are of class C^k (of course then the coordinate functions of $\mathbf{v}$ in any parameterization are of class C^k). Unless otherwise stated, our vector fields are C^∞ .

The equation $X = \mathbf{v}(X)$, when X is contained in the image of $\Phi(u, v)$, is equivalent to the equation $(u, v) = (\alpha(u, v), \beta(u, v))$ defined in an open subset of $\mathbb{R}^2$: in fact, $\varphi(t)$ is a trajectory of the latter equation if and only if $\Phi \circ \varphi(t)$ is a trajectory of the former. Theorem 3.3.1 is then carried over without changes to surfaces, since it concerns local properties of trajectories.

Example 3.3.2 As an example of the application of these ideas, let us verify that through each **non-umbilical** point p pass exactly two lines of curvature, corresponding to the two principal directions at p : to prove this, it suffices to show that there are two differentiable vector fields $\mathbf{w}_1$ and $\mathbf{w}_2$, defined in a neighborhood of p , such that, for every q , the vectors $\mathbf{w}_1(q)$ and $\mathbf{w}_2(q)$ define the two principal directions in q : the two lines of curvature through p are then the integral curves of $\mathbf{w}_1$ and $\mathbf{w}_2$ that pass through p . (The precise choice of the $\mathbf{w}_i$ does not matter, since any other suitable vector fields would have the same trajectories up to reparametrization.)

Let us consider a parameterization $\Phi(u, v)$ in a neighborhood of the non-umbilical point $p = \Phi(0, 0)$. By the formulas (3.6)–(3.9), §2, the entries a_{ij} of the matrix of $-DN_{\Phi(u, v)}$ relative to the basis (Φ_u, Φ_v) are differentiable functions of (u, v) , and so are the principal curvatures $k_1 < k_2$ in a neighborhood of $(0, 0)$. The eigenvectors of $-DN_{\Phi(u, v)}$ are $\alpha_i \Phi_u + \beta_i \Phi_v$, where

$$\begin{bmatrix} a_{11} - k_i & a_{12} \\ a_{21} & a_{22} - k_i \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad i = 1, 2. \quad (3.14)$$

If $a_{12}|_{(0,0)} \neq 0$, then nonzero solutions of (3.14) in a neighborhood of $(0, 0)$ are given by $(\alpha_i, \beta_i) = (-a_{12}, a_{11} - k_i)$ for $i = 1, 2$. If $a_{21}|_{(0,0)} \neq 0$, then we

take $(\alpha_i, \beta_i) = (a_{22} - k_i, -a_{21})$. If $a_{12}|_{(0,0)} = a_{21}|_{(0,0)} = 0$, then the $k_i|_p$ will be precisely $a_{11}|_{(0,0)}$ and $a_{22}|_{(0,0)}$, and therefore $a_{11}|_{(0,0)} \neq a_{22}|_{(0,0)}$; assuming $k_1|_p = a_{11}|_{(0,0)}$ and $k_2|_p = a_{22}|_{(0,0)}$, we see that $(\alpha_1, \beta_1) = (a_{22} - k_1, -a_{21})$ and $(\alpha_2, \beta_2) = (-a_{12}, a_{11} - k_2)$ are nonzero solutions of (3.14) in some neighborhood of $(0, 0)$.

In any case, we find non-trivial solutions of (3.14) that depend differentiably on (u, v) , and the promised vector fields are given by $\mathbf{w}_i = \alpha_i \Phi_u + \beta_i \Phi_v$, $i = 1, 2$. The local character of this construction is unavoidable, because there may not exist differentiable vector fields, defined at all non-umbilical points in the surface, which always point to a principal direction. $\square$

What happens in the above example is that the principal directions do not define *vector* fields, but *direction* fields: a *direction field* is a mapping $\mathcal{D}$, defined in an open subset $U \subseteq S$, such that, for every $p \in U$, $\mathcal{D}(p)$ is a one-dimensional linear subspace of $T_p S$. The direction field $\mathcal{D}$ is called differentiable if every point of U has an open neighborhood V in which a differentiable vector field $\mathbf{v}$ is defined such that $\mathbf{v}(p)$ generates $\mathcal{D}(p)$ for all $p \in V$. All direction fields we consider are differentiable. An *integral curve* of the direction field is a curve whose velocity at each point has the same direction as the field at that point.

Locally, the study of direction fields amounts to that of vector fields, and so it readily follows that the integral curve of a direction field through a given point exists and is unique (up to reparametrization). But, by example of the “hay fork” in Fig. 3.4, we see that there are direction fields which globally do not arise from a vector field: it is impossible to choose a coherent (continuous) orientation for all the trajectories shown in the picture.



Figure 3.4

In local coordinates the direction fields are given by a linear relation $A(u, v)\alpha + B(u, v)\beta = 0$ between the coordinates α and β of a vector in the basis (Φ_u, Φ_v) , so that the integral curves have the form $\Phi(u(t), v(t))$, where $A(u, v)u' + B(u, v)v' \equiv 0$.

For example, the asymptotic directions at a point of negative Gaussian curvature satisfy the equation $e\alpha^2 + 2f\alpha\beta + g\beta^2 = 0$. Since $eg - f^2 < 0$, the left-hand side of

the equation is the product of two linear factors: so if $e \neq 0$ then $e\alpha^2 + 2f\alpha\beta + g\beta^2 = e(\alpha - \lambda_1\beta)(\alpha - \lambda_2\beta)$ with $\lambda_1 \neq \lambda_2$. Hence, the two asymptotic directions at each point are given by the relations $\alpha - \lambda_1\beta = 0$, $\alpha - \lambda_2\beta = 0$, which define two direction fields. The two families of asymptotic lines are the line integrals of these vector fields, having equations $u' - \lambda_1 v' = 0$, $u' - \lambda_2 v' = 0$.

Example 3.3.3 The hyperbolic paraboloid parametrized by $\Phi(u, v) = (u, v, u^2 - v^2)$ has coefficients

$$e = \frac{2}{\sqrt{1 + 4u^2 + 4v^2}}, \quad f = 0, \quad g = \frac{-2}{\sqrt{1 + 4u^2 + 4v^2}},$$

so that the equation of the asymptotic lines is

$$\frac{2}{\sqrt{1 + 4u^2 + 4v^2}} (u'^2 - v'^2) = 0$$

which is equivalent to

$$(u' - v')(u' + v') = 0.$$

Thus, the two families of asymptotic lines are $u - v = c$ and $u + v = c$ for $c \in \mathbb{R}$, and they can be parametrized by $t \mapsto (t + c, t, 2ct + c^2)$ and $t \mapsto (-t + c, t, -2ct + t^2)$. $\square$

We say that two direction fields $\mathcal{D}_1$ and $\mathcal{D}_2$, defined in the same open subset $U \subseteq S$, are *independent* if, for every $p \in U$, the linear subspaces $\mathcal{D}_1(p)$ and $\mathcal{D}_2(p)$ of $T_p S$ are distinct. Our next result, the most important one in this section, establishes the existence of parameterizations in which tangent lines to coordinate curves have directions fixed beforehand.

Theorem 3.3.4 *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two independent direction fields in $U \subseteq S$. Then each point of U is contained in a parametrized neighborhood whose coordinate curves are the integral curves of $\mathcal{D}_1$ and $\mathcal{D}_2$.*

Proof Given $p_0 \in U$, consider a parameterization $\Phi(u, v)$ with $\Phi(0, 0) = p_0$. One of the vectors $\Phi_u|_{(0,0)}$ and $\Phi_v|_{(0,0)}$ is not in $\mathcal{D}_2(p_0)$, and we assume that it is the first one. We can then find, in a neighborhood of p_0 , a differentiable vector field of the form $\mathbf{w} = \alpha\Phi_u + \Phi_v$ that generates $\mathcal{D}_2$. We seek a parameterization of the form

$$\Psi(x, t) = \Phi(\varphi(t, x), t)$$

and such that the coordinate curves $x = \text{constant}$ are line integrals of $\mathcal{D}_2$; the curves $t = \text{constant}$ of such a parameterization coincide with the curves $v = \text{constant}$ of $\Phi(u, v)$. We then have

$$\begin{aligned} \Psi_x &= \frac{\partial \varphi}{\partial x} \Phi_u, \\ \Psi_t &= \frac{\partial \varphi}{\partial t} \Phi_u + \Phi_v, \end{aligned}$$

We thus see that, in order to be $\Psi_t = \mathbf{w}$, φ has to satisfy the equation

$$\frac{\partial \varphi}{\partial t} = \alpha(\varphi(t, x), t). \quad (3.15)$$

Consider now the differential equation

$$y = \alpha(y, t). \quad (3.16)$$

By the version for non-autonomous equations of Theorem 3.3.1, there exists a differentiable mapping $\varphi:]-\varepsilon, \varepsilon[\times J \rightarrow \mathbb{R}$ (where $J \subseteq \mathbb{R}$ is an open interval containing 0) such that for $x \in J$ the curve $t \mapsto \varphi(t, x)$ is the solution of (3.16) with initial condition $\varphi(0, x) = x$. Of course, such a function φ satisfies (3.15); furthermore, we have $\frac{\partial \varphi}{\partial x}|_{(0,0)} = \frac{\partial \varphi}{\partial x}|_{(0,x)} = 1$, so that $\Psi_x|_{(0,0)}$ and $\Psi_t|_{(0,0)}$ are independent. Hence, the Jacobian of $\Psi(x, t)$ at $(0, 0)$ has rank two and Ψ is a parameterization in a neighborhood of p_0 .

In conclusion: keeping one of the families of coordinate curves, we modified the other one so that it now corresponds to the integral curves of the $\mathcal{D}_2$ field; applying this method again (fixing now the coordinate curves corresponding to $\mathcal{D}_2$), we obtain a parameterization whose coordinate curves are the integral curves of $\mathcal{D}_1$ and of $\mathcal{D}_2$.
□

The next two results are an immediate consequence of 3.3.4 and the preceding discussion.

Corollary 3.3.5 *Every point in a surface has a neighborhood covered by an orthogonal parameterization.*

Corollary 3.3.6 *In the neighborhood of any non-umbilical point (hyperbolic point) there exists a parameterization whose coordinate curves are lines of curvature (asymptotic lines).*

Exercises

70. Show that in a neighborhood of a non-umbilical point (hyperbolic point) the coordinate curves of $\Phi(u, v)$ are lines of curvature (asymptotic lines) if and only if the coefficients of this parameterization satisfy the condition $F \equiv 0 \equiv f$ ($e \equiv 0 \equiv g$).

71. Consider $\mathbb{T}^2$ parametrized by

$$\Phi(u, v) = ((2 + \cos v) \cos u, (2 + \cos v) \sin u, \sin v).$$

(a) Show that the curvature of $\mathbb{T}^2$ in $\Phi(u, v)$ is $\frac{\cos v}{2 + \cos v}$. What are the regions of $\mathbb{T}^2$ with positive, zero and negative curvature?

(b) Show that the curves $v = \frac{\pi}{2}$ and $v = \frac{3\pi}{2}$ are asymptotic lines, and that the other asymptotic lines admit the parameterization $\alpha(t) = \Phi(u(t), v(t))$, where

$$(u(t), v(t)) = \left(u_0 + \int_0^t \frac{dx}{\sqrt{\cos x(2 - \cos x)}}, \pi + t \right), \quad t \in \left] -\frac{\pi}{2}, \frac{\pi}{2} \right[$$

or

$$(u(t), v(t)) = \left(u_0 - \int_0^t \frac{dx}{\sqrt{\cos x(2 - \cos x)}}, \pi + t \right), \quad t \in \left] -\frac{\pi}{2}, \frac{\pi}{2} \right[.$$

(c) Show that $\lim_{t \rightarrow -\frac{\pi}{2}} \alpha(t)$ and $\lim_{t \rightarrow \frac{\pi}{2}} \alpha(t)$ exist and are points of the parallels $v = \frac{\pi}{2}$ and $v = \frac{3\pi}{2}$. Deduce that all asymptotic lines of $\mathbb{T}^2$ have finite length.

72. Let S be a surface of constant zero Gaussian curvature and let p be a **parabolic** point of S . Show that there exists a line segment through p which is entirely contained in S .



Chapter 4

The Intrinsic Geometry of Surfaces

By intrinsic geometry of a surface we mean those properties that depend exclusively on measurements made on the surface but not on how the surface is embedded in the ambient space. Hence, the length of a curve on a sheet of paper is an intrinsic property, because it doesn't change when we fold the sheet; but the distance in space between two points on that same sheet is not intrinsic. In this chapter we study the intrinsic properties of surfaces and the mappings that preserve such properties (isometries).

All sections of this chapter are basic and should be studied consecutively, except for 4.4, which can be omitted because it is only needed for sections 5.4 and 5.5; the Rotation Index Theorem, proved in the Appendix, is used in Section 4.5. (Sections 4.5 and 4.6 are independent of each other and, from a logical point of view, can be studied in any order; but the Gauss-Bonnet theorem should be included as early as possible.)

4.1 Conformal Mappings and Isometries

It is impossible to draw an entirely accurate map of the terrestrial globe on a plane: all known maps distort the relative size of regions — making those further away from the equator appear larger than they are in reality — and distort the shape of continents. Still, a map is an approximate depiction of the real world, the more accurate as the region so represented becomes smaller. In the terminology we now introduce, this means that there exist conformal mappings of spherical regions into planes, but there are no such mappings which are isometries.

A mapping $f: U \subseteq S_1 \rightarrow S_2$ (for an open subset U of S_1) is *conformal* if for every $p \in U$ the derivative $Df_p: T_p S_1 \rightarrow T_{f(p)} S_2$ at p is an isomorphism that preserves angles — i.e., if, for all $\mathbf{v}, \mathbf{w}$ in $T_p S_1$, the angles $\angle(Df_p(\mathbf{v}), Df_p(\mathbf{w}))$ and $\angle(\mathbf{v}, \mathbf{w})$ are equal. This means that two regular curves that intersect at p at a certain angle are

sent by the conformal mapping f into curves that intersect at $f(p)$ at the same angle. (Note that we are referring to *non-oriented* angles).

We now give a brief interlude of linear algebra showing several characterizations of linear maps which preserve angles.

Lemma 4.1.1 *Let $(E_1, \langle \cdot, \cdot \rangle_1)$ and $(E_2, \langle \cdot, \cdot \rangle_2)$ be spaces of the same dimension n equipped with an inner product and let $L: E_1 \rightarrow E_2$ be a linear isomorphism. Then the following conditions are equivalent:*

- i. L preserves angles;
- ii. L is a similarity — that is, there exists $\lambda > 0$ such that $|L(\mathbf{v})|_2 = \lambda|\mathbf{v}|_1$ for all $\mathbf{v} \in E_1$;
- iii. there exists $\lambda > 0$ such that $\langle L(\mathbf{v}), L(\mathbf{w}) \rangle_2 = \lambda^2 \langle \mathbf{v}, \mathbf{w} \rangle_1$ for all $\mathbf{v}, \mathbf{w} \in E_1$;
- iv. there exist $\lambda > 0$ and a basis $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ of E_1 such that $\langle L(\mathbf{v}_i), L(\mathbf{v}_j) \rangle_2 = \lambda^2 \langle \mathbf{v}_i, \mathbf{v}_j \rangle_1$ for all $1 \leq i, j \leq n$.

Proof The equivalence between iii. and iv. is simple to verify, and ii. is included for informational purposes only; finally, to show that ii. $\Leftrightarrow$ iii. may be left as an exercise.

i. $\Rightarrow$ iii. Let $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ be an orthonormal basis of E_1 . The vectors $L(\mathbf{e}_1), \dots, L(\mathbf{e}_n)$ are then pairwise orthogonal. Given $i \neq j$, the angle $\theta \in [0, \pi]$ between $\mathbf{e}_i$ and $\mathbf{e}_i + \mathbf{e}_j$ is given by

$$\cos \theta = \frac{\langle \mathbf{e}_i, \mathbf{e}_i + \mathbf{e}_j \rangle_1}{|\mathbf{e}_i|_1 |\mathbf{e}_i + \mathbf{e}_j|_1} = \frac{1}{\sqrt{2}}$$

so that $\theta = \frac{\pi}{4}$. Since L preserves angles, we also have

$$\frac{\langle L(\mathbf{e}_i), L(\mathbf{e}_i + \mathbf{e}_j) \rangle_2}{|L(\mathbf{e}_i)|_2 |L(\mathbf{e}_i + \mathbf{e}_j)|_2} = \frac{1}{\sqrt{2}},$$

and from this

$$|L(\mathbf{e}_i)|_2 = \frac{\langle L(\mathbf{e}_i), L(\mathbf{e}_i) \rangle_2}{|L(\mathbf{e}_i)|_2} = \frac{1}{\sqrt{2}} |L(\mathbf{e}_i + \mathbf{e}_j)|_2.$$

Since the roles of $\mathbf{e}_i$ and $\mathbf{e}_j$ are interchangeable, we also have

$$|L(\mathbf{e}_j)|_2 = \frac{1}{\sqrt{2}} |L(\mathbf{e}_i + \mathbf{e}_j)|_2,$$

and so we deduce that $|L(\mathbf{e}_i)|_2 = |L(\mathbf{e}_j)|_2$ for all $1 \leq i, j \leq n$. We now easily conclude that iii. holds for $\lambda = |L(\mathbf{e}_i)|_2$.

iii. $\Rightarrow$ i. Just note that if iii. holds then we have, for all vectors $\mathbf{v}, \mathbf{w} \in E_1$

$$\frac{\langle L(\mathbf{v}), L(\mathbf{w}) \rangle_2}{|L(\mathbf{v})|_2 |L(\mathbf{w})|_2} = \frac{\lambda^2 \langle \mathbf{v}, \mathbf{w} \rangle_1}{\lambda |\mathbf{v}|_1 \lambda |\mathbf{w}|_1} = \frac{\langle \mathbf{v}, \mathbf{w} \rangle_1}{|\mathbf{v}|_1 |\mathbf{w}|_1}. \quad \square$$

From condition iii. of the lemma, it follows that $f: U \subseteq S_1 \rightarrow S_2$ is conformal if and only if there exists a function $\lambda: U \rightarrow \mathbb{R}^+$ such that, for all $p \in U$ and $\mathbf{v}, \mathbf{w} \in T_p S$,

$$\langle Df_p(\mathbf{v}), Df_p(\mathbf{w}) \rangle = \lambda(p)^2 \langle \mathbf{v}, \mathbf{w} \rangle. \quad (4.1)$$

By iv., condition (4.1) only needs to be checked for the vectors of a given basis of $T_p S$. This has the consequence that in local coordinates $\Phi(u, v)$ the condition that f is conformal is equivalent to the combination of the three equalities

$$\begin{aligned} \langle (f \circ \Phi)_u, (f \circ \Phi)_u \rangle &= \lambda(u, v)^2 \langle \Phi_u, \Phi_u \rangle \\ \langle (f \circ \Phi)_u, (f \circ \Phi)_v \rangle &= \lambda(u, v)^2 \langle \Phi_u, \Phi_v \rangle \\ \langle (f \circ \Phi)_v, (f \circ \Phi)_v \rangle &= \lambda(u, v)^2 \langle \Phi_v, \Phi_v \rangle, \end{aligned}$$

that express that the coefficients of the first fundamental form of the parameterization $\Phi(u, v)$ of S_1 are proportional to those of the parametrization $\Psi(u, v) = f \circ \Phi(u, v)$ of S_2 . (There is a certain abuse of language here, since only locally can it be guaranteed that Ψ is a parameterization.) It further follows from these formulas that $\lambda: U \rightarrow \mathbb{R}^+$ is differentiable, because its expression $\lambda(u, v)$ in local coordinates is differentiable.

Example 4.1.2 Let us verify that the stereographic projection (Exercise 30, Section 2.1) $\pi: \mathbf{S}^1 \setminus \{(0, 0, 1)\} \rightarrow \mathbb{R}^2$ is conformal. We recall that

$$\pi(x, y, z) = \left(\frac{2x}{1-z}, \frac{2y}{1-z} \right);$$

so, using local coordinates

$$\Phi(u, v) = \left(\sqrt{1-v^2} \cos u, \sqrt{1-v^2} \sin u, v \right), \quad v \in]-1, 1[,$$

in $\mathbf{S}^2$, the coordinates $\Psi = \pi \circ \Phi$ are given by

$$\Psi(u, v) = 2 \left(\sqrt{\frac{1+v}{1-v}} \cos u, \sqrt{\frac{1+v}{1-v}} \sin u \right).$$

The coefficients of $\Phi(u, v)$ are

$$E = 1 - v^2, \quad F = 0, \quad G = \frac{1}{1 - v^2},$$

and those of $\Psi(u, v)$ are

$$\bar{E} = \frac{4(1+v)}{1-v}, \quad \bar{F} = 0, \quad \bar{G} = \frac{4}{(1-v)^3(1+v)}.$$

Noting that

$$\bar{E} = \lambda(u, v)^2 E, \quad \bar{F} = \lambda(u, v)^2 F, \quad \bar{G} = \lambda(u, v)^2 G,$$

where $\lambda(u, v) = \frac{2}{1-v}$, we conclude that the restriction of π to the image of $\Phi(u, v)$ is conformal. Such a parameterization excludes a meridian, but any point of $\mathbf{S}^2$, except the poles, can be covered by $\Phi(u, v)$ if we adjust the domain of u . We have thus shown that π is conformal at all points of $\mathbf{S}^2 \setminus \{(0, 0, 1), (0, 0, -1)\}$. But since $\lim_{v \rightarrow -1} \lambda(u, v) = 1$, the function λ extends continuously to $\mathbf{S}^2 \setminus \{(0, 0, 1)\}$, and therefore π is conformal on its entire domain. $\square$

If $f: U \subseteq S_1 \rightarrow S_2$ is conformal, we have already observed that the function $\lambda: U \rightarrow \mathbb{R}^+$ given by (4.1) is differentiable, and hence continuous. This means that in a neighborhood of $p_0 \in U$, the mapping f is “approximately” a similarity of ratio $\lambda(p_0)$, which is the reason why cartographic mappings based on conformal mappings are more faithful to reality when representing small regions.

Given the importance of the concept, it is natural to ask whether between two arbitrary surfaces there exist conformal mappings, at least locally. Since the inverse and composite of conformal mappings are still conformal, the (affirmative) answer to this question is a consequence of the fact that any point on a surface has a neighborhood that can be parametrized by a conformal mapping. Such a parameterization is called *isothermal*.

The condition for $\Phi(u, v)$ to be isothermal is that its coefficients E, F, G are proportional to the coefficients of the parameterization $(u, v) \mapsto (u, v)$ of the plane, which are 1, 0, 1 — that is, $E \equiv G, F \equiv 0$.

The existence of isothermal parameterizations for arbitrary surfaces is a deep result, and we therefore omit its proof (which you can find, in all generality, in [25], vol.IV, p. 455 ff.). Note that by 4.1.2 the inverse of the stereographic projection is an isothermal parameterization of the sphere; in the exercises we give other examples.

Let us now talk about isometries. We say that $f: U \subseteq S_1 \rightarrow S_2$ is a *local isometry* if for every $p \in U$ the derivative $Df_p: T_p S_1 \rightarrow T_{f(p)} S_2$ is a linear isometry — that is, if $\langle Df_p(\mathbf{v}), Df_p(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\mathbf{v}, \mathbf{w} \in T_p S_1$. An *isometry* is a **diffeomorphism** $f: S_1 \rightarrow S_2$ which is also a local isometry. Two surfaces are *isometric* if there exists an isometry between them. They are *locally isometric* if every point of each surface has an open neighborhood which is isometric to an open subset of the other surface.

A *local isometry* is thus a mapping that preserves the first fundamental form, and therefore preserves all the quantities that depend on it: the length of curves, the angle between two curves, and the area of small regions (i. e., such that the mapping, restricted to them, is injective). In local coordinates, f is a local isometry when the coefficients of the first fundamental form of the parameterizations $\Phi(u, v)$ and $\Psi(u, v) = f \circ \Phi(u, v)$ coincide.

Examples 4.1.3 A. The linear isometries of $\mathbb{R}^2$ are isometries in the sense just defined. The converse, which is more interesting, appears in the exercises of this section: any isometry of $\mathbb{R}^2$ is the composite of a linear isometry with a translation.

B. Let S be a surface of revolution with axis r and let $L: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be a rotation around r . Then $L|_S: S \rightarrow S$ is an isometry of S .

C. The plane and cylinder $\mathcal{C} = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1\}$ are locally isometric: a local isometry $\mathbb{R}^2 \rightarrow \mathcal{C}$ is $f(u, v) = (\cos u, \sin u, v)$, and consists simply of wrapping the plane around the cylinder. By identifying $\mathbb{R}^2$ with $\mathbb{R}^2 \times \{0\} \subseteq \mathbb{R}^3$, this example shows that not all isometries or local isometries between surfaces are restrictions of isometries of $\mathbb{R}^3$.

D. The cone $\mathfrak{C}_k = \{(x, y, z) : z = k\sqrt{x^2 + y^2}, z \neq 0\}$, for all $k > 0$, is locally isometric to the plane: if we cut $\mathfrak{C}_k$ along a generatrix and unroll it, then we obtain a circular sector U bounded by two half-lines with origin O , in which the arcs of the circle with center O correspond to the parallels of $\mathfrak{C}_k$. If $\alpha \in]0, \frac{\pi}{2}[$ is the angle that the cone's generatrix makes with the z -axis, and $\beta \in]0, 2\pi[$ is the angle defined by U , then

$$\operatorname{tg} \alpha = \frac{1}{k}, \quad \beta = 2\pi \sin \alpha,$$

and also $U = \{(\rho \cos \varphi, \rho \sin \varphi) \in \mathbb{R}^2 : \rho > 0, 0 < \varphi < \beta\}$. (See Fig. 4.1.)

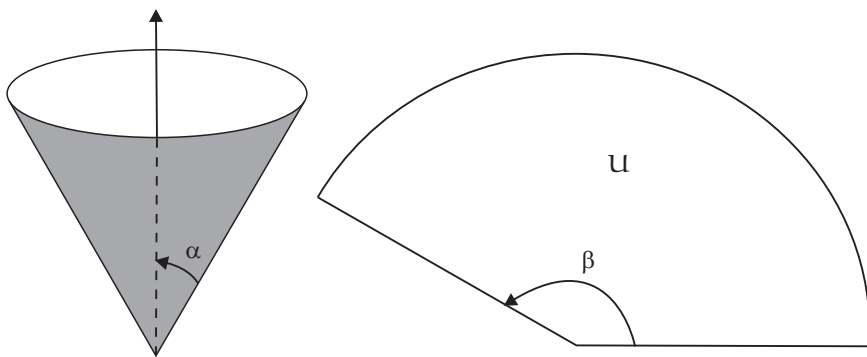


Figure 4.1

We define a mapping $f: U \rightarrow \mathfrak{C}_k$ by sending the point with polar coordinates (ρ, φ) to the point

$$\left(\frac{\rho\beta}{2\pi} \cos\left(\frac{2\pi\varphi}{\beta}\right), \frac{\rho\beta}{2\pi} \sin\left(\frac{2\pi\varphi}{\beta}\right), \rho \cos \alpha \right):$$

f is a local isometry because the parameterizations

$$\Phi(\rho, \varphi) = (\rho \cos \varphi, \rho \sin \varphi)$$

and

$$f \circ \Phi(\rho, \varphi) = \left(\frac{\rho\beta}{2\pi} \cos\left(\frac{2\pi\varphi}{\beta}\right), \frac{\rho\beta}{2\pi} \sin\left(\frac{2\pi\varphi}{\beta}\right), \rho \cos \alpha \right)$$

have the same coefficients: $E = 1$, $F = 0$, $G = \rho^2$. □

Exercises

73. The *inversion* of the plane with respect to the circle with center O and radius r is the mapping that sends each point $p \neq O$ to the point p' of the half-line Op that verifies the condition $\overline{Op'}\overline{Op} = r^2$. Show that the inversion is a conformal mapping and that it reverses orientation of the plane.

74. Consider $\mathbf{S}^2$ parametrized by two angles: the longitude $\varphi \in]-\pi, \pi[$ and the latitude $\tilde{\theta} \in]-\frac{\pi}{2}, \frac{\pi}{2}[$. The *Mercator projection* of the sphere into the plane (u, v) is defined by the following conditions: (i) the parallels $\tilde{\theta} = \text{constant}$ are sent to the lines $v = \text{constant}$, and the meridians $\varphi = \text{constant}$ to the lines $u = \text{constant}$; (ii) it is a conformal mapping; (iii) the distances along the equator $\tilde{\theta} = 0$ are converted into proportional distances; (iv) the point $\tilde{\theta} = 0, \varphi = 0$ is sent to the origin.

Show that the point with coordinates $(\varphi, \tilde{\theta})$ is sent to the point $u = \lambda\varphi, v = \lambda g(\tilde{\theta})$, where λ is a constant and

$$g'(\tilde{\theta}) = \frac{1}{\cos \tilde{\theta}} \left[\text{and therefore } g(\tilde{\theta}) = \log \operatorname{tg} \left(\frac{\pi}{4} + \frac{\tilde{\theta}}{2} \right) \right]$$

75. Find a conformal mapping of $\mathbb{T}^2$ into the plane.

76. The *catenoid* is the surface generated by the *catenary*

$$v \mapsto (a \cosh v, 0, av) \quad (a > 0)$$

around the z -axis, and it can be parameterized by

$$\Phi(u, v) = (a \cosh v \cos u, a \cosh v \sin u, av)$$

(see [Figure 4.2](#)). Consider also the helicoid parametrized by

$$\Psi(\tilde{u}, \tilde{v}) = (\tilde{v} \cos \tilde{u}, \tilde{v} \sin \tilde{u}, a\tilde{u}).$$

Using the change of coordinates $\tilde{u} = u, \tilde{v} = a \sinh v$, show that the helicoid is locally isometric to the catenoid. What are the images under this local isometry of the helices $\tilde{v} = \text{constant}$ and the straight lines $\tilde{u} = \text{constant}$?

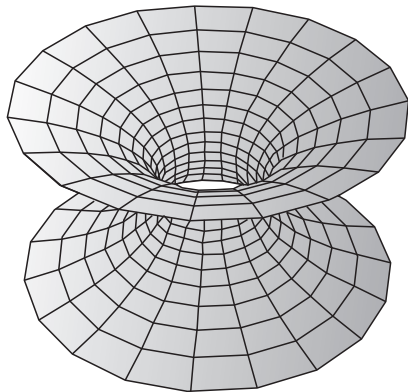


Figure 4.2

77. Let $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be an isometry such that $f(0, 0) = (0, 0)$.

(a) Show that $|f(p) - f(q)| \leq |p - q|$ for all points p, q of $\mathbb{R}^2$. Conclude that f preserves the distance between points.

(b) Show that f is linear. **Hint:** (i) use the equality $\langle f(p), f(q) \rangle = \frac{1}{2} \{ |f(p)|^2 + |f(q)|^2 - |f(p) - f(q)|^2 \}$ to show that f preserves the inner product; (ii) computing $|f(p + q) - (f(p) + f(q))|^2$, show that $f(p + q) = f(p) + f(q)$.

(c) Deduce that any isometry of $\mathbb{R}^2$ is the composite of a linear isometry and a translation.

78. (a) Let $E(\tilde{v}), G(\tilde{v})$ be C^∞ functions defined on a compact interval I . Show that there exists a parameterization

$$\Psi(\tilde{u}, \tilde{v}) = (\rho(\tilde{v}) \cos(a\tilde{u}), \rho(\tilde{v}) \sin(a\tilde{u}), z(\tilde{v}))$$

of a surface of revolution whose coefficients of the first fundamental form are $E = E(\tilde{v}), F \equiv 0, G = G(\tilde{v})$.

(b) Consider the parametrized surface $\Phi(u, v) = (3u + 3uv^2 - u^3, -3v - 3u^2v + v^3, 3u^2 - 3v^2)$, where $(u, v) \neq (0, 0)$, and compute the coefficients E, F, G of this parameterization. Changing to the coordinates $\tilde{u}, \tilde{v}$ given by $u = \tilde{v} \cos \tilde{u}, v = \tilde{v} \sin \tilde{u}$, conclude that the given surface is locally isometric to a surface of revolution.

79. Construct an isometry between the surface defined by the equation $z = y^2$ and the plane $z = 0$ which sends the parabolas $z = y^2, x = x_0$ to the lines $z = 0, x = x_0$.

80. Show that if $f: S_1 \rightarrow S_2$ is a diffeomorphism that preserves lengths of curves then f is an isometry.

4.2 Gauss's Theorema Egregium

In this section we give a necessary condition for two surfaces to be locally isometric, from which it will follow, for example, the non-existence of a local isometry between the plane and the sphere. In the proof we make use of Christoffel symbols, which are functions of local parameters that are invariant under isometries.

Given a parameterization $\Phi(u, v)$ of an oriented surface S , we consider at each point the trihedron (Φ_u, Φ_v, N) , where $N(u, v)$ is the normal field to S compatible with the orientation. [It suffices that N be a **local** orientation, since all the considerations we make are local in character — and therefore all the results in this section hold for non-orientable surfaces.] We can express the second-order derivatives of Φ in terms of this trihedron:

$$\begin{aligned}\Phi_{uu} &= \Gamma_{11}^1 \Phi_u + \Gamma_{11}^2 \Phi_v + \lambda_1 N, \\ \Phi_{uv} &= \Gamma_{12}^1 \Phi_u + \Gamma_{12}^2 \Phi_v + \lambda_2 N, \\ \Phi_{vu} &= \Gamma_{21}^1 \Phi_u + \Gamma_{21}^2 \Phi_v + \lambda_3 N, \\ \Phi_{vv} &= \Gamma_{22}^1 \Phi_u + \Gamma_{22}^2 \Phi_v + \lambda_4 N.\end{aligned}\tag{4.2}$$

The λ_i in (4.2) are just the coefficients of the second fundamental form: indeed, by the formulas in Section 3.2, we have $\lambda_1 = \langle \Phi_{uu}, N \rangle = e$, $\lambda_2 = \lambda_3 = \langle \Phi_{uv}, N \rangle = f$, $\lambda_4 = \langle \Phi_v, N \rangle = g$. The coefficients Γ_{ij}^k are the *Christoffel symbols* of the parameterization $\Phi(u, v)$, and it follows from the definition that they are symmetric with respect to the lower indices: thus, $\Gamma_{ij}^k = \Gamma_{ji}^k$. To compute them, we form the inner product of the equalities (4.2) with Φ_u and Φ_v , obtaining

$$\begin{aligned}\begin{cases} \Gamma_{11}^1 E + \Gamma_{11}^2 F = \langle \Phi_{uu}, \Phi_u \rangle = \frac{1}{2} E_u \\ \Gamma_{11}^1 F + \Gamma_{11}^2 G = \langle \Phi_{uu}, \Phi_v \rangle = F_u - \frac{1}{2} E_v \end{cases} \\ \begin{cases} \Gamma_{12}^1 E + \Gamma_{12}^2 F = \langle \Phi_{uv}, \Phi_u \rangle = \frac{1}{2} E_v \\ \Gamma_{12}^1 F + \Gamma_{12}^2 G = \langle \Phi_{uv}, \Phi_v \rangle = \frac{1}{2} G_u \end{cases} \\ \begin{cases} \Gamma_{22}^1 E + \Gamma_{22}^2 F = \langle \Phi_{vv}, \Phi_u \rangle = F_v - \frac{1}{2} G_u \\ \Gamma_{22}^1 F + \Gamma_{22}^2 G = \langle \Phi_v, \Phi_v \rangle = \frac{1}{2} G_v. \end{cases}\end{aligned}\tag{4.3}$$

Each of the equation systems in (4.3) has determinant $EG - F^2 > 0$, which shows that they suffice for computing the Christoffel symbols, and that these are functions of the coefficients E, F, G and their derivatives. In particular, and since local isometries preserve E, F, G , they also preserve the coefficients Γ_{ij}^k — more precisely, if f is a local

isometry then the Christoffel symbols of the parameterization $\Psi(u, v) = f \circ \Phi(u, v)$ coincide with the Christoffel symbols of $\Phi(u, v)$.

It is more sensible to solve the systems of equations (4.3) in each case instead of determining general expressions for the Γ_{ij}^k . However, we observe that when we deal with orthogonal parameterizations ($F \equiv 0$), the calculation of the Christoffel symbols is greatly simplified. For example, the symbols of the parameterization $(\rho(v) \cos u, \rho(v) \sin u, z(v))$ of a surface of revolution, where $E = \rho^2$, $F = 0$, $G = \rho^2 + z^2$, are

$$\begin{aligned} \Gamma_{11}^1 &= 0, & \Gamma_{11}^2 &= \frac{-\rho\rho'}{\rho^2 + z^2}, & \Gamma_{12}^1 &= \frac{\rho}{\rho}, \\ \Gamma_{12}^2 &= 0, & \Gamma_{22}^1 &= 0, & \Gamma_{22}^2 &= \frac{\rho\rho' + zz'}{\rho^2 + z^2} \end{aligned} \quad (4.4)$$

We now look for relations between the Christoffel symbols, starting from the identity

$$(\Phi_{uu})_v = (\Phi_{uv})_u. \quad (4.5)$$

Differentiating the first and second equalities of (4.2) with respect to v and to u , respectively, we obtain

$$\begin{aligned} (\Phi_{uu})_v &= (\Gamma_{11}^1)_v \Phi_u + (\Gamma_{11}^2)_v \Phi_v + \Gamma_{11}^1 \Phi_{uv} + \Gamma_{11}^2 \Phi_{vv} + e_v N + e N_v, \\ (\Phi_{uv})_u &= (\Gamma_{12}^1)_u \Phi_u + (\Gamma_{12}^2)_u \Phi_v + \Gamma_{12}^1 \Phi_{uu} + \Gamma_{12}^2 \Phi_{uv} + f_u N + f N_u. \end{aligned}$$

Using again (4.2) and expressing N_v and N_u in terms of Φ_u and Φ_v , we can write each of the terms of (4.5) as a linear combination of the vectors Φ_u, Φ_v, N . Equating the coefficients of Φ_v in the two linear combinations, we obtain

$$(\Gamma_{11}^2)_v + \Gamma_{11}^1 \Gamma_{12}^2 + \Gamma_{11}^2 \Gamma_{22}^2 - e a_{22} = (\Gamma_{12}^1)_u + \Gamma_{12}^1 \Gamma_{11}^2 + \Gamma_{12}^2 \Gamma_{12}^2 - f a_{21}. \quad (4.6)$$

By the formulas (3.6) and (3.7) in Section 3.2, we have

$$\begin{aligned} e a_{22} - f a_{21} &= \frac{e(-Ff + Eg)}{EG - F^2} - \frac{f(-Fe + Ef)}{EG - F^2} \\ &= E \frac{eg - f^2}{EG - F^2} = EK. \end{aligned} \quad (4.7)$$

Combining (4.6) and (4.7), we obtain the following expression for the Gaussian curvature:

$$K = \frac{1}{E} \left((\Gamma_{11}^2)_v - (\Gamma_{12}^2)_u + \Gamma_{12}^1 (\Gamma_{11}^1 - \Gamma_{12}^2) + \Gamma_{11}^1 (\Gamma_{22}^2 - \Gamma_{12}^1) \right). \quad (4.8)$$

Although its utility is questionable, formula (4.8) has this consequence: *it is possible to compute the Gaussian curvature knowing only the coefficients of the first fundamental form and their derivatives*. In the calculation of the principal curvatures k_1 and k_2 , the Gauss map and its derivative (i.e., the second fundamental form) are crucial

ingredients. It is remarkable that their product $K = k_1 k_2$ is after all an intrinsic concept, for it depends only on the first fundamental form.

As a consequence of this discussion we have the following theorem, which the author himself has called “egregious”.

Theorem 4.2.1 (Gauss). *The Gaussian curvature is invariant under local isometries. More precisely, if $f: U \subseteq S_1 \rightarrow S_2$ is a local isometry then, for all p in U , the curvature of S_1 at p is equal to the curvature of S_2 at $f(p)$.*

This theorem already shows, in many cases, that two given surfaces — such as the sphere (which has curvature > 0) and the plane (curvature 0) — are not locally isometric.

The deduction of formula (4.8) seems like magic: how did we pull such a rabbit out of such a hat? Later on we will deduce, by more transparent means, other formulas that also establish the invariance of the curvature by local isometries.

Exercises

81. Let $S_1 = \{(x, y, z): z = 0\}$ and $S_2 = \{(x, y, z): x^2 + y^2 = z\}$.

- (a) Check that $f: S_1 \rightarrow S_2$ given by $f(x, y, 0) = (x, y, x^2 + y^2)$ is not an isometry.
- (b) Show that S_1 and S_2 are not locally isometric.

82. Consider the surfaces S_1 and S_2 given by

$$\Phi(u, v) = (u \cos v, u \sin v, \log u), \quad (u \in \mathbb{R}^+, 0 < v < 2\pi)$$

and

$$\Psi(u, v) = (u \cos v, u \sin v, v) \quad (u, v \in \mathbb{R}).$$

Show that $\Psi \circ \Phi^{-1}: S_1 \rightarrow S_2$ is not a local isometry, although the curvature of S_2 at $\Psi(u, v)$ is equal to that of S_1 at $\Phi(u, v)$.

83. Consider $\mathbb{T}^2$ with the usual parameterization. Show that any isometry of $\mathbb{T}^2$ sends the equator $v = 0$ to itself (use Exercise 71-a) in Section 3.3).

84. Show that a conformal mapping that preserves areas (see Exercise 71, in Section 2.5) is necessarily a local isometry (and therefore all maps of the Earth’s surface have to choose one of two defects: either they distort areas, or they distort shapes).

4.3 Covariant Derivative, Parallel Transport, Geodesic Curvature

All the concepts we define in this section belong to intrinsic geometry, in the sense that they are invariant under local isometries. The method, used here repeatedly, to establish the intrinsic character of a certain concept consists of showing that it is a function solely of the coefficients E, F, G and their derivatives — which is the case whenever we can express it by means of these quantities and the Christoffel symbols only.

Let $\mathbf{v}$ be a vector field defined on an open subset $U \subseteq S$. Given a regular curve $\alpha: I \rightarrow U$, the *covariant derivative of $\mathbf{v}$ along α* is the orthogonal projection of $(\mathbf{v} \circ \alpha)'(t)$ on the tangent plane $T_{\alpha(t)}S$, denoted by $\frac{D\mathbf{v}}{dt}(t)$. So, to compute $\frac{D\mathbf{v}}{dt}(t)$, we just have to subtract from $(\mathbf{v} \circ \alpha)'(t)$ its component with respect to the normal $N \circ \alpha(t)$ to the surface.

Example 4.3.1 Consider a unit vector field $\mathbf{v}$ in $\mathbf{S}^2 \setminus \{(0, 0, 1), (0, 0, -1)\}$ having constant angle $\frac{\pi}{4}$ with the parallels of the sphere. In spherical coordinates

$$\Phi(\varphi, \theta) = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta),$$

we put

$$\begin{aligned} \mathbf{v}(\varphi, \theta) &= \frac{1}{\sqrt{2}} \left(-\Phi_\theta + \frac{1}{\sin \theta} \Phi_\varphi \right) \\ &= \frac{1}{\sqrt{2}} (-\cos \theta \cos \varphi - \sin \varphi, -\sin \theta \sin \varphi + \cos \varphi, \sin \theta). \end{aligned}$$

Let us compute the covariant derivative of $\mathbf{v}$ along the parallel $\theta = \theta_0$, parametrized by $\alpha_{\theta_0}(t) = \Phi(t, \theta_0)$. We now have

$$\mathbf{v}'(t) = \frac{1}{\sqrt{2}} (\cos \theta_0 \sin t - \cos t, -\sin \theta_0 \cos t - \sin t, 0),$$

and we obtain $\frac{D\mathbf{v}}{dt}(t)$ by subtracting from $\mathbf{v}'(t)$ its normal component:

$$\begin{aligned} \frac{D\mathbf{v}}{dt}(t) &= \mathbf{v}'(t) - \langle \mathbf{v}'(t), N(t, \theta_0) \rangle N(t, \theta_0) \\ &= \frac{1}{\sqrt{2}} (\cos \theta_0 \sin t - \cos t, -\sin \theta_0 \cos t - \sin t, 0) + \\ &\quad + \frac{1}{\sqrt{2}} \sin \theta_0 (\sin \theta_0 \cos t, \sin \theta_0 \sin t, \cos \theta_0) \\ &= \frac{-\cos \theta_0}{\sqrt{2}} \left(\frac{1}{\sin \theta_0} \Phi_\varphi|_{(t, \theta_0)} + \Phi_\theta|_{(t, \theta_0)} \right). \quad \square \end{aligned}$$

It is worthy to note that the covariant derivative of $\mathbf{v}$ along α depends on the parameterization of α , but not on the orientation of the surface. Let us now establish its intrinsic character. In local coordinates, if we write $\mathbf{v}(u, v) = a(u, v)\Phi_u + b(u, v)\Phi_v$, $\alpha(t) = \Phi(u(t), v(t))$, $a(t) = a(u(t), v(t))$ and $b(t) = b(u(t), v(t))$, we have

$$\begin{aligned} \mathbf{v}(t) &= a(t)\Phi_u + b(t)\Phi_v, \\ \mathbf{v}'(t) &= a'\Phi_u + a(u'\Phi_{uu} + v'\Phi_{uv}) + b'\Phi_v + b(u'\Phi_{uv} + v'\Phi_{vv}) \\ &= (a' + au'\Gamma_{11}^1 + av'\Gamma_{12}^1 + bu'\Gamma_{12}^1 + bv'\Gamma_{22}^1)\Phi_u \\ &\quad + (b' + au'\Gamma_{11}^2 + av'\Gamma_{12}^2 + bu'\Gamma_{12}^2 + bv'\Gamma_{22}^2)\Phi_v \\ &\quad + (au'e + av'f + bu'f + bv'g)N. \end{aligned} \tag{4.9}$$

Ignoring the normal component, the components of $\mathbf{v}'(t)$ with respect to Φ_u and Φ_v give the expression of the covariant derivative $\frac{D\mathbf{v}}{dt}(t)$ in local coordinates. If we look at the coefficients of Φ_u and Φ_v , we see that they have “intrinsic character” and are therefore preserved by isometries. To formalize this observation, let us take a mapping $f: U \rightarrow S_2$ that is a diffeomorphism onto the image, and consider the vector field $\mathbf{u} = Df(\mathbf{v})$ which is the transfer of $\mathbf{v}$ by f , defined on $f(U)$ by $\mathbf{u}(f(p)) = Df_p(\mathbf{v}(p))$. Let us denote by $\frac{D\mathbf{u}}{dt}$ the covariant derivative of $\mathbf{u}$ along the curve $\beta(t) = f \circ \alpha(t)$. We claim that if f is an isometry, then

$$\frac{D\mathbf{u}}{dt} = Df_{\alpha(t)} \left(\frac{D\mathbf{v}}{dt} \right). \quad (4.10)$$

In fact, using coordinates $\Phi(u, v)$ on S and $\Psi(u, v) = f \circ \Phi(u, v)$ on S_2 , we see by (4.9) that the expressions of $\frac{D\mathbf{u}}{dt}$ with respect to the basis (Ψ_u, Ψ_v) and of $\frac{D\mathbf{v}}{dt}$ with respect to the basis (Φ_u, Φ_v) have the same coefficients, which proves (4.10).

Another property of the covariant derivative $\frac{D\mathbf{v}}{dt}$ of $\mathbf{v}$ along α is that at each instant it only depends on the velocity vector of the curve at that instant.

Proposition 4.3.2 $\frac{D\mathbf{v}}{dt}(t)$ only depends on $\alpha'(t) = u'(t)\Phi_u + v'(t)\Phi_v$ and not on the curve α .

Proof In fact, noting that

$$\begin{aligned} a'(t) &= \frac{\partial a}{\partial u} u'(t) + \frac{\partial a}{\partial v} v'(t), \\ b'(t) &= \frac{\partial b}{\partial u} u'(t) + \frac{\partial b}{\partial v} v'(t), \end{aligned}$$

we recognize that the functions $u(t)$, $v(t)$ only enter in (4.9) through the value they and their first derivatives u' and v' take at time t . $\square$

We further observe that the expression of $\frac{D\mathbf{v}}{dt}$ is linear in u' , v' . Thanks to Proposition 4.3.2, for every $p \in U$, a linear mapping $D\mathbf{v}_p: T_p S \rightarrow T_p S$ is defined as follows: given $\mathbf{w} \in T_p S$, let $\alpha(t)$ be a curve in U such that $\alpha(t_0) = p$ and $\alpha'(t_0) = \mathbf{w}$; then $D\mathbf{v}_p(\mathbf{w})$ is the covariant derivative of $\mathbf{v}$ along α computed at t_0 . Intuitively, $D\mathbf{v}_p(\mathbf{w})$ is a kind of directional derivative of $\mathbf{v}$ in the direction of $\mathbf{w}$.

Given a field $\mathbf{v}$ in $U \subseteq S$, the *divergence* of $\mathbf{v}$ is the mapping $\text{Div } \mathbf{v}: U \subseteq S \rightarrow \mathbb{R}$ such that, at every $p \in U$, its value $\text{Div } \mathbf{v}(p)$ is the trace of the linear mapping $D\mathbf{v}_p$.

Example 4.3.3 If $\mathbf{v}(u, v) = (f_1(u, v), f_2(u, v))$ is a vector field on an open subset $U \subseteq \mathbb{R}^2$, its covariant derivative along a curve $\alpha: I \rightarrow U$ is the usual derivative $(\mathbf{v} \circ \alpha)'(t)$, and therefore the linear mapping $d\mathbf{v}_p$ is just the derivative of $\mathbf{v}$ as the

mapping $U \rightarrow \mathbb{R}^2$. The matrix of $D\mathbf{v}_p$ with respect to the canonical basis of $\mathbb{R}^2$ is the Jacobian

$$\begin{bmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \end{bmatrix} \Big|_p$$

and its trace $\frac{\partial f_1}{\partial u} \Big|_p + \frac{\partial f_2}{\partial v} \Big|_p$ is the divergence of $\mathbf{v}$ computed at p . $\square$

Proposition 4.3.4 *Divergence is invariant under isometries: if $f: U \subseteq S \rightarrow S_2$ is an isometry into its image and $\mathbf{v}$ is a vector field on U , then $\text{Div}(Df\mathbf{v})(f(p)) = \text{Div } \mathbf{v}(p)$.*

Proof We put $\mathbf{u} = Df\mathbf{v}$, and start by rewriting formula (4.10) using the linear mappings

$$D\mathbf{u}_{f(p)}: T_{f(p)}S_2 \rightarrow T_{f(p)}S_2 \quad \text{and} \quad D\mathbf{v}_p: T_pS \rightarrow T_pS.$$

Let us fix a curve $\alpha(t)$ such that $\alpha(0) = p$, and let us put $\beta(t) = f \circ \alpha(t)$, $\mathbf{w}_1 = \alpha'(0)$ and $\mathbf{w}_2 = \beta'(0)$. Then we have $\frac{D\mathbf{u}}{dt}(0) = D\mathbf{u}_{f(p)}(\mathbf{w}_2)$ and $\frac{D\mathbf{v}}{dt}(0) = D\mathbf{v}_p(\mathbf{w}_1) = D\mathbf{v}_p(Df_p^{-1}(\mathbf{w}_2))$. By (4.10) we can write

$$D\mathbf{u}_{f(p)}(\mathbf{w}_2) = Df_p \circ D\mathbf{v}_p \circ Df_p^{-1}(\mathbf{w}_2) \text{ for every } \mathbf{w}_2 \in T_{f(p)}S_2,$$

i.e.

$$D\mathbf{u}_{f(p)} = Df_p \circ D\mathbf{v}_p \circ Df_p^{-1}, \quad (4.11)$$

which means that $D\mathbf{u}_{f(p)}$ and $D\mathbf{v}_p$ are conjugate linear mappings, and therefore have the same trace. $\square$

We will revisit the notion of divergence in the next section, where we prove a theorem that adapts the divergence theorem known from vector calculus to surfaces. For now, we return to the covariant derivative.

The reader must have noticed that in the computation of the covariant derivative of $\mathbf{v}$ along α , only the way $\mathbf{v}$ is defined on the trace of the curve plays a role, and so it is not necessary that the field be defined on other points. We thus define a *vector field along a regular curve* $\alpha(t)$ as a function $\mathbf{v}(t)$ such that, for every t , the vector $\mathbf{v}(t)$ belongs to $T_{\alpha(t)}S$; if $\mathbf{v}$ is differentiable (i.e., if, by writing $\mathbf{v}(t) = a(t)\Phi_u + b(t)\Phi_v$, the functions $a(t)$ and $b(t)$ are differentiable), we can as before compute the covariant derivative of $\mathbf{v}$ along α .

Lemma 4.3.5 *If $\mathbf{v}(t)$, $\mathbf{w}(t)$ are two vector fields along the curve $\alpha(t)$ then*

$$\frac{d}{dt} \langle \mathbf{v}, \mathbf{w} \rangle = \left\langle \frac{D\mathbf{v}}{dt}, \mathbf{w} \right\rangle + \left\langle \mathbf{v}, \frac{D\mathbf{w}}{dt} \right\rangle.$$

Proof Denoting by the superscript n the normal component of a vector, we have

$$\begin{aligned}
\frac{d}{dt} \langle \mathbf{v}, \mathbf{w} \rangle &= \langle \mathbf{v}', \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w}' \rangle \\
&= \left\langle \frac{D\mathbf{v}}{dt} + (\mathbf{v}')^n, \mathbf{w} \right\rangle + \left\langle \mathbf{v}, \frac{D\mathbf{w}}{dt} + (\mathbf{w}')^n \right\rangle \\
&= \left\langle \frac{D\mathbf{v}}{dt}, \mathbf{w} \right\rangle + \left\langle \mathbf{v}, \frac{D\mathbf{w}}{dt} \right\rangle. \quad \square
\end{aligned}$$

We now introduce the concept of *geodesic curvature*. Given a regular curve $\alpha: I \rightarrow S$ on an oriented surface, we can consider the unit tangent vector field along α , given by $\tau_1(t) = \frac{1}{|\alpha'(t)|} \alpha'(t)$, and also the vector field $\tau_2 = N \times \tau_1$. This means that at each instant t , the pair $(\tau_1(t), \tau_2(t))$ is an orthonormal and positively oriented basis of $T_{\alpha(t)}S$. By 4.3.5 we have

$$\left\langle \tau_1(t), \frac{D\tau_1}{dt}(t) \right\rangle = \frac{d}{dt} \left(\frac{1}{2} |\tau_1|^2 \right) = 0,$$

and so the vectors $\frac{D\tau_1}{dt}(t)$ and $\tau_2(t)$ are collinear. The *geodesic curvature* $k_g(t)$ of α at the point $\alpha(t)$ is defined by the equality

$$\frac{D\tau_1}{dt}(t) = |\alpha'(t)| k_g(t) \tau_2(t). \quad (4.12)$$

So the geodesic curvature gives us the tangential component of the curvature of α , and is also the generalization, for curves on oriented surfaces, of the signed curvature for planar curves (1.3). Like the signed curvature, the geodesic curvature depends only (up to sign) on the point of the curve at which it is computed and not on the parameterization. In fact, if $\alpha(t)$ and $\alpha(s)$ are two parameterizations of the same curve and $\mathbf{v}$ a vector field along α , then

$$\frac{D\mathbf{v}}{dt} = \frac{ds}{dt} \frac{D\mathbf{v}}{ds}, \quad (4.13)$$

$$|\alpha'(t)| = \left| \frac{ds}{dt} \right| |\alpha'(s)|. \quad (4.14)$$

Assuming that $\alpha(t)$ and $\alpha(s)$ have the same orientation (so that $\frac{ds}{dt} > 0$), we apply (4.13) to the vector field τ_1 and use (4.12) and (4.14), obtaining $k_g(t) = k_g(s)$, which proves our claim. For purposes of computation, we note that from (4.12) it follows that

$$\begin{aligned}
k_g &= \frac{1}{|\alpha'|} \left\langle \frac{D\tau_1}{dt}, \tau_2 \right\rangle = \frac{1}{|\alpha'|} \langle \tau_1', N \times \tau_1 \rangle \\
&= \frac{1}{|\alpha'|^3} \langle \alpha'', N \times \alpha' \rangle,
\end{aligned} \quad (4.15)$$

and formula (4.15) further simplifies to $k_g = \langle \alpha'', N \times \alpha' \rangle$ when α is parametrized by arc length.

From formula (4.15) it follows that the sign of the geodesic curvature is changed when we reverse orientation of the curve or change the orientation of the surface. In our next proposition we gather some simple properties of the geodesic curvature.

Proposition 4.3.6 (i) *The geodesic curvature is invariant under local isometries that preserve orientation — that is, if $f: U \subseteq S_1 \rightarrow S_2$ is such a mapping and $\alpha: I \rightarrow U$ is a regular curve, then the geodesic curvatures of α in S_1 (computed at $\alpha(t)$) and of $f \circ \alpha$ in S_2 (computed at $f \circ \alpha(t)$) are equal.*

(ii) *If $\alpha(t)$ is a regular curve in S , then*

$$k(t)^2 = k_g(t)^2 + k_n(t)^2,$$

where k , k_n , k_g are respectively the curvature, the normal curvature and the geodesic curvature of α .

Proof The derivative Df transforms the “moving” orthonormal frame $(\tau_1(t), \tau_2(t))$ that appears in the calculation (4.12) of the geodesic curvature of α into another “moving” orthonormal frame $\tau_1(t), \tau_2(t)$. Furthermore, τ_1 is the unit tangent vector field along the curve $f \circ \alpha$ and, if f preserves orientation, this second frame is also positively oriented. Statement (i) is then an immediate consequence of (4.10).

To prove (ii), it suffices to note that from the decomposition

$$\tau_1' = \frac{D\tau_1}{dt} + (\tau_1')^n$$

into the tangential and normal components, it follows that

$$\begin{aligned} |\tau_1'|^2 &= \left| \frac{D\tau_1}{dt} \right|^2 + |(\tau_1')^n|^2, \\ \text{i.e.} \quad |\alpha'(t)|^2 k(t)^2 &= |\alpha'(t)|^2 (k_g(t)^2 + k_n(t)^2). \quad \square \end{aligned}$$

Now let $\mathbf{v}$ be any vector field along α . We say that $\mathbf{v}$ is *parallel along α* if its covariant derivative along α is constantly zero. Assume that α is defined on $[a, b]$, that $\mathbf{w}_1 \in T_{\alpha(a)}S$, and let $\mathbf{v}$ be a parallel vector field along α such that $\mathbf{v}(a) = \mathbf{w}_1$. Then the vector $\mathbf{w}_2 = \mathbf{v}(b)$ in $T_{\alpha(b)}S$ is called the *parallel transport* of $\mathbf{w}_1$ along α from $\alpha(a)$ to $\alpha(b)$. To justify the use of the definite article **the**, we have to show that there is only one parallel vector field along α with initial position $\mathbf{w}_1$; this is done in the next proposition. However, we note first that by formula (4.13) the fact that $\mathbf{v}$ is parallel along α does not depend on the parameterization of the curve. Therefore the notion of parallel transport is also independent of the parameterization.

Proposition 4.3.7 *Let $\alpha: [a, b] \rightarrow S$ be a regular curve in S . Then:*

(i) *if $\mathbf{v}$ and $\mathbf{w}$ are parallel vector fields along α , the norms $|\mathbf{v}(t)|$, $|\mathbf{w}(t)|$ and the angle between $\mathbf{v}(t)$ and $\mathbf{w}(t)$ are constant;*

(ii) *given $\mathbf{w}_1 \in T_{\alpha(a)}S$, there exists one and only one parallel vector field $\mathbf{v}(t)$ along α such that $\mathbf{v}(a) = \mathbf{w}_1$.*

Proof (i) If $\mathbf{v}$ is a parallel vector field then by 4.3.5 we have

$$\frac{d}{dt} |\mathbf{v}|^2 = 2 \left\langle \frac{D\mathbf{v}}{dt}, \mathbf{v} \right\rangle = 0,$$

and so $|\mathbf{v}|$ is constant, and the same holds for $|\mathbf{w}|$ and for the inner product $\langle \mathbf{v}, \mathbf{w} \rangle$. Therefore each of these vector fields is either always or never zero. Assuming they are both nonzero, the angle $\theta(t) = \angle(\mathbf{v}(t), \mathbf{w}(t))$ is constant, since $\cos \theta(t) = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{|\mathbf{v}| |\mathbf{w}|}$.

(ii) It follows from (i) that two parallel vector fields $\mathbf{v}$ and $\mathbf{u}$ such that $\mathbf{v}(a) = \mathbf{u}(a)$ must be identical, since they both have constant norm equal to $|\mathbf{v}(a)|$, and the angle between them is always zero. Thus there exists at most one parallel vector field $\mathbf{v}(t)$ such that $\mathbf{v}(a) = \mathbf{w}_1$.

Assuming that $\mathbf{w}_1 \neq 0$, it remains to prove the existence of $\mathbf{v}$. For this, consider the vector fields $\boldsymbol{\tau}_1(t) = \frac{1}{|\alpha'(t)|} \alpha'(t)$ and $\boldsymbol{\tau}_2 = N \times \boldsymbol{\tau}_1$. We write $\mathbf{v}(t)$ in the form

$$\mathbf{v}(t) = a(t)\boldsymbol{\tau}_1(t) + b(t)\boldsymbol{\tau}_2(t).$$

Since $\mathbf{v}$ is parallel, $\mathbf{v}$ has constant norm $r = |\mathbf{w}_1|$. We can thus try to find $\varphi(t)$ such that $a(t) = r \cos \varphi(t)$ and $b(t) = r \sin \varphi(t)$. Note that

$$\left\langle \frac{D\boldsymbol{\tau}_2}{dt}, \boldsymbol{\tau}_1 \right\rangle = - \left\langle \frac{D\boldsymbol{\tau}_1}{dt}, \boldsymbol{\tau}_2 \right\rangle = -k_g |\alpha'|$$

and therefore $\frac{D\boldsymbol{\tau}_2}{dt} = -k_g |\alpha'| \boldsymbol{\tau}_1$. Noting that the covariant derivative obeys the usual rules of differentiation, we then have

$$\begin{aligned} \frac{1}{r} \frac{D\mathbf{v}}{dt} &= -\varphi \sin \varphi \boldsymbol{\tau}_1 + \cos \varphi \frac{D\boldsymbol{\tau}_1}{dt} + \varphi \cos \varphi \boldsymbol{\tau}_2 + \sin \varphi \frac{D\boldsymbol{\tau}_2}{dt} \\ &= (\varphi + k_g |\alpha'|)(-\sin \varphi \boldsymbol{\tau}_1 + \cos \varphi \boldsymbol{\tau}_2), \end{aligned}$$

so that $\mathbf{v}$ is a parallel vector field along α if and only if $\varphi'(t) = -k_g(t)|\alpha'(t)|$. To finish, it is therefore sufficient to define $\varphi(t) = \varphi_0 - \int_a^t k_g(s)|\alpha'(s)| ds$, where φ_0 satisfies $\mathbf{w}_1 = r \cos \varphi_0 \boldsymbol{\tau}_1(a) + r \sin \varphi_0 \boldsymbol{\tau}_2(a)$. $\square$

Observations and Examples 4.3.8 A. Every local isometry f sends a parallel vector field along α to a parallel vector field along $f \circ \alpha$. This is a consequence of formula (4.10).

B. In the plane a vector field is parallel along a certain curve if and only if it is constant along that curve, which shows that in this case the parallel transport only depends on the initial and final points of the curve, not on the path covered. This property is shared (locally) by those surfaces which are locally isometric to the plane, such as the cylinder and the cone: thus, each point of these surfaces has an open neighborhood U such that the parallel transport along any curve **contained in** U depends only on the initial and final points of the curve. Later on we will conclude that it is only on surfaces of zero curvature that this holds true.

C. If the surfaces S_1 and S_2 are tangent along the regular curve $\gamma(t)$, the covariant derivative of a vector field $\mathbf{v}$ along γ is the same computed with respect to S_1 or S_2 (since the normal component we subtract from $\mathbf{v}'(t)$ is the same in both cases), and therefore also the parallel transport along γ is the same on the two surfaces. Using this observation, let us determine the parallel transport on the sphere $\mathbf{S}^2$ along the parallel of colatitude θ_0 , which we denote by γ .

Consider the cone of revolution $\mathfrak{C}$ tangent to the sphere along γ . The angle that the generatrices of $\mathfrak{C}$ make with the axis is $\frac{\pi}{2} - \theta_0$, and therefore (see example 4.1.3 D) $\mathfrak{C}$ is isometric to a planar region U that defines an angle $\beta = 2\pi \cos \theta_0$ (see Figure 4.3). By such an isometry, the parallel transport along γ corresponds to the parallel transport in U along the arc with center O and radius equal to $\text{tg } \theta_0$. However, in U parallel transport is simply translation; but a vector $\mathbf{w}_1$ which is tangent to the generatrix of the cone is translated to a vector $\mathbf{w}_2$ which makes an angle of β with that same generatrix (the generatrices of $\mathfrak{C}$ correspond in U to the half-lines of origin O , and the two half-lines bounding U are identified with the same generatrix).

In conclusion: parallel transport in $\mathbf{S}^2$ along one complete turn of the parallel $\theta = \theta_0$ makes each vector rotate through an angle of $2\pi \cos \theta_0$ at the end of a complete turn. $\square$

We end this section by defining one of the most important concepts of Differential Geometry: a *geodesic* of the surface S is a regular curve $\alpha(t)$ on S whose geodesic curvature is constantly zero. By

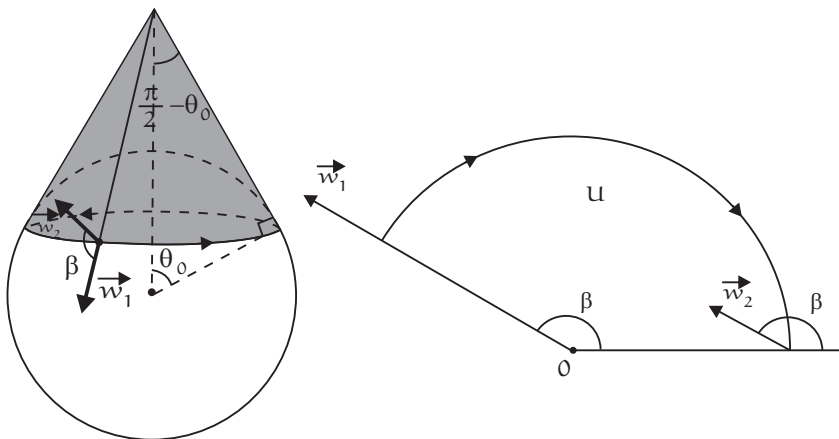


Figure 4.3

(4.12), this is equivalent to the unit vector field $\tau_1 = \frac{1}{|\alpha'|} \alpha'$ tangent to α being parallel. This simply means that at points where the curvature of α is nonzero, the principal normal to the curve is also normal to the surface.

We leave the in-depth study of geodesics to a later section; to finish this one, we give some simple examples.

Examples 4.3.9 Geodesics on the plane are simply straight lines, since in the plane geodesic curvature is the same as signed curvature. On the sphere any maximal circle is a geodesic, because the principal normal to a maximal circle passes through the center of the sphere, and therefore coincides with the normal to the sphere itself (later on we will see that there are no other geodesics on the sphere). More generally, the meridians of any surface of revolution are geodesics. $\square$

Exercises

85. Consider on the sphere $\mathbf{S}^2$ the meridian $\varphi = 0$, which we denote by γ , and the points $\mathcal{N} = (0, 0, 1)$ and $\mathcal{S} = (0, 0, -1)$. We define a mapping $L_\gamma: T_{\mathcal{N}}\mathbf{S}^2 \rightarrow T_{\mathcal{S}}\mathbf{S}^2$ as follows: $L_\gamma(\mathbf{v})$ is the parallel transport of the vector $\mathbf{v}$ from $\mathcal{N}$ to $\mathcal{S}$ along γ .

- (a) Compute $L_\gamma(1, 0, 0)$ and $L_\gamma(0, 1, 0)$.
- (b) Show that L_γ is linear.
- (c) Does L_γ depend on the choice of meridian?

86. (a) Describe all the geodesics of the cylinder $x^2 + y^2 = 1$.

(b) Compare the geodesic curvatures of the same helix in the cylinder and in the helicoid.

87. (a) Let $S = \{\alpha(t) + \lambda \mathbf{v}(t): t \in I, \lambda \in \mathbb{R}\}$ be a ruled surface. Show that the lines $r_t(\lambda) = \alpha(t) + \lambda \mathbf{v}(t)$ are all geodesics, and find out whether the curves $c_\lambda(t) = \alpha(t) + \lambda \mathbf{v}(t)$ are geodesics.

(b) Show that on any surface an asymptotic line that is also a geodesic is necessarily a (segment of) straight line.

88. Show that on the surface of revolution

$$\Phi(u, v) = (\rho(v) \cos u, \rho(v) \sin u, z(v)),$$

the parallel $v = v_0$ is a geodesic if and only if the tangent line to the generating curve at v_0 is parallel to the axis of revolution.

89. (a) Check that in the cone $\mathcal{C}_k = \{(x, y, z): z \neq 0, z = k\sqrt{x^2 + y^2}\}$ there is always some geodesic connecting each pair of points.

(b) The generatrices $t \mapsto t(x_0, y_0, k\sqrt{x_0^2 + y_0^2})$ are geodesics of $\mathcal{C}_k$ that are not defined for $t \leq 0$. Show that any other geodesic of $\mathcal{C}_k$ parametrized with constant scalar velocity extends to all values of the parameter.

90. Let $\mathcal{E}$ be the intersection of the cylinder $x^2 + y^2 = 1$ with a plane which passes through the x -axis and makes an angle θ with the xy plane. Show that $\mathcal{E}$ is an ellipse, and compute the absolute value of the geodesic curvature of $\mathcal{E}$ (relative to the cylinder) at the points where $\mathcal{E}$ intersects its axes.

91. Show that the geodesic curvature of the curve α at $p \in S$ is equal to the curvature at p of the curve obtained by projecting α orthogonally on $p + T_p S$.

4.4 The Divergence Theorem. First Variation of Area

In this section we give a simple expression for the divergence of a vector field and prove the divergence theorem. We also prove a formula for the variation of the area of a surface when it undergoes a perturbation induced by a vector field (and mention how this is related to minimal surfaces). These results are used in Section 5.4 in the proof of the Minkowski integral formulas (which in turn we use to establish the rigidity of the sphere and in the proof of Theorem 5.5.3) and otherwise are not used in the remainder of this book. This being said, we must acknowledge that some subjects treated here do not pertain to intrinsic geometry, but we could find no better place to fit them in.

Let $\mathbf{v}$ be a vector field on an open subset U of an oriented surface S . In the previous section we defined the linear mapping $D\mathbf{v}_p$ as follows: given $\mathbf{w} \in T_p S$, $D\mathbf{v}_p(\mathbf{w})$ is the covariant derivative of $\mathbf{v}$, computed at p , along any curve in S that passes through p with velocity $\mathbf{w}$. Using coordinates $\Phi(u, v)$, the vectors $D\mathbf{v}_{\Phi(u, v)}(\Phi_u)$ and $D\mathbf{v}_{\Phi(u, v)}(\Phi_v)$ are then the tangential components of $\frac{\partial \mathbf{v}}{\partial u}$ and $\frac{\partial \mathbf{v}}{\partial v}$ respectively — that is, if we write

$$\begin{aligned}\frac{\partial \mathbf{v}}{\partial u} &= b_{11}\Phi_u + b_{21}\Phi_v + \lambda_1 N \\ \frac{\partial \mathbf{v}}{\partial v} &= b_{12}\Phi_u + b_{22}\Phi_v + \lambda_2 N,\end{aligned}\tag{4.16}$$

then the matrix of $D\mathbf{v}_{\Phi(u, v)}$ relative to the basis (Φ_u, Φ_v) of $T_{\Phi(u, v)}S$ is

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

Divergence is the trace of this matrix. That is,

$$\text{Div } \mathbf{v}(\Phi(u, v)) = \text{tr } B = b_{11} + b_{22}.$$

Our goal is to obtain an explicit formula for divergence. Using (4.16), we have

$$\frac{\partial \mathbf{v}}{\partial u} \times \Phi_v + \Phi_u \times \frac{\partial \mathbf{v}}{\partial v} = (b_{11} + b_{22})(\Phi_u \times \Phi_v) + \lambda_1(N \times \Phi_v) + \lambda_2(\Phi_u \times N),$$

and therefore

$$\begin{aligned}\left\langle \frac{\partial \mathbf{v}}{\partial u} \times \Phi_v + \Phi_u \times \frac{\partial \mathbf{v}}{\partial v}, N \right\rangle &= (b_{11} + b_{22})\langle \Phi_u \times \Phi_v, N \rangle \\ &= (\text{Div } \mathbf{v})\sqrt{EG - F^2},\end{aligned}\tag{4.17}$$

provided that $\Phi(u, v)$ is compatible with the orientation of S .

The expression (4.17) for $\text{Div } \mathbf{v}$ can still be simplified. It is convenient, for the calculations below, to work with a vector field $\mathbf{v}$ which is **not necessarily** tangent to S . By letting $\mathbf{v} = \alpha\Phi_u + \beta\Phi_v + \gamma N$, and using the matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

of $-DN_{\Phi(u,v)}$ relative to the basis (Φ_u, Φ_v) to express the vectors N_u and N_v (see (3.6) in Section 3.2), we have

$$\begin{aligned} \Phi_v \times N_u + N_v \times \Phi_u &= -\Phi_v \times (a_{11}\Phi_u + a_{21}\Phi_v) - (a_{12}\Phi_u + a_{22}\Phi_v) \times \Phi_u \\ &= (a_{11} + a_{22})(\Phi_u \times \Phi_v) \\ &= 2H\sqrt{EG - F^2}N \end{aligned} \quad (4.18)$$

$$\begin{aligned} \langle \mathbf{v}, \Phi_v \times N \rangle &= \langle \alpha\Phi_u + \beta\Phi_v + \gamma N, \Phi_v \times N \rangle \\ &= \langle \alpha\Phi_u, \Phi_v \times N \rangle = \alpha \langle \Phi_u \times \Phi_v, N \rangle \\ &= \alpha\sqrt{EG - F^2} \end{aligned} \quad (4.19)$$

$$\langle \mathbf{v}, N \times \Phi_u \rangle = \beta\sqrt{EG - F^2} \quad (4.20)$$

Pursuing our calculations, we further have

$$\begin{aligned} \left\langle \frac{\partial \mathbf{v}}{\partial u} \times \Phi_v + \Phi_u \times \frac{\partial \mathbf{v}}{\partial v}, N \right\rangle &= \left\langle \frac{\partial \mathbf{v}}{\partial u}, \Phi_v \times N \right\rangle + \left\langle \frac{\partial \mathbf{v}}{\partial v}, N \times \Phi_u \right\rangle \\ &= \left\langle -\mathbf{v}, \frac{\partial}{\partial u}(\Phi_v \times N) + \frac{\partial}{\partial v}(N \times \Phi_u) \right\rangle + \\ &\quad + \frac{\partial}{\partial u} \langle \mathbf{v}, \Phi_v \times N \rangle + \frac{\partial}{\partial v} \langle \mathbf{v}, N \times \Phi_u \rangle \\ &= -\langle \mathbf{v}, \Phi_v \times N_u + N_v \times \Phi_u \rangle + \\ &\quad + \frac{\partial}{\partial u} \langle \mathbf{v}, \Phi_v \times N \rangle + \frac{\partial}{\partial v} \langle \mathbf{v}, N \times \Phi_u \rangle \\ &= -\sqrt{EG - F^2} \langle \mathbf{v}, 2HN \rangle + \\ &\quad + \frac{\partial}{\partial u} (\alpha\sqrt{EG - F^2}) + \frac{\partial}{\partial v} (\beta\sqrt{EG - F^2}), \end{aligned}$$

where we apply (4.18), (4.19), and (4.20). Of course, when the vector field $\mathbf{v}$ is tangent to S , the first summand of the last expression vanishes. Combining this formula with (4.17), we obtain the desired formula for divergence:

Proposition 4.4.1 *Let $\mathbf{v}$ be a (not necessarily tangent) vector field on an open subset U covered by a parameterization $\Phi(u, v)$ compatible with the orientation of S . Then:*

(i) *if we denote by $\mathbf{v}^\top$ the tangential component of $\mathbf{v}$, then we have*

$$\frac{1}{\sqrt{EG - F^2}} \left\langle \frac{\partial \mathbf{v}}{\partial u} \times \Phi_v + \Phi_u \times \frac{\partial \mathbf{v}}{\partial v}, N \right\rangle = -\langle \mathbf{v}, 2HN \rangle + \text{Div}(\mathbf{v}^\top);$$

(ii) if $\mathbf{v}$ is a tangent field to S and we write $\mathbf{v} = \alpha\Phi_u + \beta\Phi_v$, then we have

$$\operatorname{Div} \mathbf{v} = \frac{1}{\sqrt{EG - F^2}} \left\{ \frac{\partial}{\partial u} (\alpha \sqrt{EG - F^2}) + \frac{\partial}{\partial v} (\beta \sqrt{EG - F^2}) \right\}.$$

The divergence theorem reduces the calculation of a certain surface integral on a simple region to a line integral along its boundary. By *simple region* (or *Jordan region*) $\Omega \subseteq S$ we mean a connected region whose boundary is a simple closed curve and whose closure (in S) is homeomorphic to a closed disk (in the plane); if the boundary of Ω is piecewise regular, then Ω is also called a *polygonal region*. We now state the most important result in this section.

Divergence Theorem 4.4.2 *Let $\mathbf{v}$ be a field of tangent vectors on S , and let $\Omega \subseteq S$ be a polygonal region. Then*

$$\int_{\Omega} \operatorname{Div} \mathbf{v} d\sigma = \int_{\gamma} \langle \mathbf{v}, -\boldsymbol{\tau}_2 \rangle ds, \quad (4.21)$$

where $\gamma(s)$ is the boundary of Ω and $\boldsymbol{\tau}_2(s) \in T_{\gamma(s)}S$ is the unit vector orthogonal to $\gamma'(s)$ that points to the interior of Ω .

Proof We assume that $\gamma(s)$ is parametrized by arc length and runs in the positive direction — i.e., in such a way that Ω is always to the left of γ (in other words, we have $\boldsymbol{\tau}_2(s) = N(s) \times \gamma'(s)$) (see Fig. 4.4). Of course, $\boldsymbol{\tau}_2$ is not defined at the vertices of γ , but this happens only for a finite number of values of s , which do not affect the integration.

It is sufficient to prove the theorem under the hypothesis that the closure of Ω is contained in some parametrized neighborhood. For if this is not the case, we can decompose Ω into a finite number of sufficiently small polygonal regions $(\Omega_i)_{i=1}^k$. Supposing that (4.21) holds for the Ω_i , and

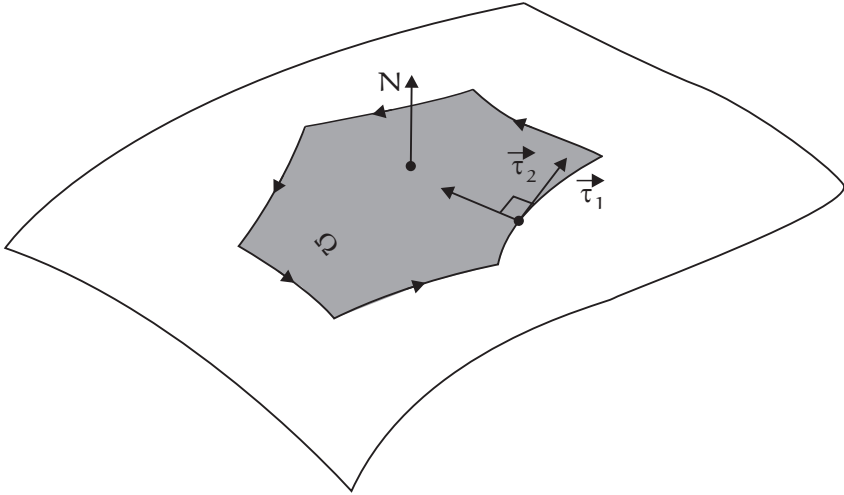


Figure 4.4

denoting by γ_i the boundary of Ω_i moving in the positive direction and by τ_2^i the associated unit vector, we obtain

$$\begin{aligned} \int_{\Omega} \text{Div } \mathbf{v} \, d\sigma &= \sum_{i=1}^k \int_{\Omega_i} \text{Div } \mathbf{v} \, d\sigma = \\ &= \sum_{i=1}^k \int_{\gamma_i} \langle \mathbf{v}, -\tau_2^i \rangle \, ds = \int_{\gamma} \langle \mathbf{v}, -\tau_2 \rangle \, ds \end{aligned}$$

– since any “edge” (segment of some γ_i) inside Ω is run through twice, with the integrand function $\langle \mathbf{v}, -\tau_2^i \rangle$ taking up opposite signs in both instances, and therefore we are left with only the integrals relative to the edges that are part of γ .

We can thus assume that Ω is covered by the parameterization $\Phi(u, v)$, and we put $\gamma(s) = \Phi(u(s), v(s))$ for every $s \in [0, L]$. Furthermore, we require that $\Phi(u, v)$ be compatible with the orientation of S . We thus have, with the notation of 4.4.1 and using Green’s theorem

$$\begin{aligned} \int_{\Omega} \text{Div } \mathbf{v} \, d\sigma &= \iint_{\Phi^{-1}(\Omega)} \left\{ \frac{\partial}{\partial u} (\alpha \sqrt{EG - F^2}) + \frac{\partial}{\partial v} (\beta \sqrt{EG - F^2}) \right\} \, du \, dv \\ &= \int_0^L \sqrt{EG - F^2} (\alpha(s)v'(s) - \beta(s)u'(s)) \, ds. \end{aligned}$$

(Note that, since Φ is compatible with the orientation of S , the curve $(u(s), v(s))$ moves along the boundary of $\Phi^{-1}(\Omega)$ in the positive direction, and we have used this fact when we applied Green’s theorem.) To simplify the latter integral, we form the vector product of the two equalities

$$\mathbf{v}(s) = \alpha(s)\Phi_u + \beta(s)\Phi_v, \quad \gamma'(s) = u'(s)\Phi_u + v'(s)\Phi_v,$$

obtaining

$$\begin{aligned}\mathbf{v}(s) \times \gamma'(s) &= (\alpha(s)v'(s) - \beta(s)u'(s))(\Phi_u \times \Phi_v) \\ &= \sqrt{EG - F^2}(\alpha(s)v'(s) - \beta(s)u'(s))N.\end{aligned}$$

The above integral can then be rewritten as

$$\begin{aligned}\int_0^L \langle \mathbf{v}(s) \times \gamma'(s), N \rangle ds &= \int_0^L \langle \mathbf{v}(s), \gamma'(s) \times N \rangle ds \\ &= \int_0^L \langle \mathbf{v}(s), -\boldsymbol{\tau}_2(s) \rangle ds = \int_\gamma \langle \mathbf{v}, -\boldsymbol{\tau}_2 \rangle ds. \quad \square\end{aligned}$$

The divergence theorem can be generalized to regions bounded by more than one closed curve, making the sum of all integrals relative to each curve constituting the boundary of Ω appear in the right-hand side of (4.21). However, it is more interesting, in the case of a compact surface S , to consider the integral of $\text{Div } \mathbf{v}$ over the whole surface. Choosing a “polygonal decomposition” $(\Omega_i)_{i=1}^k$ of S , and applying 4.4.2 to each of the Ω_i , what happens, when adding up the results, is that all the terms vanish, because now none of the edges is run through only once. Thus we have just proved (i) of the corollary below.

Corollary 4.4.3 *Let $\mathbf{v}$ be a vector field on a compact surface S . Then:*

- (i) $\int_S \text{Div } \mathbf{v} d\sigma = 0$;
- (ii) *there exists $p \in S$ such that $\text{Div } \mathbf{v}(p) = 0$.*

To prove (ii) we can assume that S is connected, because in any case its connected components are compact. Hence, if $\text{Div } \mathbf{v}$ has no zeros, then it has constant sign and therefore $\int_S \text{Div } \mathbf{v} d\sigma \neq 0$, in contradiction to (i). Therefore $\text{Div } \mathbf{v}$ has some zero, which proves (ii).

Given a differentiable vector field $\mathbf{v}$, not necessarily tangent, on a **compact** surface S , let $S_t = \{p + t\mathbf{v}(p) : p \in S\}$. The family $(S_t)_t$ is called a *variation* of S (when $\mathbf{v}$ is a normal field to S , not necessarily unitary, $(S_t)_t$ is called a *normal variation* of S); see Fig. 4.5. Denoting by $A(t)$ the area of S_t , our goal is to prove a formula for $A'(0)$ as an integral over S . First, however, we must guarantee that S_t is a surface for sufficiently small t , which we do in the proposition below.

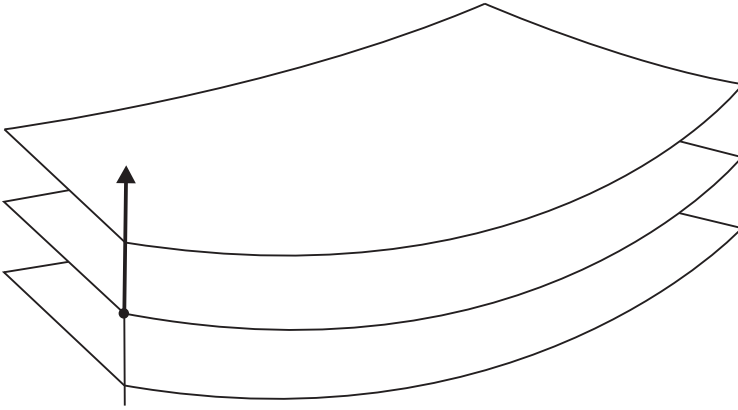


Figure 4.5

Proposition 4.4.4 *There exists $\eta > 0$ such that, for all $t \in]-\eta, \eta[$, S_t is a compact surface.*

Proof We define the mapping $L_t: S \rightarrow \mathbb{R}^3$ by $L_t(p) = p + t\mathbf{v}(p)$, and so S_t is the image of S under L_t . Using this notation, let us start by proving two auxiliary claims:

Claim 1. *For every point p of S , there exist a parametrized neighborhood $\Phi(u, v)$ of p and $\varepsilon > 0$ such that, for all $t \in]-\varepsilon, \varepsilon[$, the mapping*

$$L_t \circ \Phi(u, v) = \Phi(u, v) + t\mathbf{v}(u, v)$$

is injective and its Jacobian matrix has rank two.

Taking a parameterization $\Phi(u, v) = (x(u, v), y(u, v), z(u, v))$ with $\Phi(0, 0) = p$, we can assume that

$$\frac{\partial(x, y)}{\partial(u, v)} \Big|_{(0,0)} \neq 0.$$

Denoting by x_t, y_t, z_t the component functions of $L_t \circ \Phi$, we define

$$F(u, v, t) = (x_t(u, v), y_t(u, v), t) = (G_t(u, v), t).$$

It is easily seen that $\det(JF_{(u,v,t)}) = \det(F(G_t)_{(u,v)})$. In particular, we have

$$\det(JF_{(0,0,0)}) = \det(J(G_0)_{(0,0)}) = \frac{\partial(x, y)}{\partial(u, v)} \Big|_{(0,0)} \neq 0.$$

Hence, by the inverse mapping theorem, there exists an open neighborhood $U \times]-\varepsilon, \varepsilon[$ of $(0, 0, 0)$ such that the restriction of F to this neighborhood is a diffeomorphism onto its image. The local inverse of F has the form $F^{-1}(x, y, t) = ((G_t)^{-1}(x, y), t)$, and so, for every $t \in]-\varepsilon, \varepsilon[$, the mapping $G_t|_U$ has a differentiable inverse, which guarantees that it is injective and its Jacobian matrix has rank two. Since $L_t \circ \Phi(u, v) =$

$(G_t(u, v), z_t(u, v))$, the same holds for $L_t \circ \Phi|_U$. The parameterization (U, Φ) thus satisfies the conditions of Claim 1.

Claim 2. For $|t|$ sufficiently small, L_t is injective.

By compactness of S , we can find $\varepsilon > 0$ and parameterizations (U_i, Φ^i) , with $i = 1, \dots, k$, which cover S and such that each of them satisfies Claim 1 for the given ε . Next, we take $\delta > 0$ such that, for every $p \in S$, the ball $B_\delta(p, S) = \{q \in S : |q - p| < \delta\}$ is contained in one of the open subsets $\Phi^i(U_i)$, and we put $M = \max_{p \in S} |\mathbf{v}(p)|$ and $\eta = \min\{\frac{\delta}{4M}, \varepsilon\}$. We claim that L_t is injective for $|t| < \eta$. In fact, if $|q - p| < \delta$ it is clear that $L_t(p) \neq L_t(q)$, since in this case p and q both belong to some $\Phi^i(U_i)$ and, by Claim 1, the restriction of L_t to $\Phi^i(U_i)$ is injective. If $|q - p| \geq \delta$ then

$$\begin{aligned} |L_t(p) - L_t(q)| &= |(p - q) - t(\mathbf{v}(p) - \mathbf{v}(q))| \\ &\geq |p - q| - |t|(|\mathbf{v}(p)| + |\mathbf{v}(q)|) \\ &\geq \delta - 2\eta M \geq \frac{\delta}{2}, \end{aligned}$$

which concludes the proof of Claim 2.

Finally we show that S_t is a surface when $|t| < \eta$. Since L_t is continuous and injective and S is compact, L_t is a homeomorphism $S \rightarrow S_t$ (in particular, S_t is also compact). Now, since the images of the mappings $L_t \circ \Phi^i$, $i = 1, \dots, k$, cover S_t , it suffices to show that each of them is a parameterization. By Claim 1, the Jacobian of $L_t \circ \Phi^i$ has rank two; furthermore, $(L_t \circ \Phi^i)^{-1}$ is continuous since it is given by the composite $(\Phi^i)^{-1} \circ L_t^{-1}$ of continuous functions. This shows that $L_t \circ \Phi^i$ is a parameterization and concludes the proof of the proposition. $\square$

It deserves mention that the mapping L_t is a diffeomorphism $S \rightarrow S_t$. In fact, and as we have seen, it is a homeomorphism. Its expression in local coordinates (U_i, Φ^i) on S and $(U_i, L_t \circ \Phi^i)$ on S_t is clearly a diffeomorphism, since it is just the identity $U_i \rightarrow U_i$. This means that all surfaces of the variation $(S_t)_{-\eta < t < \eta}$ are diffeomorphic to S .

Recalling that $A(t)$ denotes the area of the surface S_t , let us now present the formula for $A'(0)$.

Theorem 4.4.5 (First variation of area). $A'(0) = - \int_S \langle \mathbf{v}, 2HN \rangle d\sigma$.

Proof We can consider a polygonal decomposition $(\Omega_j)_{1 \leq j \leq r}$ of S such that the closure of each Ω_j is contained in the image of one of the parameterizations (U_i, Φ^i) defined in the proof of 4.4.4. Thus, for $|t| < \eta$, the sets $\Omega_j^t = L_t(\Omega_j)$ form a decomposition of S_t and each of them is covered by some parameterization

$$L_t \circ \Phi^i(u, v) = \Phi^i(u, v) + t\mathbf{v}(u, v).$$

We fix j and the corresponding i , and we let $W_j = (\Phi^i)^{-1}(\Omega_j^t)$. The area of Ω_j^t is given by

$$A_j(t) = \iint_{W_j} |(L_t \circ \Phi^i)_u \times (L_t \circ \Phi^i)_v| du dv$$

and therefore

$$A'_j(0) = \iint_{W_j} \frac{\partial}{\partial t} |(L_t \circ \Phi^i)_u \times (L_t \circ \Phi^i)_v| \Big|_{t=0} dudv.$$

Now we have

$$|(L_t \circ \Phi^i)_u \times (L_t \circ \Phi^i)_v| = \left| (\Phi_u^i \times \Phi_v^i) + t \left(\frac{\partial \mathbf{v}}{\partial u} \times \Phi_v^i + \Phi_u^i \times \frac{\partial \mathbf{v}}{\partial v} \right) + t^2 \left(\frac{\partial \mathbf{v}}{\partial u} \times \frac{\partial \mathbf{v}}{\partial v} \right) \right|.$$

Therefore, using 4.4.1 (i), we have

$$\begin{aligned} \frac{\partial}{\partial t} |(L_t \circ \Phi^i)_u \times (L_t \circ \Phi^i)_v| \Big|_{t=0} &= \frac{\langle \frac{\partial}{\partial t} ((L_t \circ \Phi^i)_u \times (L_t \circ \Phi^i)_v) \Big|_{t=0}, \Phi_u^i \times \Phi_v^i \rangle}{|\Phi_u^i \times \Phi_v^i|} \\ &= \left\langle \frac{\partial \mathbf{v}}{\partial u} \times \Phi_v^i + \Phi_u^i \times \frac{\partial \mathbf{v}}{\partial v}, N \right\rangle \\ &= \{ -\langle \mathbf{v}, 2HN \rangle + \text{Div}(\mathbf{v}^\top) \} |\Phi_u^i \times \Phi_v^i|, \end{aligned}$$

where $\mathbf{v}^\top$ denotes the tangential component of $\mathbf{v}$. We thus conclude that

$$A'_j(0) = \int_{\Omega_i} \{ -\langle \mathbf{v}, 2HN \rangle + \text{Div}(\mathbf{v}^\top) \} d\sigma$$

– and, using 4.4.3 (i), we finally obtain

$$\begin{aligned} A'(0) &= \sum_{j=1}^r A'_j(0) = \int_S \{ -\langle \mathbf{v}, 2HN \rangle + \text{Div}(\mathbf{v}^\top) \} d\sigma \\ &= - \int_S \langle \mathbf{v}, 2HN \rangle d\sigma. \quad \square. \end{aligned}$$

Observation 4.4.6 If S is not compact but the vector field $\mathbf{v}$ has compact support (the *support* of $\mathbf{v}$ is the closure in S of the set $\{p \in S: \mathbf{v}(p) \neq 0\}$), then the sets S_t are still surfaces for $|t|$ small and, suitably interpreted, formula 4.4.5 remains valid. We choose a compact region $\Omega \subseteq S$ whose boundary is made up of a finite number of piecewise regular closed curves and whose closure contains the support of $\mathbf{v}$, and we take a polygonal decomposition $(\Omega_j)_{1 \leq j \leq r}$ of Ω . We denote by $A(t)$ the area of $L_t(\Omega) \subseteq S_t$. Proceeding as in 4.4.5, and using the divergence theorem, we obtain

$$\begin{aligned} A'(0) &= - \int_{\Omega} \langle \mathbf{v}, 2HN \rangle d\sigma + \int_{\partial\Omega} \langle \mathbf{v}^\top, -\boldsymbol{\tau}_2 \rangle ds \\ &= - \int_S \langle \mathbf{v}, 2HN \rangle d\sigma, \end{aligned}$$

where $\partial\Omega$ indicates the boundary of Ω (restricted to which $\mathbf{v}^\top$ vanishes).

In particular, if $H \equiv 0$ then $A'(0) = 0$. A *minimal surface* is a surface whose mean curvature H is constantly zero (by Exercise 57 in Section 3.1, no such surface can be compact). This name arose as follows: assume that S has the property that, for every

region $\Omega \subseteq S$ as above, the area of Ω is less than or equal to that of any other portion $\tilde{\Omega}$ of the surface such that $\partial\tilde{\Omega} = \partial\Omega$. Then S is a minimal surface: indeed, denoting by $A(t)$ the area of $L_t(\Omega)$ (for a variation with support contained in Ω), $A(t)$ has a minimum at 0 and therefore $A'(0) = 0$. This implies that for every vector field $\mathbf{v}$ with compact support we have $-\int_S \langle \mathbf{v}, 2HN \rangle d\sigma = 0$ — which is only possible with $H \equiv 0$.

Not all minimal surfaces minimize area in the sense just stated, but the name has stuck. Besides the plane, the reader can verify (using formula (3.8) in the Section 3.2) that the helicoid and the catenoid are minimal surfaces (Exercise 76). In the exercises of this section we prove that besides the plane the catenoid is the only minimal surface of revolution.

For more examples and an in-depth study of minimal surfaces we recommend Osserman's book [19].

Exercises

92. (a) Compute the divergence of the vector field $\mathbf{v}(\varphi, \theta)$ in $\mathbf{S}^2$ of example 4.3.1. For that example, directly compute $\int_{\mathbf{S}^2} \text{Div } \mathbf{v} d\sigma$. Could you use corollary 4.4.3?

(b) Let S be a compact surface, let $F \subseteq S$ be a **finite** set, and let $\mathbf{v}$ be a field of tangent vectors defined on $S \setminus F$ such that the set $\sup \{|\mathbf{v}(p)| : p \in S \setminus F\}$ is bounded. Show that $\int_S \text{Div } \mathbf{v} d\sigma = 0$.

93. If $\mathbf{v}$ is a tangent field to S and $f: S \rightarrow \mathbb{R}$ a differentiable function, then

$$\text{Div}(f\mathbf{v})(p) = Df_p(\mathbf{v}) + f(p) \text{Div } \mathbf{v}(p).$$

94. Given a differentiable function $f: S \rightarrow \mathbb{R}$, the *gradient* of f is the tangent vector field ∇f defined on S as follows: for each $p \in S$ and all $\mathbf{v} \in T_p S$, we have $Df_p(\mathbf{v}) = \langle \nabla f(p), \mathbf{v} \rangle$. Equivalently, ∇f is the gradient of f if we have $\alpha: I \rightarrow S$ for every differentiable curve $(f \circ \alpha)'(t) = \langle \nabla f(\alpha(t)), \alpha'(t) \rangle$.

(a) Show that in local coordinates $\Phi(u, v)$ we have

$$\begin{aligned} \nabla f &= \frac{1}{\sqrt{EG - F^2}} N \times \left(\frac{\partial f}{\partial v} \Phi_u - \frac{\partial f}{\partial u} \Phi_v \right) \\ &= \frac{1}{EG - F^2} \left\{ \left(-\frac{\partial f}{\partial v} F + \frac{\partial f}{\partial u} G \right) \Phi_u + \left(-\frac{\partial f}{\partial v} E - \frac{\partial f}{\partial u} F \right) \Phi_v \right\}. \end{aligned}$$

(b) Conclude that the vector field $\mathbf{w} = N \times \nabla f$ has zero divergence. Conversely, show that if $\text{Div } \mathbf{w} \equiv 0$ then $\mathbf{v} = \mathbf{w} \times N$ is locally a gradient field (i.e., each point of S has a neighborhood U such that $\mathbf{v}|_U$ is the gradient of some function $U \rightarrow \mathbb{R}$).

(c) Show that a vector field $\mathbf{v}$ on S is a gradient field if and only if, for every curve piecewise differentiable $\alpha: [a, b] \rightarrow S$, the line integral $\int_a^b \langle \mathbf{v} \circ \alpha(t), \alpha'(t) \rangle dt$ depends only on the initial and final points of α .

95. Let $\Phi(u, s) = (\rho(s) \cos u, \rho(s) \sin u, z(s))$ be a parameterization of a surface of revolution S , where s is the arc length of the generating curve (i.e., $\rho^2 + z^2 \equiv 1$) and $\rho(s) > 0$. Assume that S is not a plane (so that z is not constantly zero). Show that:

(a) S is a minimal surface if and only if

$$\frac{z}{\rho} = \rho z - \rho z \quad (\text{apply 3.2.1})$$

(b) assuming that $z(s_0) \neq 0$, there exists $\varepsilon > 0$ such that, for $s \in]s_0 - \varepsilon, s_0 + \varepsilon[$, the above equation is equivalent to

$$\begin{aligned} \rho\rho' + \rho^2 &= 1 \Leftrightarrow \frac{d^2}{ds^2} \left(\frac{1}{2} \rho^2 \right) = 1 \\ &\Leftrightarrow \exists A, B \in \mathbb{R} : \rho(s) = \sqrt{s^2 + As + B}; \end{aligned}$$

(c) by shifting the domain of s , we can guarantee that there exist $a > 0$ and a non-empty open interval I such that $\rho(s) = \sqrt{s^2 + a^2}$ for all $s \in I$;

(d) $\rho(s) = \sqrt{s^2 + a^2}$ for all $s \in \mathbb{R}$;

(e) assuming that $z(0) = 0$ (if not, apply a vertical translation to S) and that $z > 0$,

$$z(s) = \int_0^s \frac{a}{\sqrt{t^2 + a^2}} dt = a \sinh^{-1} \left(\frac{s}{a} \right);$$

(f) by letting $v = \frac{z}{a} = \sinh^{-1} \left(\frac{s}{a} \right)$, we have $\rho(v) = a \cosh v$, and therefore S is the catenoid.

4.5 The Gauss-Bonnet Theorem

The Gauss-Bonnet Theorem is one of the deepest results in the Differential Geometry of surfaces establishing an unexpected connection between the Euler characteristic of a compact surface (a purely topological concept) and its Gaussian curvature. Moreover, it provides a general context for a seemingly rather particular result as Girard's formula for spherical triangles (example 2.5.2).

We work with an oriented surface S . Let Ω be a polygonal region (as defined in Section 4.4) and $\alpha(s)$, $s \in [0, L]$, a parameterization of $\partial\Omega$ by arc length with positive orientation. Let $0 = s_0 < s_1 < \dots < s_k = L$ be the k vertices of α ; $\gamma_i \in [-\pi, \pi]$, for $i = 1, \dots, k-1$, the oriented angle of $\alpha'(s_i^-)$ and $\alpha'(s_i^+)$; and $\gamma_k \in [-\pi, \pi]$ the angle between $\alpha'(s_k^-)$ and $\alpha'(s_0^+)$. These angles γ_i are called the *exterior angles* of Ω .

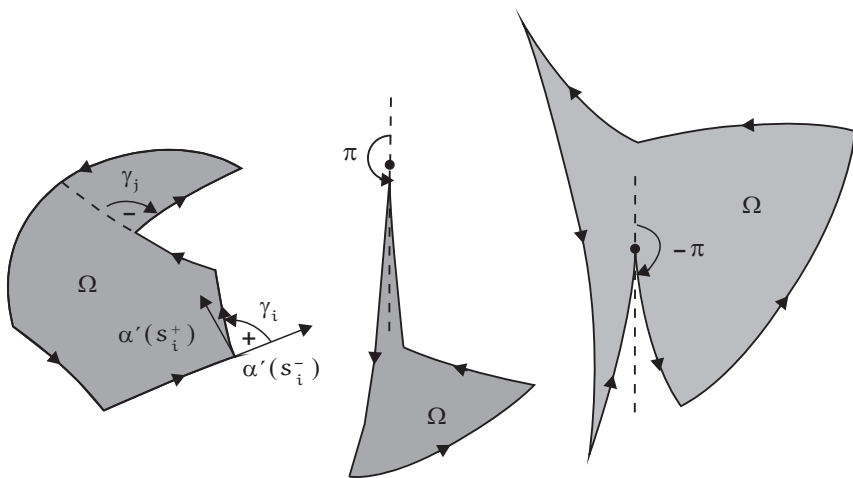


Figure 4.6

(Note: if $\alpha'(s_i^+) = -\alpha'(s_i^-)$, we have to decide which of the values π or $-\pi$ we choose for γ_i . We take $\gamma_i = \pi$ when, for $s \rightarrow s_i^-$, the trace of $\alpha|_{[s_i, s_i + \varepsilon]}$ (for small $\varepsilon > 0$) stays, like Ω , to the left of the curve; otherwise, we take $\gamma_i = -\pi$. We can make this criterion rigorous with the help of local coordinates, but Fig. 4.6 is more illuminating).

Let us state the first version of our theorem:

Gauss-Bonnet Theorem 4.5.1 (Local version). *Let $\Omega \subseteq S$ be a polygonal region whose closure is covered by some parameterization. Let γ_i ($i = 1, \dots, k$) be the exterior angles of Ω . Then*

$$\int_{\Omega} K d\sigma + \int_{\partial\Omega} k_g(s) ds + \sum_{i=1}^k \gamma_i = 2\pi.$$

Let $\Phi(u, v)$ be a parameterization compatible with the orientation of S containing the closure of Ω . Consider the unit tangent field $\mathbf{v}_1 = \frac{1}{\sqrt{E}} \Phi_u$ and define, for each interval $[s_{i-1}, s_i]$, a continuous choice $\theta_i(s)$ of the oriented angle between $\mathbf{v}_1$ and $\alpha'(s)$. (See the Appendix to this chapter for the existence of such continuous choices.) The *Rotation Index theorem* (Theorem A.4 of the Appendix) states that

$$\sum_{i=1}^k \{\theta_i(s_i) - \theta_i(s_{i-1})\} + \sum_{i=1}^k \gamma_i = 2\pi. \quad (4.22)$$

Let $\mathbf{w}(s)$, $s \in [0, L]$, be a **unit** vector field along α : the vector field $\mathbf{w}$ is therefore continuous, and is parallel along each regular arc $\alpha([s_{i-1}, s_i])$. Let $\varphi_i(s)$ ($s \in [s_{i-1}, s_i]$) be a continuous choice of the angle between $\alpha'(s)$ and $\mathbf{w}(s)$. The proof of Proposition 4.3.7 (ii) shows that

$$\varphi'_i(s) = -k_g(s). \quad (4.23)$$

Further, let $\psi(s)$, $s \in [0, L]$, be a choice of the angle between $\mathbf{v}_1$ and $\mathbf{w}(s)$. By the equality between oriented angles

$$\begin{aligned}\angle(\mathbf{v}_1, \alpha'(s)) &= \angle(\mathbf{v}_1, \mathbf{w}(s)) + \angle(\mathbf{w}(s), \alpha'(s)) \\ &= \angle(\mathbf{v}_1, \mathbf{w}(s)) - \angle(\alpha'(s), \mathbf{w}(s)),\end{aligned}$$

we see that we can assume, for $i = 1, \dots, k$, that

$$\theta_i(s) = \psi(s) - \varphi_i(s) \quad \text{for } s \in [s_{i-1}, s_i]. \quad (4.24)$$

In view of (4.24), equality (4.22) can then be rewritten as

$$\psi(L) - \psi(0) - \sum_{i=1}^k \{\varphi_i(s_i) - \varphi_i(s_{i-1})\} + \sum_{i=1}^k \gamma_i = 2\pi,$$

or, using (4.23),

$$\psi(L) - \psi(0) + \int_0^L k_g(s) ds + \sum_{i=1}^k \gamma_i = 2\pi.$$

In view of this formula, the next lemma concludes the proof of the Gauss-Bonnet theorem.

Lemma 4.5.2

$$\psi(L) - \psi(0) = \int_{\Omega} K d\sigma. \quad (4.25)$$

Proof Let $\mathbf{v}_2 = N \times \mathbf{v}_1$, so that the pair $(\mathbf{v}_1, \mathbf{v}_2)$ forms at each point a direct orthonormal frame. Since $\mathbf{v}_1$ is a unit vector field, $\frac{D\mathbf{v}_1}{ds}$ is orthogonal to $\mathbf{v}_1$, and is therefore collinear with $\mathbf{v}_2$. We thus have $\frac{D\mathbf{v}_1}{ds} = a(s)\mathbf{v}_2$, where $a(s) = \left\langle \frac{D\mathbf{v}_1}{ds}, \mathbf{v}_2 \right\rangle = \langle \mathbf{v}_1', \mathbf{v}_2 \rangle - \langle \mathbf{v}_1, \mathbf{v}_2' \rangle$, from which we conclude that

$$\frac{D\mathbf{v}_1}{ds} = \langle \mathbf{v}_1', \mathbf{v}_2 \rangle \mathbf{v}_2, \quad \frac{D\mathbf{v}_2}{ds} = -\langle \mathbf{v}_1', \mathbf{v}_2 \rangle \mathbf{v}_1. \quad (4.26)$$

From the equality

$$\mathbf{w}(s) = \cos \psi(s) \mathbf{v}_1 + \sin \psi(s) \mathbf{v}_2$$

we obtain, by taking the covariant derivative of both sides and using (4.26),

$$\frac{D\mathbf{w}}{ds} = (\psi'(s) + \langle \mathbf{v}_1', \mathbf{v}_2 \rangle)(-\sin \psi(s) \mathbf{v}_1 + \cos \psi(s) \mathbf{v}_2).$$

But since $\mathbf{w}$ is parallel, we have $\frac{D\mathbf{w}}{ds} \equiv 0$ and therefore

$$\psi'(s) = -\langle \mathbf{v}_1'(s), \mathbf{v}_2(s) \rangle.$$

Hence

$$\begin{aligned}
 \psi(L) - \psi(0) &= - \int_0^L \langle \mathbf{v}_1'(s), \mathbf{v}_2(s) \rangle ds \\
 &= - \int_0^L \left(\left\langle \frac{\partial \mathbf{v}_1}{\partial u}, \mathbf{v}_2 \right\rangle u'(s) + \left\langle \frac{\partial \mathbf{v}_1}{\partial v}, \mathbf{v}_2 \right\rangle v'(s) \right) ds \\
 &= - \iint_{\Phi^{-1}(\Omega)} \left\{ \frac{\partial}{\partial u} \left\langle \frac{\partial \mathbf{v}_1}{\partial v}, \mathbf{v}_2 \right\rangle - \frac{\partial}{\partial v} \left\langle \frac{\partial \mathbf{v}_1}{\partial u}, \mathbf{v}_2 \right\rangle \right\} dudv \\
 &\quad \text{(by Green's theorem)} \\
 &= \iint_{\Phi^{-1}(\Omega)} \left\{ \left\langle \frac{\partial \mathbf{v}_1}{\partial u}, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle - \left\langle \frac{\partial \mathbf{v}_1}{\partial v}, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle \right\} dudv.
 \end{aligned}$$

To complete the proof of the lemma, it is now sufficient to show that

$$\left\langle \frac{\partial \mathbf{v}_1}{\partial u}, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle - \left\langle \frac{\partial \mathbf{v}_1}{\partial v}, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle = K \sqrt{EG - F^2}. \quad (4.27)$$

We begin by observing that from the equalities

$$\begin{aligned}
 N_u &= \langle N_u, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle N_u, \mathbf{v}_2 \rangle \mathbf{v}_2 = - \left\langle N, \frac{\partial \mathbf{v}_1}{\partial u} \right\rangle \mathbf{v}_1 - \left\langle N, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle \mathbf{v}_2, \\
 N_v &= - \left\langle N, \frac{\partial \mathbf{v}_1}{\partial v} \right\rangle \mathbf{v}_1 - \left\langle N, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle \mathbf{v}_2,
 \end{aligned}$$

it ensures, taking into account the identity $\mathbf{v}_1 \times \mathbf{v}_2 = N$, that

$$N_u \times N_v = \left\{ \left\langle N, \frac{\partial \mathbf{v}_1}{\partial u} \right\rangle \left\langle N, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle - \left\langle N, \frac{\partial \mathbf{v}_1}{\partial v} \right\rangle \left\langle N, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle \right\} N. \quad (4.28)$$

If we write

$$\frac{\partial \mathbf{v}_1}{\partial u} = a \mathbf{v}_2 + \left\langle N, \frac{\partial \mathbf{v}_1}{\partial u} \right\rangle N, \quad \frac{\partial \mathbf{v}_2}{\partial v} = b \mathbf{v}_1 + \left\langle N, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle N,$$

and form the scalar product, then we obtain

$$\left\langle \frac{\partial \mathbf{v}_1}{\partial u}, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle = \left\langle N, \frac{\partial \mathbf{v}_1}{\partial u} \right\rangle \left\langle N, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle,$$

and analogously

$$\left\langle \frac{\partial \mathbf{v}_1}{\partial v}, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle = \left\langle N, \frac{\partial \mathbf{v}_1}{\partial v} \right\rangle \left\langle N, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle.$$

Together with (4.28), the last two equalities give

$$N_u \times N_v = \left\{ \left\langle \frac{\partial \mathbf{v}_1}{\partial u}, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle - \left\langle \frac{\partial \mathbf{v}_1}{\partial v}, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle \right\} N,$$

and therefore

$$\begin{aligned} \left\langle \frac{\partial \mathbf{v}_1}{\partial u}, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle - \left\langle \frac{\partial \mathbf{v}_1}{\partial v}, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle &= \langle N_u \times N_v, N \rangle \\ &= K \langle \Phi_u \times \Phi_v, N \rangle = K \sqrt{EG - F^2}. \end{aligned}$$

This concludes the proof of (4.27) and hence that of the lemma. $\square$

Observations 4.5.3 A. The difference $\psi(L) - \psi(0)$ in Lemma 4.5.2 is the angle between the initial and final positions of a vector that is carried in parallel along a closed curve. If K has constant (nonzero) sign on Ω , Lemma 4.5.2 shows that this angle is nonzero and becomes smaller as the region bounded by the curve becomes smaller. It follows that a surface has constant zero curvature if parallel transport along any curve on the surface depends only on the starting and ending points of the curve (and not on the path it covered). The reader is invited to elaborate the argument in the exercises in this section.

B. The formula (4.27) can be rewritten in the form

$$K = \frac{1}{\sqrt{EG - F^2}} \left\{ \frac{\partial}{\partial u} \left\langle \mathbf{v}_1, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle - \frac{\partial}{\partial v} \left\langle \mathbf{v}_1, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle \right\} \quad (4.29)$$

$$= \frac{1}{\sqrt{EG - F^2}} \left\{ \frac{\partial}{\partial u} \left\langle \mathbf{v}_1, \frac{D\mathbf{v}_2}{\partial v} \right\rangle - \frac{\partial}{\partial v} \left\langle \mathbf{v}_1, \frac{D\mathbf{v}_2}{\partial u} \right\rangle \right\}, \quad (4.30)$$

and formula (4.30) gives another proof of Gauss's Theorema Egregium, for it expresses K via intrinsic quantities: in fact, any isometry transforms $(\mathbf{v}_1, \mathbf{v}_2)$ into another orthonormal dihedron and, by (4.10) in Section 4.3, preserves the covariant derivative. We can further apply (4.29) to express K as a function of the coefficients E , F and G . For simplicity, we assume that $\Phi(u, v)$ is an orthogonal parameterization ($F \equiv 0$), so $\mathbf{v}_1 = \frac{1}{\sqrt{E}}\Phi_u$ and $\mathbf{v}_2 = \frac{1}{\sqrt{G}}\Phi_v$. We then have

$$\frac{\partial \mathbf{v}_2}{\partial v} = \frac{\partial}{\partial v} \left(\frac{1}{\sqrt{G}} \right) \Phi_v + \frac{1}{\sqrt{G}} \Phi_v, \quad \frac{\partial \mathbf{v}_2}{\partial u} = \frac{\partial}{\partial u} \left(\frac{1}{\sqrt{G}} \right) \Phi_v + \frac{1}{\sqrt{G}} \Phi_{vu},$$

so that

$$\begin{aligned} \left\langle \mathbf{v}_1, \frac{\partial \mathbf{v}_2}{\partial v} \right\rangle &= \frac{1}{\sqrt{EG}} \langle \Phi_u, \Phi_{vv} \rangle = \frac{-1}{\sqrt{EG}} \langle \Phi_{uv}, \Phi_v \rangle = \frac{-G_u}{2\sqrt{EG}}, \\ \left\langle \mathbf{v}_1, \frac{\partial \mathbf{v}_2}{\partial u} \right\rangle &= \frac{1}{\sqrt{EG}} \langle \Phi_u, \Phi_{uv} \rangle = \frac{E_v}{2\sqrt{EG}}. \end{aligned}$$

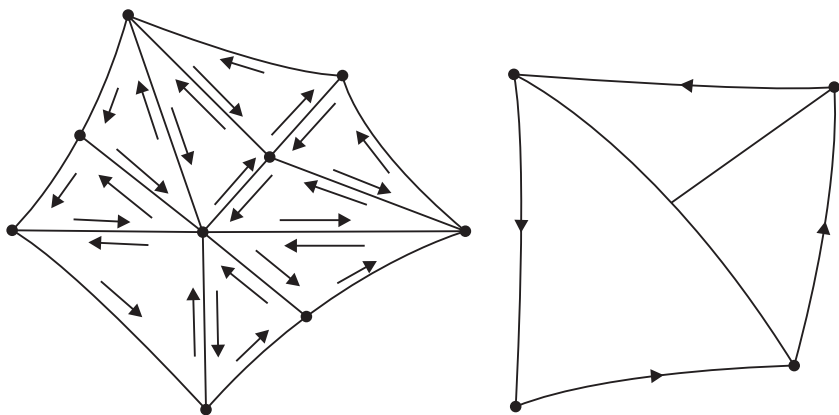
Finally, by (4.29), we have

$$K = \frac{-1}{2\sqrt{EG}} \left\{ \frac{\partial}{\partial u} \left(\frac{G_u}{\sqrt{EG}} \right) + \frac{\partial}{\partial v} \left(\frac{E_v}{\sqrt{EG}} \right) \right\}, \quad (4.31)$$

which is the promised formula. $\square$

The condition, in Theorem 4.5.1, that Ω is contained in a parametrized neighborhood, is dispensable, since it holds for every polygonal region. To prove this, we consider a *triangulation* $(\Delta_j)_{1 \leq j \leq r}$ of Ω such that each Δ_j is a **closed** polygonal region that is contained in some parametrized neighborhood (so that we can apply Theorem 4.5.1 to it). Moreover:

- (i) the closure of Ω is equal to the union of the Δ_j ;
- (ii) for each j the simple closed curve $\partial\Delta_j$ has three *vertices*; the portions of $\partial\Delta_j$ between each pair of consecutive vertices are regular curves called *edges*. The sets Δ_j are called the *faces*;
- (iii) the intersection of two distinct faces is either empty, or reduces to a vertex, or is an edge common to both.



a triangulation

not a triangulation

Figure 4.7

We denote by $\mathcal{V}$ the set of vertices of the triangulation, and by V, A, F the numbers of vertices, edges and faces, respectively. Denoting by γ_j^l ($l = 1, 2, 3$) the exterior angles of Δ_j , Theorem 4.5.1 gives, for each $1 \leq j \leq F$,

$$\int_{\Delta_j} K d\sigma + \int_{\partial\Delta_j} k_g(s) ds + \sum_{l=1}^3 \gamma_j^l = 2\pi.$$

Each interior edge of Ω is run through twice in opposite directions (the geodesic curvature showing up with opposed sign), and by summing the above formulas for $j = 1, \dots, F$ only the edges that make up $\partial\Omega$ are left. We thus obtain

$$\int_{\Omega} K d\sigma + \int_{\partial\Omega} K_g(s) ds + \sum_{j=1}^F \sum_{l=1}^3 \gamma_j^l = 2\pi F. \quad (4.32)$$

For each vertex $v \in \mathcal{V}$, let:

$$\mathcal{I}_v = \{(j, l): \gamma_j^l \text{ is adjacent to } v\};$$

$$\mathcal{S}(v) = \sum_{(j, l) \in \mathcal{I}_v} \gamma_j^l;$$

$$A(v) = \text{number of edges starting from } v.$$

We break $\mathcal{V}$ into three subsets $\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3$ defined as follows: $\mathcal{V}_1$ is the set of vertices inside Ω ; $\mathcal{V}_2$ contains the vertices that belong to $\partial\Omega$ but are not vertices of Ω ; $\mathcal{V}_3$ contains the vertices of Ω . Denoting by $\eta_j^l = \pi - \gamma_j^l$ the interior angles of Δ_j , we have:

- if $v \in \mathcal{V}_1$ then

$$\mathcal{S}(v) = \sum_{(j, l) \in \mathcal{I}_v} (\pi - \eta_j^l) = \pi \# \mathcal{I}_v - \sum_{(j, l) \in \mathcal{I}_v} \eta_j^l = \pi A(v) - 2\pi,$$

- if $v \in \mathcal{V}_2$ then

$$\mathcal{S}(v) = \pi \mathcal{I}_v - \sum_{(j, l) \in \mathcal{I}_v} \eta_j^l = \pi(A(v) - 1) - \pi = \pi A(v) - 2\pi,$$

- if $v \in \mathcal{V}_3$ then v is one of the k vertices of Ω and, if γ_i is the corresponding exterior angle, we have

$$\mathcal{S}(v) = \pi \# \mathcal{I}_v - \sum_{(j, l) \in \mathcal{I}_v} \eta_j^l = \pi(A(v) - 1) - (\pi - \gamma_i) = \pi A(v) - 2\pi + \gamma_i.$$

Adding up all these formulas, we obtain

$$\begin{aligned} \sum_{j=1}^F \sum_{l=1}^3 \gamma_j^l &= \sum_{v \in \mathcal{V}} \mathcal{S}(v) = \pi \sum_{v \in \mathcal{V}} A(v) - 2\pi V + \sum_{i=1}^k \gamma_i \\ &= 2\pi(A - V) + \sum_{i=1}^k \gamma_i \end{aligned}$$

— because each edge is counted twice (once for each one of its endpoints). Replacing in (4.32) and using Lemma 4.5.4 below, we obtain

$$\int_{\Omega} K d\sigma + \int_{\partial\Omega} k_g(s) ds + \sum_{i=1}^k \gamma_i = 2\pi(V - A + F) = 2\pi,$$

which establishes the Gauss-Bonnet formula for any polygonal region $\Omega \subseteq S$.

Lemma 4.5.4 (Euler's formula). $V - A + F = 1$.

Proof We have to show that for every triangulation of a polygonal region we have $V - A + F = 1$. We proceed by induction on the number of faces F . If $F = 1$, then there are three vertices and three edges and the formula is true. If $F > 1$, then let Δ_j be a face where at least one of the edges is part of $\partial\Omega$, and such that $\tilde{\Omega} = \Omega \setminus \Delta_j$ is a polygonal region. Consider in $\tilde{\Omega}$ the triangulation induced by the triangulation of Ω . Denoting by $\tilde{V}$, $\tilde{A}$ and $\tilde{F}$ the numbers of vertices, edges and faces of the triangulation of $\tilde{\Omega}$, we have:

- if $\partial\Omega$ contains a single edge of Δ_j , then $\tilde{V} = V$, $\tilde{A} = A - 1$, $\tilde{F} = F - 1$;
- if $\partial\Omega$ contains two edges of Δ_j , then $\tilde{V} = V - 1$, $\tilde{A} = A - 2$, $\tilde{F} = F - 1$.

In both cases, $V - A + F = \tilde{V} - \tilde{A} + \tilde{F}$ and the proof by induction is complete. $\square$

There is a special case of the Gauss-Bonnet formula that is worthy to note: if the boundary of Ω consists of geodesic arcs ($k_g \equiv 0$), we are left with

$$\int_{\Omega} K d\sigma + \sum_{i=1}^k \gamma_i = 2\pi,$$

or, denoting by $\eta_i = \pi - \gamma_i$ the interior angles of Ω ,

$$\sum_{i=1}^k \eta_i - (k - 2)\pi = \int_{\Omega} K d\sigma.$$

When $k = 3$, Ω is called a *geodesic triangle*, and we have just obtained the promised generalization of Girard's formula:

Corollary 4.5.5 *The difference between the sum of the interior angles of a geodesic triangle Δ and π is given by the integral, extended to Δ , of the Gaussian curvature:*

$$(\eta_1 + \eta_2 + \eta_3) - \pi = \int_{\Delta} K d\sigma.$$

If in particular the curvature of the surface is constant, then this difference is proportional to the area of the triangle, equal to $(\eta_1 + \eta_2 + \eta_3) - \pi = K \cdot \text{area}(\Delta)$.

Let us now assume that S is a compact surface, and consider a triangulation $\mathcal{T} = (\Delta_i)_{1 \leq i \leq r}$ of S : conditions (i) – (iii) are satisfied, but it is not required that each Δ_j is contained in a parametrized neighborhood. (The existence of triangulations for arbitrary surfaces is a deep result, and a proof is given in [17]; for regular compact surfaces — the case we are concerned with — we will give a proof in Exercise 115.) The *Euler characteristic* of S is $\chi(S) = V - A + F$, where V , A and F are the numbers of vertices, edges and faces of $\mathcal{T}$.

Gauss-Bonnet Theorem 4.5.6 (Global Version). *If S is a compact surface then*

$$\int_S K d\sigma = 2\pi\chi(S).$$

Proof Writing down the Gauss-Bonnet formula for each of the faces of a triangulation $\mathcal{T}$ and summing them up, the integrals of the geodesic curvatures all cancel, because

each edge is run through twice in opposite directions. Furthermore, the vertices of $\mathcal{T}$ are all inside S (of type $\mathcal{V}_1$) and the sum of the interior angles adjacent to each of them is 2π . By the above calculations, we are left with

$$\int_S K d\sigma = 2\pi(V - A + F),$$

which is the desired formula. $\square$

The integral $\int_S K d\sigma$ is called the *total curvature* of S . The equation now obtained shows that $\chi(S)$ is welldefined, independent of the triangulation of S chosen to compute it. Furthermore, $\chi(S)$ is invariant under diffeomorphisms (because a diffeomorphism $f: S \rightarrow S_2$ maps any triangulation $\mathcal{T}$ of S to another triangulation $f(\mathcal{T})$ of S_2 with the same number of vertices, edges and faces). We thus obtain the following result.

Corollary 4.5.7 *Any two diffeomorphic compact surfaces have the same total curvature.*

For example, any surface S diffeomorphic to the sphere has total curvature 4π . This is not surprising if the curvature of S is positive at all points (as in the ellipsoid), because in this situation we will see later on that $N: S \rightarrow \mathbf{S}^2$ is a diffeomorphism, and $\int_S K d\sigma$ is nothing but the area of the image of S under N (cf. Exercise 64 in Section 3.2), which in this case is $\mathbf{S}^2$. But if we call to mind that S can also have regions of negative curvature we more readily appreciate the strength of the result.

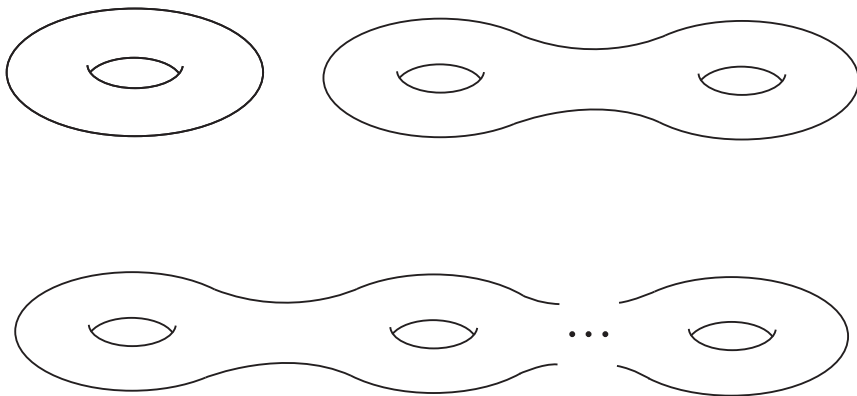


Figure 4.8

The Euler characteristic (and hence the total curvature) of the torus is zero; that of the double torus is -2 . That of the n -torus ($n \geq 1$) is $2 - 2n$. Together with the sphere, and up to diffeomorphisms, this list exhausts all orientable compact surfaces (see Fig. 4.8); this is a classical result whose proof you can find, for example, in [17]. In particular, among the compact surfaces, only those which are diffeomorphic to the sphere have non-negative total curvature.

Example 4.5.8 A non-compact surface may have finite total curvature. Consider, on a surface of revolution S given by $\rho = \rho(z)$, $z \in \mathbb{R}$, the region $\Omega(z_0, z_1)$ bounded by the two parallels $z = z_0$ and $z = z_1$ ($z_0 < z_1$). We can break $\Omega(z_0, z_1)$ into two “four-sided polygons” by two meridians and add up the two resulting Gauss-Bonnet formulas. Since there are four vertices and the sum of the two exterior angles adjacent to each vertex is π , we obtain

$$\int_{\Omega(z_0, z_1)} K d\sigma + \int_{\partial\Omega(z_0, z_1)} k_g ds + 4\pi = 4\pi,$$

i.e.,

$$\int_{\Omega(z_0, z_1)} K d\sigma = \int_{\partial\Omega(z_0, z_1)} -k_g ds.$$

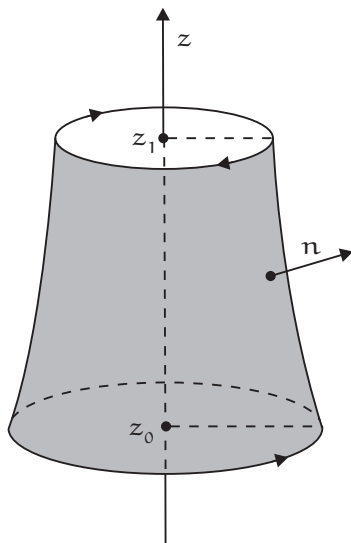


Figure 4.9

Taking in S the orientation given by

$$N(\varphi, z) = \frac{1}{\sqrt{1 + \rho^2}} (\cos \varphi, \sin \varphi, -\rho),$$

the parallel $z = z_0$ is run through counterclockwise, and $z = z_1$ clockwise. A quick calculation then shows that their geodesic curvatures are

$$k_g(z_0) = \frac{-\rho(z_0)}{\rho(z_0)\sqrt{1 + \rho(z_0)^2}}, \quad k_g(z_1) = \frac{\rho(z_1)}{\rho(z_1)\sqrt{1 + \rho(z_1)^2}},$$

and therefore

$$\int_{\Omega(z_0, z_1)} K d\sigma = -2\pi \left(\frac{\rho(z_1)}{\sqrt{1 + \rho(z_1)^2}} - \frac{\rho(z_0)}{\sqrt{1 + \rho(z_0)^2}} \right).$$

This formula guarantees that in many cases the limit

$$\int_S K d\sigma = \lim_{\substack{z_1 \rightarrow +\infty \\ z_0 \rightarrow -\infty}} \int_{\Omega(z_0, z_1)} K d\sigma$$

exists and is finite; we will call it the *total curvature* of S .

For example, the total curvature of the hyperboloid $x^2 + y^2 = 1 + z^2$ is $-2\sqrt{2}\pi$, since $\rho(z) = \sqrt{1 + z^2}$ and $\lim_{z \rightarrow \pm\infty} \rho(z) = \pm 1$.

It would not be difficult to compute directly $\int_{\Omega(z_0, z_1)} K d\sigma$ (see example 3.2.1) — but, besides illustrating a use of the Gauss-Bonnet Theorem, this method is applicable to surfaces other than surfaces of revolution (e.g., to those non-compact surfaces S for which $S \setminus \mathfrak{C}$ is a surface of revolution for some compact $\mathfrak{C} \subseteq S$). $\square$

Exercises

96. Assume that S has non-positive curvature at all its points. Show that if two geodesics start from the same point in S , they cannot meet again in such a way that their traces constitute the boundary of a simple region of S (in particular, no closed geodesic of S can be the boundary of a simple region).

97. Show that if γ is a regular closed simple curve in S^2 then $\left| \int_{\gamma} k_g ds \right| < 2\pi$.

98. Let $p \in S$ be such that $K(p) > 0$, and let (U, Φ) be a parameterization such that $p \in \Phi(U)$ and $K \circ \Phi(u, v) > 0$ for all $(u, v) \in U$. Consider a family of circles $(\mathcal{C}_r)_{0 < r < \delta}$ such that:

- i. each $\mathcal{C}_r$ has radius r ;
- ii. for every r , $\Phi^{-1}(p) \in \mathcal{C}_r$ and the closed disk bounded by $\mathcal{C}_r$ is contained in U ;
- iii. if $r < \tilde{r}$ then $\mathcal{C}_r$ is inside $\mathcal{C}_{\tilde{r}}$.

Further denote by $\psi_r \in [-\pi, \pi]$ the oriented angle between the initial and final positions of a vector carried parallel from p to p along the closed curve $\Phi(\mathcal{C}_r)$ in S .

(a) Show that there exists $\varepsilon \in]0, \delta[$ such that

$$0 < r < \tilde{r} < \varepsilon \Rightarrow 0 < |\psi_r| < |\psi_{\tilde{r}}| < \pi.$$

(b) Conclude that any surface on which parallel transport depends only on the initial and final points of the curve has constant zero curvature:

Give an example of a surface with constant zero curvature that does not have this property.

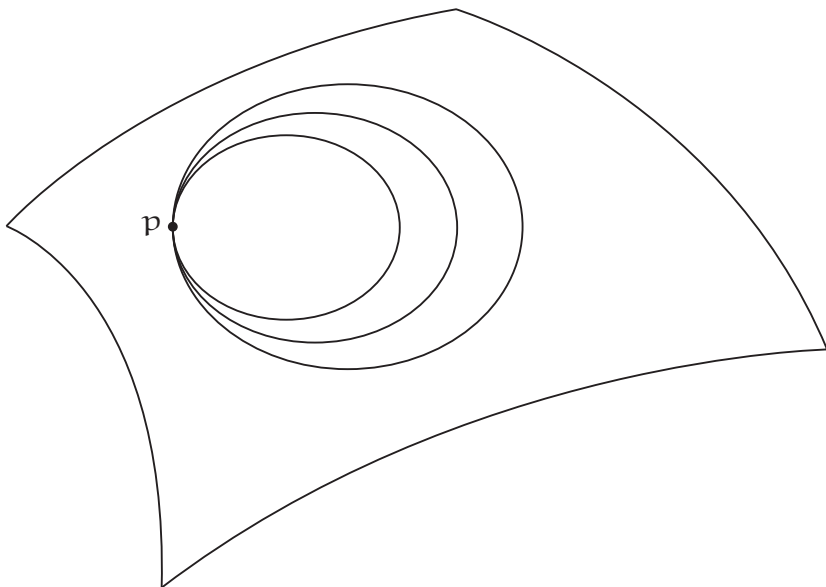


Figure 4.10

99. Is it true that two non-compact diffeomorphic surfaces necessarily have the same total curvature?

4.6 Minimizing Properties of Geodesics

We begin this section by establishing the local existence of geodesics on arbitrary surfaces: to this end we will observe that, in local coordinates, geodesics are characterized by second-order differential equations. We then introduce polar geodesic coordinates that allow us to show that the shortest path *on the surface* between any two points is, *when it exists*, given by a geodesic.

Recall that a curve α of S is a geodesic if the unit vector field tangent to the curve $\tau_1 = \frac{1}{|\alpha'|} \alpha'$ is a parallel vector field. Taking coordinates (U, Φ) and assuming that $\alpha(t) = \Phi(u(t), v(t))$ is parametrized with constant scalar velocity $m > 0$, we have

$$\tau_1 = \frac{1}{m} \alpha' = \frac{1}{m} (u' \Phi_u + v' \Phi_v).$$

From formula (4.9), we obtain

$$\begin{aligned} \frac{D\tau_1}{dt} = \frac{1}{m} \{ & (u'' + u'^2\Gamma_{11}^1 + 2u'v'\Gamma_{12}^1 + v'^2\Gamma_{22}^1)\Phi_u \\ & + (v'' + u'^2\Gamma_{11}^2 + 2u'v'\Gamma_{12}^2 + v'^2\Gamma_{22}^2)\Phi_v \} \end{aligned}$$

— and, if α is a geodesic, then $\frac{D\tau_1}{dt} \equiv 0$, so that

$$\begin{cases} u'' + u'^2\Gamma_{11}^1 + 2u'v'\Gamma_{12}^1 + v'^2\Gamma_{22}^1 = 0, \\ v'' + u'^2\Gamma_{11}^2 + 2u'v'\Gamma_{12}^2 + v'^2\Gamma_{22}^2 = 0. \end{cases} \quad (4.33)$$

Conversely, if $(u(t), v(t))$ is a non-constant solution of (4.33) then $\alpha(t) = \Phi(u(t), v(t))$ is a regular curve and its scalar velocity is constant, since in this case the covariant derivative of the vector field α' is zero and $\frac{d}{dt}|\alpha'|^2 = 2\left\langle \alpha', \frac{D\alpha'}{dt} \right\rangle = 0$.

It follows that $\tau_1 = \frac{1}{m} \alpha'$ is a parallel vector field and therefore α is a geodesic.

We define *parametrized geodesics* to be either a parameterization $\alpha(t)$ of a geodesic with constant scalar velocity $|\alpha'(t)|$, or a constant curve. This means that the solutions of the system give us, locally, all parametrized geodesics of the surface: we therefore call the equations in (4.33) the *differential equations of the geodesics*.

The system of equations (4.33) can be restated in the form

$$\begin{cases} u'' = H_1(u, v, u', v'), \\ v'' = H_2(u, v, u', v'), \end{cases} \quad (4.34)$$

or otherwise

$$\begin{cases} u' = a, \\ v' = b, \\ a' = H_1(u, v, a, b), \\ b' = H_2(u, v, a, b) \end{cases} \quad (4.35)$$

— where H_1, H_2 are differentiable functions defined on $U \times \mathbb{R}^2 \subseteq \mathbb{R}^4$ where U is the domain of the parameterization Φ . Theorem 3.3.1 guarantees that every $(u_0, v_0, a_0, b_0) \in U \times \mathbb{R}^2$ has an open neighborhood W such that there exist $\varepsilon > 0$ and a differentiable mapping

$$]-\varepsilon, \varepsilon[\times W \rightarrow U \times \mathbb{R}^2, \quad (t; u, v, a, b) \mapsto \tilde{\varphi}(t; u, v, a, b)$$

enjoying the following conditions: for every $(u, v, a, b) \in W$, the curve

$$t \mapsto \tilde{\varphi}(t; u, v, a, b)$$

is the only solution of the system (4.35) with initial condition $\tilde{\varphi}(0; u, v, a, b) = (u, v, a, b)$.

Writing $\tilde{\varphi} = (\varphi_1, \varphi_2, \varphi_3, \varphi_4)$, it is clear that φ_3 and φ_4 are redundant, since $\varphi_3 = \frac{\partial \varphi_1}{\partial t}$ and $\varphi_4 = \frac{\partial \varphi_2}{\partial t}$. Letting $\varphi = (\varphi_1, \varphi_2)$, the conclusion we obtain for the system (4.34), and hence for (4.33), is as follows. The mapping $\varphi:]-\varepsilon, \varepsilon[\times W \rightarrow U$, where W is a neighborhood of (u_0, v_0, a_0, b_0) , is such that, for $(u, v, a, b) \in W$, the curve $t \mapsto \varphi(t; u, v, a, b)$ is the only solution of (4.34) with initial conditions

$$\varphi(0; u, v, a, b) = (u, v), \quad \frac{d}{dt} \varphi(t; u, v, a, b) \Big|_{t=0} = (a, b).$$

In summary, determining $(u(0), v(0))$ and $(u'(0), v'(0))$ completely determines a solution $(u(t), v(t))$ of (4.33). The solutions depend differentiably on these initial conditions. Let us now define

$$\begin{aligned} D\Phi: U \times \mathbb{R}^2 &\rightarrow \bigcup_{p \in \Phi(U)} (\{p\} \times T_p S) \\ (u, v, a, b) &\mapsto (\Phi(u, v), (a\Phi_u + b\Phi_v) \Big|_{(u,v)}) \end{aligned} \quad (4.36)$$

and note that $D\Phi$ is a continuous bijection. Given $p_0 = \Phi(u_0, v_0) \in \Phi(U)$, consider the corresponding neighborhood W associated with $(u_0, v_0, 0, 0)$ (which we can assume to be of the form $W = W_1 \times W_2$), $\varepsilon > 0$ and the mapping φ . Let us take $V \subseteq S$, an open neighborhood of p_0 whose closure is compact and contained in $\Phi(W_1)$, and choose $\delta > 0$ such that the set

$$B_\delta(V) = \{(p, \mathbf{v}): p \in V, \mathbf{v} \in T_p S, |\mathbf{v}| < \delta\}$$

is contained in $D\Phi(W)$ (that such a choice of δ is possible is an easy exercise). We then define the mapping

$$\gamma:]-\varepsilon, \varepsilon[\times B_\delta(V) \rightarrow S, \quad \gamma(t; p, \mathbf{v}) = \Phi \circ \varphi(t; (D\Phi)^{-1}(p, \mathbf{v})),$$

and it follows from our analysis that for every $(p, \mathbf{v}) \in B_\delta(V)$, the curve $t \mapsto \gamma(t; p, \mathbf{v})$ is the only parametrized geodesic that at time 0 passes through p with velocity $\mathbf{v}$; the constant parametrized geodesics are those of the form $\gamma(t; p, 0)$. Furthermore, for $\lambda \in \mathbb{R}$ we have

$$\gamma(\lambda t; p, \mathbf{v}) = \gamma(t; p, \lambda \mathbf{v}), \quad (4.37)$$

because the two terms of (4.37) are parametrized geodesics satisfying the same initial conditions: at time 0, they both pass through p with velocity $\lambda \mathbf{v}$. It follows that, up to reparametrization, *there is exactly one geodesic whose tangent line at a given point has a given direction*.

Observation 4.6.1 At this point we can already state that there are no geodesics on the sphere other than the maximal circles (see example 4.3.9), since through each point passes a maximal circle tangent to each given direction. The sphere thus has the particularity that all its geodesics are closed. (A non-constant parametrized geodesic $\gamma(t)$ is called *closed* if it is periodic — that is, if there exists $T > 0$ such that $\gamma(t + T) = \gamma(t)$ for all $t \in \mathbb{R}$. A necessary and sufficient condition for γ to be

closed is that there exist $t_1 < t_2$ such that $\gamma(t_1) = \gamma(t_2)$ and $\gamma'(t_1) = \gamma'(t_2)$ — i.e., γ is closed if and only if it passes again through the same point with the same velocity vector.) Surprisingly, there are other surfaces with the same property, as shown in [4].
□

Let us now apply (4.37) to show that, choosing $|\mathbf{v}|$ sufficiently small, $\gamma(t; p, \mathbf{v})$ is defined for $|t| < 2$. In fact, since $\gamma(t; p, \mathbf{v}) = \gamma(\frac{\varepsilon}{2}t; p, \frac{2}{\varepsilon}\mathbf{v})$, we can state that, for $|\frac{2}{\varepsilon}\mathbf{v}| < \delta$ (i.e., for $|\mathbf{v}| < \frac{\varepsilon\delta}{2}$), $\gamma(t; p, \mathbf{v})$ is defined whenever $|\frac{\varepsilon}{2}t| < \varepsilon$ — i.e., whenever $|t| < 2$. In summary:

For every $p_0 \in S$ there exist $\eta > 0$ and a neighborhood V of p_0 such that whenever $p \in V$ and $\mathbf{v} \in B_\eta(p) = \{\mathbf{v} \in T_p S : |\mathbf{v}| < \eta\}$ the geodesic $t \mapsto \gamma(t; p, \mathbf{v})$ is defined for $t \in]-2, 2[$.

Given $p \in S$, the *exponential mapping* $\exp_p$ is defined by $\exp_p(\mathbf{v}) = \gamma(1; p, \mathbf{v})$. By the above reasoning, there exists some $\eta > 0$ such that $\exp_p$ is defined on $B_\eta(p)$; one can choose such an η suitable for all points in a neighborhood of p .

The geodesic $t \mapsto \gamma(t; p, \mathbf{v})$ has constant scalar velocity $|\mathbf{v}|$, and therefore its arc length in the interval $[0, 1]$ is also $|\mathbf{v}|$. The geometric meaning of the exponential mapping is therefore as follows: $\exp_p(\mathbf{v})$ is the point that travels a distance of $|\mathbf{v}|$ on the geodesic that begins at p and whose direction and orientation is given by $\mathbf{v}$. Note that, by (4.37), we have $\exp_p(t\mathbf{v}) = \gamma(1; p, t\mathbf{v}) = \gamma(t; p, \mathbf{v})$ — which means that the geodesics starting from p are the image of the lines (or line segments) in $T_p S$ that **pass through the origin** under $\exp_p$.

To make full use of the exponential mapping, we need the next proposition:

Proposition 4.6.2 *Given $p_0 \in S$, there exist $\delta > 0$ and an open neighborhood $W \subseteq S$ of p_0 such that, for all $p \in W$, $\exp_p|_{B_\delta(p)}$ is a diffeomorphism onto the image.*

Proof We know that there exist $\eta > 0$ and an open neighborhood $V \subseteq S$ of p_0 such that, for all p on V , the exponential mapping $\exp_p$ is defined on $B_\eta(p)$. We can thus consider the differentiable mapping

$$\begin{aligned} F: B_\eta(V) &\longrightarrow S \times S \\ (p, \mathbf{v}) &\longmapsto (p, \exp_p(\mathbf{v})). \end{aligned}$$

(A caveat: both $B_\eta(V)$ and $S \times S$ are spaces of dimension four, since $S \times S$ is the product of two spaces of dimension two and $B_\eta(V)$ can be identified, via $D\Phi$ as defined in (4.36), with an open subset of $\mathbb{R}^4$. Our proof can be made rigorous by applying the inverse mapping theorem to $(\Phi \times \Phi)^{-1} \circ F \circ D\Phi$, which is the expression of F in “local coordinates”.)

The tangent spaces to $B_\eta(V)$ at $(p_0, 0)$ and to $S \times S$ at $(p_0, p_0) = F(p_0, 0)$ coincide: both are $T_{p_0} S \times T_{p_0} S$. Given $\mathbf{v} \in T_{p_0} S$, let α be a curve in V such that $\alpha(0) = p_0$ and $\alpha'(0) = \mathbf{v}$. Then $t \mapsto (\alpha(t), 0)$ is a curve in $B_\eta(V)$ that passes through $(p_0, 0)$ with velocity $(\mathbf{v}, 0)$. Thus

$$DF_{(p_0, 0)}(\mathbf{v}, 0) = \frac{d}{dt} F(\alpha(t), 0) \Big|_{t=0} = \frac{d}{dt} (\alpha(t), \alpha(t)) \Big|_{t=0} = (\mathbf{v}, \mathbf{v}).$$

On the other hand, if $\mathbf{w} \in T_{p_0}S \setminus \{0\}$ then the curve $t \mapsto (p_0, t\mathbf{w})$, $|t| < \frac{\eta}{|\mathbf{w}|}$, is contained in $B_\eta(V)$ and passes through $(p_0, 0)$ with velocity $(0, \mathbf{w})$, and therefore

$$DF_{(p_0,0)}(0, \mathbf{w}) = \frac{d}{dt}F(p_0, t\mathbf{w})\big|_{t=0} = \left(0, \frac{d}{dt}\exp_{p_0}(t\mathbf{w})\big|_{t=0}\right) = (0, \mathbf{w}).$$

We thus conclude that

$$DF_{(p_0,0)}(\mathbf{v}, \mathbf{w}) = (\mathbf{v}, \mathbf{v} + \mathbf{w}) \text{ for every } (\mathbf{v}, \mathbf{w}) \in T_{p_0}S \times T_{p_0}S$$

– which shows that $DF_{(p_0,0)}$ is a linear isomorphism. The inverse mapping theorem then guarantees that the restriction of F to some neighborhood of $(p_0, 0)$ in $B_\eta(V)$ is a diffeomorphism onto the image, and we can choose such a neighborhood of the form $B_\delta(W)$, where W is an open subset of S and $\delta > 0$; it is easily verified that these choices of W and δ satisfy the desired condition. $\square$

Given $p \in S$, a neighborhood $V \subseteq S$ of p is called a *normal neighborhood* of p if there exists $\delta > 0$ such that $\exp_p|_{B_\delta(p)}$ is a diffeomorphism onto the image and $V \subseteq \exp_p(B_\delta(p))$. For such a δ , we write $D_\delta(p) = \exp_p(B_\delta(p))$; hence, $D_\delta(p)$ is the neighborhood of p covered by the geodesic rays of length δ starting from p , but note that for now we only define $D_\delta(p)$ for sufficiently small δ .

Let $(\mathbf{v}_1, \mathbf{v}_2)$ be an orthonormal basis of T_pS . $\exp_p$ gives rise to several coordinate systems in $D_\delta(p)$:

$$\Phi(u, v) = \exp_p(u\mathbf{v}_1 + v\mathbf{v}_2) \text{ (set on disk } u^2 + v^2 < \delta^2);$$

$$\Psi(\rho, \varphi) = \exp_p(\rho \cos \varphi \mathbf{v}_1 + \rho \sin \varphi \mathbf{v}_2) \text{ (} 0 < \rho < \delta, \varphi \in]\varphi_0 - \pi, \varphi_0 + \pi[).$$

The coordinates $\Psi(\rho, \varphi)$ are known as *geodesic polar coordinates*, and we call $\Phi(u, v)$ *geodesic Cartesian coordinates*. Of course, these coordinates depend on the choice of the basis $(\mathbf{v}_1, \mathbf{v}_2)$. If we want them to respect the orientation of S , it suffices that $(\mathbf{v}_1, \mathbf{v}_2)$ are positively oriented. In the case of geodesic polar coordinates, we obtain different parameterizations by restricting φ to intervals of length 2π , and each of these parameterizations excludes a *radial geodesic* ($\varphi = \text{constant}$). However, since the excluded radial geodesic is arbitrary, the conclusions we draw with one of these parameterizations are valid throughout the “punctured disc” $D_\delta(p) \setminus \{p\}$.

The next lemma says, among other things, that $\Psi(\rho, \varphi)$ is an orthogonal parameterization. Geometrically, this means that the radial geodesics ($\varphi = \text{constant}$) starting from p and the *circumference geodesics* ($\rho = \text{constant}$) with center p intersect orthogonally (see Fig. 4.11.).

Lemma 4.6.3 *The coefficients E, F, G of the geodesic polar coordinates satisfy*

$$E \equiv 1, \quad F \equiv 0, \quad \lim_{\rho \rightarrow 0} \sqrt{G} = 0, \quad \lim_{\rho \rightarrow 0} (\sqrt{G})_\rho = 1.$$

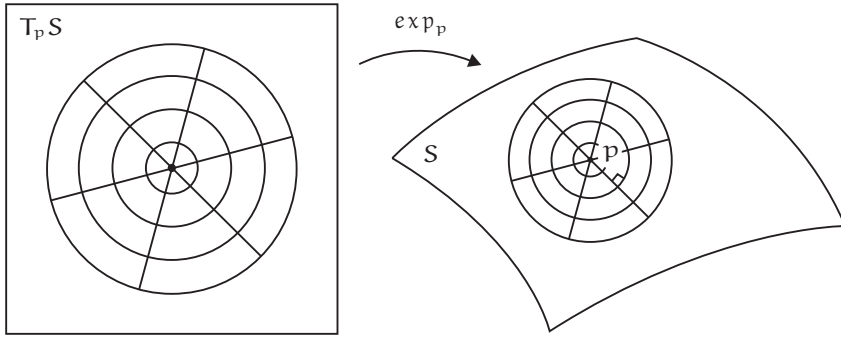


Figure 4.11

Proof To shorten the notation, we write $\mathbf{v}_\varphi = \cos \varphi \mathbf{v}_1 + \sin \varphi \mathbf{v}_2$ and $\mathbf{w}_\varphi = -\sin \varphi \mathbf{v}_1 + \cos \varphi \mathbf{v}_2$. We have

$$E = |\Psi_\rho|^2 = \left| \frac{\partial}{\partial \rho} \exp_p(\rho \mathbf{v}_\varphi) \right|^2 = |\mathbf{v}_\varphi|^2 = 1,$$

since $\rho \mapsto \exp_p(\rho \mathbf{v}_\varphi)$ is a parametrized geodesic, and so has constant scalar velocity. Furthermore, denoting by $\frac{D\Psi_\rho}{d\rho}$ the derivative covariant of Ψ_ρ along $\rho \mapsto \Psi(\rho, \varphi)$ — which is zero because Ψ_ρ is the velocity field of a parametrized geodesic — we have

$$\begin{aligned} F_\rho &= \frac{\partial}{\partial \rho} \langle \Psi_\rho, \Psi_\varphi \rangle = \left\langle \frac{D\Psi_\rho}{d\rho}, \Psi_\varphi \right\rangle + \langle \Psi_\rho, \Psi_{\varphi\rho} \rangle \\ &= \frac{\partial}{\partial \varphi} \left(\frac{1}{2} |\Psi_\rho|^2 \right) = 0 \quad (\text{because } E \equiv 1). \end{aligned}$$

We thus conclude that F does not depend on ρ . Since

$$\begin{aligned} F &= \langle D(\exp_p)_{\rho \mathbf{v}_\varphi}(\mathbf{v}_\varphi), D(\exp_p)_{\rho \mathbf{v}_\varphi}(\rho \mathbf{w}_\varphi) \rangle \\ &= \rho \langle D(\exp_p)_{\rho \mathbf{v}_\varphi}(\mathbf{v}_\varphi), D(\exp_p)_{\rho \mathbf{v}_\varphi}(\mathbf{w}_\varphi) \rangle, \end{aligned}$$

we have

$$|F| \leq \rho |D(\exp_p)_{\rho \mathbf{v}_\varphi}(\mathbf{v}_\varphi)| |D(\exp_p)_{\rho \mathbf{v}_\varphi}(\mathbf{w}_\varphi)|,$$

which, together with the facts that $D(\exp_p)_O$ is the identity (as we saw in the proof of Proposition 4.6.2) and $\mathbf{v}_\varphi$ and $\mathbf{w}_\varphi$ are unit vectors, implies $\lim_{\rho \rightarrow 0} F = 0$ and therefore $F \equiv 0$.

It remains to prove the last two equalities. For this let us consider the geodesic Cartesian coordinates $\Phi(u, v)$, whose coefficients we denote by $\bar{E}, \bar{F}, \bar{G}$. Noting that for $(u, v) \neq (0, 0)$, we have $u = \rho \cos \varphi$, $v = \rho \sin \varphi$ and that at $(0, 0)$ the coefficients $\bar{E}, \bar{F}, \bar{G}$ are equal to 1, 0, 1 (because $\Phi_u|_{(0,0)} = \mathbf{v}_1$ and $\Phi_v|_{(0,0)} = \mathbf{v}_2$), we have

$$\sqrt{G} = \sqrt{EG - F^2} = \left| \frac{\partial(u, v)}{\partial(\rho, \varphi)} \right| \sqrt{E\bar{G} - \bar{F}^2} = \rho \sqrt{E\bar{G} - \bar{F}^2}$$

and therefore $\lim_{\rho \rightarrow 0} \sqrt{G} = 0$. Finally, differentiating the last equality, we obtain

$$\begin{aligned} (\sqrt{G})_\rho &= \sqrt{E\bar{G} - \bar{F}^2} + \rho \frac{\partial}{\partial \rho} \left(\sqrt{E\bar{G} - \bar{F}^2} \right) \\ &= \sqrt{E\bar{G} - \bar{F}^2} + \rho \left\{ \frac{\partial}{\partial u} \left(\sqrt{E\bar{G} - \bar{F}^2} \right) \cos \varphi + \frac{\partial}{\partial v} \left(\sqrt{E\bar{G} - \bar{F}^2} \right) \sin \varphi \right\}, \end{aligned}$$

and thus

$$\lim_{\rho \rightarrow 0} (\sqrt{G})_\rho = \lim_{(u, v) \rightarrow (0, 0)} \sqrt{E\bar{G} - \bar{F}^2} = 1. \quad \square$$

Example 4.6.4 Let us make explicit a geodesic coordinate system in $\mathbf{S}^2$ with origin at the North Pole $\mathcal{N} = (0, 0, 1)$. Fixing the orthonormal basis $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$, the geodesic $\varphi = \text{constant}$ is the meridian that forms angle φ with the plane $y = 0$; its parameterization by arc length is s ($0 \leq s < \pi$) $\mapsto (\sin s \cos \varphi, \sin s \sin \varphi, \cos s)$, and we find the point $\Psi(\rho, \varphi)$ at distance ρ from $\mathcal{N}$. Thus $\Psi(\rho, \varphi) = (\sin \rho \cos \varphi, \sin \rho \sin \varphi, \cos \rho)$ — which shows that these geodesic coordinates are just the spherical coordinates. Let us also point out that $\mathbf{S}^2 \setminus \{(0, 0, -1)\} = D_\pi(\mathcal{N})$ is a normal neighborhood of $\mathcal{N}$, because it is the diffeomorphic image of the disk $B_\pi(\mathcal{N})$ under $\exp_{\mathcal{N}}$. $\square$

We can now almost specify under what conditions a geodesic describes a shortest path on the surface. But first, let us define the *intrinsic distance* on the surface, which is the distance “experienced” by those who move on it. For a **connected** surface S and $p, q \in S$, we define $d(p, q)$ to be the least of the lengths of the curves in S that connect p to q :

$$\begin{aligned} d(p, q) &= \inf \{ l(\alpha) \mid \alpha: [a, b] \rightarrow S \text{ is piecewise regular,} \\ &\quad \alpha(a) = p \text{ and } \alpha(b) = q \}. \end{aligned}$$

The intrinsic distance d is a true distance, since it enjoys the following three properties:

- (i) $d(p, q) \geq 0$ and $d(p, q) = 0$ if and only if $p = q$ (*positivity*);
- (ii) $d(p, q) = d(q, p)$ (*symmetry*);
- (iii) $d(p, q) \leq d(p, r) + d(r, q)$ (*triangular inequality*).

These properties are easily verified; we prove property (iii) via an example. If $\alpha_1: [a, b] \rightarrow S$ is a curve from p to r , and $\alpha_2: [c, d] \rightarrow S$ another from r to q , then their juxtaposition $\alpha_1 * \alpha_2$, defined on $[a, b + d - c]$ by

$$(\alpha_1 * \alpha_2)(t) = \begin{cases} \alpha_1(t) & \text{if } a \leq t \leq b \\ \alpha_2(t + c - b) & \text{if } b \leq t \leq b + d - c, \end{cases}$$

is a piecewise regular curve from p to q , and $l(\alpha_1 * \alpha_2) = l(\alpha_1) + l(\alpha_2)$. We then have

$$\begin{aligned} d(p, q) &\leq \inf_{\alpha_1, \alpha_2} l(\alpha_1 * \alpha_2) = \inf_{\alpha_1, \alpha_2} \{l(\alpha_1) + l(\alpha_2)\} \\ &= \inf_{\alpha_1} l(\alpha_1) + \inf_{\alpha_2} l(\alpha_2) = d(p, r) + d(r, q) \end{aligned}$$

— which proves (iii).

We say that the curve α in S from p to q , which is piecewise regular, *minimizes distance* (or is *minimizing*) if $l(\alpha) = d(p, q)$ (indeed, by Proposition 4.6.8 below, any minimizing curve is a geodesic and is therefore regular).

Example 4.6.5 There do not always exist minimizing curves in S . It is sufficient to take for S a punctured plane (i.e., a plane from which a point O has been removed) and consider for S two points p, q such that $O \in [p, q]$: the intrinsic distance in S between p and q is $|p - q|$, but there is in S no curve from p to q with length $|p - q|$. $\square$

Proposition 4.6.6 *Geodesics locally minimize distance. More exactly, if $D_\delta(p)$ is a normal neighborhood of p , and $q \in D_\delta(p)$, then the radial geodesic from p to q is the only minimizing curve between p and q .*

Proof Take geodesic polar coordinates $\Psi(\rho, \varphi)$ centered at p , and let (ρ_0, φ_0) be such that $\Psi(\rho_0, \varphi_0) = q$. Given a curve $\alpha: [a, b] \rightarrow S$, piecewise regular, such that $\alpha(a) = p, \alpha(b) = q$, our goal is to show that $l(\alpha) \geq \rho_0$, unless α is a reparametrization of the radial geodesic ρ ($0 \leq \rho \leq \rho_0$) $\mapsto \Psi(\rho, \varphi_0)$.

We first deal with the case where $\alpha([a, b]) \subseteq D_\delta(p)$. We may assume, truncating the curve if necessary, that $\alpha(t) \neq p$ for all $t \in]a, b]$. Under this hypothesis, there exist functions $\rho(t)$ and $\varphi(t)$, piecewise differentiable, such that $\varphi(b) = \varphi_0$ and $\alpha(t) = \Psi(\rho(t), \varphi(t))$ for $t \in]a, b]$. Using Lemma 4.6.3, we then have

$$\begin{aligned} l(\alpha) &= \int_a^b \sqrt{E\rho^2 + 2F\rho\varphi + G\varphi^2} dt \\ &= \int_a^b \sqrt{\rho^2 + G\varphi^2} dt \geq \int_a^b \sqrt{\rho^2} dt \geq \int_a^b \rho dt = \rho_0, \end{aligned}$$

and equality holds only if $\varphi \equiv 0$ (i.e., if φ is constant) and $\rho \geq 0$, which implies that α is the radial geodesic from p to q .

If $\alpha([a, b])$ is not contained in $D_\delta(p)$ then, given $0 < r < \delta$, let t_r be the first point such that $\alpha(t_r)$ belongs to the geodesic circumference $\rho = r$. By the above conclusion, we have $l(\alpha) \geq l(\alpha|_{[a, t_r]}) \geq r$. This inequality, valid for all $r < \delta$, implies that $l(\alpha) \geq \delta > \rho_0$. $\square$

Observations 4.6.7 A. From the proof of Proposition 4.6.6 it also follows that, for every point $q \in S \setminus D_\delta(p)$, we have $d(p, q) \geq \delta$. In short, if we put, for arbitrary $\delta > 0$, $D_\delta(p) = \{q \in S : d(p, q) < \delta\}$, this agrees with our previous definition when $D_\delta(p)$ is a normal neighborhood of p .

B. We should stress that we can only guarantee locally that geodesics are minimizing. On the cylinder $x^2 + y^2 = 1$, for example, the geodesic $\alpha(t) = (\cos t, \sin t, t)$ does not minimize the distance between $\alpha(0) = (1, 0, 0)$ and $\alpha(2\pi) = (1, 0, 2\pi)$: the line segment $\{(1, 0)\} \times [0, 2\pi]$ is shorter. $\square$

Proposition 4.6.8 *Every minimizing curve is a geodesic.*

Proof Let $\alpha: [0, L] \rightarrow S$ be a minimizing curve, piecewise regular, parametrized by arc length. It suffices to see that, given any point $p_0 = \alpha(s_0)$ on the curve, there exists $\varepsilon > 0$ such that the restriction from α to $[s_0 - \varepsilon, s_0 + \varepsilon]$ ($[0, \varepsilon]$ or $[L - \varepsilon, L]$, if $s_0 = 0$ or $s_0 = L$) is a geodesic. We take $\delta > 0$ and the neighborhood W of p_0 given in Proposition 4.6.2, and we choose $\varepsilon > 0$ so that $\varepsilon < \frac{\delta}{2}$ and the points $p = \alpha(s_0 - \varepsilon)$ and $q = \alpha(s_0 + \varepsilon)$ are both in W . Since $\alpha|_{[s_0 - \varepsilon, s_0 + \varepsilon]}$ is minimizing, we have $d(p, q) = 2\varepsilon$ and, by note 4.6.7, **A**, q belongs to the normal neighborhood $D_\delta(p)$ of p . It follows then from Proposition 4.6.6 that $\alpha|_{[s_0 - \varepsilon, s_0 + \varepsilon]}$ is a geodesic. $\square$

Note that in the above proof, $V = W \cap D_{\delta/2}(p_0)$ is a normal neighborhood of all its points, therefore enjoying the following property: any two points of V are joined by a single minimizing curve, which is a geodesic of length δ .

In note 4.5.3 **B** we gave a formula (4.31) for the curvature K as a function of the coefficients of an orthogonal parameterization. This formula simplifies substantially for geodesic polar coordinates $\Psi(\rho, \varphi)$, where $E \equiv 1$:

$$K = \frac{-1}{2\sqrt{G}} \frac{\partial}{\partial \rho} \left(\frac{G_\rho}{\sqrt{G}} \right) = \frac{-(\sqrt{G})_{\rho\rho}}{\sqrt{G}}. \quad (4.38)$$

The formula (4.38) becomes even more interesting when K is constant. We now need an elementary result of Calculus:

Lemma 4.6.9 *Let $f:]a, b[\rightarrow \mathbb{R}$ be a continuous function, differentiable on $]a, b[$, and such that $\lim_{x \rightarrow a^+} f'(x)$ exists and is finite. Then f is differentiable at a and $f'(a) = \lim_{x \rightarrow a^+} f'(x)$.*

Fixing φ , Lemmas 4.6.3 and 4.6.9 say that the continuous extension of $\rho \mapsto \sqrt{G}(\rho, \varphi)$ defined by $\sqrt{G}(0, \varphi) = 0$ is differentiable and its derivative at 0 has value 1. Another application of Lemma 4.6.9 (together with the equality $(\sqrt{G})_{\rho\rho} + K\sqrt{G} = 0$) further ensures that it is twice differentiable at 0. We thus have $\rho \mapsto \sqrt{G}(\rho, \varphi)$ as solution of the differential equation

$$x(\rho) + Kx(\rho) = 0 \quad (4.39)$$

with initial conditions $x(0) = 0$ and $x'(0) = 1$. There are three cases to consider:

- i. if $K < 0$ then $\sqrt{G} = \frac{1}{\sqrt{-K}} \sinh(\sqrt{-K}\rho)$, and therefore $G = -\frac{1}{K} \sinh^2(\sqrt{-K}\rho)$;
- ii. if $K = 0$ then $\sqrt{G} = \rho$, $G = \rho^2$;

iii. if $K > 0$ then $\sqrt{G} = \frac{1}{\sqrt{K}} \sin(\sqrt{K}\rho)$, $G = \frac{1}{K} \sin^2(\sqrt{K}\rho)$.

We thus conclude that if K is constant then the coefficients E, F, G of the coordinates $\Psi(\rho, \varphi)$ only depend on K . As a consequence we have the following result.

Theorem 4.6.10 *Any two surfaces of equal constant Gaussian curvature are locally isometric.*

Proof Assume that S_1 and S_2 have the same constant curvature. Given $p \in S_1$ and $q \in S_2$, let $\delta > 0$ be such that $D_\delta(p)$ and $D_\delta(q)$ are normal neighborhoods of p and q . Fixing orthonormal bases $(\mathbf{v}_1, \mathbf{v}_2)$ on $T_p S_1$ and $(\mathbf{w}_1, \mathbf{w}_2)$ on $T_q S_2$, let $L: T_p S_1 \rightarrow T_q S_2$ be the linear isometry such that $L(\mathbf{v}_i) = \mathbf{w}_i$ ($i = 1, 2$). We shall see that $f = \exp_q \circ L \circ \exp_p|_{B_\delta(p)}^{-1}$ is an isometry of $D_\delta(p)$ onto $D_\delta(q)$: in fact, f is clearly a diffeomorphism. Furthermore, the geodesic coordinates $\Psi(\rho, \varphi)$ in $D_\delta(p)$ associated with $(\mathbf{v}_1, \mathbf{v}_2)$ are sent by f to the geodesic coordinates $\tilde{\Psi}(\rho, \varphi)$ in $D_\delta(q)$ associated with $(\mathbf{w}_1, \mathbf{w}_2)$. From what we have seen above, Ψ and $\tilde{\Psi}$ have the same coefficients, and therefore $f|_{D_\delta(p) \setminus \{p\}} = \tilde{\Psi} \circ \Psi^{-1}$ is an isometry. Thus, and since $Df_q = L$, we see that Df_r is a linear isometry for all $r \in D_\delta(p)$, which completes the proof. $\square$

Every surface of constant Gaussian curvature is thus locally isometric to the pseudosphere ($K < 0$), to the plane ($K = 0$), or to the sphere ($K > 0$), but these three examples do not exhaust all surfaces of constant curvature (see, in Section 3.2, example 3.2.1 and Exercise 63). Later on, however, we will see that any **compact** surface in $\mathbb{R}^3$ of constant Gaussian curvature is a sphere.

Exercises

100. (a) Check that the differential equations of the geodesics of a surface of revolution parametrized by

$$\Phi(u, v) = (\rho(v) \cos u, \rho(v) \sin u, z(v))$$

are

$$\begin{cases} u'' + \frac{2\rho}{\rho} u' v' = 0, \\ v'' - \frac{\rho \rho'}{\rho^2 + z^2} u'^2 + \frac{\rho \rho' + z z'}{\rho^2 + z^2} v'^2 = 0 \end{cases}$$

— where ρ, z , etc. denote the derivatives with respect to v , and u', v' , etc. the derivatives with respect to the curve parameter.

(b) Use these equations to conclude again that any meridian $u = \text{constant}$ is a geodesic, and that the parallel $v = v_0$ is a geodesic if and only if the tangent line to the generating curve at v_0 is parallel to the z -axis.

(c) Show that if $\gamma(s) = \Phi(u(s), v(s))$ is a geodesic parametrized by arc length then $(\rho(s))^2 u'(s)$ is constant. Check that $(\rho(s))^2 u'(s) = \rho(s) \cos \theta(s)$, where

$\rho(s) = \rho \circ v(s)$ is the radius of the parallel where γ lies at time s and $\theta(s)$ is the angle of intersection of the curve with that parallel. (The equality $\rho \cos \theta = \text{constant}$ is the *Clairaut Equation*, and plays a fundamental role in the study of geodesics on surfaces of revolution.)

101. (In this exercise use the Clairaut Equation from Exercise 100.) Consider the geodesic γ that starts from a point p on the upper half ($z > 0$) of the hyperboloid of revolution $x^2 + y^2 - z^2 = 1$ and makes angle θ given by $\cos \theta = \frac{1}{\rho}$ with the parallel (with radius ρ) passing through p . Put $\gamma(s) = (\rho(s) \cos \varphi(s), \rho(s) \sin \varphi(s), z(s))$, $\gamma(0) = p$. Show that:

- (a) while $\gamma(s)$ stays in the upper half $z > 0$, however $z'(s) \neq 0$;
- (b) if $z'(0) < 0$ then

$$z(s) > 0 \quad \forall s \in \mathbb{R}, \quad \lim_{s \rightarrow +\infty} z(s) = 0, \quad \lim_{s \rightarrow +\infty} z'(s) = 0, \quad \lim_{s \rightarrow +\infty} \varphi'(s) = 1.$$

102. Consider, on the paraboloid of revolution $z = x^2 + y^2$, the geodesic $\alpha(t) = (0, t, t^2)$, $t \in \mathbb{R}$.

(a) Show that there exists $\varepsilon > 0$ such that for all $0 < t_0 \leq \varepsilon$, $\alpha|_{[-t_0, t_0]}$ minimizes the intrinsic distance between $\alpha(-t_0)$ and $\alpha(t_0)$.

(b) Show that, for t_0 sufficiently large, $\alpha|_{[-t_0, t_0]}$ no longer minimizes the distance between $\alpha(-t_0)$ and $\alpha(t_0)$.

103. Let S be a connected surface on which the sum of the interior angles of any geodesic triangle is equal to π . Show that S is locally isometric to a plane.

104. Show that if all geodesics of a connected surface are planar curves then that surface is contained in a sphere or in a plane.

105. Let S be a surface of constant curvature, and let Δ_1, Δ_2 be geodesic triangles of S . Assuming that Δ_1, Δ_2 are “sufficiently small”, show that:

(a) if two of the sides of Δ_1 are equal then the angles opposite to those sides are also equal;

(b) if Δ_1 and Δ_2 have pairwise equal sides then there exist open subsets $W_1 \supseteq \Delta_1$, $W_2 \supseteq \Delta_2$ and an isometry $f: W_1 \rightarrow W_2$ for which $f(\Delta_1) = \Delta_2$. **Hint:** consider first the case where the Δ_i have the same angle and sides adjacent to that angle.

106. Show that on the surface of revolution $(\rho(z) \cos \varphi, \rho(z) \sin \varphi, z)$ (where $\rho(z) > 0$ for all $z \in \mathbb{R}$), the only minimizing geodesic between two points on the same meridian $\varphi = \text{constant}$ is precisely that meridian.

107. Let S be a connected surface, p a point on S , and let $\alpha:]a, b[\rightarrow S$ be a regular curve that does not pass through p . Consider the intrinsic distance on S .

Show that if $\alpha(t_0)$ is a point of α at the minimizing distance from p and γ a minimizing geodesic starting from p to $\alpha(t_0)$ then γ intersects α orthogonally.

108. Let S_C ($C \neq 1$) be the surface of revolution of constant curvature 1 given in example 3.2.1 and $\gamma \subseteq S_C$ the meridian $u = \pi$.

(a) Explicitly define a local isometry $f: S_C \setminus \gamma \rightarrow \mathbf{S}^2$. **Hint:** send the equator of S_C to the equator of $\mathbf{S}^2$ and use spherical coordinates.

(b) The equator $v = 0$ is a closed (i.e., periodic) geodesic of S_C . Show that S_C has other closed geodesics if and only if C is rational.

109. Consider a geodesic polar coordinate system (ρ, φ) centered at a point $p_0 \in S$ with curvature $K(p_0)$. Prove that:

$$(a) \quad \sqrt{G} = \rho - \frac{K(p_0)}{6} \rho^3 + o(\rho^3), \text{ where } \lim_{\rho \rightarrow 0} \frac{o(\rho^3)}{\rho^3} = 0 \text{ uniformly on } \varphi.$$

Hint: Lemma 4.6.9 and formula (4.38) should show that, given φ , the function $\rho \mapsto \sqrt{G}(\rho, \varphi)$ admits a Taylor polynomial expansion around 0 as above — the problem lies in proving that the said limit is uniform on φ .

(b) by denoting by $l(\rho)$ the perimeter of the geodesic circle with radius ρ centered at p_0 at $K(p_0) = \lim_{\rho \rightarrow 0} \frac{6\pi\rho - 3l(\rho)}{\pi\rho^3}$.

110. (a) Show that in geodesic polar coordinates (ρ, φ) the geodesic curvature of the geodesic circumferences $\rho = \text{constant}$ is given by $\frac{G_\rho}{2G}$.

(b) Conclude that these circles all have constant geodesic curvature if and only if there exist differentiable and strictly positive functions $\beta(\rho)$ and $\lambda(\varphi)$ such that

$$G(\rho, \varphi) = \beta(\rho)\lambda(\varphi). \quad (4.40)$$

(c) Show that if G is of the form (4.40) then the Gaussian curvature along each geodesic circumference $\rho = \text{constant}$ is constant.

(d) Conclude that the only oriented connected surfaces on which any geodesic circumference has constant geodesic curvature are surfaces of constant Gaussian curvature.

111. Show that a conformal mapping $f: S_1 \rightarrow S_2$ that sends geodesics of S_1 to geodesics of S_2 is necessarily a similarity — i.e., there exists $\lambda > 0$ such that $|Df_p(\mathbf{v})| = \lambda|\mathbf{v}|$ for all $p \in S_1$ and $\mathbf{v} \in T_p S_1$.

112. Let p be a point of S and $\alpha(u)$, $|u| < \varepsilon$, a regular curve in S such that $\alpha(0) = p$. Choose along α a unit vector field $\mathbf{w}(u)$ of tangent vectors orthogonal to $\alpha'(u)$ and write $\Phi(u, v) = \gamma(v - v_0; \alpha(u), \mathbf{w}(u))$ (i.e., $\Phi(u, \cdot)$ is the geodesic that at time v_0 passes through $\alpha(u)$ with velocity $\mathbf{w}(u)$). Show that:

(a) $\Phi(u, v)$ is a parameterization in a neighborhood of p (the coordinates $\Phi(u, v)$ obtained this way are called *semi-geodesic*);

(b) $\Phi(u, v)$ is an orthogonal parameterization;

(c) geodesic polar coordinates are an example of semi-geodesic coordinates.

113. Let S be a connected oriented surface and let $f: S \rightarrow S$ be a local isometry. Show that:

(a) if there exists $p \in S$ such that $f(p) = p$ and $Df_p = \text{id}|_{T_p S}$ then f is the identity (**Hint:** what are the geodesics that start from p mapped to?);

- (b) if f is not the identity and if there exists a regular curve $\alpha: I \rightarrow S$ such that $f \circ \alpha = \alpha$ then f reverses orientation of S ;
 (c) the curve α in b) is a geodesic.

114. Let S_1, S_2 be connected surfaces, $f: S_1 \rightarrow S_2$ be a diffeomorphism, and d_1, d_2 be the intrinsic distances on S_1 and S_2 . Show that the following conditions are equivalent:

- (1) f is an isometry;
 (2) $d_1(p, q) = d_2(f(p), f(q))$ for all $p, q \in S_1$.

(Hint for (2) $\Rightarrow$ (1): a) show that f transforms geodesics of S_1 into geodesics of S_2 , preserving scalar velocity; b) use the equivalence between (ii) and (iii) for $\lambda = 1$ in Lemma 4.1.1, §1.)

115. In this exercise we show that any compact surface has a triangulation (a fact used when establishing the global version of the Gauss-Bonnet theorem). Fix $\delta > 0$ so that, for all $p \in S$, $D_\delta(p)$ is a normal neighborhood of p , and consider a family $(\mathcal{R}_k)_{1 \leq k \leq k}$ of geodesic triangles such that:

- the interiors of $\mathcal{R}_j$ ($j = 1, \dots, k$) cover S ;
- the diameter of each $\mathcal{R}_j$ is δ (i.e., if $p, q \in \mathcal{R}_j$ then $d(p, q) < \delta$).

(a) Show that any two edges of two distinct $\mathcal{R}_j$ are either disjoint, or intersect at a single point, or intersect along an arc common to both. Hence, the intersections of the $\mathcal{R}_j$ form a finite number of regions Ω_l .

(b) By properly triangulating each region Ω_l , obtain a triangulation of S .

Appendix: Rotation Index

In this appendix we prove the Rotation Index theorem that we used in the proof of the Gauss-Bonnet theorem. First however let us explain what we mean by *continuous choice of the angle between two vector fields*, an expression used repeatedly in Section 4.5. Let us take two unit vector fields $\mathbf{v}_1(s)$ and $\mathbf{w}(s)$ along a certain curve $\alpha(s)$, for $s \in [a, b]$, and consider an auxiliary vector field $\mathbf{v}_2(s)$ so that $(\mathbf{v}_1, \mathbf{v}_2)$ is, for every s , an orthonormal and positively oriented basis of $T_{\alpha(s)}S$. We can then write $\mathbf{w}(s) = \lambda_1(s)\mathbf{v}_1(s) + \lambda_2(s)\mathbf{v}_2(s)$, where

$$[\lambda_1(s)]^2 + [\lambda_2(s)]^2 = |\mathbf{w}(s)|^2 = 1.$$

This shows that $s \mapsto (\lambda_1(s), \lambda_2(s))$ is a mapping of $[a, b]$ into the unit circle $\mathbf{S}^1$. If the vector fields considered are differentiable, then also this mapping is differentiable, and therefore (see note 1.3.1 and Exercise 7 in Section 1.3) there exists a differentiable mapping $\varphi(s)$ such that $(\lambda_1(s), \lambda_2(s)) = (\cos \varphi(s), \sin \varphi(s))$ — that is, such that

$$\mathbf{w}(s) = \cos \varphi(s)\mathbf{v}_1(s) + \sin \varphi(s)\mathbf{v}_2(s).$$

It is this function $\varphi(s)$ that we call the continuous choice of the oriented angle between $\mathbf{v}_1(s)$ and $\mathbf{w}(s)$. As we noted in note 1.3.1, any other choice of the same angle is the sum of $\varphi(s)$ with an integer multiple of 2π .

For later use, it is convenient to obtain a description like the one we saw above for functions $[a, b] \rightarrow \mathbf{S}^1$ that are not necessarily differentiable:

A.1 Lemma. *Let $F: [a, b] \rightarrow \mathbf{S}^1$ be a continuous function. Then there is a lift of F , that is, a continuous function $\varphi: [a, b] \rightarrow \mathbb{R}$ such that $F(s) = (\cos \varphi(s), \sin \varphi(s))$ for all s on $[a, b]$. Every other lift of F is the sum of φ with a constant integer multiple of 2π .*

Proof Consider the mapping $\Pi(t) = (\cos t, \sin t)$ which wraps the line $\mathbb{R}$ into $\mathbf{S}^1$. Note that the restriction of Π to any interval $[t_1, t_2]$ with $t_2 - t_1 < 2\pi$ is a homeomorphism onto its image, since $[t_1, t_2]$ is compact and $\Pi|_{[t_1, t_2]}$ is continuous and injective. What we are looking for is a continuous function φ such that $F = \Pi \circ \varphi$. The idea is to restrict F to small intervals where we can apply a local inverse of Π to both sides of this equality.

By uniform continuity of F , there exists $\delta > 0$ such that for $|s - t| < \delta$, the points $F(s)$ and $F(t)$ are never diametrically opposite in $\mathbf{S}^1$. If we take a partition $s_0 < s_1 < \dots < s_k$ of $[a, b]$ with $s_i - s_{i-1} < \delta$, then, for every $1 \leq i \leq k$, the image $F([s_{i-1}, s_i])$ is contained in a semi-circle. Let us now define φ recursively, starting at the interval $[s_0, s_1]$. By construction, there exists an interval J_1 , of amplitude π , such that the arc $\Pi(J_1)$ contains $F([s_0, s_1])$: thus, for $s \in [s_0, s_1]$, we define $\varphi(s) = \Pi|_{J_1}^{-1} \circ F(s)$. Assuming we have defined $\varphi(s)$ for all $s \in [s_0, s_{i-1}]$, we take J_i , of amplitude π , such that $\varphi(s_{i-1}) \in J_i$ and $F([s_{i-1}, s_i]) \subseteq \Pi(J_i)$, for $s \in [s_{i-1}, s_i]$, we define $\varphi(s) = \Pi|_{J_i}^{-1} \circ F(s)$. This ends the construction of φ . It is clear that $F = \Pi \circ \varphi$. Since φ is continuous (because it is continuous on every interval $[s_{i-1}, s_i]$), φ is a lift of F .

Regarding uniqueness of φ (minus a constant), the proof is given in note 1.3.1. $\square$

In general, a *lift(ing)* of a continuous function $F: \Omega \rightarrow \mathbf{S}^1$, where Ω is a domain of $\mathbb{R}^n$, is a continuous function $\varphi: \Omega \rightarrow \mathbb{R}$ such that $F = \Pi \circ \varphi$. If we do not impose conditions on Ω , it is not true that all such functions F have a lifting. But we can guarantee the existence of a lifting if, for example, Ω is a rectangle (i.e., the Cartesian product of compact intervals): the next lemma proves this in the two-dimensional case, the only one we need besides the case $n = 1$ treated in A.1.

A.2. Lemma. *Let $F(s, t)$ be a continuous function $[a, b] \times [c, d] \rightarrow \mathbf{S}^1$. Then F has a lifting — that is, there exists $\varphi(s, t)$ continuous such that $F(s, t) = (\cos \varphi(s, t), \sin \varphi(s, t))$.*

Proof Let $\theta(t)$ be a lifting of the mapping $t \mapsto F(0, t)$, whose existence is guaranteed by A.1. Again, using A.1, let us next take, for every $t \in [c, d]$, a lifting $s \mapsto \varphi(s, t)$ of $s \mapsto F(s, t)$ that satisfies $\varphi(0, t) = \theta(t)$. The function φ obtained this way satisfies the equality $F = \Pi \circ \varphi$, and it remains to show that φ is continuous. By construction,

its restriction to each of the horizontal segments $[a, b] \times \{t\}$, and to the vertical line segment $\{0\} \times [c, d]$, is continuous.

Given $\varepsilon > 0$ with $\varepsilon < \frac{\pi}{2}$, the uniform continuity of F gives us $\delta > 0$ such that the angle between $F(s, t)$ and $F(\tilde{s}, \tilde{t})$ has amplitude ε whenever $|(s, t) - (\tilde{s}, \tilde{t})| < \delta$. Thus, given (s_0, t_0) , and under the assumption that $|(s, t) - (s_0, t_0)| < \delta$, we can write

$$\varphi(s, t) - \varphi(s_0, t_0) = \varepsilon(s, t) + 2\pi k(s, t),$$

where $k(s, t)$ is an integer and $|\varepsilon(s, t)| < \varepsilon$. If we show that $k(s, t) = 0$, the continuity of φ on (s_0, t_0) follows. Now, since $\varphi|_{\{0\} \times [c, d]}$ is continuous, we have $|\varphi(0, t) - \varphi(0, t_0)| < \varepsilon < \frac{\pi}{2}$. Furthermore, since the difference $\varphi(\cdot, t) - \varphi(\cdot, t_0)$ is continuous and takes neither of the values $\pm \frac{\pi}{2}$, it takes values in the interval $] -\frac{\pi}{2}, \frac{\pi}{2} [$; we conclude that $|\varphi(s, t) - \varphi(s, t_0)| < \frac{\pi}{2}$. But as also $|\varphi(s, t_0) - \varphi(s_0, t_0)| < \frac{\pi}{2}$, it follows that

$$\pi > |\varphi(s, t) - \varphi(s_0, t_0)| = |\varepsilon(s, t) + 2\pi k(s, t)| \geq 2\pi |k(s, t)| - |\varepsilon(s, t)|,$$

and therefore $k(s, t) = 0$. This concludes the proof of the continuity of φ . $\square$

Here, in a simplified version, is the Rotation Index theorem:

A.3. Rotation Index theorem (first version). *If $\alpha: [a, b] \rightarrow \mathbb{R}^2$ is a closed, regular, simple curve, then $\mathfrak{R}(\alpha) = \pm 1$, where the sign depends on the orientation of the curve.*

Proof We suppose, as usual, that α is parametrized by arc length, and define a continuous function $F: [a, b] \times [a, b] \rightarrow \mathbf{S}^1$ as follows: if $0 < |s - t| < b - a$, we put

$$F(s, t) = \frac{\alpha(s) - \alpha(t)}{|\alpha(s) - \alpha(t)|};$$

in the other cases, we put $F(s, s) = \alpha'(s)$ and $F(a, b) = -F(b, a) = \alpha'(a)$. Denoting by $\varphi(s, t)$ a lifting of $F(s, t)$, the restriction of φ to the diagonal $\{(s, s): a \leq s \leq b\}$ is a lifting of $s \mapsto \alpha'(s)$, and therefore the rotation index is

$$\mathfrak{R}(\alpha) = \frac{1}{2\pi} (\varphi(b, b) - \varphi(a, a)).$$

Suppose that the initial point $\alpha(a)$ is chosen so that the curve α is all on the same side of its tangent line at that point (such a choice is always possible: see for example Section 1.6). Then the image of $F(s, a)$, for $s \in [a, b]$, is contained in a semi-circle, and therefore its lift $\varphi(s, a)$ covers at most an interval of amplitude π . But since $F(a, a) = -F(b, a)$, we have

$$\varphi(b, a) - \varphi(a, a) = \varepsilon\pi,$$

where $\varepsilon = 1$ if α is positively oriented, and $\varepsilon = -1$ otherwise. Since $F(b, t)$, $t \in [a, b]$, runs exactly along the curve at $\mathbf{S}^1$ diametrically opposite to $F(s, a)$, we also have

$$\varphi(b, b) - \varphi(b, a) = \varepsilon\pi.$$

Adding these two equalities, we obtain $\Re(\alpha) = \varepsilon$, which concludes the proof. $\square$

Before we move to a generalization of Theorem A.3, we note that the above proof works under the assumption that α is only of class C^1 .

We will now show that A.3 remains valid for curves on surfaces, provided such curves are contained in parametrized neighborhoods. Let $\alpha(s)$, for $s \in [a, b]$, be a closed, simple, regular curve of class C^1 , contained in the image of the parameterization $\Phi(u, v)$ that we assume to be compatible with the orientation of the surface. Let $\mathbf{v}_1 = \frac{1}{\sqrt{E}}\Phi_u$, and let $\theta(s)$ be a continuous choice of the oriented angle between $\mathbf{v}_1$ and $\alpha'(s)$. We claim that *if α is positively oriented, then $\theta(b) - \theta(a) = 2\pi$.*

The idea, of course, is to apply Theorem A.3 to the planar curve $\beta = \Phi^{-1} \circ \alpha$. If Φ were an isothermal parameterization, our statement would be an immediate consequence of A.3, but it is unnecessary to invoke such a strong result as the existence of isothermal parameterizations. Let $\varphi(s)$ be a continuous choice of the angle between $(1, 0)$ and $\tau(s) = \frac{\beta'(s)}{|\beta'(s)|}$. For every instant s , and since Φ preserves orientation and $D\Phi_{\beta(s)}$ sends $(1, 0)$ and $\tau(s)$ to vectors that are positive multiples of $\mathbf{v}_1$ and $\alpha'(s)$, the dihedra $((1, 0), \tau(s))$ and $(\mathbf{v}_1, \alpha'(s))$ are both positively oriented or both negatively oriented. Hence, $\varphi(s) - \theta(s) \neq \pm\pi$ for all s ; choosing $\theta(a)$ and $\varphi(a)$ with $|\varphi(a) - \theta(a)| < \pi$, we also have $|\varphi(b) - \theta(b)| < \pi$. It follows that

$$|\{\varphi(b) - \varphi(a)\} - \{\theta(b) - \theta(a)\}| < 2\pi$$

or, using A.3,

$$|2\pi - (\theta(b) - \theta(a))| < 2\pi,$$

and therefore, since $\theta(b) - \theta(a)$ is an integer multiple of 2π ,

$$\theta(b) - \theta(a) = 2\pi,$$

which proves our assertion.

Let us now state the more general version of Theorem A.3. used in the proof of the Gauss-Bonnet theorem. Consider a curve $\alpha(s)$, for $s \in [a, b]$, closed, simple, piecewise regular, which is the boundary of a polygonal region Ω contained in a neighborhood parametrized by $\Phi(u, v)$. Let $a = s_0 < s_1 < \dots < s_k = b$ be the $k + 1$ vertices of α .

A.4. Rotation Index theorem (second version). *Let γ_i , for $i = 1, \dots, k$, be the exterior angles of Ω and $\theta_i(s)$, for $s \in [s_{i-1}, s_i]$, a continuous choice of the angle between $\mathbf{v}_1 = \frac{1}{\sqrt{E}}\Phi_u$ and $\alpha'(s)$. Then, if α runs through $\partial\Omega$ with positive orientation, we have*

$$\sum_{i=1}^k \{\theta_i(s_i) - \theta_i(s_{i-1})\} + \sum_{i=1}^k \gamma_i = 2\pi.$$

For the proof, we approximate α by curves α_ε of class C^1 , obtained by rounding each of the vertices of α . For sufficiently small $\varepsilon > 0$, $\alpha_\varepsilon: [0, L_\varepsilon] \rightarrow S$ is thus a closed, simple, regular curve, parametrized by arc length, for which there exist disjoint and consecutive intervals $[a_i^\varepsilon, b_i^\varepsilon] \subseteq [0, L_\varepsilon]$ ($i = 1, \dots, k$) such that each arc $\alpha_\varepsilon([a_i^\varepsilon, b_i^\varepsilon])$ is a segment of the regular arc $\alpha([s_{i-1}, s_i])$, traversed with the same orientation. Furthermore, the total length of the arcs of α and α_ε that are not common to both curves, i.e., in the complement of

$$\bigcup_{i=1}^k \alpha_\varepsilon([a_i^\varepsilon, b_i^\varepsilon]),$$

is ε .

Let us denote by $\theta_\varepsilon(s)$ a continuous choice of the angle between $\mathbf{v}_1$ and $\alpha'_\varepsilon(s)$. It is clear then that each of the differences

$$\theta_\varepsilon(b_i^\varepsilon) - \theta_\varepsilon(a_i^\varepsilon)$$

comes arbitrarily close to $\theta_i(s_i) - \theta_i(s_{i-1})$, taking ε sufficiently small, and in this way also

$$\theta_\varepsilon(a_{i+1}^\varepsilon) - \theta_\varepsilon(b_i^\varepsilon), \text{ for } i = 1, \dots, k-1,$$

$$\text{and } \theta_\varepsilon(a_1^\varepsilon) - \theta_\varepsilon(0) + \theta_\varepsilon(L_\varepsilon) - \theta_\varepsilon(b_k^\varepsilon), \text{ for } i = k,$$

lie close to γ_i for $i = 1, \dots, k$. We thus conclude that

$$\begin{aligned} 2\pi &= \theta_\varepsilon(L_\varepsilon) - \theta_\varepsilon(0) \\ &= \{\theta_\varepsilon(L_\varepsilon) - \theta_\varepsilon(b_k^\varepsilon)\} + \sum_{i=1}^k \{\theta_\varepsilon(b_i^\varepsilon) - \theta_\varepsilon(a_i^\varepsilon)\} + \\ &\quad + \sum_{i=1}^{k-1} \{\theta_\varepsilon(a_{i+1}^\varepsilon) - \theta_\varepsilon(b_i^\varepsilon)\} + \{\theta_\varepsilon(a_1^\varepsilon) - \theta_\varepsilon(0)\} \end{aligned}$$

is arbitrarily close to

$$\sum_{i=1}^k \{\theta_i(s_i) - \theta_i(s_{i-1})\} + \sum_{i=1}^k \gamma_i,$$

which proves the theorem.

Exercises

116. Lemma A.2 applies not only to rectangles, but also, obviously, to any regions that are homeomorphic to rectangles. For example, any continuous function $F: \mathbb{D}^2 \rightarrow \mathbf{S}^1$, where $\mathbb{D}^2$ is the closed disk $\{p \in \mathbb{R}^2: |p| \leq 1\}$, has a lifting. Using this fact, we next give a proof of Brouwer's fixed point theorem: *any continuous function $f: \mathbb{D}^2 \rightarrow \mathbb{D}^2$ has some fixed point* (i.e., a point p such that $f(p) = p$).

Assuming that f has no fixed points, we define $F: \mathbb{D}^2 \rightarrow \mathbf{S}^1$ by

$$F(p) = \frac{f(p) - p}{|f(p) - p|},$$

and consider a lifting $\varphi: \mathbb{D}^2 \rightarrow \mathbb{R}$ of F .

(a) Write $\gamma(t) = F(\cos t, \sin t)$, for $t \in [0, 2\pi]$, and let $\theta(t)$ be a lifting of γ such that $\theta(0) \in]\frac{\pi}{2}, \frac{3\pi}{2}[$. Show that $\theta(t) \in]t + \frac{\pi}{2}, t + \frac{3\pi}{2}[$ for all t , and conclude that $\theta(2\pi) - \theta(0) = 2\pi$.

(b) Note that $\tilde{\theta}(t) = \varphi(\cos t, \sin t)$ is also a lift of $\gamma(t)$, but that $\tilde{\theta}(2\pi) = \tilde{\theta}(0)$. What follows?

117. Let $\Omega \subseteq \mathbb{R}^2$ be an open disk and let $\mathbf{w}: \Omega \rightarrow \mathbb{R}^2$ be a differentiable field of unit vectors. Given a regular closed curve $\alpha: [a, b] \rightarrow \Omega$, show that the rotation index of α is also equal to

$$\frac{1}{2\pi} (\theta(b) - \theta(a)),$$

where θ is a continuous choice of the (oriented) angle of $\mathbf{w}(\alpha(t))$ and $\alpha'(t)$.

118. Consider, in $\mathbb{T}^2$ with the parameterization of Exercise 31, the vector field $\mathbf{w}(u, v) = (-\sin v \cos u, -\sin v \sin u, \cos v)$. Let the *rotation index* of a regular closed curve $\alpha: [a, b] \rightarrow \mathbb{T}^2$ be the integer

$$\Delta(\alpha) = \frac{1}{2\pi} (\theta(b) - \theta(a)),$$

where θ is a continuous choice of the oriented angle of $\mathbf{w}$ and $\alpha'(t)$.

(a) Compute $\Delta(\alpha)$ for the curve $t \mapsto \Phi(t, nt)$, where $n \in \mathbb{Z}$ is a constant and $t \in [0, 2\pi]$.

(b) Does the result of (a) depend on the vector field $\mathbf{w}$?



Chapter 5

The Global Geometry of Surfaces

Global geometry deals with those results that concern the surface as a whole. In Chapter 4 we have already seen examples of global theorems, such as the divergence theorem (Corollary 4.4.3.i) and the Gauss-Bonnet theorem (Theorem 4.5.6). Another example is the sphere theorem that we will present in this chapter: any compact surface of constant curvature in $\mathbb{R}^3$ is a sphere. As we already observed, the assumption that the curvature is constant is insufficient, and a global condition (in this case, the compactness of the surface) is needed to draw such a conclusion.

Compact surfaces are inextensible, in the following sense: if S_1 and S_2 are connected surfaces such that S_1 is compact and $S_1 \subseteq S_2$ then $S_1 = S_2$. The global results must naturally deal with inextensible surfaces. In Section 5.1 we will define the notion of a complete surface, which is a sufficient but not necessary condition for a surface to be inextensible.

This chapter includes a mixed bag of topics unusual in introductory texts of Differential Geometry: for instance, a Blaschke formula for surfaces of constant width (Theorem 5.5.3) and the description of all complete surfaces of constant non-negative curvature (Theorems 5.7.7 and 5.7.10). To help the readers in their choice of topics, we mention that Sections 5.1 and 5.2 form the main body of the chapter, from which two independent branches emerge, one consisting of 5.4 and 5.5 and the other one of 5.3, 5.6 and 5.7.

All surfaces in this chapter are connected.

5.1 Complete Surfaces

In Chapter 4 we gave an example (4.6.5) of a surface on which no minimizing geodesic exists between given two points. Even though this example seems disingenuous (the surface in question is a plane from which a point has been removed “unduly”), it points out a possible problem: although there is a geodesic γ that starts from p in the direction of q , γ is not defined for some value of the parameter and therefore does not reach q .

We say that a surface S is *complete* if any parametrized geodesic $\gamma(t; p, \mathbf{v})$ of S is defined for all $t \in \mathbb{R}$. Equivalently, S is complete if $\exp_p$ is defined on the entire tangent space $T_p S$ for all $p \in S$. The next result gives us many examples of complete surfaces.

Proposition 5.1.1 *If a surface $S \subseteq \mathbb{R}^3$ is closed in $\mathbb{R}^3$ then S is complete.*

Proof Given $p \in S$ and $\mathbf{v} \in T_p S$ with $|\mathbf{v}| = 1$, we verify that the geodesic $\gamma(s) = \gamma(s; p, \mathbf{v})$ is defined for every $s \in [0, +\infty[$ (replacing $\mathbf{v}$ by $-\mathbf{v}$, it follows that $\gamma(s)$ is also defined on the entire interval $]-\infty, 0]$). For this, it suffices to check that if γ is defined on $[0, s_0[$ then it is defined on some interval $[0, s_0 + \delta[$ for $\delta > 0$ as well.

Take a monotone sequence $(s_n)_{n \geq 1}$ on $[0, s_0[$ that converges to s_0 . Since

$$|\gamma(s_n) - \gamma(s_m)| \leq d(\gamma(s_n), \gamma(s_m)) \leq |s_n - s_m|,$$

it follows that $(\gamma(s_n))_{n \geq 1}$ is a Cauchy sequence in $\mathbb{R}^3$. Since S is closed, the limit p of this sequence is a point of S . We can then choose $\eta > 0$ and a neighborhood U of p in S such that, for all $q \in U$, the radial geodesics starting from q have length at least η . We fix n so that $s_n > s_0 - \frac{\eta}{2}$ and $\gamma(s_n) \in U$, and let $q = \gamma(s_n)$, $\mathbf{w} = \gamma'(s_n)$. The geodesic $\tilde{\gamma}(t) = \gamma(t; q, \mathbf{w})$ is defined for $t \in]-\eta, \eta[$, and $\gamma(s) = \tilde{\gamma}(s - s_n)$. Thus $\gamma(s)$ is extensible to the interval $[0, s_n + \eta[\supseteq [0, s_0 + \frac{\eta}{2}[$. $\square$

In particular, it follows from the above proposition that all compact surfaces are complete. However, we observe that there are complete surfaces other than those closed in $\mathbb{R}^3$:

Example 5.1.2 Let S be the surface given by

$$\Phi(u, v) = ((1 + e^{-u}) \cos u, (1 + e^{-u}) \sin u, v),$$

$u, v \in \mathbb{R}$. S is complete because it is isometric to the plane, which is a complete surface. However, S is not a closed subset of $\mathbb{R}^3$: a point of the cylinder $x^2 + y^2 = 1$ lies outside of S , and is moreover the limit of a sequence of points in S (see Fig. 5.1).

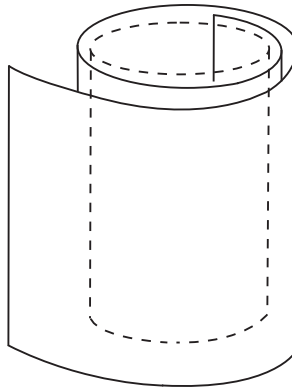


Figure 5.1

The intrinsic distance d in S and the distance $|\cdot|$ in $\mathbb{R}^3$ are *equivalent*, in the sense that a sequence (p_n) in S converges to $p \in S$ for one of these distances if and only if it converges for the other one. However, this example shows that they do not have to be *uniformly equivalent*: there can be a sequence of points on S that is Cauchy for $|\cdot|$ but is not so for d . $\square$

The major advantage of complete surfaces is that there exist minimizing geodesics between any two points on them. Before we prove this, we note that it follows from the triangle inequality (see Section 4.6) that, for every $p \in S$, the mapping $q \mapsto d(p, q)$ is continuous (since $|d(p, q) - d(p, r)| \leq d(q, r) \quad \forall q, r \in S$) — and therefore its restriction to any compact S attains a maximum and a minimum.

Theorem 5.1.3 *Let p, q be two points of a complete surface S . Then there is on S some minimizing geodesic from p to q .*

Proof Let $D_\delta(p)$ be a normal neighborhood of p . If $q \in D_\delta(p)$, there is nothing to show. Otherwise, letting $l = d(p, q)$, we have $l \geq \delta$. Consider now the geodesic circumference $\mathbf{S}^1(p, \frac{\delta}{2})$ with center p and radius $\frac{\delta}{2}$, and let r be a point of $\mathbf{S}^1(p, \frac{\delta}{2})$ at the shortest possible distance from q . By the triangle inequality, we have

$$d(r, q) \geq d(p, q) - d(p, r) = l - \frac{\delta}{2}. \quad (5.1)$$

Let $\alpha(t)$, for $t \in [a, b]$, be any piecewise regular curve from p to q , and let t_0 be the first instant at which $\alpha(t_0) \in \mathbf{S}^1(p, \frac{\delta}{2})$. Then

$$l(\alpha) = l(\alpha|_{[a, t_0]}) + l(\alpha|_{[t_0, b]}) \geq d(p, \alpha(t_0)) + d(\alpha(t_0), q) \geq \frac{\delta}{2} + d(r, q),$$

which implies $l \geq \frac{\delta}{2} + d(r, q)$. Together with (5.1), this gives $d(r, q) = l - \frac{\delta}{2}$. Let $\gamma(s)$ be the radial geodesic, parametrized by arc length, such that $\gamma(0) = p$ and $\gamma(\frac{\delta}{2}) = r$. Note that for $s \geq 0$ we have $d(\gamma(s), q) \geq d(p, q) - d(p, \gamma(s)) \geq l - s$. Defining

$$I = \{s \in [0, l] : d(\gamma(s), q) = l - s\},$$

the preceding inequality shows that $s \in I$ if and only if $s \geq 0$ and $d(\gamma(s), q) \leq l - s$. It follows that I is a (necessarily closed) interval: in fact, if $s \in I$, and $0 \leq t < s$, then $d(\gamma(t), q) \leq d(\gamma(t), \gamma(s)) + d(\gamma(s), q) \leq (s - t) + (l - s) = l - t$, and therefore $t \in I$. We have already seen that $[0, \frac{\delta}{2}] \subseteq I$; let us now show that $I = [0, l]$. To this end, it suffices to show that if $s_0 \in]0, l[$ is in I then also $s_0 + \eta \in I$ for some $\eta > 0$ — because then necessarily $\sup I = l$. Let us fix a normal neighborhood $D_{2\eta}(\gamma(s_0))$ of $\gamma(s_0)$ so that $d(\gamma(s_0), q) > \eta$, and let $\tilde{r}$ be a point of $\mathbf{S}^1(\gamma(s_0), \eta)$ at the shortest possible distance from q : the above argument shows that

$$d(\tilde{r}, q) = d(\gamma(s_0), q) - \eta = l - (s_0 + \eta) \quad (\text{because } s_0 \in I), \quad (5.2)$$

and therefore

$$d(\gamma(s_0 - \eta), \tilde{r}) \geq d(\gamma(s_0 - \eta), q) - d(\tilde{r}, q) = 2\eta. \quad (5.3)$$

The restriction $\gamma|_{[s_0 - \eta, s_0]}$, followed by the radial geodesic from $\gamma(s_0)$ to $\tilde{r}$, is a piecewise regular curve from $\gamma(s_0 - \eta)$ to $\tilde{r}$ whose length is exactly 2η . By (5.3) the curve is minimizing and therefore a geodesic

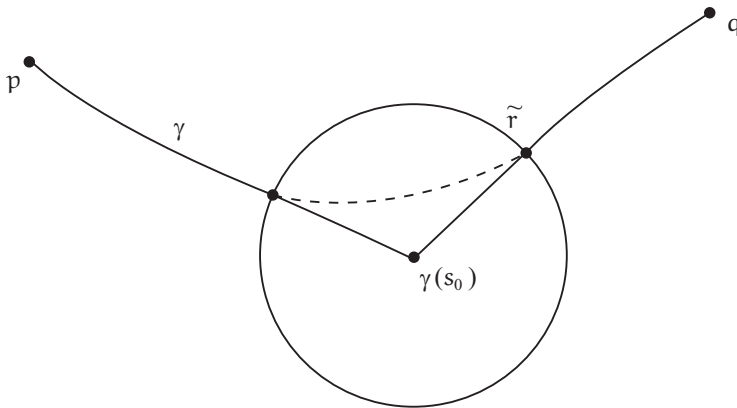


Figure 5.2

(Proposition 4.6.8), so that it coincides with $\gamma|_{[s_0 - \eta, s_0 + \eta]}$. It follows that $\gamma(s_0 + \eta) = \tilde{r}$ and, by (5.2), $s_0 + \eta \in I$. We then have $I = [0, l]$; in particular, $d(\gamma(l), q) = 0$ and therefore $\gamma(l) = q$. Thus, and since $d(p, q) = l$ as well, $\gamma|_{[0, l]}$ is a minimizing geodesic from p to q . $\square$

If we read the above proof carefully, we obtain: if $p \in S$ is such that all geodesics starting from p extend to all values of the parameter then, for every point q of S , there is some minimizing geodesic from p to q . This observation lets us easily assert that any complete surface S is *inextensible*: if S is contained in another surface S_1 then $S_1 = S$. In fact, S is necessarily open in S_1 , and therefore the geodesics of S are also geodesics of S_1 . Let us fix $p \in S$: the geodesics of S_1 that start from p , because they are geodesics of S , are defined for all values of the parameter. This means that, given $q \in S_1$, there exists some minimizing geodesic γ from p to q . But $\gamma \subseteq S$ and therefore $q \in S$ — which shows that $S_1 \subseteq S$.

Example 5.1.4 Besides the punctured plane, also the cone $\mathcal{C}$ given by the equation $z = \sqrt{x^2 + y^2}$ ($z > 0$) is a non-complete surface, since the generating lines are geodesics that are not defined for all parameter values. Nevertheless, between every pair of points of $\mathcal{C}$ there exists a minimizing geodesic (Exercise 89 of Section 4.3) — and $\mathcal{C}$ is inextensible. To prove the latter statement let us suppose, to the contrary, that there exists a connected surface S such that $\mathcal{C} \subseteq S$ but $\mathcal{C} \neq S$. Let p be a point on the boundary of $\mathcal{C}$ in S : p does not belong to $\mathcal{C}$ because $\mathcal{C}$ is open in S . Let $(p_n)_{n \geq 1}$ be a sequence of points of $\mathcal{C}$ that converges to p .

Lemma Given $\varepsilon > 0$, there exists n_0 such that p_n is **below** the plane $z = \varepsilon$ for all $n \geq n_0$.

If the lemma is false, then some infinite subsequence of $(p_n)_{n \geq 1}$ is contained in $\mathcal{C} \cap \{z \geq \varepsilon\}$ — which is a closed subset of $\mathbb{R}^3$ and therefore contains all the accumulation points of the mentioned subsequence; then $p \in \mathcal{C} \cap \{z \geq \varepsilon\} \subseteq \mathcal{C}$, which is absurd. We have thus proved the lemma.

It immediately follows that $\lim_{n \rightarrow +\infty} p_n = (0, 0, 0)$. Thus the boundary of $\mathcal{C}$ on S consists only of the point $(0, 0, 0)$, which implies $S = \mathcal{C} \cup \{(0, 0, 0)\}$, otherwise $S \setminus \{(0, 0, 0)\}$ would not be connected. But $\mathcal{C} \cup \{(0, 0, 0)\}$ is not a surface, which proves the non-existence of S . $\square$

Exercises

119. (a) Show that S is a complete surface if and only if (S, d) (where d is the intrinsic distance on S) is a complete metric space. (*Complete* means that any Cauchy sequence converges. Try to prove the stronger statement that any bounded sequence has a convergent subsequence.)

(b) Show that if there exists $p \in S$ such that any geodesic passing through p is defined in $\mathbb{R}$, then S is complete.

120. If S is a non-compact, complete surface, and p a point of S , then there exists a geodesic $\gamma(s)$ of S such that $\gamma(0) = p$ and that minimizes the distance between p and $\gamma(s)$ for all $s \in \mathbb{R}$.

121. Let S be a complete surface, p a point of S , and $\mathbf{v} \in T_p S$ a unit vector. Write $\gamma(s) = \exp_p(s\mathbf{v})$, and suppose that there exists $s > 0$ such that $\gamma|_{[0,s]}$ does not minimize the distance between p and $\gamma(s)$. Consider the set $I = \{s \geq 0 : d(p, \gamma(s)) = s\}$. Show that:

- (a) I is a closed interval $[0, s_0]$;
- (b) for $s > s_0$ there exists a geodesic $\tilde{\gamma}$ of length s connecting p to $\gamma(s)$;
- (c) for $0 < s < s_0$, $\gamma|_{[0,s]}$ is the only minimizing geodesic connecting p with $\gamma(s)$;
- (d) for $s = s_0$ two cases could apply:
 - there is another minimizing geodesic connecting p with $\gamma(s_0)$;
 - $s_0 \mathbf{v}$ is not a regular point of $\exp_p$.

122. Consider in $\mathbb{T}^2$, parametrized by

$$\Phi(u, v) = ((2 + \cos v) \cos u, (2 + \cos v) \sin u, \sin v),$$

the points $p = \Phi(0, 0)$, $q = \Phi(u_0, v_0)$ and $r = \Phi(u_0, \pi)$, where $0 < u_0 < \pi$, for $0 < v_0 < \pi$. Let $\gamma(s) = \Phi(u(s), v(s))$, for $s \in [0, a]$, be a minimizing geodesic from p to q with $u(0) = v(0) = 0$. Show that:

- (a) $0 \leq u(s) \leq u_0$, $0 \leq v(s) < \pi$ for all $s \in [0, a]$;
- (b) there are two and only two minimizing geodesics from p to r (use the Clairaut Equation, Exercise 100, to conclude that there is only one such geodesic in the region $0 \leq u \leq u_0$, $0 \leq v \leq \pi$);
- (c) if $\tilde{\gamma}(s)$ is a geodesic such that $\tilde{\gamma}(0) = p$ and $\tilde{\gamma}(b) = r$, then $\tilde{\gamma}(2b) = \Phi(2u_0, 0)$;
- (d) if $\frac{\pi}{u_0}$ is rational then the geodesic $\tilde{\gamma}$ is periodic, otherwise it is dense in $\mathbb{T}^2$;

- (e) all geodesics (except the parallel $v = \pi$) intersect the parallel $v = 0$;
- (f) there exist in $\mathbb{T}^2$ geodesics that are neither periodic nor dense.

123. Consider in $\mathbb{T}^2$ the parameterization $\Phi(u, v)$ from Exercise 122. Let Δ be the geodesic triangle of vertices $p = \Phi(0, 0)$, $q = \Phi(\frac{\pi}{2}, 0)$, $r = \Phi(\frac{\pi}{2}, v_0)$, where $0 < v_0 < \pi$, and where $\widehat{pq} = \{\Phi(u, 0) : 0 \leq u \leq \frac{\pi}{2}\}$, $\widehat{qr} = \{\Phi(\frac{\pi}{2}, v) : 0 \leq v \leq v_0\}$, and $\widehat{pr}$ is a minimizing geodesic from p to r .

Use the Clairaut Equation to show that the sum of the interior angles of Δ is greater than π . Can you draw the same conclusion from the Gauss-Bonnet theorem?

124. Let S be a connected surface such that for every p on S , there exists an isometry $\xi_p: S \rightarrow S$ with $\xi_p(p) = p$ and $D(\xi_p)_p = -\text{id}$. Show that:

- (a) the sphere satisfies that condition;
- (b) such a surface is complete and has constant curvature.

125. Let $\mathcal{C}$ be the cone $z = \sqrt{x^2 + y^2}$, $z > 0$. Show that there is no mapping $f: \mathcal{C} \rightarrow S$ such that S is a complete surface and f an isometry onto the image (i.e., there is no isometric *embedding* of $\mathcal{C}$ into a complete surface).

5.2 Coverings

The notion of a covering is one of the most fruitful in topology — but, to save time, we will restrict ourselves to coverings of surfaces, though they prove useful in more general topological spaces. The theory developed here will allow us to show that a good number of surfaces are (globally) images of local diffeomorphisms of standard surfaces like the plane or the sphere.

A mapping $f: S_1 \rightarrow S_2$ between two surfaces is called a *covering (map)* of S_2 if each $q \in S_2$ has an open neighborhood U with the following property: $f^{-1}(U)$ is a collection of disjoint open subsets $(U_i)_i$ such that, for every i , the restriction $f|_{U_i}$ is a diffeomorphism of U_i on U . Such an open subset U is called an *evenly covered neighborhood* of q — or simply an *evenly covered open subset*; see Fig. 5.3.

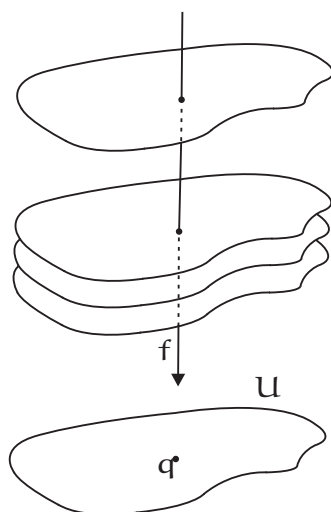


Figure 5.3

Example 5.2.1 The mapping $\mathbb{R}^2 \rightarrow \{x^2 + y^2 = 1\}$ given by $(u, v) \rightarrow (\cos u, \sin u, v)$ is a covering; as is the usual mapping $\mathbb{R}^2 \rightarrow \mathbb{T}^2$ given by

$$\Phi(u, v) = ((2 + \cos v) \cos u, (2 + \cos v) \sin u, \sin v).$$

It follows from the definition that any covering is a locally surjective diffeomorphism, but it is worthy to note that not all locally surjective diffeomorphisms are coverings. An example is the restriction of Φ to the square $]-2\pi, 2\pi[\times]-2\pi, 2\pi[$: in Figure 5.4, the preimage of the marked open subset U joins nine disjoint open subsets, but only one of them is surjectively sent onto U under Φ . $\square$

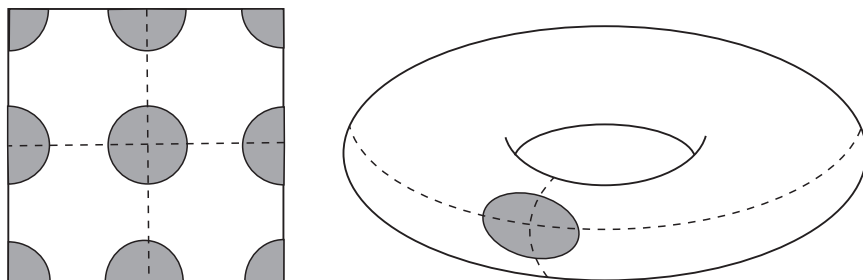


Figure 5.4

There is however an additional condition under which a local diffeomorphism is guaranteed to be a covering:

Proposition 5.2.2 *If S_1 is compact and $f: S_1 \rightarrow S_2$ a local diffeomorphism then f is a covering.*

Proof Let us first show that f is surjective. Since f is a local diffeomorphism, its image $f(S_1)$ is open in S_2 . Given a sequence $(f(p_n))_{n \geq 1}$ on $f(S_1)$ that converges to $q \in S_2$, if $p \in S_1$ is an accumulation point of $(p_n)_{n \geq 1}$ (which exists because S_1 is compact), then $f(p) = q$. Thus $f(S_1)$ is also closed in S_2 and, since S_2 is connected, necessarily $f(S_1) = S_2$.

We now observe that each point q of S_2 has a finite number of preimages, otherwise they would accumulate in the neighborhood of some point of S_1 , in contradiction to f being a local diffeomorphism. Let $p_1, \dots, p_k$ be the preimages of q , and let us choose open neighborhoods V_i of p_i such that each $f|_{V_i}$ is a diffeomorphism onto the image and $V_i \cap V_j$ is empty if $i \neq j$.

Claim: *There is $\varepsilon > 0$ such that $f^{-1}(\{r \in S_2: |r - q| < \varepsilon\}) \subseteq \bigcup_{i=1}^k V_i$.*

In fact, the negation of this statement implies the existence of a sequence $(\tilde{r}_n)_{n \geq 1}$ in $S_1 \setminus \left(\bigcup_{i=1}^k V_i\right)$ such that $\lim_{n \rightarrow +\infty} f(\tilde{r}_n) = q$, and any accumulation point of $(\tilde{r}_n)_{n \geq 1}$ is a preimage of q distinct from $p_1, \dots, p_k$. Finally, for the ε given by the claim, the open subset $\{r \in S_2: |r - q| < \varepsilon\} \cap \left(\bigcap_{i=1}^k f(V_i)\right)$ is an evenly covered neighborhood of q . $\square$

We say that a differentiable mapping $f: S_1 \rightarrow S_2$ *lifts curves* if, for every curve $\alpha: [a, b] \rightarrow S_2$ and $p \in S_1$ such that $f(p) = \alpha(a)$, there exists a single curve $\tilde{\alpha}: [a, b] \rightarrow S_1$ such that $\tilde{\alpha}(a) = p$ and $f \circ \tilde{\alpha}(t) = \alpha(t)$ for all $t \in [a, b]$. That is, for every preimage p of the final point $\alpha(a)$ of α , there exists a single curve in S_1 that starts from p and whose image under f is α . Note that any mapping that lifts curves is necessarily surjective.

Proposition 5.2.3 *Every covering lifts curves.*

Proof Let $f: S_1 \rightarrow S_2$ be a covering, $\alpha: [a, b] \rightarrow S_2$ a curve, and $p \in S_1$ a preimage of $\alpha(a)$. By compactness of $[a, b]$ there exists a partition

$$a = t_0 < t_1 < \dots < t_k = b$$

such that each $\alpha([t_{i-1}, t_i])$ is contained in an evenly covered open subset W_i . We construct $\tilde{\alpha}$ step by step, starting with the interval $[t_0, t_1]$: denoting by $\tilde{W}_1$ the component of $f^{-1}(W_1)$ that contains p , we define $\tilde{\alpha}|_{[t_0, t_1]} = f|_{\tilde{W}_1}^{-1} \circ \alpha|_{[t_0, t_1]}$: $\tilde{\alpha}$ is continuous and obviously satisfies $\tilde{\alpha}(t_0) = p$ and $f \circ \tilde{\alpha} = \alpha$.

Assuming that we constructed $\tilde{\alpha}|_{[t_0, t_{j-1}]}$ and defined $\tilde{W}_i$ for $i = 0, \dots, j-1$, we denote by $\tilde{W}_j$ the component of $f^{-1}(W_j)$ that contains $\tilde{\alpha}(t_{j-1})$ and we put $\tilde{\alpha}|_{[t_{j-1}, t_j]} = f|_{\tilde{W}_j}^{-1} \circ \alpha|_{[t_{j-1}, t_j]}$. Repeating this until $j = k$, we complete the construction of $\tilde{\alpha}$.

Regarding uniqueness of $\tilde{\alpha}$, it follows from the following general observation: *if $f: S_1 \rightarrow S_2$ is a local diffeomorphism and $\widehat{\alpha}, \tilde{\alpha}: [a, b] \rightarrow S_1$ are curves such that $\widehat{\alpha}(a) = \tilde{\alpha}(a)$ and $f \circ \widehat{\alpha} = f \circ \tilde{\alpha}$, then $\widehat{\alpha} = \tilde{\alpha}$.*

To see this, we observe that $I = \{t \in [a, b]: \widehat{\alpha}(t) = \tilde{\alpha}(t)\}$ is open and closed on $[a, b]$, by which $I = [a, b]$ (I is closed because $\widehat{\alpha}$ and $\tilde{\alpha}$ are continuous, and is open because f is a local diffeomorphism). $\square$

A remarkable property of coverings $f: S_1 \rightarrow S_2$ is that *all points of S_2 have the same number of preimages*, which is a simple consequence of Proposition 5.2.3. Indeed, given p and q in S_2 , consider a curve α in S_2 from p to q . For each preimage p_i of p , let α_i be the lifting of α that starts from p_i . The endpoint q_i of α_i is a preimage of q . By Proposition 5.2.3, if $i \neq j$ then $q_i \neq q_j$ — since, obviously, there is a unique lifting for a given endpoint. Thus $\#f^{-1}(\{p\}) \leq \#f^{-1}(\{q\})$, and swapping the roles of p and q we obtain the opposite inequality.

The number of preimages of each point of S_2 is usually called the *number of sheets* of the covering.

By Proposition 5.2.3, we see that any covering $f: S_1 \rightarrow S_2$ enjoys the following properties:

- (i) f is a local diffeomorphism,
- (ii) f lifts curves,

— and the converse is also true:

Proposition 5.2.4 *If $f: S_1 \rightarrow S_2$ has properties (i) and (ii) then f is a covering.*

Note: In topological spaces other than surfaces — where, instead of diffeomorphisms, we would speak of local homeomorphisms — (i) and (ii) no longer guarantee that a mapping is a covering.

We say that two curves $\alpha_0, \alpha_1: [a, b] \rightarrow S_2$ with $\alpha_0(a) = \alpha_1(a) = q_0$ and $\alpha_0(b) = \alpha_1(b) = q_1$ are *homotopic with fixed endpoints* if there exists a continuous function

$$H: [0, 1] \times [a, b] \rightarrow S_2$$

such that $H(0, t) = \alpha_0(t)$, $H(1, t) = \alpha_1(t)$ for all $t \in [a, b]$, and $H(s, a) = q_0$, $H(s, b) = q_1$ for all $s \in [0, 1]$. Letting $\alpha_s = H(s, \cdot)$, the family of curves $(\alpha_s)_{0 \leq s \leq 1}$ is called a *continuous deformation of α_0 with fixed endpoints*. Note that if α_0 is closed then all α_s curves are also closed.

Lemma 5.2.5 *With the above notation, let $\tilde{\alpha}_s$, for $0 \leq s \leq 1$, be a lifting of α_s in S_1 . If all the curves $\tilde{\alpha}_s$ have the same initial point p_0 , then they also all have the same endpoint.*

Proof It suffices to show that, given $s_0 \in [0, 1]$, there exists $\delta > 0$ such that if $|s - s_0| < \delta$, for all $s \in [0, 1]$ then $\tilde{\alpha}_s(b) = \tilde{\alpha}_{s_0}(b)$. Let $a = t_0 < t_1 < \cdots < t_l = b$ be a partition of $[a, b]$ such that $\tilde{\alpha}_{s_0}([t_{i-1}, t_i]) \subseteq \tilde{D}_i$, where for each i :

- $\tilde{D}_i$ is an open subset diffeomorphic to a disk of the plane;
- if $i < k$ then there exists an open subset including the closure of $\tilde{D}_i \cup \tilde{D}_{i+1}$ such that the restriction of f to it is a diffeomorphism onto the image.

Every $D_i = f(\tilde{D}_i)$ is an open subset of S_2 diffeomorphic to a disk. By (uniform) continuity of $H(s, t) = \alpha_s(t)$, there exists $\delta > 0$ such that if $|s - s_0| < \delta$, for all $s \in [0, 1]$, then $\alpha_s(t) \in D_i$ for all $t \in [t_{i-1}, t_i]$ and $i = 1, \dots, k$.

Fixing $s \in [0, 1] \cap]s_0 - \delta, s_0 + \delta[$, let us show that $\tilde{\alpha}_s([t_{i-1}, t_i]) \subseteq \tilde{D}_i$ for $i = 1, \dots, k$. For $i = 1$, we have $\tilde{\alpha}_s(t_0) = p_0 \in \tilde{D}_1$; if not $\tilde{\alpha}_s([t_0, t_1]) \subseteq \tilde{D}_1$, there is $t \in]t_0, t_1]$ such that $\tilde{\alpha}_s(t)$ is on the boundary of $\tilde{D}_1$. For this t , the point $\alpha_s(t)$ is on the boundary of D_1 , which contradicts the choice of s . We then have $\tilde{\alpha}_s([t_0, t_1]) \subseteq \tilde{D}_1$. Assuming now that $\tilde{\alpha}_s([t_{i-1}, t_i]) \subseteq \tilde{D}_i$ ($i < k$), we observe that $\alpha_s(t_i) = f(\tilde{\alpha}_s(t_i))$ is in $D_i \cap D_{i+1}$. Since, by the second condition above, the restriction of f to $\tilde{D}_i \cup \tilde{D}_{i+1}$ is a bijection onto $D_i \cup D_{i+1}$, we have $f|_{\tilde{D}_i \cup \tilde{D}_{i+1}}^{-1}(D_i \cap D_{i+1}) = \tilde{D}_i \cap \tilde{D}_{i+1}$ and therefore $\tilde{\alpha}_s(t_i) \in \tilde{D}_{i+1}$. From this, as above, we conclude that $\tilde{\alpha}_s([t_i, t_{i+1}]) \subseteq \tilde{D}_{i+1}$ — and the proof by induction is finished. Finally, from the fact that $\tilde{\alpha}_s(b) \in \tilde{D}_k$ and $f(\tilde{\alpha}_s(b)) = q_1 = f(\tilde{\alpha}_{s_0}(b))$, it follows, just as desired, that $\tilde{\alpha}_s(b) = \tilde{\alpha}_{s_0}(b)$. $\square$

We now prove Proposition 5.2.4. Let us fix $q_0 \in S_2$ and an open neighborhood V of it that is diffeomorphic to a disk. We say that two points of $f^{-1}(V)$ are in the same *connected component* if there exists a curve in $f^{-1}(V)$ that connects them, and we write

$$f^{-1}(V) = \bigcup_j V_j,$$

where the V_j are the connected components. Each V_j is open: in fact, if $p \in V_j$ and $W \subseteq f^{-1}(V)$ is a neighborhood of p diffeomorphic to a disk, then $W \subseteq V_j$.

Let us now verify that $f|_{V_j}: V_j \rightarrow V$ is bijective. Take $p \in V_j$ and let $q = f(p)$. Given $r \in V$, let $\alpha: [a, b] \rightarrow V$ be a curve with $\alpha(a) = q$ and $\alpha(b) = r$, and let $\tilde{\alpha}$ be the lifting of α that starts from p . By definition, $\tilde{\alpha}(b)$ is in the same component V_j as p and $f(\tilde{\alpha}(b)) = r$. Therefore $f|_{V_j}$ is surjective. Regarding injectivity, assume that $\tilde{p}, \hat{p} \in V_j$ have the same image q , and $\tilde{\alpha}_0: [a, b] \rightarrow V_j$ is a curve from $\tilde{p}$ to $\hat{p}$, and $\alpha_0 = f \circ \tilde{\alpha}_0$. This curve α_0 is then closed. Since V is diffeomorphic to a disk, there exists a continuous deformation $(\alpha_s)_{0 \leq s \leq 1}$ of α_0 with fixed endpoints such that the trace of each α_s is in V and α_1 is the constant curve equal to q . Denoting by $\tilde{\alpha}_s$ the lifting of α_s that starts from $\tilde{p}$, Lemma 5.2.5 guarantees that the curves $\tilde{\alpha}_s$ all have the same endpoints $\tilde{p}$ and $\hat{p}$. But $\tilde{\alpha}_1$ is constant, and therefore $\tilde{p} = \tilde{\alpha}_1(a) = \tilde{\alpha}_1(b) = \hat{p}$. Thus $f|_{V_j}$ is injective.

In conclusion: each component V_j is diffeomorphically sent into V , and therefore V is an evenly covered neighborhood of q_0 . This shows that f is a covering and concludes the proof of Proposition 5.2.4.

In this proof we used that, for every open subset U diffeomorphic to a disk, every closed curve in U is homotopic (with fixed endpoints) to a constant curve by a homotopy that only takes values in U . A set U with such a property is called *simply connected*. Examples of simply connected surfaces are the plane and the sphere (recall that $\mathbf{S}^2$ minus one point is diffeomorphic to the plane via stereographic projection). If S_2 is simply connected and $f: S_1 \rightarrow S_2$ a covering then $f^{-1}(S_2)$ has only one connected component, which is all of S_1 — and, by the proof of Proposition 5.2.4, f is injective, hence a diffeomorphism. This proves our next result.

Proposition 5.2.6 *Every covering of a simply connected surface is a diffeomorphism.*

We now have gathered all necessary tools about coverings and it is good to start at once to make interesting use of it.

Theorem 5.2.7 (Hadamard)

Let $S \subseteq \mathbb{R}^3$ be a compact surface with positive curvature at all points. Then S is diffeomorphic to the sphere.

In fact, such a surface is orientable (e.g. Example 61-b in Section 3.2), and the hypothesis implies that the normal mapping $N: S \rightarrow \mathbf{S}^2$ is a local diffeomorphism, hence (by Proposition 5.2.2 and the compactness of S) a covering, hence (by Proposition 5.2.6 and $\mathbf{S}^2$ being simply connected) a diffeomorphism.

In Section 5.4 we will discuss in more depth the compact surfaces of positive curvature in $\mathbb{R}^3$ (called *ovals*) and show that ovals are convex (in the sense that they bound convex regions of $\mathbb{R}^3$).

Exercises

126. Show that in Lemma 5.2.5 the function defined by $\tilde{H}(s, t) = \tilde{\alpha}_s(t)$ is continuous.

127. Let S_1 and S_2 be two connected surfaces and $f: S_1 \rightarrow S_2$ a covering. Further, let $\mathfrak{I}(f)$ be the set of diffeomorphisms $g: S_1 \rightarrow S_1$ such that $f \circ g = f$. Show that:

- (a) $\mathfrak{I}(f)$ is a group with respect to composition of functions;
- (b) if S_1 is simply connected, then for every pair of points p_0, p_1 in S_1 such that $f(p_0) = f(p_1)$, there exists a unique $g \in \mathfrak{I}(f)$ such that $g(p_0) = p_1$;
- (c) if S_1 is not simply connected, then the property in (b) may fail.

5.3 Complete Surfaces of Non-Positive Curvature

In this section, S is a complete surface of non-positive curvature. For every $p \in S$, the exponential mapping $\exp_p$ is then defined on the entire tangent space $T_p S$. We shall show the assumption that the curvature of S is non-positive implies that $\exp_p$ is a covering.

Fixing $p \in S$ and an orthonormal basis $(\mathbf{v}_1, \mathbf{v}_2)$ of $T_p S$, we note that the coefficients of the geodesic polar coordinates $\Psi(\rho, \varphi)$ centered at p have a well-defined meaning for all (ρ, φ) with $\rho > 0$, even if globally $\Psi(\rho, \varphi)$ is not a parameterization. For example,

$$G(\rho, \varphi) = |D(\exp_p)_{\rho \mathbf{v}_\varphi}(\rho \mathbf{w}_\varphi)|^2, \quad (5.4)$$

where $\mathbf{v}_\varphi = \cos \varphi \mathbf{v}_1 + \sin \varphi \mathbf{v}_2$, $\mathbf{w}_\varphi = -\sin \varphi \mathbf{v}_1 + \cos \varphi \mathbf{v}_2$. Furthermore, the equalities $E \equiv 1$, $F \equiv 0$ are valid for all (ρ, φ) without having to change anything in the proof of Lemma 4.6.3.

Ψ is a true parameterization in the neighborhood of any point (ρ, φ) such that $G(\rho, \varphi) \neq 0$, and hence formula (4.38) is valid. That is,

$$(\sqrt{G})_{\rho\rho} + K\sqrt{G} = 0 \quad (5.5)$$

whenever $G \neq 0$. We know that with φ fixed the function $\rho \mapsto \sqrt{G}(\rho, \varphi)$ extends continuously to $\rho = 0$, and such an extension is twice differentiable at 0 (see Lemma 4.6.9) — with $\sqrt{G}(0, \varphi) = 0$ and $(\sqrt{G})_\rho(0, \varphi) = 1$. Assuming that $K \leq 0$ on S , we obtain from (5.5) that $(\sqrt{G})_{\rho\rho} \geq 0$; thus

$$\begin{aligned} (\sqrt{G})_\rho(\rho, \varphi) &\geq (\sqrt{G})_\rho(0, \varphi) = 1 \\ \Rightarrow \sqrt{G}(\rho, \varphi) &\geq \sqrt{G}(0, \varphi) + \rho = \rho. \end{aligned} \quad (5.6)$$

Both these inequalities hold, as does (5.5), up to the first $\rho_0 > 0$ where $G(\rho_0, \varphi) = 0$. But (5.6) ensures that such a ρ_0 does not exist, and therefore the inequalities hold for all $\rho \geq 0$. Since $G(\rho, \varphi)$ is nonzero for all $\rho > 0$, it follows that $\exp_p$ is a local diffeomorphism. This proves one part of the following theorem:

Theorem 5.3.1 *Let S be a complete surface of non-positive curvature at all points. Then, for every $p \in S$, the local diffeomorphism $\exp_p: T_p S \rightarrow S$ is a covering.*

In view of Proposition 5.2.4, to complete the proof of Theorem 5.3.1, it is only left to show that $\exp_p$ lifts curves. The next result says that $\exp_p$ increases the length of curves:

Lemma 5.3.2 *If $\beta: [a, b] \rightarrow T_p S$ is a differentiable curve then $l(\beta) \leq l(\exp_p \circ \beta)$.*

Proof It suffices to show that for every $r \in T_p S$, and vector $\mathbf{v}$, we have $|D(\exp_p)_r(\mathbf{v})| \geq |\mathbf{v}|$, since then

$$l(\beta) = \int_a^b |\beta'(t)| dt \leq \int_a^b |D(\exp_p)_{\beta(t)}(\beta'(t))| dt = l(\exp_p \circ \beta).$$

Since $D(\exp_p)_O$ is the identity, we can assume that $r = \rho \mathbf{v}_\varphi$ with $\rho > 0$. The vectors $\mathbf{v}_\varphi$ and $\mathbf{w}_\varphi$ form an orthonormal basis of $T_p S$, and the vectors $D(\exp_p)_r(\mathbf{v}_\varphi)$ and $D(\exp_p)_r(\mathbf{w}_\varphi)$ are also orthogonal ($F \equiv 0$). Moreover,

$$|D(\exp_p)_r(\mathbf{v}_\varphi)| = 1 = |\mathbf{v}_\varphi|,$$

$$|D(\exp_p)_r(\mathbf{w}_\varphi)| = \frac{\sqrt{G}}{\rho} \geq 1 = |\mathbf{w}_\varphi| \quad [\text{by (5.4) and (5.6)}]$$

— and therefore

$$\begin{aligned} |D(\exp_p)_r(\lambda \mathbf{v}_\varphi + \mu \mathbf{w}_\varphi)|^2 &= \lambda^2 |D(\exp_p)_r(\mathbf{v}_\varphi)|^2 + \mu^2 |D(\exp_p)_r(\mathbf{w}_\varphi)|^2 \\ &\geq \lambda^2 + \mu^2 = |\lambda \mathbf{v}_\varphi + \mu \mathbf{w}_\varphi|^2. \square \end{aligned}$$

Returning to the proof of Theorem 5.3.1, consider a differentiable curve $\alpha: [a, b] \rightarrow S$, and let r be a point of $T_p S$ such that $\exp_p(r) = \alpha(a)$. Taking a neighborhood of r that is diffeomorphically sent onto the image, we see that there exists $c \in]a, b]$ such that the lifting $\tilde{\alpha}$ of α starting at r is defined on $[a, c]$ (in the proof of Proposition 5.2.3 we saw that for a local diffeomorphism $S_1 \rightarrow S_2$ the lifting in S_1 of a curve in S_2 for a given fixed initial point is unique, whenever it exists). Let $\tilde{t}$ be the supremum of the set

$$I = \{t \in [a, b] : \tilde{\alpha} \text{ is defined on } [a, t]\} :$$

$\tilde{\alpha}$ is therefore defined on $[a, \tilde{t}[$. If $a \leq s < t < \tilde{t}$ then, using Lemma 5.3.2,

$$|\tilde{\alpha}(t) - \tilde{\alpha}(s)| \leq l(\tilde{\alpha}|_{[s, t]}) \leq l(\alpha|_{[s, t]})$$

— and therefore

$$\lim_{s, t \rightarrow \tilde{t}^-} |\tilde{\alpha}(t) - \tilde{\alpha}(s)| = 0,$$

which implies the existence of $\lim_{t \rightarrow \tilde{t}^-} \tilde{\alpha}(t)$. We thus conclude that $\tilde{t} \in I$, and it follows that $\tilde{t} = b$, otherwise the fact that $\exp_p$ is a local diffeomorphism would allow $\tilde{\alpha}$ to be extended beyond $\tilde{t}$. Hence $\exp_p$ lifts curves, which concludes the proof of Theorem 5.3.1.

Combining Theorem 5.3.1 with Proposition 5.2.6 we obtain the following corollary.

Theorem 5.3.3 (Hadamard)

If S is a complete, simply connected surface of curvature $K \leq 0$, then, for every $p \in S$, the local diffeomorphism $\exp_p: T_p S \rightarrow S$ is a diffeomorphism.

We saw in Chapter 4 how to construct local isometries between two surfaces of equal **constant** curvature K (Theorem 4.6.10): given $p \in S_1$, $q \in S_2$, and a linear isometry $L: T_p S_1 \rightarrow T_q S_2$, the mapping $\exp_q \circ L \circ \exp_p^{-1}$ is defined in a neighborhood of p and is an isometry onto the image. Let us now assume that $K \leq 0$, that both S_1 and S_2 are complete, and that S_1 is simply connected. Then $f = \exp_q \circ L \circ \exp_p^{-1}$ is defined on the entire surface S_1 (by Theorem 5.3.3), is a local isometry, and is still a covering (by Theorem 5.3.1). In case S_2 is also simply connected, f is a (global) isometry between S_1 and S_2 . To summarize, we proved the following theorem:

Theorem 5.3.4 *Let S_1 be a complete, simply connected surface of constant curvature $K \leq 0$, and let S_2 be another complete surface with the same curvature K . Then:*

- i. there exists a covering $f: S_1 \rightarrow S_2$ which is a local isometry;
 ii. if S_2 is simply connected, the covering f is an isometry.

For every $K \leq 0$, there is therefore, up to isometry, only one complete surface, simply connected, of constant curvature K , and all other complete surfaces of equal curvature are images of this one under some *isometric covering*. The cylinder $x^2 + y^2 = 1$, for example, is (as we already knew) the image of the plane under an isometric covering: the function that wraps the plane around the cylinder.

We now face a problem: we do not yet know any **complete** surface on $\mathbb{R}^3$, simply connected or not, with negative constant curvature. Indeed, a famous theorem by Hilbert (of which a proof is found in [6]) states that such a surface does not exist. Do we have to conclude that Theorem 5.3.4 is void when $K < 0$? The solution is to consider, as we will do in Section 5.6, abstract surfaces whose metric structure is not induced by any ambient space.

Theorem 5.3.4 also holds for $K > 0$, with a slightly different proof; but in this case, for surfaces in $\mathbb{R}^3$, Theorem 5.3.4 is a triviality, since S_1 and S_2 have to be compact (see Exercise 129), and therefore are spheres of equal radius (Section 5.4). Again we have to consider abstract surfaces for the result to be interesting.

Exercises

128. Let S be a complete surface, of non-positive curvature, and p a point of S .

- (a) Show that the equations of geodesics in geodesic polar coordinates are

$$\rho - \frac{1}{2} G_\rho \varphi^2 = 0, \quad \varphi + \frac{G_\rho}{G} \rho \varphi + \frac{G_\varphi}{2G} \varphi^2 = 0.$$

(b) Let $\gamma(s)$ be a geodesic of S that does not pass through p , and denote by $\tilde{\gamma}: \mathbb{R} \rightarrow T_p S$ a lifting of γ . Show that the function $\rho(s) = |\tilde{\gamma}(s)|$ is convex (i.e., $\rho \geq 0$) and has at most one local minimum.

(c) Assume now that S is simply connected. Show that the trace of any geodesic of S is a closed set on S and that, given a geodesic $\gamma(s)$ that does not pass through p , there exists a single point of γ at the minimal distance from p .

(d) Compare the result of (c) with the case of the sphere. Give an example of a complete surface of non-positive curvature where not all geodesics are closed sets.

129. Let S be a complete surface of constant curvature $K > 0$. Using polar geodesic coordinates, show that any minimizing geodesic of S has length $\leq \frac{\pi}{\sqrt{K}}$. Conclude that S is compact.

5.4 Ovals (First Part): The Rigidity of the Sphere

The *oval surfaces* (or simply *ovals*) are the compact surfaces in $\mathbb{R}^3$ with positive curvature at all points. Such surfaces are, as we saw in Section 5.2 (Theorem 5.2.7),

diffeomorphic to the sphere, the field of normal vectors $N: S \rightarrow \mathbf{S}^2$ being a diffeomorphism.

A surface S is *strictly convex* if, for every $p \in S$, the intersection of S with the tangent plane $\{p\} + T_p S$ with S reduces to the point p ; equivalently, if $S \setminus \{p\}$ is entirely in one of the two half-spaces into which $\{p\} + T_p S$ divides $\mathbb{R}^3$.

Proposition 5.4.1 *Every oval S is strictly convex.*

Proof Fixing $p \in S$, we must check that the mapping $S \rightarrow \mathbb{R}$ given by $h_p(q) = \langle N(p), q - p \rangle$ has constant sign, vanishing only at p . Otherwise h_p reaches a minimum and a maximum at points $q_0, q_1 \in S \setminus \{p\}$. Each of the tangent spaces $T_{q_i} S$ ($i = 0, 1$) is then orthogonal to $N(p)$, which implies that two of the vectors $N(p)$, $N(q_0)$ and $N(q_1)$ are equal — in contradiction to N being a diffeomorphism.

Now suppose that $p_0 \in S$ satisfies $\langle N(p_0), q - p_0 \rangle > 0$ for all $q \in S \setminus \{p_0\}$. Let us verify that one also has $\langle N(p), q - p \rangle > 0$ for all distinct points $p, q \in S$. Since h_p has constant sign, we can assume $q \neq p_0$ and consider a curve $\alpha: [a, b] \rightarrow S$ such that $\alpha(a) = p_0$, $\alpha(b) = p$, and $\alpha(t) \neq q$ for all $t \in [a, b]$. The function $t \mapsto \langle N(\alpha(t)), q - \alpha(t) \rangle$ never vanishes, and therefore takes the same sign for $t = a$ and $t = b$, which proves that $\langle N(p), q - p \rangle > 0$. $\square$

From now on we fix an orientation N of S such that $\langle N(p), q - p \rangle > 0$ for all $p \neq q \in S$. With this orientation the principal curvatures (and the mean curvature) at each point are positive.

Our goal now is to show that the region Ω of $\mathbb{R}^3$ bounded by S is convex in the usual sense: the line segment joining each pair of points in Ω is also contained in Ω . Let us consider the sets

$$\mathcal{C} = \{q \in \mathbb{R}^3 : \langle N(p), q - p \rangle > 0 \ \forall p \in S\},$$

$$\mathcal{D} = \{q \in \mathbb{R}^3 : \exists p \in S \text{ such that } \langle N(p), q - p \rangle < 0\}.$$

Proposition 5.4.2 *The sets $\mathcal{C}$ and $\mathcal{D}$ are connected open, $\mathcal{C}$ is convex and bounded, S is the boundary of both sets $\mathcal{C}$ and $\mathcal{D}$, and $\mathbb{R}^3 \setminus S = \mathcal{C} \cup \mathcal{D}$.*

Proof Each of the sets $\mathcal{D}_p = \{q \in \mathbb{R}^3 : \langle N(p), q - p \rangle < 0\}$, for $p \in S$, is open, and $\mathcal{D}$ is the union of them all. Therefore $\mathcal{D}$ is open, and it is easily seen to be connected. The set $\mathcal{C}$ is the intersection of the convex sets $\mathcal{C}_p = \{q \in \mathbb{R}^3 : \langle N(p), q - p \rangle > 0\}$ ($p \in S$), and is therefore convex; and any convex set is connected.

Given $q \in \mathcal{C}$, let us consider $\delta = \min_{p \in S} \langle N(p), q - p \rangle > 0$. If $|r - q| < \delta$ then we have, for every $p \in S$,

$$\langle N(p), r - p \rangle = \langle N(p), q - p \rangle - \langle N(p), q - r \rangle \geq \delta - |q - r| > 0,$$

so that $r \in \mathcal{C}$ — which proves that $\mathcal{C}$ is open.

We shall show that $S \subseteq \partial \mathcal{C} \cap \partial \mathcal{D}$. Let us fix $p_0 \in S$ and a **unit** vector $\mathbf{v}$ such that $\langle N(p_0), \mathbf{v} \rangle > 0$, and let $p_t = p_0 + t\mathbf{v}$. When $t < 0$, we have $\langle N(p_0), p_t - p_0 \rangle = t \langle N(p_0), \mathbf{v} \rangle < 0$ and therefore $p_t \in \mathcal{D}$. This shows that $p_0 \in \partial \mathcal{D}$, and we thus conclude that $S \subseteq \partial \mathcal{D}$.

Now let $U \subseteq S$ be an open neighborhood of p_0 such that $\langle N(p), \mathbf{v} \rangle > 0$ whenever $p \in U$, and let us put $\varepsilon = \min_{p \in S \setminus U} \langle N(p), p_0 - p \rangle$. It follows immediately that if $p \in U$ and $t > 0$ then $\langle N(p), p_t - p \rangle > 0$. On the other hand, if $p \in S \setminus U$ and $0 < t < \varepsilon$ then

$$\langle N(p), p_t - p \rangle = \langle N(p), p_0 - p \rangle + t \langle N(p), \mathbf{v} \rangle \geq \varepsilon - t |N(p)| |\mathbf{v}| > 0$$

— which shows that $p_t \in \mathcal{C}$ for $t \in]0, \varepsilon[$. Thus $p_0 \in \partial \mathcal{C}$, and therefore $S \subseteq \partial \mathcal{C}$.

We now show that $\mathbb{R}^3 \setminus (\mathcal{C} \cup \mathcal{D}) \subseteq S$. Indeed, if $q \in \mathbb{R}^3 \setminus (\mathcal{C} \cup \mathcal{D})$ then there exists some $p_0 \in S$ such that $\langle N(p_0), q - p_0 \rangle = 0$. If $q \neq p_0$ then we can consider a curve $\alpha:]-\eta, \eta[\rightarrow S$ such that $\alpha(0) = p_0$ and $\alpha'(0) = DN_{p_0}^{-1}(q - p_0)$, where

$$\frac{d}{dt} \langle N \circ \alpha(t), q - \alpha(t) \rangle \Big|_{t=0} = \langle DN_{p_0}(\alpha'(0)), q - p_0 \rangle = |q - p_0|^2 > 0.$$

This implies that $\langle N \circ \alpha(t), q - \alpha(t) \rangle < 0$ for $t < 0$ near 0, and therefore $q \in \mathcal{D}$, in contradiction to our hypothesis. We thus conclude that $p = p_0 \in S$ and $\mathbb{R}^3 \setminus (\mathcal{C} \cup \mathcal{D}) \subseteq S$ — which, together with what we have shown above, also proves $\partial \mathcal{C} = \partial \mathcal{D} = S$.

Finally, let us take $\lambda > 0$ such that $S \subseteq D_\lambda = \{q \in \mathbb{R}^3 : |q| \leq \lambda\}$. The complement of D_λ , being connected and disjoint to S , is contained in either of the sets $\mathcal{C}$ or $\mathcal{D}$; but as $\mathcal{D}$ is not bounded, $\mathcal{D} \cap (\mathbb{R}^3 \setminus D_\lambda)$ is non-empty and therefore $\mathbb{R}^3 \setminus D_\lambda \subseteq \mathcal{D}$. Thus $\mathcal{C} \subseteq D_\lambda$ and $\mathcal{C}$ is bounded. $\square$

The next result is used immediately afterwards in the proof of the rigidity of the sphere, and in the proof of a Blaschke formula in the next section (Theorem 5.5.3).

Theorem 5.4.3 (Minkowski integral formulas)

Let S be an oval and $p_0 \in \mathbb{R}^3$. Then, denoting by A the area of S , we have

$$A = - \int_S H(p) \langle p - p_0, N(p) \rangle d\sigma, \quad (5.7)$$

$$\int_S H(p) d\sigma = - \int_S K(p) \langle p - p_0, N(p) \rangle d\sigma. \quad (5.8)$$

Both formulas (5.7) and (5.8) are valid for every compact surface $S \subseteq \mathbb{R}^3$, and indeed we prove (5.7) [but not (5.8)] in this generality. We will make use of the results of Section 4.4, in particular of Theorem 4.4.5 (first variation of area).

Let us consider the vector field on S given by $\mathbf{v}(p) = p - p_0$, and let S_t be the variation of S induced by $\mathbf{v}$. Since S_t is the image of S under the homothety with center p_0 and ratio $1 + t$, its area is $A(t) = (1 + t)^2 A$, which implies $A'(0) = 2A$. Comparing with Theorem 4.4.5, we obtain

$$2A = - \int_S \langle \mathbf{v}(p), 2H(p)N(p) \rangle d\sigma = -2 \int_S H(p) \langle p - p_0, N(p) \rangle d\sigma,$$

which proves (5.7).

The proof of (5.8) follows the steps of the previous one. We denote by $M = \int_S H(p) d\sigma$ the *total mean curvature* of S , and we denote by $M(t)$ the

analogous quantity for S_t . Note that the parameterization $\Phi(u, v)$ of S is transformed, when composed with the homothety just mentioned, into the parameterization $(1+t)\Phi(u, v) - tp_0$ of S_t ; the coefficients of the first fundamental form are multiplied by $(1+t)^2$ and the coefficients of the second by $1+t$. Hence the area element is multiplied by $(1+t)^2$ and, by formula (3.8) in Section 3.2, H is divided by $1+t$. We hence have $M(t) = (1+t)M$ and therefore $M'(0) = M$; (5.8) is an immediate consequence of the lemma below:

Lemma 5.4.4 *Let S be an oval and let S_t be the variation of S induced by a vector field $\mathbf{v}$. Then, if $M(t)$ denotes the total mean curvature of S_t , we have*

$$M'(0) = - \int_S K(p) \langle \mathbf{v}(p), N(p) \rangle d\sigma.$$

Proof Given a parameterization (U, Φ) of S , compatible with the orientation, let

$$\Phi^t(u, v) = \Phi(u, v) + t\mathbf{v}(u, v);$$

(U, Φ^t) is a parameterization of S_t for sufficiently small $|t|$. Let us denote by $N^t(u, v)$ the field of normal vectors in $\Phi^t(U)$. By formula (4.18) of Section 4.4 we have

$$H^t |\Phi_u^t \times \Phi_v^t| = -\frac{1}{2} \langle \Phi_u^t \times N_v^t + N_u^t \times \Phi_v^t, N^t \rangle. \quad (5.9)$$

We now look at the derivative of the second term in (5.9). We have

$$\begin{aligned} & \frac{d}{dt} \langle \Phi_u^t \times N_v^t + N_u^t \times \Phi_v^t, N^t \rangle \Big|_{t=0} \\ &= \left\langle \Phi_u \times N_v + N_u \times \Phi_v, \frac{d}{dt} N^t \Big|_{t=0} \right\rangle + \left\langle \frac{d}{dt} \{ \Phi_u^t \times N_v^t + N_u^t \times \Phi_v^t \} \Big|_{t=0}, N \right\rangle \\ &= \left\langle \frac{\partial \mathbf{v}}{\partial u} \times N_v + N_u \times \frac{\partial \mathbf{v}}{\partial v}, N \right\rangle + \left\langle \Phi_u \times \left(\frac{d}{dt} N_v^t \Big|_{t=0} \right) + \left(\frac{d}{dt} N_u^t \Big|_{t=0} \right) \times \Phi_v, N \right\rangle \end{aligned} \quad (5.10)$$

(because $\frac{d}{dt} N^t \Big|_{t=0}$ is orthogonal to N , and $\Phi_u \times N_v + N_u \times \Phi_v$ is collinear with N).

If we consider the field of tangent vectors $\mathbf{w} = \frac{d}{dt} N^t \Big|_{t=0}$ and use Proposition 4.4.1, the right-hand side of (5.10) becomes

$$\left\langle \Phi_u \times \frac{\partial \mathbf{w}}{\partial v} + \frac{\partial \mathbf{w}}{\partial u} \times \Phi_v, N \right\rangle = (\text{Div } \mathbf{w}) |\Phi_u \times \Phi_v|. \quad (5.11)$$

Regarding the left-hand side in (5.9), we note that $N: S \rightarrow \mathbf{S}^2$ induces in $\mathbf{S}^2$ the orientation for which the mean curvature is negative, equal to -1 . Using Proposition 4.4.1 we obtain

$$\begin{aligned} \left\langle \frac{\partial \mathbf{v}}{\partial u} \times N_v + N_u \times \frac{\partial \mathbf{v}}{\partial v}, N \right\rangle &= |N_u \times N_v| \{ \text{Div}^*(\mathbf{v}^\top) + 2\langle \mathbf{v}, N \rangle \} \\ &= \text{Div}^*(\mathbf{v}^\top) |N_u \times N_v| + 2K \langle \mathbf{v}, N \rangle |\Phi_u \times \Phi_v|, \end{aligned} \quad (5.12)$$

where $\text{Div}^*(\mathbf{v}^\top)$ indicates the divergence of $\mathbf{v}^\top$ as the tangent vector field to $\mathbf{S}^2$ — which is fine since N is a diffeomorphism and the tangent spaces $T_p S$ and $T_{N(p)} \mathbf{S}^2$ are parallel. Now the total mean curvature of $\Phi^t(U) \subseteq S_t$ is $M_\Phi(t) = \int_U H^t |\Phi_u^t \times \Phi_v^t| du dv$, so that

$$M'_\Phi(0) = \int_U \frac{d}{dt} \{H^t |\Phi_u^t \times \Phi_v^t|\}_{t=0} du dv. \quad (5.13)$$

Combining formulas (5.9)–(5.13) and using Corollary 4.4.3, we finally obtain

$$\begin{aligned} M'(0) &= -\frac{1}{2} \int_S \text{Div } \mathbf{w} d\sigma - \frac{1}{2} \int_{\mathbf{S}^2} \text{Div}^*(\mathbf{v}^\top) d\sigma - \int_S K \langle \mathbf{v}, N \rangle d\sigma \\ &= - \int_S K \langle \mathbf{v}, N \rangle d\sigma. \quad \square \end{aligned}$$

Theorem 5.4.5 (Rigidity of the Sphere)

Let $S \subseteq \mathbb{R}^3$ be a compact surface with constant curvature K . Then S is a sphere.

Proof Being compact, S has some point of positive curvature (see Exercise 57 in Section 3.1), and therefore $K > 0$. S is then an oval, and we can fix the orientation $N: S \rightarrow \mathbf{S}^2$ such that $\langle N(p), q - p \rangle > 0$ for every pair of points $p \neq q$ on S . With this orientation we have $H(p) > 0$ at every point $p \in S$.

Let us put $k = \sqrt{K}$. By the inequality between the arithmetic and geometric means, we have $kH(p) - K \geq 0$, and the equality holds at p if and only if p is an umbilical point. Let $p_0 \in \mathbb{R}^3$ (whose existence is guaranteed by Proposition 5.4.2) be such that $\langle N(p), p_0 - p \rangle > 0$ for all $p \in S$, and let us put

$$\begin{aligned} \Delta_0 &= \int_S (kH - K) d\sigma = kM - k^2 A, \\ \Delta_1 &= \int_S (kH - K) \langle N, p_0 - p \rangle d\sigma \\ &= k \int_S H \langle N, p_0 - p \rangle d\sigma - \int_S K \langle N, p_0 - p \rangle d\sigma \\ &= kA - M \quad (\text{by Minkowski's formulas}). \end{aligned}$$

We thus have $\Delta_0 = -k\Delta_1$, but their definitions ensure that Δ_0 and Δ_1 are both non-negative. Hence, we have $\Delta_0 = \Delta_1 = 0$, and therefore the integrand function $kH - K$ is constantly zero. All points of S are therefore umbilical and, by Proposition 3.1.5, S is a sphere. $\square$

In particular, all surfaces in $\mathbb{R}^3$ that are isometric to the sphere are themselves spheres, and hence the theorem speaks of the rigidity of the sphere. The theorem has the following generalization (see [13]): *if $f: S_1 \rightarrow S_2$ is an isometry between two oval surfaces then f is the restriction of an isometry of $\mathbb{R}^3$ (which, as is known, is the composite of a linear isometry with a translation).*

We now give another result of the same kind as Theorem 5.4.5.

Theorem 5.4.6 *Let $S \subseteq \mathbb{R}^3$ be an oval surface of constant mean curvature H . Then S is a sphere.*

We choose the orientation N and the point p_0 as above, and notice that $H^2 - K \geq 0$, with equality only at the umbilical points. Then

$$\begin{aligned} \int_S (H^2 - K) \langle N, p_0 - p \rangle d\sigma &= H \int_S H \langle N, p_0 - p \rangle d\sigma - \int_S K \langle N, p_0 - p \rangle d\sigma \\ &= HA - \int_S H d\sigma = 0 \end{aligned}$$

— and from this, since the integrand function is non-negative, it follows that $H^2 - K \equiv 0$, and therefore S is a sphere.

Let us point out that Theorem 5.4.6 remains valid without the assumption that S has positive curvature: *any compact surface $S \subseteq \mathbb{R}^3$ with constant mean curvature is a sphere*. An accessible proof of this result appears in Osserman's paper [20].

Exercise

130. Show that the open subset $\mathcal{C}$ bounded by the oval surface S (Proposition 5.4.2) is given by $\{(1-t)p + tq : p, q \in S, 0 < t < 1\}$.

5.5 Ovals: Areas and Volumes; Surfaces of Constant Width

In this section we prove a number of formulas involving the area, the total mean curvature, and the volume bounded by an oval surface. Some of these formulas concern convex bodies in $\mathbb{R}^3$ — which include not only solids bounded by oval surfaces but also by convex polyhedra. As we noted in Section 1.1, the length of a regular curve γ can be computed by considering polygonal lines inscribed in γ with increasing numbers of segments. The generalization to surfaces is not so obvious: there are approximations of the cylindrical surface $\mathcal{C} = \{x^2 + y^2 = 1, 0 \leq z \leq 1\}$ by polyhedra whose vertices are all in $\mathcal{C}$ but whose areas do not converge to that of $\mathcal{C}$ (see Exercise 131).

The solution, for a compact convex surface S , is to approximate S by **convex** polyhedra: if $\mathcal{P}_1$ and $\mathcal{P}_2$ are convex polyhedra such that $\mathcal{P}_1 < S < \mathcal{P}_2$, where the sign $<$ means “is in”, then we have the expected inequality of areas

$$A(\mathcal{P}_1) < A(S) < A(\mathcal{P}_2). \quad (5.14)$$

Knowing that we can find $\mathcal{P}_1$ and $\mathcal{P}_2$ such that $\mathcal{P}_1 < S < \mathcal{P}_2$ and with $A(\mathcal{P}_2) - A(\mathcal{P}_1)$ as small as we want, it follows from (5.14) that

$$A(S) = \sup_{\mathcal{P}_1 < S} A(\mathcal{P}_1) = \inf_{S < \mathcal{P}_2} A(\mathcal{P}_2). \quad (5.15)$$

Formula (5.15) suggests that for certain formulas involving areas of convex surfaces, it is sufficient to prove them for polyhedra; moreover, it provides a definition of area that does not depend on whether the surface in question is regular or not.

The interested reader can find the details of this construction in several books on convexity (e.g., [9]). For now, we will make use of (5.15) to prove a Cauchy formula. Let S be a compact convex surface. Given $p \in \mathbf{S}^2$, let us denote by $A_S(p)$ the area of the orthogonal projection of S onto $T_p\mathbf{S}^2$.

Theorem 5.5.1 (Cauchy)

The area of S is given by

$$A = \frac{1}{\pi} \int_{\mathbf{S}^2} A_S(p) d\sigma. \quad (5.16)$$

Proof Let F be a polygon included in a plane Π , and let $\mathbf{v}$ be a unit vector orthogonal to Π . Let $\theta \in [0, \pi]$ be the angle between $\mathbf{v}$ and $\mathbf{Op}$ ($p \in \mathbf{S}^2$); the area of the orthogonal projection of F onto $T_p\mathbf{S}^2$ is $A_F(p) = A(F)|\cos \theta|$. Using spherical coordinates (φ, θ) relative to an orthonormal frame in which the third vector is $\mathbf{v}$, we then have

$$\begin{aligned} \int_{\mathbf{S}^2} A_F(p) d\sigma &= \int_0^{2\pi} \left(\int_0^\pi A(F)|\cos \theta| \sin \theta d\theta \right) d\varphi \\ &= 2\pi A(F) \int_0^\pi |\cos \theta| \sin \theta d\theta = 2\pi A(F). \end{aligned}$$

Thus

$$A(F) = \frac{1}{2\pi} \int_{\mathbf{S}^2} A_F(p) d\sigma. \quad (5.17)$$

Now let $\mathcal{P}$ be a convex polyhedron and $F_1, \dots, F_k$ its faces. Except for the points on the boundary (which are negligible), each point of the orthogonal projection of $\mathcal{P}$ in the plane $T_p\mathbf{S}^2$ is the image of exactly two points of $\mathcal{P}$, belonging to two distinct faces. We then have

$$A_{\mathcal{P}}(p) = \frac{1}{2} \sum_{i=1}^k A_{F_i}(p);$$

and, using (5.17), we obtain

$$\begin{aligned} A(\mathcal{P}) &= \sum_{i=1}^k A(F_i) = \frac{1}{\pi} \int_{\mathbf{S}^2} \left(\frac{1}{2} \sum_{i=1}^k A_{F_i}(p) \right) d\sigma \\ &= \frac{1}{\pi} \int_{\mathbf{S}^2} A_{\mathcal{P}}(p) d\sigma. \end{aligned} \quad (5.18)$$

Noting that if $\mathcal{P}_1 < S < \mathcal{P}_2$ then $A_{\mathcal{P}_1}(p) \leq A_S(p) \leq A_{\mathcal{P}_2}(p)$ for every $p \in \mathbf{S}^2$, it follows from (5.18) that

$$A(\mathcal{P}_1) \leq \frac{1}{\pi} \int_{\mathbf{S}^2} A_S(p) d\sigma \leq A(\mathcal{P}_2),$$

which, together with (5.15), proves the theorem. $\square$

As in the case of curves (see Section 1.6), we say that an oval surface S has *constant curvature* $\mathfrak{L}$ if, for every pair of parallel planes tangent to S at two distinct points, the distance between them is equal to $\mathfrak{L}$.

Corollary 5.5.2 *If S has constant curvature $\mathfrak{L}$ and area A then*

$$A \leq \pi \mathfrak{L}^2,$$

with equality if and only if S is a sphere.

Sketch of the proof. The orthogonal projection of S onto the plane $T_p \mathbf{S}^2$, which we denote by $S(p)$, is bounded by a curve of constant width $\mathfrak{L}$, having (by Theorem 1.6.3) perimeter $\pi \mathfrak{L}$. By the isoperimetric inequality (Theorem 1.8.2), the area of $S(p)$ satisfies the inequality

$$A_S(p) \leq \frac{\pi \mathfrak{L}^2}{4} \quad (5.19)$$

— with equality only in the case where $S(p)$ is a disk with radius $\mathfrak{L}/2$. Combining (5.19) and (5.16), we obtain

$$A \leq \frac{1}{\pi} \int_{\mathbf{S}^2} \frac{\pi \mathfrak{L}^2}{4} d\sigma = \pi \mathfrak{L}^2,$$

with equality if and only if every orthogonal projection $S(p)$ of S is a disk with radius $\mathfrak{L}/2$ — and this last condition implies that S is a sphere (see Exercise 132). $\square$

Let us further consider an oval surface S , and on it the field of normal vectors N pointing into S . The variation of S induced by $-N$ is, as we have already defined, the family $S_t = \{p - tN(p) : p \in S\}$. But in this case, the surfaces S_t are *parallel* to S , either surrounding it (for $t > 0$) or being surrounded by it (for $t < 0$) at a fixed distance equal to $|t|$. Furthermore S_t is a surface for all $t \geq 0$ (Exercise 135), though not for all $t < 0$.

Given a parameterization $\Phi(u, v)$ of S , let us put

$$\Phi^t(u, v) = \Phi(u, v) - tN(u, v).$$

Now we have

$$\begin{aligned} \Phi_u^t \times \Phi_v^t &= \Phi_u \times \Phi_v - t\{\Phi_u \times N_v + N_u \times \Phi_v\} + t^2\{N_u \times N_v\} \\ &= (1 + 2tH + t^2K)\{\Phi_u \times \Phi_v\} \quad [\text{per (4.18), Section 4.4}]. \end{aligned} \quad (5.20)$$

For $t > 0$ (and for negative t near 0), the expression $1 + 2tH + t^2K$ is positive. As a consequence, and taking absolute values on both sides of (5.20), we conclude that the area of S_t is given by

$$A(t) = \int_S (1 + 2tH + t^2K) d\sigma = A + 2Mt + 4\pi t^2, \quad (5.21)$$

where A and M are the area and the total mean curvature of S (where we recall that as S is diffeomorphic to the sphere its total curvature $\int_S K d\sigma$ is 4π). Formula (5.21) thus expresses the remarkable fact that *the area of a surface parallel to S is a polynomial function of its distance to S .*

We can apply (5.21) to deduce an analogous formula involving volumes. Let us denote by $V(t)$ the volume of the region bounded by S_t , where $V = V(0)$. Assuming that $t > 0$, the difference $V(t) - V$ is the volume of the region between S and S_t . This region is the union of the images of the functions $\Psi(u, v, s) = \Phi^s(u, v)$, with $(u, v, s) \in U \times [0, t]$, where (U, Φ) is a parameterization of S . Now

$$|\det J\Psi| = \left| \left\langle \frac{\partial \Psi}{\partial u} \times \frac{\partial \Psi}{\partial v}, \frac{\partial \Psi}{\partial s} \right\rangle \right| = |\langle \Phi_u^s \times \Phi_v^s, N \rangle| = |\Phi_u^s \times \Phi_v^s|,$$

and therefore the volume of the image of Ψ is

$$\begin{aligned} V(\Psi) &= \int_0^t \left(\iint_U |\det J\Psi| du dv \right) ds \\ &= \int_0^t \left(\iint_U |\Phi_u^s \times \Phi_v^s| du dv \right) ds = \int_0^t A(\Phi^s) ds, \end{aligned} \quad (5.22)$$

where $A(\Phi^s)$ denotes the area of $\Phi^s(U)$. Using (5.22) and (5.21), we finally obtain

$$V(t) = V + \int_0^t A(s) ds = V + At + Mt^2 + \frac{4\pi}{3} t^3, \quad (5.23)$$

and here it is, the promised formula, which is also valid for negative t near 0. There is a formula analogous to (5.23) for the area of planar regions bounded by parallel curves as well, which we will give in Exercise 133.

To conclude the section, we again assume that S has constant width $\mathfrak{L}$. As in the case of curves, the antipode $\mathcal{A}(p)$ of $p \in S$ is given by

$$\mathcal{A}(p) = p + \mathfrak{L}N(p),$$

and $\mathcal{A}: S \rightarrow S$ is an involutive diffeomorphism (i.e., $\mathcal{A} \circ \mathcal{A} = \text{id}$). Since S is inside the sphere with center $\mathcal{A}(p)$ and radius $\mathfrak{L}$, and is tangent to the sphere at the point p , the principal curvatures at p are both bounded by $1/\mathfrak{L}$ (see Exercise 57 in Section 3.1). If $\Phi(u, v)$ is a parameterization of S then, since $\mathcal{A}$ is a diffeomorphism, $\tilde{\Phi} = \mathcal{A} \circ \Phi$ is another parameterization. But from the expression of $\mathcal{A}$, we see that with the above notation $\tilde{\Phi} = \Phi^{-\mathfrak{L}}$, and from this, using (5.20), we take

$$\tilde{\Phi}_u \times \tilde{\Phi}_v = (1 - 2H\mathfrak{L} + K\mathfrak{L}^2) \{ \Phi_u \times \Phi_v \} \quad (5.24)$$

— a formula that ensures that the expression $1 - 2H\mathfrak{L} + K\mathfrak{L}^2$ never vanishes. In terms of the principal curvatures k_1, k_2 , we have

$$1 - 2H\mathfrak{L} + K\mathfrak{L}^2 = (k_1\mathfrak{L} - 1)(k_2\mathfrak{L} - 1) > 0,$$

since both factors are non-negative. It then follows from (5.24) that *for every continuous function* $f: S \rightarrow \mathbb{R}$ *we have*

$$\int_S f \, d\sigma = \int_S (f \circ \mathcal{A})(1 - 2H\mathfrak{L} + K\mathfrak{L}^2) \, d\sigma. \quad (5.25)$$

The formula (5.25) has interesting consequences. For example, taking $f \equiv 1$, and using the fact that the total curvature of S is 4π , we obtain

$$A = A - 2M\mathfrak{L} + 4\pi\mathfrak{L}^2 \Rightarrow M = 2\pi\mathfrak{L}$$

— that is, *the total mean curvature of a surface of constant width $\mathfrak{L}$ is $M = 2\pi\mathfrak{L}$.*

For another application of (5.25), we will prove a Blaschke formula that relates the area A of a surface of constant width to the volume V of the region bounded by it.

Theorem 5.5.3 (Blaschke)

If S has constant width $\mathfrak{L}$ then $V = \frac{1}{2} A\mathfrak{L} - \frac{1}{3} \pi\mathfrak{L}^3$.

Proof Let p_0 be a point of the region Ω bounded by S . The volume of Ω is given by

$$V = \frac{1}{3} \int_S \langle N(p), p_0 - p \rangle \, d\sigma(p)$$

(this formula corresponds to regarding Ω as a union of “infinitesimal cones” of vertex p_0 and base in S , and is a particular case of the divergence theorem in $\mathbb{R}^3$ — see e.g. [16], p. 493). By (5.25), denoting by $\tilde{p}$ the antipode of p , we also have

$$V = \frac{1}{3} \int_S \langle N(\tilde{p}), p_0 - \tilde{p} \rangle (1 - 2H\mathfrak{L} + K\mathfrak{L}^2) \, d\sigma(p).$$

Since $N(\tilde{p}) = -N(p)$ and $\tilde{p} = p + \mathfrak{L}N(p)$, we obtain from this, using Minkowski’s formulas (Theorem 5.4.3) and the fact that $M = 2\pi\mathfrak{L}$,

$$\begin{aligned} V &= \frac{1}{3} \int_S (-\langle N(p), p_0 - p \rangle + \mathfrak{L})(1 - 2H\mathfrak{L} + K\mathfrak{L}^2) \, d\sigma(p) \\ &= -V + \frac{1}{3} A\mathfrak{L} + \frac{2}{3} \left(\int_S H \langle p_0 - p, N \rangle \, d\sigma \right) \mathfrak{L} - \frac{1}{3} \left(\int_S K \langle p_0 - p, N \rangle \, d\sigma \right) \mathfrak{L}^2 \\ &= -V + \frac{1}{3} A\mathfrak{L} + \frac{2}{3} A\mathfrak{L} - \frac{1}{3} M\mathfrak{L}^2 = -V + A\mathfrak{L} - \frac{2}{3} \pi\mathfrak{L}^3, \end{aligned}$$

which concludes the proof of the theorem. $\square$

Theorem 5.5.3 shows that among surfaces with a certain constant width $\mathfrak{L}$ those with the largest (or smallest) area are also those that enclose the largest (or smallest) volume. From Corollary 5.5.2 it then follows that *the maximal volume is that of the sphere*. The problem of finding the surface of a given constant width with minimal area (or minimal volume) is still open (see [7]).

Exercises

131. The segment $\mathcal{C}$ of the cylinder $x^2 + y^2 = 1$ bounded by the planes $z = 0$ and $z = 1$ has area 2π . Given $k \geq 1$ and $n \geq 3$, consider the $k + 1$ circles in $\mathcal{C}$ given by

$$z = 0, \frac{1}{k}, \frac{2}{k}, \dots, \frac{k-1}{k}, 1.$$

Divide each of these circles into n equal arcs, so that the ends of the arcs in each circle are vertically at the midpoints of the arcs of the preceding circle. The ends of these arcs define regular polygons of n sides inscribed in each of the circles. Joining each vertex to the two vertices closest to the neighboring polygons, we obtain a polyhedron $\mathcal{P}(k, n)$ whose faces are triangles, all of which are congruent.

(a) Show that the area of $\mathcal{P}(k, n)$ is

$$A(k, n) = 2n \sin\left(\frac{\pi}{n}\right) \sqrt{1 + 4k^2 \sin^4\left(\frac{\pi}{2n}\right)}.$$

(b) Compute $\lim_{n \rightarrow +\infty} A(n^r, n)$ for $r = 1, 2, 3$. What conclusion can you draw?

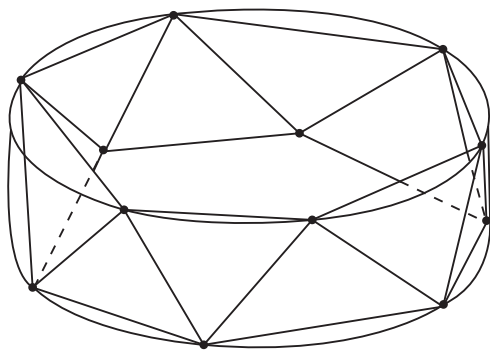


Figure 5.5

132. Let S be an oval surface and $r > 0$ such that the orthogonal projection of S onto each plane $\mathcal{P} \subseteq \mathbb{R}^3$ is a disk with radius r . Show that:

(a) for every plane $\mathcal{P}$, the surface S is inscribed in a straight circular cylinder $\mathcal{C}$ of height $2r$ whose base has radius r and is parallel to $\mathcal{P}$;

(b) if γ is the equator of $\mathcal{C}$ then $\gamma \subseteq S$ and the normals to $\mathcal{C}$ and to S along γ coincide;

(c) each normal section of S is a circle with radius r , and therefore S is a sphere.

133. Let α be a convex, regular, closed planar curve. Denote by α_t the curve parallel to α at distance t from α , and by $\mathfrak{U}(t)$ the area of the region bounded by α_t . Show that

$$\mathfrak{U}(t) = \mathfrak{U} + l(\alpha)t + \pi t^2 \quad [\text{where } \mathfrak{U} = \mathfrak{U}(0)].$$

Hint: verify that $l(\alpha_t) = l(\alpha) + 2\pi t$ and $\mathfrak{U}(t) = \mathfrak{U} + \int_0^t l(\alpha_s) ds$.

Note: since any convex curve can be approximated by a regular convex curve, the obtained formula is valid for any convex curve.

134. Let S be an oval surface, and denote by $l_S(p)$ the perimeter of the orthogonal projection of S onto $T_p\mathbb{S}^2$. Show that the total mean curvature of S is given by

$$M = \frac{1}{2\pi} \int_{\mathbb{S}^2} l_S(p) d\sigma.$$

Hint: use Exercise 133 and Theorem 5.5.1 to obtain an expression for the area $A(t)$ of the surface S_t parallel to S , and compare (5.21) with the formula obtained.

135. Let S be an oval surface and N be the normal field pointing into S . Show that, for all $t \geq 0$, the surface $S_t = \{p - tN(p) : p \in S\}$ is diffeomorphic to S .

5.6 Abstract Surfaces. The Hyperbolic Plane

All measures intrinsic to a surface depend, as we know, on the inner product defined on the tangent space (i.e., the first fundamental form) — which, so far, has simply been the usual scalar product restriction on $\mathbb{R}^3$. But nothing prevents us from considering other inner products, obtaining in this way surfaces whose metric structure is not inherited from $\mathbb{R}^3$. The next step is to dispense with the ambient space, defining abstract surfaces that are not even diffeomorphic to surfaces of $\mathbb{R}^3$.

Let $U \subseteq \mathbb{R}^2$ be an open and connected region, and let $E, F, G: U \rightarrow \mathbb{R}$ be C^∞ -functions such that

- E and G are strictly positive;
- $EG - F^2 > 0$ at all points of U .

For every point $(u, v) \in U$ we define an inner product $\langle \cdot, \cdot \rangle_{(u,v)}$ on $T_{(u,v)}U = \mathbb{R}^2$ as follows:

$$\langle (x_1, y_1), (x_2, y_2) \rangle_{(u,v)} = x_1x_2E(u, v) + (x_1y_2 + y_1x_2)F(u, v) + y_1y_2G(u, v).$$

If $\alpha: [a, b] \rightarrow U$ is a regular curve then, in this inner product, the norm of $\alpha'(t)$ is given by

$$\sqrt{I_{\alpha(t)}(\alpha'(t))} = \sqrt{u'(t)^2E(\alpha(t)) + 2u'(t)v'(t)F(\alpha(t)) + v'(t)^2G(\alpha(t))},$$

and the length of α is

$$l(\alpha) = \int_a^b \sqrt{I_{\alpha(t)}(\alpha'(t))} dt.$$

The area of a region $\Delta \subset U$ is computed, as usual, by

$$\iint_U \sqrt{EG - F^2} du dv.$$

The inner product thus imposed on U determines therefore all intrinsic measures: lengths of curves, angles between vectors, areas of regions. We say that U has been equipped with a *Riemannian metric*. We can define, for such a metric, the notions of Gaussian curvature, covariant derivative, geodesic — and, under the assumption that U is oriented, that of geodesic curvature. To this end, we make use of the formulas in Chapter 4 that express such notions using E , F , G , Christoffel's symbols, and their derivatives [such as (4.8), (4.9), (4.31)].

Examples 5.6.1 **A.** Let us take $\mathbb{R}^2$ with the Riemannian metric given by

$$ds^2 = du^2 + \{2 + \cos(u + v)\}^2 dv^2$$

(which is a shorthand for $E \equiv 1$, $F \equiv 0$, $G(u, v) = \{2 + \cos(u + v)\}^2$). Using formula (4.38) in Section 4.6, we find that the Gaussian curvature of this metric is

$$K(u, v) = \frac{\cos(u + v)}{\{2 + \cos(u + v)\}}.$$

B. Let us consider, also in $\mathbb{R}^2$, the metric

$$ds^2 = \frac{1}{1 + (u^2 + v^2)^2} (du^2 + dv^2).$$

This is a conformal metric, in the sense that angles are measured as usual in $\mathbb{R}^2$, since $F \equiv 0$ and $E = G$. Another particularity is that with this metric $\mathbb{R}^2$ has finite area and is not a complete surface (the proof of this statement is Exercise 136). $\square$

An *abstract surface* (of class C^∞) is a **connected** topological space S equipped with an *atlas* $\mathcal{A} = ((U_\alpha, \Phi_\alpha))_{\alpha \in \mathcal{I}}$ such that:

- every U_α is an open subset of $\mathbb{R}^2$, $\Phi_\alpha(U_\alpha)$ is an open subset of S , and $S = \bigcup_{\alpha \in \mathcal{I}} \Phi_\alpha(U_\alpha)$;
- $\Phi_\alpha: U_\alpha \rightarrow S$ is a homeomorphism onto its image;
- if $\Phi_\alpha(U_\alpha) \cap \Phi_\beta(U_\beta) = V \neq \emptyset$ then $\Phi_\beta^{-1} \circ \Phi_\alpha: \Phi_\alpha^{-1}(V) \rightarrow \Phi_\beta^{-1}(V)$ is C^∞ .

To avoid pathologies (see the appendix to volume I of [25] for a catalogue of them) we further impose certain conditions on the topology of the surface S , namely: any two distinct points of S have disjoint open neighborhoods (S is *Hausdorff*) and there exists a subset of S which is dense and countable (S is *separable*).

By Proposition 2.2.1, and since the mentioned topological requirements are satisfied by any subset of $\mathbb{R}^n$, any regular connected surface of $\mathbb{R}^3$ is an abstract surface. What we have done was to require the coordinate changes on S to be C^∞ -diffeomorphisms, which allows us to define all sorts of notions using local coordinates. Note that a surface does not determine an atlas uniquely: we can add to or take away from a given atlas a few mappings, and as long as the remaining ones still cover S and coordinate changes remain C^∞ , we obtain several different atlases;

what matters is that they all define the same differentiable structure on S , and thus the same abstract surface.

Given an abstract surface S , it is not easy, now that we have no ambient space, to define a tangent space $T_p S$. Our approach, instead of defining it, is to explain how to work with it in local coordinates. So if we have a curve that in local coordinates is written $\alpha(t) = \Phi(u(t), v(t))$, we would still like it to be

$$\alpha'(t) = u'(t)\Phi_u + v'(t)\Phi_v \quad (5.26)$$

— whatever the meaning of the vectors Φ_u and Φ_v . If we use other coordinates $\Psi(\tilde{u}, \tilde{v})$, we should have

$$\begin{aligned} \Phi_u &= \frac{\partial \tilde{u}}{\partial u} \Psi_{\tilde{u}} + \frac{\partial \tilde{v}}{\partial u} \Psi_{\tilde{v}}, \\ \Phi_v &= \frac{\partial \tilde{u}}{\partial v} \Psi_{\tilde{u}} + \frac{\partial \tilde{v}}{\partial v} \Psi_{\tilde{v}}; \end{aligned}$$

and, replacing in (5.26), we obtain

$$\begin{aligned} \alpha'(t) &= \left(\frac{\partial \tilde{u}}{\partial u} u'(t) + \frac{\partial \tilde{v}}{\partial u} v'(t) \right) \Psi_{\tilde{u}} + \left(\frac{\partial \tilde{v}}{\partial u} u'(t) + \frac{\partial \tilde{v}}{\partial v} v'(t) \right) \Psi_{\tilde{v}} \\ &= \tilde{u}'(t) \Psi_{\tilde{u}} + \tilde{v}'(t) \Psi_{\tilde{v}} \end{aligned}$$

— which shows that (5.26) is compatible with the change of coordinates. Hence for us the tangent space to S at the point $\Phi(u, v)$ is simply the space generated by the two independent vectors Φ_u and Φ_v . The transition matrix from the basis $(\Psi_{\tilde{u}}, \Psi_{\tilde{v}})$ to (Φ_u, Φ_v) is the Jacobian of $\Psi^{-1} \circ \Phi$.

A *Riemannian metric* on an abstract surface S is given by an inner product $\langle \cdot, \cdot \rangle_p$ in the tangent space $T_p S$ for every $p \in S$. In local coordinates $\Phi(u, v)$, the matrix of this inner product with respect to the basis (Φ_u, Φ_v) is $\begin{bmatrix} E & F \\ F & G \end{bmatrix}$, and we usually require that E, F, G are C^∞ functions; the coefficients $\tilde{E}, \tilde{F}, \tilde{G}$ of other coordinates $\Psi(\tilde{u}, \tilde{v})$ are obtained from these by the relation

$$\begin{bmatrix} \tilde{E} & \tilde{F} \\ \tilde{F} & \tilde{G} \end{bmatrix} = \{J(\Phi^{-1} \circ \Psi)\}^\top \begin{bmatrix} E & F \\ F & G \end{bmatrix} J(\Phi^{-1} \circ \Psi),$$

where $\{J(\Phi^{-1} \circ \Psi)\}^\top$ denotes the transpose of $J(\Phi^{-1} \circ \Psi)$.

We can define, using the Riemannian metric $(\langle \cdot, \cdot \rangle_p)_{p \in S}$, intrinsic notions such as those of Gaussian curvature, covariant derivative, etc., using the formulas that express such concepts via the coefficients E, F, G and the Christoffel symbols Γ_{ij}^k . Of course, we should now check that the definitions do not depend on the coordinates used (which was previously unnecessary since such concepts had been defined without any use of local coordinates), but we just assure the reader that such a check is possible; for details, see [13].

Examples 5.6.2 A. Let us consider the equivalence relation on $\mathbb{R}^2$ given by

$$(u, v) \sim (\tilde{u}, \tilde{v})$$

if and only if both differences $u - \tilde{u}$ and $v - \tilde{v}$ are integers, and let $[u, v]$ denote the equivalence class of the pair (u, v) . Let $\mathbb{R}^2 / \sim$ be the set of equivalence classes and $\Pi: \mathbb{R}^2 \rightarrow \mathbb{R}^2 / \sim$ the *quotient mapping*, given by $\Pi(u, v) = [u, v]$. We define a topology on $\mathbb{R}^2 / \sim$ by $U \subseteq \mathbb{R}^2 / \sim$ being open if and only if $\Pi^{-1}(U)$ is an open subset of $\mathbb{R}^2$. Thus $\mathbb{R}^2 / \sim$ is a surface, since a restriction of Π to all squares of the form $]u_0, u_0 + 1[\times]v_0, v_0 + 1[$ constitutes an atlas of $\mathbb{R}^2 / \sim$.

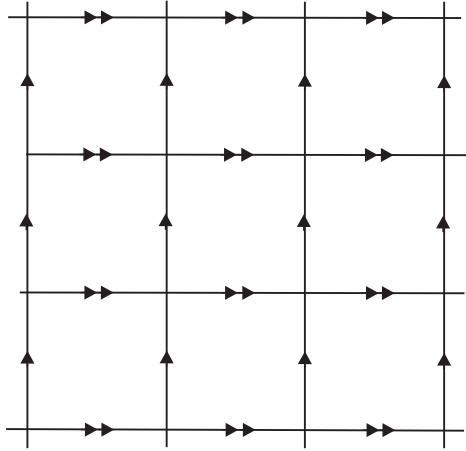


Figure 5.6

It is not difficult to verify that $\mathbb{R}^2 / \sim$ is diffeomorphic to the torus $\mathbb{T}^2$. The mapping Π induces a Riemannian metric on $\mathbb{R}^2 / \sim$, for which Π_u, Π_v constitute an orthonormal basis of $T_{[u,v]}(\mathbb{R}^2 / \sim)$ ($E \equiv G \equiv 1, F \equiv 0$). This is the only metric such that $\Pi: \mathbb{R}^2 \rightarrow \mathbb{R}^2 / \sim$ is a local isometry, and therefore $\mathbb{R}^2 / \sim$ is a surface with zero constant curvature, commonly called a *flat torus*. In fact Π is an isometric covering.

We suggest as an exercise to show that the equivalence relation

$$(u, v) \simeq (\tilde{u}, \tilde{v}) \Leftrightarrow \exists n \in \mathbb{Z} : \tilde{u} = u + n \text{ and } \tilde{v} = (-1)^n v$$

defines a surface of zero curvature diffeomorphic to the Möbius strip.

B. Let us define a complete, simply connected surface of constant curvature equal to -1 . This surface, which we denote by $\mathbb{D}$ (*hyperbolic plane*), is by Theorem 5.3.3 diffeomorphic to the plane, and it suffices therefore to find a Riemannian metric on $\mathbb{R}^2$ of curvature -1 . But this is easy if we use polar coordinates and recall the formulas given at the end of Section 4.6. Then the coefficients of the parameterization $\Phi(\rho, \varphi) = (\rho \cos \varphi, \rho \sin \varphi)$ become

$$E \equiv 1, \quad F \equiv 0, \quad G \equiv \sinh^2 \rho \quad (5.27)$$

— and we can take the coefficients of this metric in Cartesian coordinates. However, this is not the best model of $\mathbb{D}$, because the metric defined this way is not conformal. Let U be the open disk with radius λ ($0 < \lambda \leq +\infty$) centered at the origin, and $f: [0, \lambda[\rightarrow [0, +\infty[$ a differentiable, strictly monotone function such that $f(0) = 0$, and consider the mapping $\Psi: U \rightarrow \mathbb{R}^2$ which transforms the polar coordinate point (ρ, φ) into the coordinate point $(f(\rho), \varphi)$. We want to find f so that the Riemannian metric induced by Ψ on U , from the metric defined by (5.27) on $\mathbb{R}^2$, is conformal. If p has coordinates (ρ, φ) then, letting $\mathbf{v}_\varphi = (\cos \varphi, \sin \varphi)$ and $\mathbf{w}_\varphi = (-\sin \varphi, \cos \varphi)$, we have

$$D\Psi_p(\mathbf{v}_\varphi) = f'(\rho)\mathbf{v}_\varphi, \quad D\Psi_p(\mathbf{w}_\varphi) = \frac{f(\rho)}{\rho} \mathbf{w}_\varphi \quad (5.28)$$

— by which we obtain

$$\begin{aligned} \langle \mathbf{v}_\varphi, \mathbf{v}_\varphi \rangle_p &= [f'(\rho)]^2, \\ \langle \mathbf{v}_\varphi, \mathbf{w}_\varphi \rangle_p &= 0, \\ \langle \mathbf{w}_\varphi, \mathbf{w}_\varphi \rangle_p &= \frac{G(f(\rho), \varphi)}{\rho^2} = \frac{\sinh^2(f(\rho))}{\rho^2}. \end{aligned} \quad (5.29)$$

For the metric (5.29) to be conformal, necessarily

$$f'(\rho) = \frac{\sinh(f(\rho))}{\rho}.$$

The general solution of this equation is $f(\rho) = 2\operatorname{tgh}^{-1}(c\rho)$, where c is a real constant, and is defined for $\rho \in [0, \frac{1}{|c|}[$. We take $c = 1$, so that $\lambda = 1$ and U is the unit disk. In this case $f'(\rho) = \frac{2}{1-\rho^2}$, and therefore (5.29) defines in Cartesian coordinates the metric

$$ds^2 = \frac{4}{\{1 - (u^2 + v^2)\}^2} (du^2 + dv^2). \quad (5.30)$$

From now on, let $\mathbb{D}$ denote the disk $u^2 + v^2 < 1$ (or the disk $|z| < 1$ in the complex plane $\mathbb{C}$) equipped with the Riemannian metric (5.30). $\mathbb{D}$ is usually called *Poincaré's disk*. We next give some of its properties:

(1) $\mathbb{D}$ has constant curvature equal to -1 . We verify this using formula (4.31) in Section 4.5. To obtain a surface of constant curvature $K < 0$, we would take, on the same disk, the metric

$$ds^2 = \frac{4}{|K|(1 - (u^2 + v^2))^2} (du^2 + dv^2).$$

(2) The diameters of $\mathbb{D}$ are geodesic. This can be seen by checking that its geodesic curvature is zero, but we prefer to show that the diameters are minimizing curves. Indeed, if $\alpha(t) = (u(t), v(t))$ is a curve from $\alpha(0) = (0, 0)$ to $\alpha(1) = (u_1, 0)$ ($u_1 > 0$) then

$$\begin{aligned}
l(\alpha) &= \int_0^1 \sqrt{I_{\alpha(t)}(\alpha'(t))} dt = \int_0^1 \frac{2\sqrt{u'(t)^2 + v'(t)^2}}{1 - [u(t)^2 + v(t)^2]} dt \\
&\geq \int_0^1 \frac{2|u'(t)|}{1 - u(t)^2} dt \geq \int_0^1 \frac{2u'(t)}{1 - u(t)^2} dt = \log \left(\frac{1 + u_1}{1 - u_1} \right),
\end{aligned} \tag{5.31}$$

and equality holds if and only if $v(t) = 0$ and $u'(t) \geq 0$ for all $t \in [0, 1]$ — i.e., if and only if the trace of α is the line segment $[0, u_1] \times \{0\}$, which is therefore a minimizing curve. The diameter $] - 1, 1[\times \{0\}$ is thus a geodesic of $\mathbb{D}$. Since rotations around the origin are isometries of $\mathbb{D}$, all other diameters are also geodesics.

It follows from (5.31) that geodesics starting from the origin have infinite length, and therefore (see Exercise 119) $\mathbb{D}$ is complete. Furthermore, the intrinsic distance between 0 and $z \in \mathbb{D}$ is

$$d_{\mathbb{D}}(0, z) = \log \left(\frac{1 + |z|}{1 - |z|} \right).$$

(3) Let $a, b \in \mathbb{C}$ be such that $|a| > |b|$, and consider the mapping

$$h_{a,b}(z) = \frac{az + \bar{b}}{bz + \bar{a}}.$$

The reader may check that $h_{a,b}$ sends $\mathbb{D}$ bijectively onto itself. More remarkable is that $h_{a,b}$ is an isometry of $\mathbb{D}$. Indeed, for $z \in \mathbb{D}$ and $w \in \mathbb{C}$, we have (abbreviating $h_{a,b}$ to h)

$$\begin{aligned}
I_{h(z)}(Dh_z(w)) &= I_z(w) \Leftrightarrow I_{h(z)}(h'(z)w) = I_z(w) \\
&\Leftrightarrow \frac{4|h'(z)|^2}{(1 - |h(z)|^2)^2} |w|^2 = \frac{4}{(1 - |z|^2)^2} |w|^2 \\
&\Leftrightarrow |h'(z)| = \frac{1 - |h(z)|^2}{1 - |z|^2}.
\end{aligned} \tag{5.32}$$

Checking this last equality is a simple calculation.

Using Proposition 5.7.3 of the next section, we can show that all isometries of $\mathbb{D}$ that preserve orientation are of this form (one obtains the rotations by letting $b = 0$), and those that reverse it are the conjugates of these $\left(z \mapsto \frac{\bar{a}\bar{z} + b}{\bar{b}\bar{z} + a}, \text{ with } |a| > |b| \right)$.

Now given $z_0 \in \mathbb{D}$,

$$h(z) = \frac{z - z_0}{1 - \bar{z}_0 z}$$

is an isometry that maps z_0 to 0, which allows us to deduce a formula for the distance between z_0 and another point $z_1 \in \mathbb{D}$:

$$\begin{aligned}
 d_{\mathbb{D}}(z_0, z_1) &= d_{\mathbb{D}}(h(z_0), h(z_1)) = d_{\mathbb{D}}\left(0, \frac{z_1 - z_0}{1 - \overline{z_0}z_1}\right) \\
 &= \log \left(\frac{|1 - \overline{z_0}z_1| + |z_1 - z_0|}{|1 - \overline{z_0}z_1| - |z_1 - z_0|} \right).
 \end{aligned}
 \tag{5.33}$$

(4) Mappings of the type

$$z \mapsto \frac{az + b}{cz + d},$$

where a, b, c, d are complex numbers such that $ad - bc \neq 0$, are called *Möbius transformations*, and are bijections of $\mathbb{C} \cup \{\infty\}$ (the Riemann sphere) onto itself. We can assume that $ad - bc = 1$, since if we multiply each of the numbers a, b, c, d by the same nonzero factor the transformation does not change.

Möbius transformations form a group $\mathfrak{M}$ with respect to the composition of functions; if to each matrix $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with $ad - bc = 1$ we attach

$$f_M(z) = \frac{az + b}{cz + d},$$

we have $f_{M_1 M_2} = f_{M_1} \circ f_{M_2}$. We hence obtain a group homomorphism $\text{Sl}(2, \mathbb{C}) \rightarrow \mathfrak{M}$ (where $\text{Sl}(2, \mathbb{C})$ is the multiplicative group of the complex 2×2 matrices with determinant equal to 1).

We are now interested in the geometric properties of these transformations. Writing

$$\frac{az + b}{cz + d} = \frac{a}{c} + \frac{b - \frac{ad}{c}}{c(z + \frac{d}{c})} \quad (\text{if } c \neq 0), \quad \frac{az + b}{d} = \frac{a}{d}z + \frac{b}{d},$$

we recognize that any Möbius transformation is written as a composition of:

- translations $z \mapsto z + \zeta \quad (\zeta \in \mathbb{C})$
- rotations $z \mapsto e^{i\theta}z \quad (\theta \in [0, 2\pi])$
- homotheties $z \mapsto \lambda z \quad (\lambda \in]0, +\infty[)$
- inversions $z \mapsto \frac{1}{z}$

Transformations of the first three types have known properties: in particular, they transform straight lines into straight lines and circles into circles. Regarding the inversions, we have the following:

- if r is a line passing through 0, its inverse is still a line. If r does not pass through 0, its inverse is a circle passing through 0;
- the inverse of a circle $\mathcal{C}$ is a straight line if $0 \in \mathcal{C}$, otherwise it is also a circle.

In fact, the inverse of $u + iv$ is the point $\tilde{u} + i\tilde{v}$ given by

$$\tilde{u} = \frac{u}{u^2 + v^2}, \quad \tilde{v} = \frac{-v}{u^2 + v^2}.$$

If r is the line given by the equation $\alpha u + \beta v + \gamma = 0$ then, for $u + iv \in r$, we have

$$\alpha \tilde{u} - \beta \tilde{v} + \gamma(\tilde{u}^2 + \tilde{v}^2) = \frac{\alpha u + \beta v + \gamma}{u^2 + v^2} = 0$$

— which shows that $f(r)$ is a straight line if $\gamma = 0$, and is a circle otherwise. The second statement allows for an analogous verification.

We thus conclude that any Möbius transformation preserves the family of lines and circles in the plane. Moreover, every $f \in \mathfrak{M}$ is a conformal mapping and therefore keeps the orthogonality relation between two curves. Returning to the hyperbolic plane $\mathbb{D}$, the geodesics passing through a point $z_0 \neq 0$ are, in particular, images of the geodesics passing through 0 (which are line segments and intersect $\mathbf{S}^1$ orthogonally) by elements of $\mathfrak{M}$ which fix $\mathbf{S}^1$. Hence, *every geodesic of $\mathbb{D}$ is either a diameter of $\mathbf{S}^1$ or an arc of some circumference that intersects $\mathbf{S}^1$ orthogonally* (see Fig. 5.7).

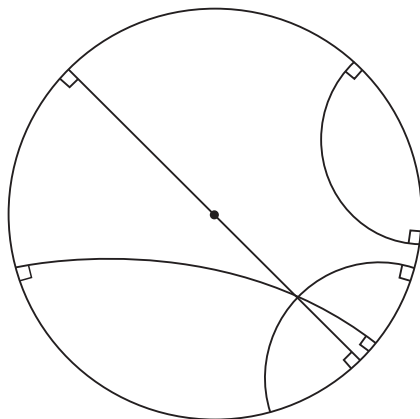


Figure 5.7

(5) If “lines” mean geodesics, Poincaré’s disc provides a model of the non-Euclidean geometry of Lobachevski and Bolyai. In this geometry all the axioms of Euclidean geometry are valid, except the axiom of parallels: in $\mathbb{D}$, through a point outside a “line” r , pass infinitely many “lines” that do not intersect r — and not only one, as in the Euclidean case.

Exercises

136. Show that $\mathbb{R}^2$, with the metric of example 5.6.1.B, has finite area and is not complete.

In the remaining exercises in this section we work with the hyperbolic plane $\mathbb{D}$.

137. Consider two geodesics $\gamma_1(s)$ and $\gamma_2(s)$ which, at time $s = 0$, start from $z_0 \in \mathbb{D}$ in directions that make an angle θ with each other. Show that

$$\delta(\theta) = \lim_{s \rightarrow +\infty} \{d_{\mathbb{D}}(\gamma_1(s), \gamma_2(s)) - 2s\}$$

exists.

138. Check that for all $z_0, z_1 \in \mathbb{D}$, the set $\{z \in \mathbb{D}: d_{\mathbb{D}}(z, z_0) = d_{\mathbb{D}}(z, z_1)\}$ is a geodesic.

139. (a) Show that the hyperbolic circle $\mathbb{S}(z_0; r) = \{z \in \mathbb{D}: d_{\mathbb{D}}(z, z_0) = r\}$ is also a Euclidean circle.

(b) Check that any equilateral triangle in $\mathbb{D}$ (geodesic triangle with all sides equal) can be inscribed into a hyperbolic circle, but that this is no longer true for every geodesic triangle.

140. (a) Check that $g(z) = -\frac{i}{2} + \frac{1}{z-i}$ sends $\mathbb{D}$ into the half-plane $\{z \in \mathbb{C}: \Im(z) > 0\}$.

(b) Let $\mathbb{H}$ be the half-plane $\{z \in \mathbb{C}: \Im(z) > 0\}$ with the metric given by $E = G = 1/\Im(z)^2$ and $F \equiv 0$. Show that g is an isometry of $\mathbb{D}$ on $\mathbb{H}$.

(c) Show that the geodesics of $\mathbb{H}$ are the vertical semicircles $\Re(z) = \text{constant}$ and the half circumferences with center on the axis $\Im(z) = 0$.

(d) Show that the isometries of $\mathbb{H}$ that preserve orientation are the functions

$$z \mapsto \frac{az + b}{cz + d},$$

where a, b, c, d are **real** numbers such that $ad - bc = 1$ (you can use Proposition 5.7.3 of the next section).

(e) Denote by $\mathfrak{I}(\mathbb{D})$ and $\mathfrak{I}(\mathbb{H})$ the groups of the isometries of $\mathbb{D}$ and $\mathbb{H}$ that preserve orientation. Note that $\mathfrak{I}(\mathbb{H}) = g \circ \mathfrak{I}(\mathbb{D}) \circ g^{-1}$, and therefore $\mathfrak{I}(\mathbb{D})$ and $\mathfrak{I}(\mathbb{H})$ are *conjugate* subgroups of the group $\mathfrak{M}$ of Möbius transformations.

141. Check that $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto f_M(z) = \frac{az + b}{cz + d}$ is a homomorphism of $\text{Sl}(2, \mathbb{C})$ on $\mathfrak{M}$, and that the kernel of this homomorphism is $\{I, -I\}$. Conclude that $f_M = f_N$ if and only if $M = \pm N$.

142. Two elements $f, g \in \mathfrak{M}$ are *conjugate* if there exists $h \in \mathfrak{M}$ such that $f = h \circ g \circ h^{-1}$.

(a) Given $f \in \mathfrak{M}$, define $\tau(f) = |\text{trace}(M)|$ if $f = f_M$ for $M \in \text{Sl}(2, \mathbb{C})$. Show that if $f, g \neq \text{id}$ then f and g are conjugate if and only if $\tau(f) = \tau(g)$.

(b) Let $\Gamma = \mathfrak{I}(\mathbb{D})$ or $\Gamma = \mathfrak{I}(\mathbb{H})$. Show that if $f, g \in \Gamma$ are conjugate by an element of $\mathfrak{M}$, then they are conjugate by some element of Γ .

143. Denote by Γ the group $\mathfrak{I}(\mathbb{D})$ or $\mathfrak{I}(\mathbb{H})$, and let $f \in \Gamma$. Show that:

(a) if $\tau(f) > 2$ then f is conjugate to $z \mapsto \lambda z$ in $\mathbb{H}$, where $\lambda > 0$ satisfies $\tau(f) = \sqrt{\lambda} + \frac{1}{\sqrt{\lambda}}$;

(b) if $\tau(f) = 2$ and $f \neq \text{id}$ then h is conjugate to $z \mapsto z + 1$ in $\mathbb{H}$;

(c) if $\tau(f) < 2$ then f is conjugate to $z \mapsto e^{i\theta} z$ on $\mathbb{D}$, where $2 \cos\left(\frac{\theta}{2}\right) = \tau(f)$.

Note: $f \neq \text{id}$ is called *hyperbolic*, *parabolic* or *elliptic* according to whether a), b) or c) of Ex. 143 holds.

144. Let $f \in \mathfrak{I}(\mathbb{D}) \setminus \text{id}$. Show that:

(a) $\inf_{z \in \mathbb{D}} d_{\mathbb{D}}(f(z), z) > 0$ if and only if f is hyperbolic.

(b) there exists z_0 such that $d_{\mathbb{D}}(f(z_0), z_0) = \inf_{z \in \mathbb{D}} d_{\mathbb{D}}(f(z), z)$ if and only if f is elliptic or hyperbolic.

5.7 Complete Surfaces of Constant Curvature

In this section we will study the isometry groups of the complete and simply connected surfaces of constant curvature, and show that the other complete surfaces of equal curvature are obtained from these as quotients by a certain subgroup of the isometry group. This approach allows us to describe all complete surfaces of non-negative constant curvature.

We begin with a lemma that will be applied repeatedly:

Lemma 5.7.1 *Let $f, g: S_1 \rightarrow S_2$ be local isometries such that for a certain $p \in S_1$ we have $f(p) = g(p)$ and $Df_p = Dg_p$. Then $f = g$.*

Proof Given $\mathbf{v} \in T_p S_1$, let us consider the geodesic $\gamma(t) = \exp_p(t\mathbf{v})$. Then $f \circ \gamma$ and $g \circ \gamma$ are parametrized geodesics with the same initial conditions, since $f \circ \gamma(0) = g \circ \gamma(0) = f(p) = g(p)$ and $(f \circ \gamma)'(0) = Df_p(\mathbf{v}) = Dg_p(\mathbf{v}) = (g \circ \gamma)'(0)$ — and therefore $f \circ \gamma(t) = g \circ \gamma(t)$ whenever $\gamma(t)$ is defined. We thus conclude that $f \circ \exp_p = g \circ \exp_p$, which implies that f and g coincide in a neighborhood of p . This shows that the set $U = \{q \in S_1: f(q) = g(q), Df_q = Dg_q\}$ is non-empty and open. But its definition ensures that U is also closed, and therefore, S_1 being connected, $U = S_1$. $\square$

We denote by S_K the complete, simply connected surface of constant curvature K . S_K is thus the sphere with radius $\frac{1}{\sqrt{K}}$ (if $K > 0$), the Euclidean plane (if $K = 0$), or the hyperbolic plane of curvature K (if $K < 0$). We can then rewrite Theorem 5.3.4 as follows:

Theorem 5.7.2 *Let S be a complete surface of constant curvature K . Then there exists an isometric covering $f: S_K \rightarrow S$.*

Proof The case $K \leq 0$ was treated in Theorem 5.3.4. Only the case $K > 0$ is left. Given $p \in S_K$, the mapping $\exp_p$ sends the disk $B_{\pi/\sqrt{K}}(p)$ diffeomorphically into $S_K \setminus \{-p\}$. Thus, if q is a point of S and $L: T_p S_K \rightarrow T_q S$ a linear isometry, the mapping $g: S_K \setminus \{-p\} \rightarrow S$ defined by $g = \exp_q \circ L \circ \exp_p^{-1}$ is a local isometry. Let us now take $\tilde{p} \in S_K \setminus \{p, -p\}$, and let $\tilde{q} = g(\tilde{p})$, $\tilde{L} = Dg_{\tilde{p}}$ and $h: S_K \setminus \{-\tilde{p}\} \rightarrow S$ be the mapping given by $h = \exp_{\tilde{q}} \circ \tilde{L} \circ \exp_{\tilde{p}}^{-1}$. The mapping h is also a local isometry, and furthermore $g(\tilde{p}) = h(\tilde{p})$ and $Dg_{\tilde{p}} = Dh_{\tilde{p}}$. By Lemma 5.7.1, g and h coincide on the intersection of their domains, and we can thus define a local isometry $f: S_K \rightarrow S$ by $f(r) = g(r)$ if $r \neq -p$, and $f(r) = h(r)$ if $r \neq -\tilde{p}$. Since S_K is compact, by Proposition 5.2.2 this local isometry is a covering. $\square$

The following result is a consequence of Lemma 5.7.1 and the proofs of Theorems 5.3.4 and 5.7.2.

Proposition 5.7.3 *Given $p, q \in S_K$ and a linear isometry $L: T_p S_K \rightarrow T_q S_K$, there exists one and only one isometry $f: S_K \rightarrow S_K$ such that $f(p) = q$ and $Df_p = L$.*

Since the composite and inverse of isometries are still isometries, the set of isometries of a given surface forms a group. By the above proposition, the group of isometries of S_K , which we denote by $\mathfrak{I}_K$, is exceptionally large.

Let us now assume that $f: S_K \rightarrow S$ is an isometric covering. Given $q \in S$, let p_0 and p_1 be two of the preimages of q , and L_i ($i = 0, 1$) the linear isometry $Df_{p_i}: T_{p_i} S_K \rightarrow T_q S$. By Proposition 5.7.3, there exists an isometry $g: S_K \rightarrow S_K$ such that $g(p_0) = p_1$ and $Dg_{p_0} = L_1^{-1} \circ L_0$. Now $f \circ g: S_K \rightarrow S$ is a local isometry such that

$$\begin{aligned} f \circ g(p_0) &= q = f(p_0), \\ D(f \circ g)_{p_0} &= Df_{p_1} \circ Dg_{p_0} = L_1 \circ (L_1^{-1} \circ L_0) = Df_{p_0} \end{aligned}$$

— it follows, by Lemma 5.7.1, that $f \circ g = f$.

It is easily checked that the set $\mathfrak{I}_K(f) = \{g \in \mathfrak{I}_K: f \circ g = f\}$ is a subgroup of $\mathfrak{I}_K$; we call it the *covering group* of f .

Proposition 5.7.4 *For every $q \in S$, the group $\mathfrak{I}_K(f)$ acts transitively on $f^{-1}(\{q\})$ — that is, for every pair of points $p_0, p_1 \in f^{-1}(\{q\})$ there exists one and only one $g \in \mathfrak{I}_K(f)$ such that $g(p_0) = p_1$.*

The action of $\mathfrak{I}_K(f)$ on S_K is discontinuous. This means that every point p of S_K has an open neighborhood V such that $g(V) \cap V$ is empty for all $g \in \mathfrak{I}_K(f) \setminus \{\text{id}\}$.

The first statement (apart from the uniqueness of g , which is easy) has already been proved. Regarding the second, let us take an evenly covered neighborhood U of $f(p)$ and let V be the component of $f^{-1}(U)$ that contains p . If $g \in \mathfrak{I}_K(f) \setminus \{\text{id}\}$ and $q \in V$ then $q \neq g(q)$, since the identity is the only element of $\mathfrak{I}_K(f)$ with some fixed point. Since $f(q) = f(g(q))$, necessarily $g(q) \notin V$. Thus $g(V) \cap V = \emptyset$, which proves the statement.

It deserves mention that what we call *discontinuous action* is called *proper discontinuous action* by most authors, who reserve the former name for a weaker condition, of which we will make no use. We further say that a subgroup Γ of $\mathfrak{I}_K$ is *discrete* if its action on S_K is discontinuous (this terminology is also not the usual one).

The covering group completely determines the surface in the following sense: if we have two isometric coverings $f_i: S_K \rightarrow S_i$ ($i = 1, 2$) such that $\mathfrak{I}_K(f_1) = \mathfrak{I}_K(f_2)$ then S_1 and S_2 are isometric. We can further ask which subgroups of $\mathfrak{I}_K$ are covering groups. The answer is simple:

Proposition 5.7.5 *A subgroup Γ of $\mathfrak{I}_K$ is a covering group if and only if it is discrete.*

Proof It remains to prove that such a subgroup is a covering group. We note that the action of Γ on S_K induces an equivalence relation whose classes are $[p] = \{g(p) : g \in \Gamma\}$. Let S_K/Γ be the set of equivalence classes and let $\Pi: S_K \rightarrow S_K/\Gamma$ be the quotient mapping. Given $p \in S$, let (U, Φ) be a parameterization in a neighborhood of p such that $g(\Phi(u)) \cap \Phi(U) = \emptyset$ for all $g \in \Gamma \setminus \{\text{id}\}$, and let us put $\Psi = \Pi \circ \Phi$: the set of mappings Ψ so defined constitutes an atlas of S_K/Γ , which is therefore a surface; moreover, $\Psi(U)$ is an evenly covered neighborhood of $[p]$, which shows that Π is a covering. With the Riemannian metric induced by Π on S_K/Γ the mapping Π is an isometric covering whose covering group is obviously Γ . $\square$

Consider, for example, the case of the sphere S_K given by the equation $x^2 + y^2 + z^2 = \frac{1}{K}$ ($K > 0$). The isometries of S_K are the restrictions of the linear isometries of $\mathbb{R}^3$, and correspond to the orthogonal 3×3 matrices over $\mathbb{R}$ — the group of which is denoted by $O(3, \mathbb{R})$. An example of a discrete subgroup of $O(3, \mathbb{R})$ is $\{-I, I\}$, where I is the identity matrix. The surface $S_K/\{-I, I\}$ is called the *projective plane*, and is the surface obtained by identifying in S_K the pairs of diametrically opposed points. The projective plane is non-orientable, and in fact contains a Möbius strip (see Fig. 5.8)

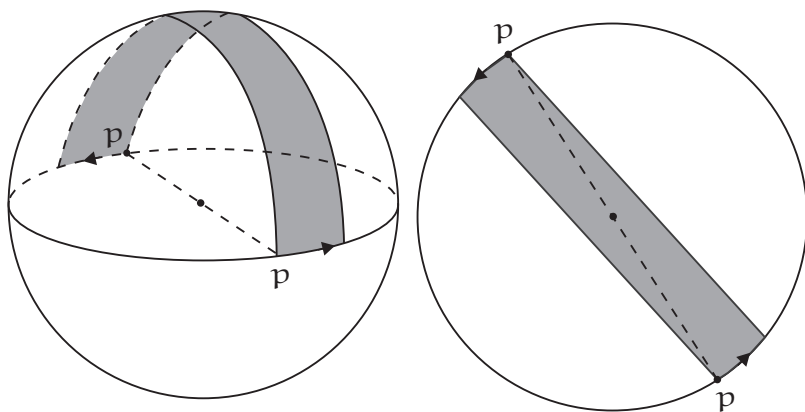


Figure 5.8

More generally, we have the following result (the proof of which is Exercise 145 of this section):

Proposition 5.7.6 *Let Γ be a discrete subgroup of $\mathfrak{S}_K$. Then S_K/Γ is orientable if and only if every isometry $g \in \Gamma$ preserves the orientation of S_K .*

We can now describe all complete surfaces of constant positive curvature.

Theorem 5.7.7 *The only complete surfaces of constant curvature $K > 0$ are, up to isometry, the sphere and the projective plane.*

Proof In view of Proposition 5.7.5, it suffices to show that the only discrete subgroups of $\mathfrak{I}_K \simeq O(3, \mathbb{R})$ are $\{I, -I\}$ and the trivial group $\{I\}$. Let Γ be such a subgroup: given $A \in \Gamma$, the isometry A has some real eigenvalue λ , which is necessarily 1 or -1 . If $\lambda = 1$, then A has some fixed point in S_K , which implies that $A = I$. If $\lambda = -1$ then A^2 has the eigenvalue $\lambda^2 = 1$ and therefore $A^2 = I$. The eigenvalues of A are thus all equal to 1 or -1 ; but none of them can be 1, and therefore $A = -I$. We thus have $\Gamma = \{I\}$ or $\Gamma = \{I, -I\}$. $\square$

In example 5.6.2.A we described the torus as the quotient of $\mathbb{R}^2$ by the group Γ of the translations $T(\mathbf{v})$ associated with vectors $\mathbf{v}$ of integer coordinates. Γ is generated by the two independent translations $T(1, 0)$ and $T(0, 1)$; as we will see below, this is a typical situation. The square $]0, 1[\times]0, 1[$ is a fundamental region for Γ . In general, we say that an open subset U of S_K is a *fundamental region* for the action of a discrete subgroup Γ of $\mathfrak{I}_K$ if

- $g(U) \cap h(U) = \emptyset$ for all $g \neq h$ in Γ ;
- S_K is the union of the closures of $g(U)$ for $g \in \Gamma$.

The most interesting fundamental regions are polygons, where the sides are geodesic segments. It is possible to reconstruct the surface S_K/Γ if one knows how to identify the sides of the polygon. In Fig. 5.9 we illustrate a fundamental region $\mathcal{P}$ for a certain discrete subset Γ of $\mathfrak{I}_K$ ($K < 0$): it is a regular polygon of eight sides whose sum of interior angles equals 2π ; the pairs of sides to be identified are designated by the same letter.

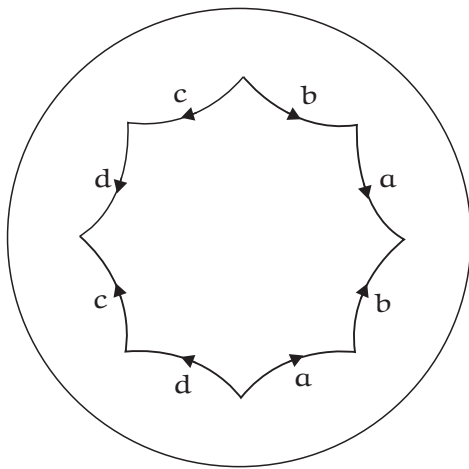


Figure 5.9

The surface S_K/Γ is, in this case, the double torus. We would obtain the n -torus, $n \geq 2$, from a regular polygon $\mathcal{P}_n$ of $4n$ sides and sum of angles equal to 2π ($n = 1$ is impossible because, by the Gauss-Bonnet theorem, the sum of the angles of any four-sided polygon in the hyperbolic plane is $< 2\pi$). A theorem of Poincaré ensures

that the elements of Γ that identify the sides of $\mathcal{P}_n$ generate Γ . The interested reader can find details in [10], [2] and [3].

To conclude this section and the book, we will determine all discrete subgroups of the group $\mathfrak{I}_0$ of the isometries of $\mathbb{R}^2$ — which is equivalent, by Proposition 5.7.5, to determining all complete surfaces of constant zero curvature.

Lemma 5.7.8 *The isometries of $\mathbb{R}^2$ without fixed points are translations and glide reflections (reflection in an axis r followed by a translation in the direction of r). Translations preserve orientation, while glide reflections reverse it.*

Proof Every isometry of $\mathbb{R}^2$ is of the form $f(p) = L(p) + \mathbf{v}$, where L is a linear isometry and $\mathbf{v}$ a vector. If $I - L$ were an isomorphism then f would have some fixed point, and so for f to have no fixed points, L must have eigenvalue 1. Let $(\mathbf{e}_1, \mathbf{e}_2)$ be an orthonormal basis of $\mathbb{R}^2$ such that $L(\mathbf{e}_1) = \mathbf{e}_1$. Since L is an isometry, necessarily $L(\mathbf{e}_2) = \pm \mathbf{e}_2$. If $L(\mathbf{e}_2) = \mathbf{e}_2$ then L is the identity and f a translation. If $L(\mathbf{e}_2) = -\mathbf{e}_2$, then L is the reflection in the straight line generated by $\mathbf{e}_1$. Letting $\mathbf{v} = \alpha \mathbf{e}_1 + \beta \mathbf{e}_2$, we recognize that f is the composite of the reflection in the line $\lambda \mapsto \lambda \mathbf{e}_1 + \frac{\beta}{2} \mathbf{e}_2$ with the translation associated with the vector $\alpha \mathbf{e}_1$ (in particular, f has **no fixed points if and only if** $\alpha \neq 0$).

Regarding the orientation, just note that it is preserved by f if and only if it is by L . □

Let $\Gamma \neq \{\text{id}\}$ be a discrete subgroup of $\mathfrak{I}_0$. Assume first that $\mathbb{R}^2/\Gamma$ is orientable: by Proposition 5.7.6, this means that all elements of Γ preserve orientation — that is, that they are all translations. To simplify notation, we identify the translation associated with $\mathbf{v}$ with the vector $\mathbf{v}$ itself.

Since Γ is discrete, we can choose $\mathbf{v}_1 \neq 0$ in Γ such that $|\mathbf{v}_1|$ is as small as possible. Then any other $\mathbf{v} \in \Gamma$ that is collinear with $\mathbf{v}_1$ is an integer multiple of $\mathbf{v}_1$ (since, if $\mathbf{v} = \lambda \mathbf{v}_1$, then, denoting by $[\lambda]$ the integer part of λ , $\{\lambda - [\lambda]\}$ is an element of Γ with norm less than that of $\mathbf{v}_1$, and is therefore zero). If Γ contains only such vectors, it is therefore of the form

$$\Gamma = \{n\mathbf{v}_1 : n \in \mathbb{Z}\} \quad (5.34)$$

— in which case $\mathbb{R}^2/\Gamma$ is a cylinder.

If Γ contains elements that are non-collinear with $\mathbf{v}_1$, then let $\mathbf{v}_2$ be one such vector with minimal norm. Then

$$\Gamma = \{n\mathbf{v}_1 + m\mathbf{v}_2 : n, m \in \mathbb{Z}\}. \quad (5.35)$$

Indeed, given $\mathbf{v} \in \Gamma$, it belongs to some parallelogram $\mathcal{P} = \{\alpha \mathbf{v}_1 + \beta \mathbf{v}_2 : n_1 \leq \alpha \leq n_1 + 1, m_1 \leq \beta \leq m_1 + 1\}$. Denoting by $\mathbf{w}$ the vertex of $\mathcal{P}$ closest to $\mathbf{v}$ and by d the length of the longest diagonal of $\mathcal{P}$, we have

$$|\mathbf{v} - \mathbf{w}| \leq \frac{d}{2} < \frac{1}{2} \{|\mathbf{v}_1| + |\mathbf{v}_2|\} \leq |\mathbf{v}_2|$$

— which implies, by the choice of $\mathbf{v}_2$, that $\mathbf{v} - \mathbf{w}$ is collinear with $\mathbf{v}_1$, and therefore an integer multiple of $\mathbf{v}_1$. We have therefore proved equality (5.35). *If Γ is of this form,*

$\mathbb{R}^2/\Gamma$ is a torus — and a fundamental region for Γ is the parallelogram of sides $\mathbf{v}_1$ and $\mathbf{v}_2$.

Then $f \circ g(p) = L_1 \circ L_2(p) + \mathbf{w}$, where $\mathbf{w} = \mathbf{u} + L_1(\mathbf{v})$. Now $f \circ g$ preserves orientation and, by Lemma 5.7.8, it must be a translation. Therefore $L_1 \circ L_2$ is the identity, hence $L_2 = L_1^{-1} = L_1$. In conclusion:

there is a linear isometry L and a set $\Gamma_R \subseteq \mathbb{R}^2$ such that $\{p \mapsto L(p) + \mathbf{v} : \mathbf{v} \in \Gamma_R\}$ is the set of the elements of Γ that reverse orientation.

We further denote by Γ_T the set of vectors associated with the translations of Γ . Γ_T is a discrete subgroup and, from the discussion above, the only two possibilities are that Γ_T is either one-dimensional (of the form $\{n\mathbf{v}_1 : n \in \mathbb{Z}\}$) or two-dimensional (of the form $\{n\mathbf{v}_1 + m\mathbf{v}_2 : n, m \in \mathbb{Z}\}$).

Lemma 5.7.9 *If $\mathbf{u}, \mathbf{v} \in \Gamma_R$ and $\mathbf{w} \in \Gamma_T$, then:*

- (i) $\mathbf{u} + L(\mathbf{v}) \in \Gamma_T$;
- (ii) $-L(\mathbf{v}) \in \Gamma_R$;
- (iii) $\mathbf{u} - \mathbf{v} \in \Gamma_T$;
- (iv) $\mathbf{v} + \mathbf{w} \in \Gamma_R$;
- (v) $L(\mathbf{w}) \in \Gamma_T$.

Proof The composite of $f(p) = L(p) + \mathbf{u}$ with $g(p) = L(p) + \mathbf{v}$ is given by $f \circ g(p) = p + \{\mathbf{u} + L(\mathbf{v})\}$, which proves (i). Regarding (ii), we observe that the inverse of $p \mapsto L(p) + \mathbf{v}$ is $p \mapsto L(p) - L(\mathbf{v})$. For (iii), we write $\mathbf{u} - \mathbf{v} = \mathbf{u} + L(-L(\mathbf{v}))$ and apply (i) and (ii). The composite of $h(p) = p + \mathbf{w}$ with $g(p) = L(p) + \mathbf{v}$ is $h \circ g(p) = L(p) + (\mathbf{v} + \mathbf{w})$, which proves (iv). Finally, we have $L(\mathbf{w}) = -L(\mathbf{v}) + L(\mathbf{v} + \mathbf{w})$ — and since, by (ii) and (iv), we have $-L(\mathbf{v}), \mathbf{v} + \mathbf{w} \in \Gamma_R$, it follows by (i) that $L(\mathbf{w}) \in \Gamma_T$, which proves (v). $\square$

Statement (v) says that $L(\Gamma_T) \subseteq \Gamma_T$. Applying L to both sides, we obtain the opposite inclusion $\Gamma_T \subseteq L(\Gamma_T)$; we thus conclude that Γ_T is invariant under L . Moreover, (iii) and (iv) say that Γ_R is the result of a translation on Γ_T : for all $\mathbf{v} \in \Gamma_R$ we have $\Gamma_R = \{\mathbf{v}\} + \Gamma_T$. This means that to obtain all the glide reflections of Γ , we just need to compose one of these transformations with each of the translations of Γ .

Let $(\mathbf{e}_1, \mathbf{v}_2)$ be an orthonormal basis of $\mathbb{R}^2$ such that $L(\mathbf{e}_1) = \mathbf{e}_1$, $L(\mathbf{e}_2) = -\mathbf{e}_2$. An important observation, contained in the proof of Lemma 5.7.8, is that if $\alpha\mathbf{e}_1 + \beta\mathbf{e}_2 \in \Gamma_R$ then $\alpha \neq 0$. With these remarks in mind, let us now look at the forms that the group Γ can take.

If Γ_T is of the form $\{n\mathbf{v}_1 : n \in \mathbb{Z}\}$, then $\mathbf{v}_1$ is an eigenvector of L , which is collinear with $\mathbf{e}_1$ or with $\mathbf{e}_2$. Take $\mathbf{v} = \alpha\mathbf{e}_1 + \beta\mathbf{e}_2 \in \Gamma_R$: then $\alpha \neq 0$ and, by (i), $2\alpha\mathbf{e}_1 = \mathbf{v} + L(\mathbf{v})$ belongs to Γ_T . We thus have $\mathbf{v}_1 = \lambda\mathbf{e}_1$, where we suppose $\lambda > 0$. Adding to $\mathbf{v}$, if necessary, an integer multiple of $\mathbf{v}_1$, we can assume that $0 < \alpha < \lambda$. Since $2\alpha\mathbf{e}_1$ is an integer multiple of $\mathbf{v}_1$, it follows that $\lambda = 2\alpha$ and $\mathbf{v}_1 = \mathbf{v} + L(\mathbf{v})$, which means that the translation associated with $\mathbf{v}_1$ is the composite of $f(p) = L(p) + \mathbf{v}$ with itself. Since $\Gamma_R = \{\mathbf{v}\} + \Gamma_T$, we conclude that Γ is the cyclic group generated by $f(p)$. Therefore:

If Γ contains elements that reverse orientation and Γ_T is one-dimensional, then Γ is generated by a single glide reflection. In this case $\mathbb{R}^2/\Gamma$ is a Möbius strip.

The case where Γ_T is two-dimensional is left. Since $L(\Gamma_T) = \Gamma_T$, there are two possibilities:

- (a) $\Gamma_T = \{n\mathbf{v}_1 + mL(\mathbf{v}_1) : n, m \in \mathbb{Z}\}$;
- (b) $\Gamma_T = \{n\lambda_1\mathbf{e}_1 + m\lambda_2\mathbf{e}_2 : n, m \in \mathbb{Z}\}$.

Suppose (a). Writing $\mathbf{v}_1 = \lambda\mathbf{e}_1 + \eta\mathbf{e}_2$ ($\lambda > 0, \eta \neq 0$), we have $\Gamma_T \cap \langle \mathbf{e}_1 \rangle = \{2n\lambda\mathbf{e}_1 : n \in \mathbb{Z}\}$. Take $\mathbf{v} = \alpha\mathbf{e}_1 + \beta\mathbf{e}_2 \in \Gamma_R$: we may suppose, by adding to $\mathbf{v}$ a multiple of $\mathbf{v}_1$, that $0 < \alpha < \lambda$. But $2\alpha\mathbf{e}_1 = \mathbf{v} + L(\mathbf{v})$ is in $\Gamma_T \cap \langle \mathbf{e}_1 \rangle$, implying $\alpha = n\lambda$, for some $n \in \mathbb{Z}$, which is absurd.

Thus (b) must be true. Repeating the argument above, we see that there exists $\mathbf{v} = \alpha\mathbf{e}_1 + \beta\mathbf{e}_2 \in \Gamma_R$ such that $2\alpha = \lambda_1$, and therefore $\mathbf{v} + L(\mathbf{v}) = \lambda_1\mathbf{e}_1$. As $\Gamma_R = \{\mathbf{v}\} + \Gamma_T$, the group Γ is generated by $f(p) = L(p) + \mathbf{v}$ and the translation associated with the vector $\lambda_2\mathbf{e}_2$. To summarize:

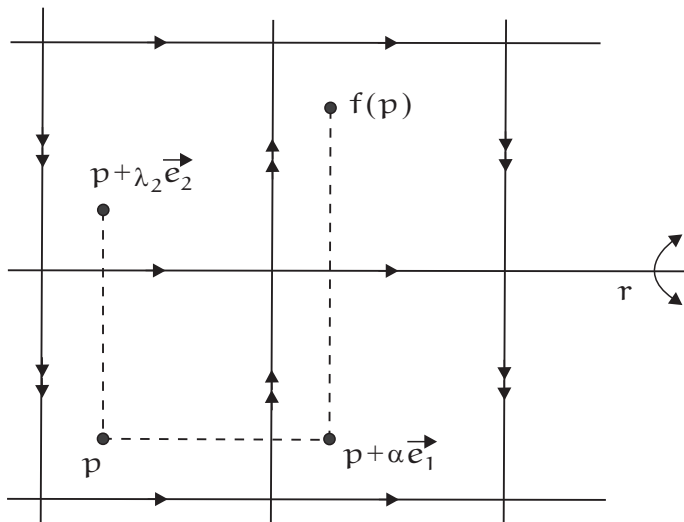


Figure 5.10

If Γ contains elements that reverse orientation and Γ_T is two-dimensional, then Γ is generated by a glide reflection, with axis on a line r say, and by a translation in the direction perpendicular to r . In this case $\mathbb{R}^2/\Gamma$ is a Klein bottle.

In Fig. 5.10 we show several copies of a fundamental region (a rectangle) for this group, marking the identifications to be made.

Let us summarize our findings:

Theorem 5.7.10 Every complete surface of constant zero curvature is diffeomorphic to one of the following surfaces:

- the plane, the cylinder, or the torus, if it is orientable;
- the Möbius strip or the Klein bottle, if non-orientable.

Exercises

145. Prove Proposition 5.7.6. **Hint:** if S_K/Γ is orientable and an orientation is fixed on it then the quotient mapping $S_K \rightarrow S_K/\Gamma$ induces an orientation on S_K .

146. (a) Let Γ_1 and Γ_2 be discrete subgroups of $\mathfrak{I}$. Show that S_K/Γ_1 and S_K/Γ_2 are isometric if and only if Γ_1 and Γ_2 are conjugate subgroups.

(b) Give an example of two tori of equal area and constant zero curvature that are not isometric.

147. Show that if Γ is a discrete subgroup of $\mathfrak{I}(\mathbb{D})$, then $\Gamma \setminus \{\text{id}\}$ contains only hyperbolic isometries (see, in Section 5.6, exercises 140, 143 and 144).

148. (a) Consider on the torus of revolution $\mathbb{T}^2 \subseteq \mathbb{R}^3$, given by $\Phi(u, v) = ((2 + \cos v) \cos u, (2 + \cos v) \sin u, \sin v)$, the equivalence relation $\sim$ that identifies (x, y, z) and $(-x, -y, -z)$. Show that $\mathbb{T}^2/\sim$ is a Klein bottle.

(b) If $\simeq$ is the equivalence relation that identifies (x, y, z) and $(-x, -y, z)$, what surface is $\mathbb{T}^2/\simeq$?

References

1. E. Barbier: *Note sur le problème de l'aiguille et le jeu du joint couvert*. J. Math. Pures Appl. (2) 5 (1860).
2. A.F. Beardon: *The Geometry of Discrete Groups*. Springer-Verlag 1983.
3. A.F. Beardon: *An introduction to Hyperbolic Geometry* [in Bedford, Keane, Series (eds.): *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces*]. Oxford U.P. 1991.
4. A.L. Besse: *Manifolds all of whose Geodesics are Closed*. Springer-Verlag 1978.
5. J.H. Cadwell: *Topics in Recreational Mathematics*. Cambridge University Press 1966.
6. M.P. do Carmo: *Differential Geometry of Curves and Surfaces*. Prentice-Hall 1976.
7. C.D. Chakerian, H. Groemer: *Convex bodies of constant width* [in P.M. Gruber, J.M. Wills (eds.): *Convexity and its applications*]. Birkhäuser 1983.
8. S.S. Chern: *Curves and Surfaces in Euclidean Space* [in S.S. Chern (ed.): *Global Differential Geometry*]. Math. Association of America 1988.
9. H.G. Eggleston: *Convexity*. Cambridge University Press 1958.
10. R. Fenn: *What is the Geometry of a Surface?* Am. Math. Monthly 90 (1983)
11. D.G. Figueiredo: *Análise de Fourier e equações diferenciais parciais*. Projeto Euclides, IMPA 1977.
12. A. Goetz: *Introduction to Differential Geometry*. Addison-Wesley 1970.
13. W. Klingenberg: *A course in Differential Geometry*. Springer-Verlag 1978.
14. E. Kreyszig: *Introduction to Differential Geometry and Riemannian Geometry*. University of Toronto Press 1968.
15. E.L. Lima: *Duas novas demonstrações do teorema de Jordan-Brouwer no caso diferenciável*. Matemática Universitária 4 (1986).
16. E.L. Lima: *Curso de Análise*, vol. 2 (3ª edição). Projeto Euclides, IMPA 1989.
17. E.E. Moise: *Geometric Topology in dimensions 2 and 3*. Springer-Verlag 1977.
18. I. Niven: *Maxima and minima without Calculus*. Math. Association of America 1981.
19. R. Osserman: *A survey of Minimal Surfaces*. Dover Publications 1986.
20. R. Osserman: *Curvature in the Eighties*. Am. Math. Monthly 97 (1990).
21. R. Osserman: *The four-or-more vertex theorem*. Am. Math. Monthly 92 (1985).
22. L.A. Santaló: *Integral Geometry* [in S.S. Chern (ed.): *Global Differential Geometry*]. Math. Association of America 1988.
23. J. Sotomayor: *Lições de equações diferenciais ordinárias*. Projeto Euclides, IMPA 1979.
24. M. Spivak: *Calculus*. Benjamin 1967.
25. M. Spivak: *A comprehensive introduction to Differential Geometry*, vol. I-IV. Publish or Perish 1970-75.

Index

- Discontinuous group Action, 174
- Angle, 1, 53
 - oriented, 10, 53
- Area, 54, 157
 - of an oval surface, 158
 - of surfaces parallel to an oval, 160
 - between parallel curves, 162
- Atlas, 164 – 165
 - oriented, 49, 49
- Klein bottle, 179
- Euler's Characteristics, 116
- Catenary, 90
- Catenoid, 90, 108
- Center of curvature, 6
- Change of coordinates, 38, 166
- Clairaut equation, 129, 143
- Clover, 2
- Cylinder, 89, 102, 143, 151, 177, 179
- Geodesic circumference, 123, 131, 170
- Connected component, 147
- Cone, 39, 89, 100, 141, 143
- Contact
 - of order n between planar curves, 12
 - of order ≥ 2 between surfaces, 73, 75
- Coordinates
 - spherical, 33
 - isothermal, 88
 - local (or parametrization), 32, 41
 - orthogonal, 54, 82, 112
 - polar, 35
 - polar geodesic, 123, 127, 130
 - semi-geodesic, 131
- Covering, 144, 145, 148
 - isometric, 151, 172
 - number of sheets, 146
- Curve
 - convex, 15 – 18
 - coordinate, 32, 57, 81
 - strictly convex, 18
 - closed, 14
 - of constant width, 19 – 24
 - integral of a vector field, 76
 - parallel, 24, 162
 - regular, 2
 - regular in parts, 23
 - simple, 14
- Curvature
 - of a curve, 5, 8
 - of Gauss, 62, 69, 93, 112, 127, 131
 - geodesic, 98, 99, 110, 131
 - mean, 62, 70, 101, 156
 - total mean, 155, 160, 161, 163
 - normal, 60, 69
 - principal, 60, 70
 - with sign, 10, 10
 - total, 116 – 118, 119
- Derivative
 - of a differentiable mapping, 44, 44
 - of the Gauss mapping, 59, 69
 - covariant, 95 – 96, 97
- Diffeomorphism, 42, 45
 - local, 46
 - orientation-preserving, 50, 52
 - area-preserving, 57, 94
- Continuous Deformation of a curve, 146
- Direction
 - asymptotic, 63, 67
 - principal, 60, 70
- Poincaré Disk, 167
 - geodesic circumferences, 170
 - distance between two points, 169

- geodesics, 170
- and non-Euclidean geometry, 170
- isometries, 168, 171, 172
- Intrinsic Distance, 126, 131
- Divergence, 96, 100, 102, 103
- Ellipsoid, 64 – 65
- Symmetric Endomorphism (or self-adjoint), 60
- Evolute, 11
- Direction Field, 80, 81
- Gradient Field, 107 – 108
- Field of vectors
 - along a curve, 97
 - divergence, 96, 100, 102, 103
 - of integral curves or paths, 76
 - of normal vectors, 47
 - of parallel vectors, 99 – 100
 - of tangent vectors on a surface, 79
- First fundamental form, 52
- Formula
 - Blaschke's, 161
 - Cauchy's, 158
 - Euler's, 115
 - Girard's, 55, 115 – 116
- Formulas
 - Frenet's, 7, 10
 - Minkowski integrals, 154
- Geodesics, 100
 - surfaces of revolution, 128
 - differential equations, 120, 128, 151
 - closed, 118, 122
 - minimizers, 126, 140
 - parametrized, 120
- Gradient, 107
- Group
 - acting discontinuously, 174
 - discrete, 174
 - isometries of a surface, 173
 - covering, 173
 - of Möbius transformations, 169
- Helix, 2, 102
- Helicoid, 35, 90, 107
- Homotopy with fixed ends, 146
- Index of Rotation, 14, 18, 110, 134, 136
- Isometry, 88, 131, 151, 173
 - elliptical, 172
 - hyperbolic, 172, 180
 - local, 88, 93, 128, 151, 172
 - parabolic, 172
- Isoperimetric inequality, 28
- Length, 3, 53
 - of an arc, 4
 - of a curve of constant width, 22, 24
 - of a parallel curve, 162
- Lifting, 132 – 133, 146
- Lifting curves, 146, 146
- Lines
 - asymptotic, 74, 75, 81, 82
 - of curvature, 74, 79, 82
 - integrals, 80
- Mapping
 - conformal, 86, 94, 131
 - differentiable, 42
 - exponential, 122, 150, 151
 - linear symmetric (or self-adjoint), 60
 - Gauss normal, 59, 148
 - quotient, 166
- Riemannian Metric, 164, 166
 - conformal, 164
- Möbius strip, 47, 52, 178, 179
- Chebyshev Net, 57
- Evenly covered Neighborhood, 144
- Normal Neighborhood, 123
- Principal Norm, 5
- Open subset of a surface, 32
 - connected, 40
 - evenly covered, 144
 - simply connected, 148
- Orientation, 47, 49, 175
- Oval, 148, 152
- Osculating Paraboloid, 74
- Parametrization: see Coordinates
- Plane, 32, 179
 - geodesics, 101
 - isometries, 88, 90
- Hyperbolic Plane: see Poincaré disk
- Osculating Plane, 6
- Projective Plane, 175
- Tangent Plane, 44
- Point
 - antipodal, 20, 55, 57, 76, 160
 - elliptical, 62, 71
 - hyperbolic, 62, 71
 - parabolic, 62, 67, 83
 - planar, 62, 67
 - umbilical, 62, 65
- Projection
 - of Archimedes, 57
 - of Mercator, 90
 - stereographic, 36, 87

- gnomonic, 57
- Pseudo-sphere, 36, 75
- Radius of curvature, 6
- Glide reflection, 176
- Fundamental region, 175
- Polygonal region, 102
- Normal Section, 61
- Second fundamental form, 67, 74
- Similarity, 86, 87, 131
- Poincaré Semi-plane, 171
- Christoffel Symbols, 91
- Conjugate Subgroups, 171, 180
- Sphere, 33, 63, 70
 - geodesics, 122
 - isometries, 174 – 175
 - rigidity theorem, 156
 - parallel transport, 100
- Tangent Space, 44, 165 – 166
- Surfaces
 - abstract, 164
 - complete, 139, 140, 142
 - connected, 40
 - convex, 152
 - of constant curvature, 70, 75, 123, 128, 130, 131, 156, 172, 175, 179
 - of non-positive curvature, 150, 151
 - inextensible, 141
 - of constant width, 75 – 76, 159 160 – 161
 - minimal, 107, 108
 - level, 40, 42
 - orientable, 46 – 47, 49, 75, 175
 - oval, 148, 152
 - parallel, 160
 - parameterized, 34
 - ruled, 75, 102
 - regular, 31, 40
 - of revolution, 34, 70, 75, 102, 128
 - simply connected, 148, 151, 151
- Spindle, 55
- Theorem
 - of Barbier, 22
 - of Brouwer's fixed point, 136
 - of divergence, 102, 103, 161 – 161
 - of Gauss, 93
 - of Gauss-Bonnet, 109, 116
 - of Hadamard, 148, 151
 - of the four vertices, 26
 - of the rigidity of the sphere, 156
 - of the rotation index, 134, 136
- Torsion, 7, 8, 14
- Torus, 36, 43, 82, 94, 143, 144, 166, 177, 179
- Double Torus, 117, 175
- Trace of a curve, 1
- Tractrix, 36
- Trajectory of a vector field, 76
- Möbius Transformation, 169
- Parallel Transport, 99, 100, 112, 119
- Triangulation, 113, 131
- Geodesic triangle, 115, 130, 143, 171
- Reuleaux triangle, 23
- Frenet's triangle, 6
- Regular Value, 40
- Variation
 - of area, 106
 - normal, 103
 - of a surface, 103, 159
- Velocity, 2
 - scalar, 3
- Volume
 - between parallel surfaces, 160
 - bounded by a surface of constant width, 161
- Width, 19
- Constant Width
 - of a curve, 19 – 24
 - of a surface, 71 – 76, 159, 160 – 161
- Wirtinger's lemma, 27